《线性代数》

杨启哲

上海师范大学信机学院计算机系

2024 年 5 月 29 日

此次课件参考上海交通大学陈翌佳教授的课程课件，感谢陈翌佳教授的分享。

目录 I

目录 II

- 向量空间 (Vector Space)
  - 向量空间的基本概念
  - 子空间
- 相关性、基和维度 (Independence, Basis and Dimension)
  - 线性相关性
  - 向量空间的基和维度
- 矩阵的列秩与行秩 (Column Rank and Row Rank of Matrices)
  - 列空间的秩与行空间的秩
  - 通过 Gauss–Jordan 消元法来求 A^{-1}
- 解线性方程组 (II)(Solving Linear Equations(II))
  - 矩阵的秩
  - $A\mathbf{x} = \mathbf{0}$ 的解
  - $A\mathbf{x} = \mathbf{b}$ 的解

目录 III

目录 IV

上海师范大学 Shanghai Normal University

- 对称矩阵和正定矩阵 (Symmetric Matrices and Positive Definite Matrices)
  - 对称矩阵
  - 正定矩阵
  - 特征空间、代数重数以及几何重数
- 线性变换 (Linear Transformation)
  - 线性变换的概念
  - 线性变换的矩阵形式
  - 线性变换的像和核
  - 对偶性 (Duality)
- 奇异值分解 (Singular Value Decomposition)
  - SVD 基础
  - 用线来拟合数据
  - 用 k 维子空间拟合数据
  - 再看 $A\mathbf{x} = \mathbf{b}$ 的近似解

线性代数课程介绍 (Introduction to Linear Algebra)

线性代数课程介绍 (Introduction to Linear Algebra)

什么是线性代数

课程目标

高中数学到大学数学的飞跃

1. 从强调计算（算术）到理解数学结构的转变。
2. 从牢记结论到掌握证明的转变。
3. 建立抽象的几何直观。

什么是线性代数?

线性代数研究:

1. 向量空间（vector space）。
2. 向量空间之间的线性变换（linear transformations or linear map）

一个线性函数

考虑如下的函数:

$$f(x) := 3x$$

这是一个 $\mathbb{R} \to \mathbb{R}$ 的**线性函数**。

- 从几何的角度来讲，这是平面 $\mathbb{R}^2$ 上的一条线。
- 从代数的角度来讲，对于任意的 $x_1, x_2, c, x \in \mathbb{R}$ 我们有:

$$\begin{aligned} f(x_1 + x_2) &= f(x_1) + f(x_2) \\ f(cx) &= cf(x) \end{aligned}$$

牛顿第二定律告诉我们:

$$F = ma \text{ 或者} a = \frac{F}{m}$$

在 $\mathbb{R}^2$ 上

- F 是一个向量 (F_x, F_y)，从而:

$$a = (a_x, a_y) = \left(\frac{F_x}{m}, \frac{F_y}{m}\right)$$

- 这是一个线性的转换。

$$\begin{aligned} \left(\frac{F_x + F_x'}{m}, \frac{F_y + F_y'}{m}\right) &= \left(\frac{F_x}{m}, \frac{F_y}{m}\right) + \left(\frac{F_x'}{m}, \frac{F_y'}{m}\right) \\ \left(\frac{cF_x}{m}, \frac{cF_y}{m}\right) &= c\left(\frac{F_x}{m}, \frac{F_y}{m}\right) \end{aligned}$$

在 $\mathbb{R}^3$ 上

$$F = (F_x, F_y, F_z) \text{ 和} a = (a_x, a_y, a_z) = \left(\frac{F_x}{m}, \frac{F_y}{m}, \frac{F_z}{m}\right)$$

- 我们有:

$$\begin{aligned} a_x &= \frac{1}{m}F_x + 0F_y + 0F_z \\ a_y &= 0F_x + \frac{1}{m}F_y + 0F_z \\ a_z &= 0F_x + 0F_y + \frac{1}{m}F_z \end{aligned}$$

- 该线性变换可以用如下的矩阵 (matrix)表示。

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

- 代数 (Algebra): 通俗来讲，代数就是将一些符号化对的对象组合起来并进行运算。比如如何简化类似下面的表达式:

$$(x+1)(x-3) = x^2 - 2x - 3$$

而对于线性代数而言，我们运算的对象并不一定时标量 (scalars)，还可能是向量 (vectors) 或者矩阵 (matrices)，抑或是线性变换 (linar transformations)。

- 线性方程组 (linear systems): 不难发现，如下的线性方程组 $A\mathbf{x} = \mathbf{b}$ 实际上是线性代数中的一个核心问题:

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 1 \\ 3x_1 + 4x_2 + 5x_3 &= 2 \\ 4x_1 + 5x_2 + 6x_3 &= 3 \end{aligned}$$

$$\underbrace{\begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{\mathbf{b}}$$

让我们从向量开始吧!

向量 (Vectors)

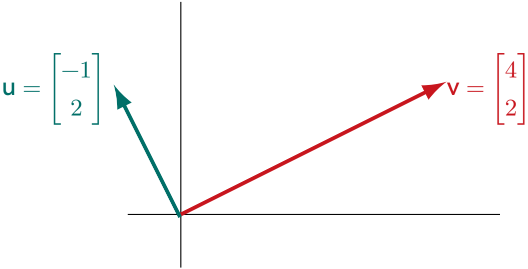
向量 (Vectors)

向量加法和数乘

向量 (Vectors)

首先我们来考察一下二维空间中的向量，用**列向量 (colomn vectors)** 来表示：

- $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$



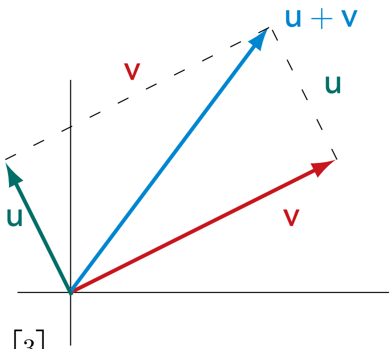
符号说明

在本课程的课件中，我们使用 $\mathbf{x}, \mathbf{y}, \mathbf{z} \ldots$ 等符号来表示向量，使用 $x, y, z, \ldots$ 等符号来表示标量的值；特别的对于一个向量 $\mathbf{u}$ 来说，我们经常使用 u_i 表示其第 i 个分量的值。

向量加法 (Vector Addition)

向量加法:

$$\underbrace{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}_{\mathbf{u}} + \underbrace{\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}}_{\mathbf{v}} = \underbrace{\begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}}_{\mathbf{u}+\mathbf{v}}$$



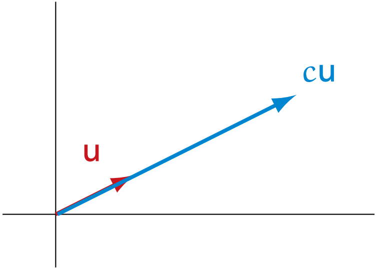
- 一个例子: $\begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

向量数乘 (Scalar Multiplication)

向量数乘

$$c \quad \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad = \quad \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}$$

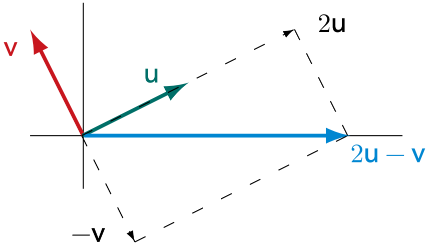
其中 c 是一个标量，也称为 scalar.



- 一个例子：$3\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$

考虑如下的例子:

$$2\mathbf{u}-\mathbf{v}=2\begin{bmatrix}2\\1\end{bmatrix}+(-1)\begin{bmatrix}-1\\2\end{bmatrix}=\begin{bmatrix}4\\2\end{bmatrix}-\begin{bmatrix}-1\\2\end{bmatrix}=\begin{bmatrix}5\\0\end{bmatrix}$$



向量的线性组合

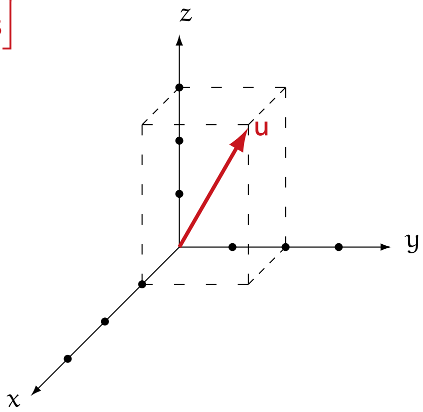
$$c\begin{bmatrix}u_1\\u_2\end{bmatrix}+d\begin{bmatrix}v_1\\v_2\end{bmatrix}=\begin{bmatrix}cu_1+dv_1\\cu_2+dv_2\end{bmatrix}$$

$$c\mathbf{u}\qquad d\mathbf{v}\qquad c\mathbf{u}+d\mathbf{v}$$

向量在三维的情况

- 二维向量 $\begin{bmatrix} x \\ y \end{bmatrix}$ 可以视作在 2 维 xOy 平面上从 $(0,0)$ 指向 (x,y) 的一个有向线段。
- 三维向量也是类似的，只不过是在 3 维空间中从 $(0,0,0)$ 指向 (x,y,z) 的一个有向线段，

  比如考察向量 $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$：



三维向量的线性组合

上海师范大学 Shanghai Normal University

给定 3 维的向量:

$$\mathbf{u}=\begin{bmatrix}1\\0\\3\end{bmatrix},\ \mathbf{v}=\begin{bmatrix}1\\2\\1\end{bmatrix},\ \mathbf{w}=\begin{bmatrix}2\\3\\-1\end{bmatrix}$$

其线性组合也是类似的:

$$\begin{aligned}
\mathbf{u}=1\mathbf{u} &= \begin{bmatrix}1\\0\\3\end{bmatrix}\\
\mathbf{u}+2\mathbf{v} &= \begin{bmatrix}1\\0\\3\end{bmatrix}+2\begin{bmatrix}1\\2\\1\end{bmatrix}=\begin{bmatrix}3\\4\\5\end{bmatrix}\\
\mathbf{u}+4\mathbf{v}-2\mathbf{w} &= \begin{bmatrix}1\\0\\3\end{bmatrix}+4\begin{bmatrix}1\\2\\1\end{bmatrix}-2\begin{bmatrix}2\\3\\-1\end{bmatrix}=\begin{bmatrix}1\\2\\9\end{bmatrix}
\end{aligned}$$

三个问题

假设 $\mathbf{u}, \mathbf{v}, \mathbf{w}$ 是三维空间的三个向量:

- 所有形如 $c\mathbf{u}$ 的线性组合对应的几何直观是什么?
- 所有形如 $c\mathbf{u}+d\mathbf{v}$ 的线性组合对应的几何直观是什么?
- 所有形如 $c\mathbf{u}+d\mathbf{v}+e\mathbf{w}$ 的线性组合对应的几何直观是什么?

答案当然依赖于 $\mathbf{u}, \mathbf{v}, \mathbf{w}$ 的具体取值，但是我们可以通过一些例子来感受一下。

上海师范大学 Shanghai Normal University

我们考虑之前的三个向量：

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$$

1. 形如 $c\mathbf{u}$ 的线性组合填满了一条通过 $(0,0,0)$ 的线。
2. 形如 $c\mathbf{u}+d\mathbf{v}$ 的线性组合填满了一个通过 $(0,0,0)$ 的平面。
   - $(2,3,-1)$ 便不在此平面上，因此 $\mathbf{w}$ 无法表示成 $c\mathbf{u}+d\mathbf{v}$ 的线性组合。
3. 形如 $c\mathbf{u}+d\mathbf{v}+e\mathbf{w}$ 的线性组合填满了整个三维空间。
   - 这意味着对于任何的 $a,b,c \in \mathbb{R}$

$$\begin{aligned} x+y+2z &= a \\ 2y+3z &= b \\ 3x+y-z &= c \end{aligned}$$

存在一个解。

一个例子

问题 1.

描述一下由向量 $\mathbf{u}=\begin{bmatrix}1\\1\\0\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}0\\1\\1\end{bmatrix}$ 的线性组合在三维空间中所组成的平面。

解 2.

形如 $c\mathbf{u}+d\mathbf{v}$ 的线性组合填满了一个通过 $(0,0,0)$ 的平面，我们有：

$$c\mathbf{u}+d\mathbf{v}=\begin{bmatrix}c\\c+d\\d\end{bmatrix}$$

c,d 是任意的，所以该平面包含了所有第二维是第一维和第三维的和的向量。

阶段总结

1. 二维向量是一个具有两个分量的有序对，可以视作二维平面原点出发的一个有向线段。
2. 向量的加法和数乘是针对对应分量进行操作。
3. 向量之间的线性组合是指形如 $c\mathbf{u} + d\mathbf{v} + e\mathbf{w}$ 的向量。
4. 在三维空间中，向量的线性组合可以填满一条线，一个平面，或者整个三维空间。

向量 (Vectors)

向量长度和点积

符号说明

为了节省空间，列向量 $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ 有时使用 (u_1, u_2, u_3) 表示。

我们来考察两维中的向量，给定两个向量 $\mathbf{u} = (u_1, u_2)$，$\mathbf{v} = (v_1, v_2)$：

1. 向量 $\mathbf{u}$ 的长度 $\|\mathbf{u}\|$ 是多少？

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2}$$

2. 向量 $\mathbf{u}$ 与 $\mathbf{v}$ 的夹角 θ 是多少？

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

这些问题我们都可以使用点积 (dot product) 的概念来解决。

点积 (Dot Products)

上海师范大学 Shanghai Normal University

定义 3 **[点积].**

向量 $\mathbf{u} = (u_1, u_2)$，$\mathbf{v} = (v_1, v_2)$ 的点积 $\mathbf{u} \cdot \mathbf{v}$ 定义为：

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2$$

一般的，对于向量 $\mathbf{u} = (u_1, u_2, \ldots, u_n)$，$\mathbf{v} = (v_1, v_2, \ldots, v_n)$，其点积定义为：

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

点积的一些性质

1. $\mathbf{u}$ 和 $\mathbf{v}$ 是垂直的 (perpendicular) 当且仅当 $\mathbf{u} \cdot \mathbf{v} = 0$。
2. 点积是可交换的，即 $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$。

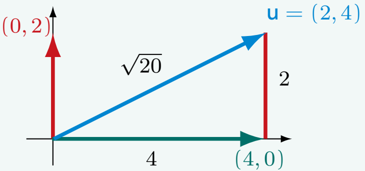
上海师范大学 Shanghai Normal University

定义 4 **[向量的长度].**

向量 $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ 的长度 $\|\mathbf{u}\|$ 定义为:

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = \sqrt{\mathbf{u}\cdot\mathbf{u}}$$

勾股定理 (Pythagorean Law)



一般来说，对于垂直的 $\mathbf{u}$ 和 $\mathbf{v}$:

$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}-\mathbf{v}\|^2$$

单位向量 (Unit Vector)

定义 5.

长度为 1 的向量 $\mathbf{u}$ 被称作为单位向量，即 $\|\mathbf{u}\| = 1$.

例 6.

如下的向量都是单位向量:

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ \mathbf{u} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

单位化 (Normalization)

给定一个向量 $\mathbf{u} = (2, 2, 1)$，其长度为:

$$\|\mathbf{u}\| = \sqrt{2^2 + 2^2 + 1^2} = \sqrt{9} = 3$$

我们可以将其缩小成一个单位向量:

$$(\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$$

引理 7.

给定一个非零向量 $\mathbf{u} = (u_1, \ldots, u_n)$，则:

$$\frac{\mathbf{u}}{\|\mathbf{u}\|} = (\frac{u_1}{\|\mathbf{u}\|}, \ldots, \frac{u_n}{\|\mathbf{u}\|})$$

是一个与 $\mathbf{u}$ 同方向的单位向量。

互相垂直的向量

现在我们来说明点积值和向量之间的夹角的关系：

引理 8.

令 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$，如果 $\mathbf{u}$ 和 $\mathbf{v}$ 是垂直的，则：$\mathbf{u} \cdot \mathbf{v} = 0$。

证明. 由勾股定理，我们有：

$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} - \mathbf{v}\|^2$$

将其展开有：

$$\begin{aligned} u_1^2 + u_2^2 + v_1^2 + v_2^2 &= (u_1 - v_1)^2 + (u_2 - v_2)^2 \\ u_1 v_1 + u_2 v_2 &= 0 \end{aligned}$$

□

说明

1. $\mathbf{u} \cdot \mathbf{v} = 0 \iff \mathbf{u}$ 和 $\mathbf{v}$ 是垂直的。
2. $\mathbf{0} \cdot \mathbf{u} = 0$，零向量 $\mathbf{0}$ 与任何向量都是垂直的。

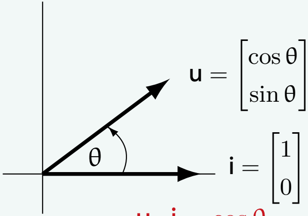
定理 9.

令 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ 是两个单位向量，θ 是 $\mathbf{u}$ 和 $\mathbf{v}$ 之间的夹角，则：

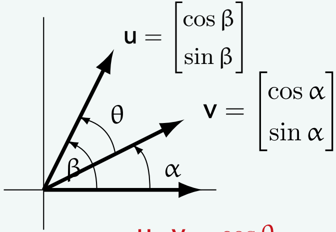
$$\cos\theta = \mathbf{u} \cdot \mathbf{v}$$

几何视角

1. $\mathbf{v} = \mathbf{i} = (1, 0)$，则有 $\cos\theta = \mathbf{u} \cdot \mathbf{v} = u_1$。
2. $\mathbf{v} \neq (1, 0)$，则可以视其旋转了 α 角度，其中 $\mathbf{v} = (\cos\alpha, \sin\alpha)$。



$\mathbf{u} \cdot \mathbf{i} = \cos\theta$



$\mathbf{u} \cdot \mathbf{v} = \cos\theta$

定理 9的证明.

- 若 $\mathbf{v} = (1, 0)$，则易得 $\mathbf{u} \cdot \mathbf{v} = u_1 = \cos\theta$.
- 若 $\mathbf{v} \neq (1, 0)$，则可以视 $\mathbf{v}$ 是由 $(1, 0)$ 旋转了 α 角度得到的单位向量，即 $\mathbf{v} = (\cos\alpha, \sin\alpha)$; 同理令 $\mathbf{u} = (\cos\beta, \sin\beta)$，则其夹角为 $\theta = \beta - \alpha$，并且:

$$\mathbf{u} \cdot \mathbf{v} = \cos\beta\cos\alpha + \sin\beta\sin\alpha = \cos(\beta - \alpha) = \cos\theta$$

□

如果 $\mathbf{u}, \mathbf{v}$ 不是单位向量，怎么求其夹角？

- 将其单位化。

定理 10.

令 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ 是两个非零向量，θ 是 $\mathbf{u}$ 和 $\mathbf{v}$ 之间的夹角，则：

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

上海师范大学 Shanghai Normal University

现在我们来看一些例子:

定理 11 [Cauchy-Schwarz-Buniakowsky].

$$|\mathbf{u}\cdot\mathbf{v}|\leqslant\|\mathbf{u}\|\|\mathbf{v}\|$$

证明. 我们用一个不用余弦定理的方法来证明。

注意到:

$$\begin{aligned}(u_1^2+u_2^2)(v_1^2+v_2^2)-(u_1v_1+u_2v_2)^2 &= u_1^2v_2^2+u_2^2v_1^2-2u_1u_2v_1v_2\\ &= (u_1v_2-u_2v_1)^2\geqslant 0\end{aligned}$$

□

定理 12 **[三角不等式].**

$$\|\mathbf{u}+\mathbf{v}\| \leqslant \|\mathbf{u}\|+\|\mathbf{v}\|$$

证明. 我们同样使用一个不用余弦定理的方法来证明。

注意到:

$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 &= \|\mathbf{u}\|^2+\|\mathbf{v}\|^2+2\mathbf{u}\cdot\mathbf{v} \\ \mathbf{u}\cdot\mathbf{v} &\leqslant \|\mathbf{u}\|\|\mathbf{v}\| \end{aligned}$$

从而我们有:

$$\|\mathbf{u}+\mathbf{v}\|^2 \leqslant (\|\mathbf{u}\|+\|\mathbf{v}\|)^2$$

□

问题

如何证明 $\|\mathbf{u}+\mathbf{v}\|^2=\|\mathbf{u}\|^2+\|\mathbf{v}\|^2+2\mathbf{u}\cdot\mathbf{v}$?

1. 向量的点积是相应部分的乘积的和，即 $\mathbf{u}\cdot\mathbf{v}=u_1v_1+\ldots u_nv_n$。
2. 向量的长度是点积的平方根，其对应的单位向量为：$\frac{\mathbf{u}}{\|\mathbf{u}\|}$，长度为 1。
3. $\mathbf{u}\cdot\mathbf{v}=0$ 意味着两个向量 $\mathbf{u},\mathbf{v}$ 是垂直的。
4. 向量夹角的余弦值等于点积除以长度的乘积，即:

$$\cos\theta=\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

5. 柯西-施瓦茨不等式和三角不等式。

$$|\mathbf{u}\cdot\mathbf{v}|\leqslant\|\mathbf{u}\|\|\mathbf{v}\|,\qquad \|\mathbf{u}+\mathbf{v}\|\leqslant\|\mathbf{u}\|+\|\mathbf{v}\|$$

向量 (Vectors)

矩阵

线性组合的矩阵视角 (I)

给定三个向量:

$$\mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

其线性组合可以表示为 $x_1\mathbf{u} + x_2\mathbf{v} + x_3\mathbf{w}$，也就是:

$$x_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}$$

矩阵表示

我们可以利用矩阵来表示上述行为:

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}$$

我们可以理解成，矩阵 A 作用在一个列向量 $\mathbf{x}$ 上，其结果是矩阵 A 中的列向量的线性组合。

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b}$$

例 13.

上述的矩阵 A 被称作差分矩阵(difference matrix)，因为其得到的向量是原向量的差分。

$$A \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 1-0 \\ 4-1 \\ 9-4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

另一个视角来看待矩阵的作用

我们现在用行的角度来看待矩阵作用在向量上的结果:

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (1,0,0)\cdot\mathbf{x} \\ (-1,1,0)\cdot\mathbf{x} \\ (0,-1,1)\cdot\mathbf{x} \end{bmatrix} = \mathbf{b}$$

补充说明

这是大多数中文教材中定义的方式。

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b}$$

- 之前我们讨论的是线性组合，即给定三个向量 $\mathbf{u}, \mathbf{v}, \mathbf{w}$ 和三个数 x_1, x_2, x_3，求其线性组合 $x_1\mathbf{u} + x_2\mathbf{v} + x_3\mathbf{w}$; 将矩阵 A 看成 $\begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{bmatrix}$ 的话，即知道了 A 和 $\mathbf{x}$，求 $\mathbf{b}$。
- 现在我们来考虑另一个问题：给定矩阵 A 和 $\mathbf{b}$，求 $\mathbf{x}$。

一个大家更为熟知的形式：线性方程组 (Linear Equations):

$$
\begin{aligned}
x_1 + 0x_2 + 0x_3 &= b_1 \\
-x_1 + x_2 + 0x_3 &= b_2 \\
0x_1 - x_2 + x_3 &= b_3
\end{aligned}
$$

不难验证，其解可以表示为：

$$
\begin{aligned}
x_1 &= b_1 \\
x_2 &= b_1 + b_2 \\
x_3 &= b_1 + b_2 + b_3
\end{aligned}
$$

上述过程也可以写成另一个矩阵：

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = A^{-1}\mathbf{b}$$

上述矩阵称为差分矩阵 A 的逆矩阵 A^{-1}。

我们再来进行一下对比:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \ A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \ A^{-1}\mathbf{b} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

另一个例子 (I)

给定三个向量:

$$\mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

其线性组合可以表示为 $x_1\mathbf{u} + x_2\mathbf{v} + x_3\mathbf{w}$，也就是:

$$x_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}$$

矩阵表示

我们将上述矩阵记为 C，也被称为循环差分矩阵 (cyclic difference matrix):

$$C\mathbf{x} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix}$$

另一个例子 (II)

与 A 的不同的是，$C\mathbf{x}=\mathbf{b}$ 不一定有解，例如：

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

没有解。

$x_1\mathbf{u}+x_2\mathbf{v}+x_3\mathbf{w}$ 的几何直观

换个角度讲，$x_1\mathbf{u}+x_2\mathbf{v}+x_3\mathbf{w}$ 的所有线性组合并没有充满了整个 3 维空间。事实上，其仅仅覆盖了如下的一个平面：

$$b_1+b_2+b_3=0$$

更多的讨论

无关性和相关性 (Independence and Dependence)

- 在上述 $A\mathbf{x}=\mathbf{b}$ 的例子中，$x\mathbf{u}+y\mathbf{v}+z\mathbf{w}=\mathbf{0}$ 当且仅当 $x=y=z=0$。
  后面的课程中我们会看到：
  1. $\mathbf{u},\mathbf{v},\mathbf{w}$ 是线性无关的。
  2. $A\mathbf{x}=\mathbf{0}$ 只有一个解，称 A 是可逆矩阵 (invertible matrix)。
- 在上述 $C\mathbf{x}=\mathbf{b}$ 的例子中，存在任意多个 x,y,z 满足 $x\mathbf{u}+y\mathbf{v}+z\mathbf{w}=\mathbf{0}$。
  后面的课程中我们会看到：
  1. $\mathbf{u},\mathbf{v},\mathbf{w}$ 是线性相关的。
  2. $C\mathbf{x}=\mathbf{0}$ 有无穷多个解，称 C 是一个奇异矩阵 (singular matrix)。

阶段总结

1. 矩阵作用在向量 $A\mathbf{x}$ = 矩阵 A 的列向量的线性组合。
2. $A\mathbf{x} = \mathbf{b}$ 的解为 $\mathbf{x} = A^{-1}\mathbf{b}$。
3. $C\mathbf{x} = \mathbf{0}$ 存在无穷多个解，C 没有逆矩阵。

注意

我们并没有给出严格的相关定义，但我们已经描述了这些关键的想法。

解线性方程组 (I)(Solving Linear Equations(I))

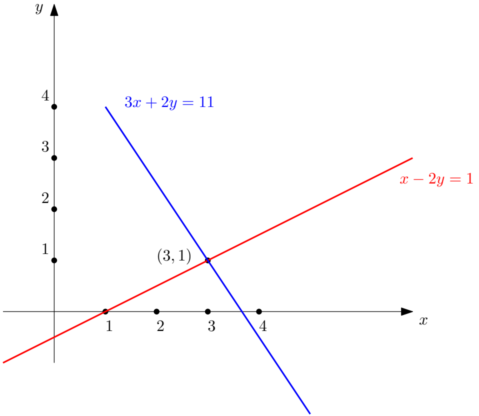
线性方程组

线性方程组的行图像

考虑如下的线性方程组:

$$x - 2y = 1$$
$$3x + 2y = 11$$

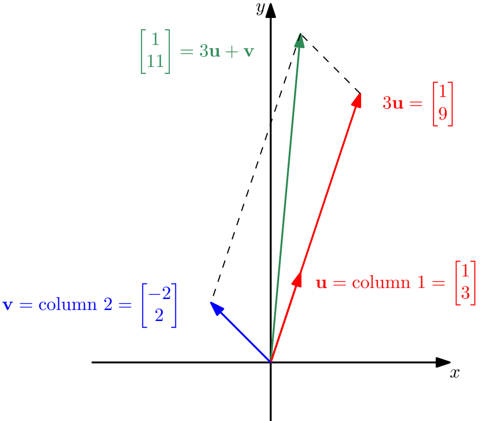
其行图像为:



该方程组也可表示为:

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}$$

其列图像为:



$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \iff x\begin{bmatrix}1\\3\end{bmatrix} + y\begin{bmatrix}-2\\2\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix} \iff \begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}1\\11\end{bmatrix}$$

上述矩阵 $\begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}$ 被称作系数矩阵 (coefficient matrix)，我们也可以用点积的形式理解:

$$\begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}(1,-2)\cdot(x,y)\\(3,2)\cdot(x,y)\end{bmatrix} = \begin{bmatrix}x-2y\\3x+2y\end{bmatrix}$$

不止 2 维!

以上所有当然可以推广到 3 维，4 维，...!

一个三元一次方程组的例子 (I)

考虑 3 个变元的例子：

$$\begin{cases} x + 2y + 3z = 6 \\ 2x + 5y + 2z = 4 \\ 6x - 3y + z = 2 \end{cases}$$

这个方程具有唯一的解 $(x, y, z) = (0, 0, 2)$。

几何视角

- 从行视角来看，这 3 个方程代表了 3 个平面，它们相交于一个点 $(0, 0, 2)$。
- 从列视角来看，这 3 个方程代表了 3 个列向量的线性组合，即：

$$0\begin{bmatrix}1\\2\\6\end{bmatrix} + 0\begin{bmatrix}2\\5\\-3\end{bmatrix} + 2\begin{bmatrix}3\\2\\1\end{bmatrix} = \begin{bmatrix}6\\4\\2\end{bmatrix}$$

同样的，我们有：

$$\begin{cases} x+2y+3z=6 \\ 2x+5y+2z=4 \\ 6x-3y+z=2 \end{cases} \iff \mathbf{A}\mathbf{x} \overset{\text{记为}}{=} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

其 $\mathbf{x}=(x,y,z)$ 是未知数，对应的解是 $(0,0,2)$。从而 $\mathbf{A}\mathbf{x}$ 的目标便是能够有如下的成立：

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

矩阵与向量的乘积

对于：

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

定义 $A\mathbf{x}$ 的第 i 部分值为：

$$a_{i1}x_1 + \cdots + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j = (a_{i1}, \ldots, a_{in}) \cdot (x_1, \ldots, x_n)$$

现在一个 n 元一次方程组已经可以自然的表示成 $A\mathbf{x} = \mathbf{b}$ 了。我们来开始解这个方程组。

高斯消元法 (Gaussian Elimination)

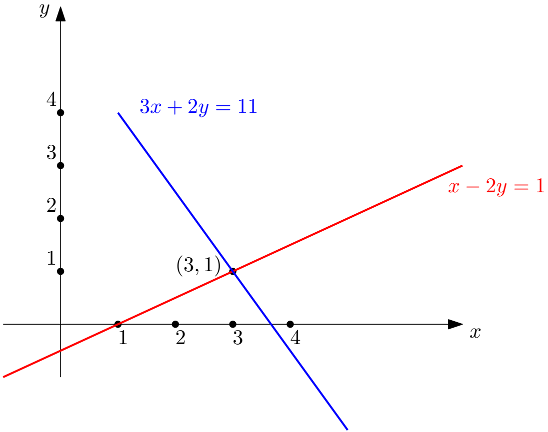
来考虑下面的方程组:

$$\begin{aligned} x - 2y &= 1 \\ 3x + 2y &= 11 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x - 2y &= 1(\text{将第一个式子乘以 } 3) \\ 8y &= 8(\text{减去上面的式子消去 } x) \end{aligned}$$

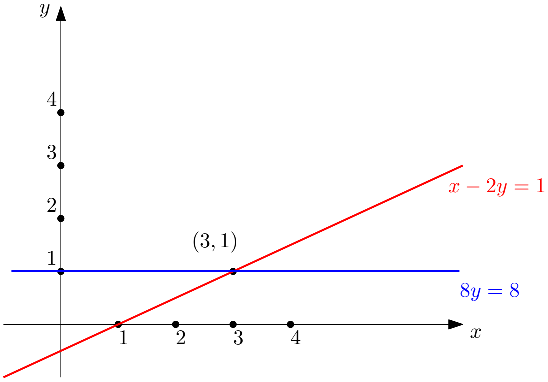
新的方程是上三角的 (upper triangular)，类似的，有:

$$\begin{aligned} 4x - 8y &= 4 \\ 3x + 2y &= 11 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \color{red}{4}x - 8y &= 4(\text{将第一个式子乘以} \frac{3}{4}) \\ \color{red}{8}y &= 8(\text{减去上面的式子消去 } x) \end{aligned}$$

- 首元 (pivot)：行中第一个做其他行消去的非零元素。在完成消去后，首元在对角线上。
- 倍数 (multiplier)：用来消去其他行的倍数。



消去前



消去后

另外一个例子 (I)

考察方程组:

$$\begin{cases} \color{red}{2}x + 4y - 2z = 2 \\ 4x + 9y - 3z = 8 \\ -2x - 3y + 7z = 10 \end{cases}$$

1 第二个等式减去 2 倍的第一个式子，得到 $y + z = 4$。

2 第三个等式减去 -1 倍的第一个式子，得到 $y + 5z = -8$。

此时 x 已经被消去了，剩下:

$$\begin{cases} \color{red}{1}y + z = 4 \\ y + 5z = 12 \end{cases}$$

3 将第二个等式乘以 -1，然后加上第三个等式，得到 $4z = 8$。

最终我们得到了：

$$\begin{aligned} 2x + 4y - 2z &= 2 \\ 4x + 9y - 3z &= 8 \\ -2x - 3y + 7z &= 10 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} 2x + 4y - 2z &= 2 \\ 1y + z &= 4 \\ 4z &= 8 \end{aligned}$$

回代 (Back Substitution)(I)

接下来我们只要从最底下的方程组依次往上便可求出对应的解:

- $$\begin{aligned} x-2y&=1 \\ 3x+2y&=11 \end{aligned} \quad\Longrightarrow\quad \begin{aligned} x-2y&=1 \\ 8y&=8 \end{aligned}$$

  由 $8y=8$ 自然而然有 $y=1$，再代入 $x-2y=1$ 便有 $x=3$。

- $$\begin{aligned} x-2y&=1 \\ 3x-6y&=3 \end{aligned} \quad\Longrightarrow\quad \begin{aligned} x-2y&=1 \\ 0y&=0 \end{aligned}$$

  由 $0y=0$ 从而有无数多个解。

- $$\begin{aligned} x - 2y &= 1 \\ 3x - \color{red}{6}y &= 11 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x - 2y &= 1 \\ 0y &= 8 \end{aligned}$$

  $0y = 8$ 没有解。

首元的个数

可以看到，消元会失败在当 n 个方程没有 n 个首元的时候：

- ”$0 \neq 0$“的方程：没有解。
- ”$0 = 0$“的方程：有无数多个解。

阶段总结

上海师范大学 Shanghai Normal University

- 线性方程组的几何视角，行图像和列图像。
- 线性方程组的矩阵表示，矩阵与向量的乘积。
- 高斯消元法 (Gaussian Elimination)。

解线性方程组 (I)(Solving Linear Equations(I))

解线性方程组的矩阵表示

现在我们用矩阵的形式研究高斯消元法:

$$\begin{aligned} 2x + 4y - 2z &= 2 \\ 4x + 9y - 3z &= 8 \\ -2x - 3y + 7z &= 10 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} {\color{red}2}x + 4y - 2z &= 2 \\ {\color{red}1}y + z &= 4 \\ {\color{red}4}z &= 8 \end{aligned}$$

矩阵形式:

$$\begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ 2 & -2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 3 \end{bmatrix} \quad \Longrightarrow \quad \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 8 \end{bmatrix}$$

记左边形式为 $A\mathbf{x} = \mathbf{b}$，右边的形式为 $U\mathbf{x} = \mathbf{c}$。矩阵 A 怎么变到 U, $\mathbf{b}$ 怎么变到 $\mathbf{c}$?

一步消去步骤的矩阵表示 (I)

上海师范大学 Shanghai Normal University

我们来考察其一个步骤:

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 2 \\ 4x_1 + 9x_2 - 3x_3 &= 8 \\ -2x_1 - 3x_2 + 7x_3 &= 10 \end{aligned}$$

$$\underbrace{\begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix}}_{\mathbf{b}}$$

第二行减去 2 倍的第一行，得到:

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 2 \\ 1x_2 + 1x_3 &= 4 \\ -2x_1 - 3x_2 + 7x_3 &= 10 \end{aligned}$$

$$\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}$$

一步消去步骤的矩阵表示 (II)

我们同样希望能够用矩阵来表示这个过程，即：

$$\begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}$$

注意到：

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}$$

自然我们希望:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix}$$

我们称上述的矩阵为初等矩阵 (elementary matrix)或者消元矩阵 (elimination matrix).

- 如何去定义? 点积。

$A\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ 的列视角:

$$A\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1(\text{colomn 1 of } A) + x_2(\text{colomn 2 of } A) + x_3(\text{colomn 3 of } A)$$

$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} A$ 的行视角:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} A = x_1(\text{row 1 of } A) + x_2(\text{row 2 of } A) + x_3(\text{row 3 of } A)$$

矩阵的乘法 AB 理解

- $AB = A\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix}$
- $AB = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} B = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \mathbf{a}_3 B \end{bmatrix}$

我们现在再回过头来看消元矩阵，我们从如下的单位矩阵 (identity matrix)开始：

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

对于任意 b 有：

$$\mathrm{I}\mathbf{b} = \mathbf{b}$$

当我们将其中某一项 0 的位置替换成了一个非零的数 $-k$，则我们有：

$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -k & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - kb_1 \\ b_3 \end{bmatrix}$$

- 我们将这样子将第 i 行第 j 列的 0 替换成一个非零的数 $-k$ 的矩阵称为 E_{ij}.
- $E_{ij}\mathbf{b}$ 的作用实际上是将 $\mathbf{b}$ 的第 j 行的 $-k$ 倍加到了第 i 行后得到的

定义 14 **[消元矩阵].**

消元矩阵 E_{ij} 是将单位矩阵 I 的第 i 行第 j 列的 0 替换成一个非零的数 $-k$ 得到的矩阵。

符号说明

同济的教材里使用了 $E(ij(-k))$ 的符号表示。

有的时候我们需要将两行进行交换，这样的矩阵称为置换矩阵 (permutation matrix)。

- 如何将 b 的第一行和第二行交换？

$$P_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad P_{12} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_2 \\ b_1 \\ b_3 \end{bmatrix}$$

定义 15 **[置换矩阵].**

置换矩阵 P_{ij} 是将单位矩阵 I 的第 i 行和第 j 行进行交换得到的矩阵。

符号说明

同济的教材里使用了 $E(i,j)$ 的符号表示。

让我们再来看刚刚的例子:

1

$$\begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}$$

其对应的消元矩阵

$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} : \qquad \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix}$$

2

$$\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 12 \end{bmatrix}$$

其对应的消元矩阵

$$E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} : \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 5 \end{bmatrix}$$

3

$$\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 12 \end{bmatrix}$$

其对应的消元矩阵

$$E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}: \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

高斯消元法的矩阵表示-总结 (I)

我们将上述消去的过程合起来，可以表示成:

$$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -1 & 1\end{bmatrix}\left(\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 1 & 0 & 1\end{bmatrix}\left(\begin{bmatrix}1 & 0 & 0\\ -2 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}2 & 4 & -2\\ 4 & 9 & -3\\ -2 & -3 & 7\end{bmatrix}\right)\right)$$

矩阵乘法的运算规则

我们先不加证明的列举出矩阵乘法的运算性质:

1. 满足结合律 (Associate Law): $(AB)C = A(BC)$。
2. 不满足交换律 (Commutative Law): $AB \neq BA$。

高斯消元法的矩阵表示-总结 (II)

通过矩阵乘法的运算规则可知:

$$\begin{bmatrix}1&0&0\\0&1&0\\0&-1&1\end{bmatrix}\left(\begin{bmatrix}1&0&0\\0&1&0\\1&0&1\end{bmatrix}\left(\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&4&-2\\4&9&-3\\-2&-3&7\end{bmatrix}\right)\right)$$

$$=\left(\left(\begin{bmatrix}1&0&0\\0&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\1&0&1\end{bmatrix}\right)\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}\right)\begin{bmatrix}2&4&-2\\4&9&-3\\-2&-3&7\end{bmatrix}$$

$$=\begin{bmatrix}1&0&0\\-2&1&0\\3&-1&1\end{bmatrix}\begin{bmatrix}2&4&-2\\4&9&-3\\-2&-3&7\end{bmatrix}=\begin{bmatrix}2&4&-2\\0&1&1\\0&0&4\end{bmatrix}.$$

回顾线性方程组的表示:

$$\begin{aligned} 2x + 4y - 2z &= 2 \\ 4x + 9y - 3z &= 8 \\ -2x - 3y + 7z &= 10 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} 2x + 4y - 2z &= 2 \\ 1y + z &= 4 \\ 4z &= 8 \end{aligned}$$

相应的消元过程可以用矩阵表示为:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix}.$$

和

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

增广矩阵 (Augmented Matrix)

注意到消元对应的矩阵是同样作用在系数矩阵 A 和 $\mathbf{b}$ 上的，我们可以将其合并成一个矩阵看待：

- 增广矩阵 (Augmented Matrix)：$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 3 \end{bmatrix}$

则上述过程可以一并表示为：

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 & 2 \\ 4 & 9 & -3 & 8 \\ -2 & -3 & 7 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 & 2 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 4 & 8 \end{bmatrix}$$

阶段总结

- 高斯消元法的矩阵表示
- 消元矩阵，置换矩阵。
- 矩阵的乘法视角。

矩阵 (Matrices)

矩阵 (Matrices)

矩阵的运算

定义 16 **[矩阵 (Matrix)].**

令 $m, n \geqslant 1$. 一个 $m \times n$ 的矩阵 A 具有如下的形式:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

特别的，我们用 $A(i, j)$ 表示矩阵 A 的第 i 行第 j 列的元素。

线性方程组:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

可以转换成如下的矩阵形式:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$m \times n \qquad n \times 1 \qquad m \times 1$$

定义 17 **[矩阵加法 (Addition)].**

$$
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
+
\begin{bmatrix}
a'_{11} & a'_{12} & \cdots & a'_{1n} \\
a'_{21} & a'_{22} & \cdots & a'_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a'_{m1} & a'_{m2} & \cdots & a'_{mn}
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
a_{11}+a'_{11} & a_{12}+a'_{12} & \cdots & a_{1n}+a'_{1n} \\
a_{21}+a'_{21} & a_{22}+a'_{22} & \cdots & a_{2n}+a'_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1}+a'_{m1} & a_{m2}+a'_{m2} & \cdots & a_{mn}+a'_{mn}
\end{bmatrix}
$$

矩阵加法 (II)

将两个线性方程组加起来:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \qquad \begin{cases} a'_{11}x_1 + a'_{12}x_2 + \cdots + a'_{1n}x_n = b'_1 \\ a'_{21}x_1 + a'_{22}x_2 + \cdots + a'_{2n}x_n = b'_2 \\ \qquad\qquad\vdots \\ a'_{m1}x_1 + a'_{m2}x_2 + \cdots + a'_{mn}x_n = b'_m \end{cases}$$

得到:

$$\begin{cases} (a_{11} + a'_{11})x_1 + (a_{12} + a'_{12})x_2 + \cdots + (a_{1n} + a'_{1n}x_n = b_1 + b'_1 \\ (a_{21} + a'_{21})x_1 + (a_{22} + a'_{22})x_2 + \cdots + (a_{2n} + a'_{2n})x_n = b_2 + b'_2 \\ \qquad\qquad\vdots \\ (a_{m1} + a'_{m1})x_1 + (a_{m2} + a'_{m2})x_2 + \cdots + (a_{mn} + a'_{mn})x_n = b_m + b'_m \end{cases}$$

定义 18 **[矩阵数乘 (scalar multiplication)].**

$$c\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix}$$

定义 19 **[矩阵乘法 (Matrix Multiplication)].**

令 $m, n, p \geqslant 1$，A 是一个 $m \times n$ 的矩阵，B 是一个 $n \times p$ 的矩阵：

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \qquad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}$$

则矩阵的乘积 AB 是一个 $m \times p$ 的矩阵：

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}$$

其中 c_{ij} 定义为：

$$c_{ij} = (a_{i1}, \dots, a_{in}) \cdot (b_{1j}, \dots, b_{nj}) = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

说明

当 $p = 1$ 的时候便是我们已经定义过的矩阵与向量的乘法。

例 20.

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 3 & 2 & 4 \end{bmatrix} \begin{bmatrix} 2 & 0 & 2 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 5 & 7 \\ 1 & 3 & 3 \\ 8 & 6 & 12 \end{bmatrix}$$

矩阵乘法视角-替换 (I)

Shanghai Normal University

考察如下的两个线性方程组:

$$(\mathrm{I})\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = y_2 \\ \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = y_m \end{cases} \quad \text{矩阵形式} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$(\mathrm{II})\begin{cases} b_{11}z_1 + b_{12}z_2 + \cdots + b_{1p}z_p = x_1 \\ b_{21}z_1 + b_{22}y_2 + \cdots + b_{2p}z_p = x_2 \\ \qquad \vdots \\ b_{n1}y_1 + b_{n2}z_2 + \cdots + b_{np}z_p = x_n \end{cases} \quad \text{矩阵形式} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

上海师范大学 Shanghai Normal University

如果我们希望用 $z_1, \dots z_p$ 来表示 $y_1, \dots, y_m$，则有:

$$\begin{aligned}
y_i &= a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \\
&= a_{i1}(b_{11}z_1 + b_{12}z_2 + \cdots + b_{1p}z_p) + \cdots + a_{in}(b_{n1}z_1 + b_{n2}z_2 + \cdots + b_{np}z_p) \\
&= (a_{i1}b_{11} + a_{i2}b_{21} + \cdots + a_{in}b_{n1})z_1 + \cdots + (a_{i1}b_{1p} + a_{i2}b_{2p} + \cdots + a_{in}b_{np})z_p \\
&= \sum_{j=1}^{p} (a_{j1}, \dots, a_{jn}) \cdot (b_{1j}, \dots, b_{nj})z_j
\end{aligned}$$

用矩阵的形式描述便是由:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \text{和} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \left(\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix} \right)$$

注意到: $y_i = \sum_{j=1}^{p} (a_{i1}, \ldots, a_{in}) \cdot (b_{1j}, \ldots, b_{nj}) z_j$, 这就是:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} \right) \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix}$$

矩阵乘法的列视角

当我们将 B 看成若干列向量组合而成的矩阵时，即:

$$B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix}$$

我们有:

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix}$$

即:

矩阵 AB 的每一列都是 A 中列向量的线性组合。

当我们将 A 看成若干行向量组合而成的矩阵时，即：

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$$

我们有：

$$AB = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B \end{bmatrix}$$

即：

矩阵 AB 的每一行都是 B 中行向量的线性组合。

矩阵乘法的第四种视角

将 A 写成列向量的形式:

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

将 B 写成行向量的形式:

$$B = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$$

则我们有:

$$AB = \mathbf{a}_1\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2 + \cdots + \mathbf{a}_n\mathbf{b}_n$$

矩阵运算的性质 (I)

下面我们给出矩阵运算的一些性质，我们不妨假定里面的矩阵总是符合运算的要求的：

引理 21.

矩阵加法和数乘满足：

1. 交换律 (Commutative Law)： $A + B = B + A$
2. 分配律 (Distributive Law)： $c(A + B) = cA + cB$
3. 结合律 (Associative Law)： $(A + B) + C = A + (B + C)$

引理 22.

矩阵乘法满足:

1. 结合律 (不需要括号): $(AB)C = A(BC)$
2. 分配律 (左分配律): $(A+B)C = AC + BC$
3. 分配律 (右分配律): $A(B+C) = AB + AC$

注意

我们再次强调，矩阵乘法不满足交换律，即一般情况下 $AB \neq BA$。

阶段总结

- 矩阵以及矩阵运算的定义。
- 矩阵乘法的四种理解方式。
- 矩阵运算的性质。

分块矩阵

我们在乘法中已经展示了矩阵的分块视角:

$$AB = A\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p\end{bmatrix} = \begin{bmatrix}A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p\end{bmatrix} \text{和} AB = \begin{bmatrix}\mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m\end{bmatrix} B = \begin{bmatrix}\mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B\end{bmatrix}$$

一般来说，我们可以将矩阵分块为若干个子矩阵，比如：

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots A_{1q} \\ A_{21} & A_{22} & \cdots A_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots A_{pq} \end{bmatrix}$$

其中每个 A_{ij} 是一个 $m_j \times n_j$ 的矩阵。

例 23.

$$A = \left[\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] = \begin{bmatrix} I & I & I \\ I & I & I \end{bmatrix}$$

如果对应的矩阵满足乘法的要求，那么分块矩阵的乘法也可以转换为对应块矩阵的乘法：

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$$

- 上述成立的要求在于 A_{ij} 的列数等于 B_{jk} 的行数。

一个例子：矩阵乘法的第四种视角

将 A 写成列向量的形式：

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

将 B 写成行向量的形式：

$$B = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}$$

则我们有：

$$AB = \mathbf{a}_1\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2 + \cdots + \mathbf{a}_n\mathbf{b}_n$$

另一个例子：消元的分块 (I)

回顾之前的消元矩阵，比如：

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

可以将其看成：

$$\left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline -2 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c|cc} 1 & x & x \\ \hline 2 & x & x \\ 5 & x & x \end{array}\right] = \begin{bmatrix} \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} & \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} x & x \end{bmatrix} + \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} x & x \\ x & x \end{bmatrix} \\ \begin{bmatrix} -2 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} & \begin{bmatrix} -2 \\ 0 \end{bmatrix} \begin{bmatrix} x & x \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x & x \\ x & x \end{bmatrix} \end{bmatrix}$$

一般来说，我们有：

$$\begin{bmatrix} I & O \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ O & D - CA^{-1}B \end{bmatrix}$$

补充说明

上述的 $D - CA^{-1}B$ 被称作矩阵 A 的舒尔补 (Schur complement)，其在图像处理、优化等领域有着重要的应用。

逆矩阵

回顾 $A\mathbf{x} = \mathbf{b}$

$$A\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b}$$

其解:

$$\begin{cases} x_1 = b_1 \\ x_2 = b_1 + b_2 \\ x_3 = b_1 + b_2 + b_3 \end{cases} \quad \text{即:} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = A^{-1}\mathbf{b}$$

A, A^{-1} 满足:

$$\begin{aligned} (AA^{-1})\mathbf{x} &= A(A^{-1}\mathbf{b}) = A\mathbf{x} = \mathbf{b} = I\mathbf{b} \\ (A^{-1}A)\mathbf{b} &= A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b} = \mathbf{x} = I\mathbf{x} \end{aligned}$$

定义 24 **[可逆矩阵 (Invertible Matrix)].**

称一个方阵 A(A 是一个 $n\times n$ 的矩阵)是可逆的 (invertbile)，如果存在一个矩阵 A^{-1}，使得：

$$AA^{-1}=A^{-1}A=I$$

例 25.

1. $A=\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}$ 是可逆的，其逆矩阵为 $A^{-1}=\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}$.
2. $C=\begin{bmatrix}1&0&-1\\-1&1&0\\0&-1&1\end{bmatrix}$ 是不可逆的。

对角矩阵 (Diagonal Matrix):

$$D = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}$$

是可逆的 $(d_1, \ldots, d_n \neq 0)$，其逆矩阵也是对角矩阵:

$$D^{-1} = \begin{bmatrix} \frac{1}{d_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{d_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{d_n} \end{bmatrix}$$

给定一个 2×2 的矩阵：

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

其是否是可逆的？是的话其逆矩阵是多少？

引理 26.

矩阵 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 是可逆的当且仅当 $ad - bc \neq 0$，此时其逆矩阵为：

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

消元矩阵 E_{ij} 的逆矩阵 (I)

我们再来看消元矩阵E_{ij}:

$$\begin{array}{cc} & \\ 第\ i\ 行 & \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 1 & \cdots & -k & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \\ & \text{第 } j \text{ 列} \end{array} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i - k\mathbf{a}_j \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

其逆矩阵是什么?

$E_{ij}A$ 得到将第 j 列的 $-k$ 倍加到第 i 列的结果，因此E_{ij}^{-1} 就是将第 i 列的 k 倍加到第 j 列的结果：

$$E_{ij}^{-1} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 1 & \cdots & k & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}$$

引理 27.

如果方阵 A 是可逆的，那么其逆矩阵是唯一的。

证明. 反设存在 B, C 使得 $BA = AC = I$，则注意到:

$$
\begin{aligned}
B(AC) &= (BA)C = IC = C \\
B(AC) &= BI = B
\end{aligned}
$$

从而 $B = C$. □

引理 28.

如果方阵 A 是可逆的，则对于任意的 $\mathbf{b}$，方程 $A\mathbf{x} = \mathbf{b}$ 有唯一解。

证明. 假设 A 是可逆的，令其逆矩阵为 A^{-1}，则：

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

□

推论 29.

如果 $A\mathbf{x} = \mathbf{0}$ 存在一个非零解，那么 A 不是可逆的。

引理 30.

如果 A 和 B 是可逆的，那么 AB 也是可逆的，且 $(AB)^{-1} = B^{-1}A^{-1}$.

证明. 注意到：

$$
\begin{aligned}
(AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} = AA^{-1} = I \\
(B^{-1}A^{-1})(AB) &= B^{-1}(A^{-1}A)B = B^{-1}B = I
\end{aligned}
$$

□

推论 31.

如果 $A_1, A_2, \ldots, A_n$ 都是可逆的，那么 $A_1A_2\cdots A_n$ 也是可逆的，且其逆矩阵为：

$$(A_1A_2\cdots A_n)^{-1} = A_n^{-1}A_{n-1}^{-1}\cdots A_2^{-1}A_1^{-1}$$

我们再来看个例子:

定义 32 [对角主导矩阵 (Diagonally Dominant Matrix)].

令 A 是一个 $n \times n$ 的矩阵:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

称 A 是对角主导的，如果对于每一个 $i \in [n]$，有:

$$|a_{ii}| \geqslant \sum_{j \neq i} |a_{ij}|$$

定理 33.

如果 A 是对角主导的，那么 A 是可逆的。

证明. 反设 A 是不可逆的，则:

$$A\mathbf{x} = \mathbf{0}$$

存在非零解 $(x_1, \ldots, x_n)$。令 $x_i \stackrel{\text{def}}{=} \max_{j\in[n]|x_j|>0}$，则有:

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ii}x_i + \cdots + a_{in}x_n = 0$$

但是:

$$|a_{ii}||x_i| = |\sum_{j\neq i} a_{ij}x_j| \leqslant \sum_{j\neq i} |a_{ij}||x_j| \leqslant \sum_{j\neq i} |a_{ij}||x_i| = |x_i|(\sum_{j\neq i} |a_{ij}|) < |a_{ii}||x_i|$$

$\square$

阶段总结

- 分块矩阵。
- 可逆矩阵。
  - 可逆矩阵的定义，性质。
  - 一些可逆矩阵的例子。

实际上，我们还有许多问题没有解决，比如：

- 如果 $AB = I$ 是否就能说明 A 是可逆的?
- 如何求 A^{-1}?
- $\dots$

我们将在之后的课程中逐一解决这些问题。

关于可逆矩阵 (II)

一个前瞻性的视角，事实上存在非常多的可逆矩阵刻画，比如:

1. A 的行列式不为 0。
2. A 的秩等于 n。
3. A 的列向量线性无关。
4. A 的列向量张成 $\mathbb{R}^n$。
5. A 的行向量线性无关。
6. A 的行向量张成 $\mathbb{R}^n$。
7. $A\mathbf{x} = \mathbf{0}$ 只有一个解 $\mathbf{x} = \mathbf{0}$。
8. $A\mathbf{x} = \mathbf{b}$ 有唯一解 $\mathbf{x} = A^{-1}\mathbf{b}$。
9. A 有 n 个首元。
10. A 的所有特征值非零。
11. $\cdots$

我们将在后续的课程一一刻画这些性质。

转置矩阵和置换矩阵

我们看到在矩阵中行和列扮演了不同的角色。我们定义一个新的矩阵：

定义 34 **[转置矩阵 (Tranpose Matrix)].**

给定一个 $m \times n$ 的矩阵 A，其转置矩阵 $A^\top$ 是一个 $n \times m$ 的矩阵，其满足：

$$(A^\top)(i,j) = A_(j,i)$$

例 35.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \text{的转置矩阵为} \quad A^\top = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

引理 36.

1. $(A^\top)^\top = A$
2. $(cA)^\top = cA^\top$
3. $(A+B)^\top = A^\top + B^\top$
4. $(AB)^\top = B^\top A^\top$
5. $(A^{-1})^\top = (A^\top)^{-1}$

理解 $(AB)^\top = B^\top A^\top$

$A\mathbf{x}$ 是 A 中列向量的线性组合，而 $\mathbf{x}^\top A^\top$ 则是 $A^\top$ 中行向量的线性组合；两者恰好是一致的。

定义 37 **[对称矩阵 (Symmetric Matrix)].**

$n \times n$ 的矩阵 S 是对称的，如果其满足：

$$S^{\top} = S$$

或者说对于任意的 $i, j \in [n]$，有 $S(i,j) = S(j,i)$.

例 38.

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$ 是对称矩阵。

引理 39.

1. 若 A 是一个可逆的对称矩阵，那么 A^{-1} 也是对称矩阵。
2. 对任何矩阵 A(不需要是方阵)，$A^\top A$ 和 $AA^\top$ 都是对称矩阵。

证明.

1. $(A^{-1})^\top = (A^\top)^{-1} = A^{-1}$.
2. $(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A$.

□

我们已经介绍了点乘（内积）· 的概念，其也可以由转置矩阵来表示：

引理 40.

- 令 $\mathbf{x}, \mathbf{y}$ 是两个 $n \times 1$ 的矩阵，则：

$$\mathbf{x}^\top \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

- 令 $\mathbf{x}$ 是 $n \times 1$ 的矩阵，$\mathbf{y}$ 是 $m \times 1$ 的矩阵 A 是一个 $m \times n$ 的矩阵，则：

$$A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^\top \mathbf{y}$$

说明

假设 $\mathbf{x}, \mathbf{y}$ 是 $n \times 1$ 的矩阵（列向量），则：

- $\mathbf{x}^\top \mathbf{y}$ 是一个值。
- $\mathbf{x}\mathbf{y}^\top$ 是一个矩阵。

我们再来看置换矩阵P_{ij}:

$$
\begin{array}{c c}
 & \\
 & \\
第\ i\ 行 & \\
 & \\
第\ j\ 行 & \\
 & \\
 &
\end{array}
\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
a_1 \\ \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \\ a_n
\end{bmatrix}
=
\begin{bmatrix}
a_1 \\ \vdots \\ a_j \\ \vdots \\ a_i \\ \vdots \\ a_n
\end{bmatrix}
$$

其将第 i 行和第 j 行进行了交换，如果我们任意交换行的顺序那?

定义 41.

置换矩阵 P 是将单位矩阵 I 的行重排列得到的矩阵。

例 42.

所有 3×3 的置换矩阵都可以由之前提到的 P_{ij} 得到:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P_{21} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P_{32}P_{21} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$P_{31} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad P_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad P_{21}P_{32} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

事实 43.

$n \times n$ 的置换矩阵一共有 $n!$ 个。

引理 44.
$P^{-1} = P^{\top}$.

证明. 首先可以发现对于任意的矩阵 P_{ij}，有：$P_{ij}^{-1} = P_{ij} = P_{ij}^{\top}$。注意到对于任意的置换矩阵 P，其可以被表述为：

$$P = P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k}$$

从而：

$$P^{\top} = (P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k})^{\top} = P_{i_kj_k}^{\top}P_{i_{k-1}j_{k-1}}^{\top}\cdots P_{i_1j_1}^{\top} = P_{i_kj_k}P_{i_{k-1}j_{k-1}}\cdots P_{i_1j_1}$$
$$P^{-1} = (P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k})^{-1} = P_{i_kj_k}^{-1}P_{i_{k-1}j_{k-1}}^{-1}\cdots P_{i_1j_1}^{-1} = P_{i_kj_k}P_{i_{k-1}j_{k-1}}\cdots P_{i_1j_1}$$

因此：

$$P^{\top} = P^{-1}$$

□

阶段总结

- 转置矩阵和对称矩阵。
- 置换矩阵及其逆矩阵。

向量空间 (Vector Space)

向量空间 (Vector Space)

向量空间的基本概念

定义 45.

空间 $\mathbb{R}^n$ 包含了所有如下的 n 维列向量 $\mathbf{v}$:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

其中对于任意的 $i \in [n]$，$v_i \in \mathbb{R}$，这里的 $\mathbb{R}$ 是实数集。

定义 46.

空间 $\mathbb{C}^n$ 包含了所有如下的 n 维列向量 $\mathbf{v}$:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

其中对于任意的 $i \in [n]$，$v_i \in \mathbb{C}$，这里的 $\mathbb{C}$ 是复数集。

向量空间的形式化定义 (I)

一个向量空间 $\mathbb{V}$ 是一个非空集合，其中的元素称之为向量，并且其满足以下两种运算：

- 向量加法：对于任意的 $\mathbf{u}, \mathbf{v} \in \mathbb{V}$，$\mathbf{u} + \mathbf{v} \in \mathbb{V}$。
- 数与向量的乘法（数乘）：对于任意的 $\mathbf{u} \in \mathbb{V}$ 和任意的实数 $c \in \mathbb{R}$，$c\mathbf{u} \in \mathbb{V}$。

向量空间的形式化定义 (II)

其中的加法满足如下的性质:

1. 加法满足交换律:
$$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$$
2. 加法满足结合律:
$$\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$$
3. 加法存在一个零元素（唯一的）$\mathbf{0}$，其满足 $\mathbf{u}+\mathbf{0}=\mathbf{u}$ 对任意的 $\mathbf{u}\in\mathbb{V}$。
4. 加法存在一个负元素（逆元），即对于任意的 $\mathbf{u}\in\mathbb{V}$，存在一个 $\mathbf{v}\in\mathbb{V}$，使得 $\mathbf{u}+\mathbf{v}=\mathbf{0}$，特别的，将 $\mathbf{v}$ 记为 $-\mathbf{u}$。

向量空间的形式化定义 (III)

其中的数乘满足如下的性质:

5. 数乘存在单位元 $\mathbf{1}$，使得 $\mathbf{1u} = \mathbf{u}$ 对于任意的 $\mathbf{u} \in \mathbb{V}$。
6. 数乘满足结合律:
$$c_1(c_2\mathbf{u}) = (c_1c_2)\mathbf{u}$$
7. 数乘是线性的，即对于任意的 $c \in \mathbb{R}$ 和 $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ 均有:
$$c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$$
8. 数乘对于加法满足分配律，即对于任意的 $c_1, c_2 \in \mathbb{R}$ 和 $\mathbf{u} \in \mathbb{V}$ 均有:
$$(c_1 + c_2)\mathbf{u} = c_1\mathbf{u} + c_2\mathbf{u}$$

例子-矩阵组成的向量空间 (I)

对于 $m, n \geqslant 1$，令$\mathbb{M}_{m\times n}(\mathbb{R})$表示所有的 $m \times n$ 的实数矩阵的集合:

- 其中的加法就定义成矩阵的加法。
- 其中的数乘就定义成矩阵的数乘。

可以验证，$\mathbb{M}_{m\times n}(\mathbb{R})$是一个向量空间。

例子-矩阵组成的向量空间 (II)

上海师范大学 Shanghai Normal University

- 矩阵的加法满足交换律和结合律。
- 其零元为全零矩阵 $\mathbf{0}_{m\times n}$, 即所有的入口都是 0。
- 对于任意的 $M \in \mathbb{M}_{m\times n}(\mathbb{R})$，其负元 $-M$ 为：$-M = (-1)M$.
- 数乘的单位元就是 $1 \in \mathbb{R}$.
- 数乘满足结合律和分配律。
- 数乘满足线性性质。

例子-只有一个向量的向量空间

有没有只有一个向量的向量空间呢?

- 有。

只有一个元素的向量空间

$$\mathbb{Z}_0 = \{\mathbf{0}\}$$

是一个向量空间。可以认为 $\mathbb{R}^0$ 是 $\mathbb{Z}_0$ 的一个特殊情况。

例子 $\mathbb{R}^2, \mathbb{R}^3, \cdots$

通过前面所叙述的向量加法和数乘，可以验证 $\mathbb{R}^n$ 是一个向量空间。

- 所有 $\begin{bmatrix} x \\ y \end{bmatrix}$ 组成的集合 $\mathbb{R}^2$。
- 所有 $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ 组成的集合 $\mathbb{R}^3$。
- $\cdots$

问题 47.

能否将 $\mathbb{R}^n$ 中推广到 $\mathbb{R}^\infty$ 中?

例子-$\mathbb{R}^\infty$(I)

假设我们遵循着 $\mathbb{R}^n$ 的例子推广，则 $\mathbb{R}^\infty$ 应该是这样的：

- $\mathbb{R}^\infty = \{(x_1, x_2, \cdots) \mid$ 对所有的 i, $x_i \in \mathbb{R}\}$。
- $c(x_1, x_2, \cdots) = (cx_1, cx_2, \cdots)$。
- $(x_1, x_2, \cdots) + (y_1, y_2, \cdots) = (x_1 + y_1, x_2 + y_2, \cdots)$。

但问题是：

$$(x_1, x_2, \cdots)$$

是什么？ 函数！

例子-$\mathbb{R}^\infty$(II)

定义集合$\mathbb{F}$:

$$\mathbb{F} = \mathbb{F}(\mathbb{N} \to \mathsf{R}) = \{f \mid f : \mathbb{N} \to \mathbb{R}\}$$

- 给定 $f_1, f_2 \in \mathbb{F}$，定义函数 $f_1 + f_2 : \mathbb{F} \to \mathbb{R}$ 为:

$$(f_1 + f_2)(x) = f_1(x) + f_2(x)$$

- 对于任意的 $f \in \mathbb{F}$ 和 $c \in \mathbb{R}$，定义函数 $cf : \mathbb{N} \to \mathbb{R}$ 为:

$$(cf)(x) = cf(x)$$

可以验证，$\mathbb{F}$是一个向量空间。

一个更奇怪的例子

我们再来看一个例子，考虑如下的集合:

$$\mathbb{V}_+ = \mathbb{R}^+ = \{x \mid x \in \mathbb{R} \text{ and } x > 0\}$$

- 对于任意的 $x, y \in \mathbb{V}_+$，定义加法运算$\oplus$: $x \oplus y$ 为 $x \oplus y = xy$。
- 对于任意的 $x \in \mathbb{V}_+$ 和 $c \in \mathbb{R}$，定义数乘运算$\otimes$: $c \otimes x$ 为 $c \otimes x = x^c$。

可以验证，$\mathbb{V}_+$ 是一个向量空间。

引理 48.
零向量 $\mathbf{0}$ 是唯一的。

证明. 反设存在两个零向量 $\mathbf{0}_1, \mathbf{0}_2$，则有:

$$\mathbf{0}_1 = \mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2 + \mathbf{0}_1 = \mathbf{0}_2$$

□

引理 49.
对于任何向量 $\mathbf{v}$，其负向量是唯一的。

证明. 反设存在两个负向量 $\mathbf{v}_1, \mathbf{v}_2$，则有:

$$\mathbf{v}_1 = \mathbf{v}_1 + \mathbf{0} = \mathbf{v}_1 + (\mathbf{v} + \mathbf{v}_2) = (\mathbf{v}_1 + \mathbf{v}) + \mathbf{v}_2 = \mathbf{0} + \mathbf{v}_2 = \mathbf{v}_2$$

□

引理 50 **[向量的消去律 (Cancellation Law)].**

如果 $\mathbf{u}+\mathbf{v}=\mathbf{u}+\mathbf{w}$，则 $\mathbf{v}=\mathbf{w}$。

引理 51.

1. $0\mathbf{v}=\mathbf{0}$。
2. $c\mathbf{0}=\mathbf{0}$。
3. $(-1)\mathbf{v}=-\mathbf{v}$。
4. $-(\mathbf{u}+\mathbf{v})=(-\mathbf{u})+(-\mathbf{v})$.
5. $c(-\mathbf{u})=(-c)\mathbf{u}=-(c\mathbf{u})$.

阶段总结

- 向量空间的概念。一些例子。
- 向量空间的性质。

接下来我们来关注向量空间的一类特殊子集。

向量空间 (Vector Space)

子空间

$\mathbb{R}^2$ 中的例子

让我们从一个简单的例子看起。考察 $\mathbb{R}^2$ 中的子集:

- $L_1 = \{(x, 0) \mid x \in \mathbb{R}\}$ 也是一个向量空间。
- $L_2 = \{(x, x+1) \mid x \in \mathbb{R}\}$ 不是一个向量空间。

显然并不是所有的子集都是向量空间。我们称这样的子集为子空间。

定义 52 **[子空间 (Subspace)].**

给定一个向量空间 $\mathbb{V}$，如果 $\mathbb{W}$ 是 $\mathbb{V}$ 的一个非空子集，并且 $\mathbb{W}$ 满足如下两个条件：

1. 对于任意的 $\mathbf{u}, \mathbf{v} \in \mathbb{W}$，$\mathbf{u} + \mathbf{v} \in \mathbb{W}$。
2. 对于任意的 $c \in \mathbb{R}$ 和 $\mathbf{u} \in \mathbb{W}$，$c\mathbf{u} \in \mathbb{W}$。

则称 $\mathbb{W}$ 是 $\mathbb{V}$ 的一个子空间。

定理 53.

如果 $\mathbb{W}$ 是向量空间 $\mathbb{V}$ 的一个子空间，则 $\mathbb{W}$ 对于 $\mathbb{V}$ 上定义的加法和数乘运算构成一个向量空间。

子空间的例子 (I)

上海师范大学 Shanghai Normal University

考察如下集合:

$$\mathbb{V} = \mathbb{R}^3$$

$$\mathbb{W} = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$$

则 $\mathbb{W}$ 是 $\mathbb{V}$ 的一个子空间，原因在于:

- 对于任意的 $\mathbf{u} = (x_1, y_1, 0), \mathbf{v} = (x_2, y_2, 0) \in \mathbb{W}$，$\mathbf{u} + \mathbf{v} = (x_1 + x_2, y_1 + y_2, 0) \in \mathbb{W}$。
- 对于任意的 $c \in \mathbb{R}$ 和 $\mathbf{u} = (x, y, 0) \in \mathbb{W}$，$c\mathbf{u} = (cx, cy, 0) \in \mathbb{W}$。

定义对角矩阵为:

定义 54 **[对角矩阵 (Diagonal Matrix)].**

令 $n \geqslant 1$, $A \in \mathbb{M}_{n\times n}(\mathbb{R})$, 如果对于任意的 $i \neq j$ 均有:

$$A(i,j) = 0$$

则称 A 是一个对角矩阵。

考虑如下集合:

$$\mathsf{D}_n(\mathbb{R}) = \{A \in \mathbb{M}_{n\times n}(\mathbb{R}) \mid A\text{是对角矩阵。}\}$$

则 $\mathsf{D}_n(\mathbb{R})$ 是 $\mathbb{M}_{n\times n}(\mathbb{R})$ 的一个子空间。

引理 55.

如果 $\mathbb{W}$ 是向量空间 $\mathbb{V}$ 的一个子空间，则 $\mathbf{0} \in \mathbb{W}$.

证明. 取 $\mathbf{w} \in \mathbb{W}$，$0\mathbf{w} = \mathbf{0} \in \mathbb{W}$. $\square$

引理 56.

令 $\mathbf{u}, \mathbf{v} \in \mathbb{W}$，则所有 $\mathbf{u}, \mathbf{v}$ 的线性组合 $c\mathbf{u} + d\mathbf{v}$ 均在 $\mathbb{W}$ 中。

证明. 由子空间的定义，对于任意的 $c, d \in \mathbb{R}$，均有：

$$c\mathbf{u}, d\mathbf{v} \in \mathbb{W}$$

从而：

$$c\mathbf{u} + d\mathbf{v} \in \mathbb{W}$$

□

引理 57.

令 $\mathbb{V}$ 是一个向量空间，$\mathbb{W}$ 是 $\mathbb{V}$ 的一个子集。则 $\mathbb{W}$ 是 $\mathbb{V}$ 的一个子空间当且仅当：对于任意的 $k \geqslant 0$，$c_1, \cdots, c_k \in \mathbb{R}$ 和 $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{W}$ 均有：

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \in \mathbb{W}$$

特别的，当 $k = 0$ 时我们令上述和为 $\mathbf{0}$.

下一步

让我们回到矩阵里看看矩阵里的向量空间。

定义 58 **[列空间 (Column Space)].**

给定一个 $m \times n$ 的矩阵 A，定义其列空间 $\mathbf{C}(A)$ 为：

$$\mathbf{C}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

即 $\mathbf{C}(A)$ 是所有由 A 的列向量线性组合而成的集合。

定理 59.

列空间 $\mathbf{C}(A)$ 是 $\mathbb{R}^m$ 的一个子空间。

引理 60.

$A\mathbf{x} = \mathbf{b}$ 有解当且仅当 $\mathbf{b} \in \mathbf{C}(A)$。

我们可以利用转置矩阵来定义 A 的行空间。

定义 61 **[行空间 (Row Space)].**

给定一个 $m \times n$ 的矩阵 A，定义其行空间为矩阵 $A^\top$ 的列空间 $\mathbf{C}(A^\top)$。

引理 62.

矩阵 A 的行空间 $\mathbf{C}(A^\top)$ 是 $\mathbb{R}^n$ 的一个子空间。

定义 63 **[零空间 (Null Space)].**

给定一个 $m \times n$ 的矩阵 A，定义其零空间 $\mathsf{N}(A)$ 为:

$$\mathsf{N}(A) = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$$

即 $\mathsf{N}(A)$ 是所有满足 $A\mathbf{x} = \mathbf{0}$ 的 $\mathbf{x}$ 的集合。

定理 64.

零空间 $\mathsf{N}(A)$ 是 $\mathbb{R}^n$ 的一个子空间。

生成一个子空间 (I)

上海师范大学 Shanghai Normal University

令 $S \subseteq \mathbb{V}$，显然我们知道:

- S 不一定是一个子空间。
- S 可能为空。

如何取构造一个包含 S 的子空间?

$$\mathrm{span}(S) = \{c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \mid k \geqslant 0,\ c_1, \cdots, c_k \in \mathbb{R},\ \mathbf{v}_1, \ldots, \mathbf{v}_k \in S\}$$

定理 65.

令 $S \subseteq \mathbb{V}$，则 $\mathrm{span}(S)$ 是 $\mathbb{V}$ 的包含 S 的最小子空间，即：

1. $\mathrm{span}(S)$ 是 $\mathbb{V}$ 的子空间。
2. 令 $\mathbb{W} \subseteq \mathbb{V}$ 是一个 $\mathbb{V}$ 的子空间，且 $S \subseteq \mathbb{W}$，则 $\mathrm{span}(S) \subseteq \mathbb{W}$。

阶段总结

上海师范大学 Shanghai Normal University

- 子空间的概念、例子以及性质。
- 矩阵的列空间和零空间。
- 生成一个子空间。

相关性、基和维度 (Independence, Basis and Dimension)

相关性、基和维度 (Independence, Basis and Dimension)

线性相关性

我们先来回顾以下线性组合的概念，固定一个向量空间 $\mathbb{V}$。

定义 66.

令 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{V}$. $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 的线性组合是一个具有如下形式的向量:

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

其中 $c_1, \ldots, c_n \in \mathbb{R}$。

说明

可以看到 $\mathrm{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_n\})$ 实际上就是 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 的所有线性组合的集合。

定义 67 **[线性无关 (Linearly Independent)].**

给定一个向量组 $\mathbf{v}_1, \ldots, \mathbf{v}_n$。如果对于任意的 $c_1, \ldots, c_n \in \mathbb{R}$，当且仅当 $c_1 = \cdots = c_n = 0$ 时有：

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

则称 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是线性无关的。

例 68.

- $\{(1,0,0),(0,1,0),(0,0,1)\}$ 是线性无关的。
- $\{(1,0,0),(1,1,0)\}$ 是线性无关的。
- $\{(1,0,0),(0,1,0),(0,0,1),(1,1,1)\}$ 不是线性无关的。

- 单个向量是线性无关的么?
- 包含 $\mathbf{0}$ 的向量组是线性无关的么?

引理 69.

线性无关的向量组的任何一个子集都是线性无关的。

我们知道，一个矩阵方程 $A\mathbf{x}$ 可以看成是其列向量 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 的线性组合，所以我们有：

引理 70.

给定一个矩阵 A，则其列向量是线性无关的当且仅当方程 $A\mathbf{x} = \mathbf{0}$ 只有唯一解 $\mathbf{x} = \mathbf{0}$。

我们也可以给出线性相关的定义：

定义 71 [线性相关 (Linearly Dependent)].

给定一个向量组 $\mathbf{v}_1, \ldots, \mathbf{v}_n$。如果存在 $c_1, \ldots, c_n \in \mathbb{R}$ 满足至少一个 $c_i \neq 0$，使得：

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

则称 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是线性相关的。

例 72.

- $\{(1,0,0),(0,1,0),(0,0,1),(1,1,1)\}$ 是线性相关的。
- $\{(2,2,0),(1,1,0)\}$ 是线性相关的。
- 包含 $\mathbf{0}$ 的向量组是线性相关的。

引理 73.

$\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是线性相关的当且仅当至少有一个 $\mathbf{v}_i$ 可以表示成其余向量的线性组合。

引理 74.

给定一个矩阵 A，则其列向量是线性相关的当且仅当方程 $A\mathbf{x} = \mathbf{0}$ 有非零解。

相关性、基和维度 (Independence, Basis and Dimension)

向量空间的基和维度

给定一个集合 S，回顾 span(S):

$$\text{span}(S)=\{c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k \mid k\geqslant 0,\ c_1,\cdots,c_k\in\mathbb{R},\ \mathbf{v}_1,\ldots,\mathbf{v}_k\in S\}$$

我们知道 span(S) 是一个向量空间，进一步的，我们称 span(S) 是由 S 生成的向量空间。

定义 75.

给定一个向量集合 S 和向量空间 $\mathbb{V}$，如果 $\mathbb{V}=\text{span}(S)$，则称 S 生成了向量空间 $\mathbb{V}$

定义 76 [向量空间的基 (A Bssis for a Vector Space)].

一组向量是一个向量空间的基 如果其满足:

1. 这组向量是线性无关的。
2. 这组向量生成了向量空间 $\mathbb{V}$。

例 77.

1. $\{(1,0,0),(0,1,0),(0,0,1)\}$ 是 $\mathbb{R}^3$ 的基。
2. $\{(1,0),(0,1)\}$ 是 $\mathbb{R}^2$ 的基。
3. $\{(1,0),(2,4)\}$ 也是 $\mathbb{R}^2$ 的基。

例 78.

1. $\mathbb{M}_{m\times n}(\mathbb{R})$ 的基是什么?
2. $\mathbb{Z}_0 = \{\mathbf{0}\}$ 的基是什么?
3. 考虑之前提到的向量空间:

$$\mathbb{V}_+ = \mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$$

其中的加法和数乘运算定义为:

$$x \oplus y = x \times y,\ c \otimes x = x^c$$

$\mathbb{V}$ 的基是什么?

我们现在来定义向量空间的维度，直观上来讲，向量空间的维度就是需要多少个向量才能生成这个向量空间。

定义 79 **[维度 (dimension)].**

给定一个向量空间 $\mathbb{V}$，其维度，记作 $\dim(\mathbb{V})$, 是指 $\mathbb{V}$ 的一个基中的向量个数。

问题 80.

上述定义会不会产生问题？

显然只有 $\mathbb{V}$ 中所有基的向量个数都相同时，上述定义才是合理的。

引理 81 **[Steinitz Exchange Lemma].**

令 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 是向量空间 $\mathbb{V}$ 的一个基，$\mathbf{v}_1, \ldots, \mathbf{v}_m$ 是 $\mathbb{V}$ 的一个线性无关的向量组，其中 $1 \leqslant m \leqslant n$。则存在 $1 \leqslant i_1 < i_2 < \cdots < i_{n-m} < n$，使得 $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{n-m}}$ 是 $\mathbb{V}$ 的一个基。

说明

当 $m = 0$ 时，上述引理是平凡的。

推论 82.

给定一个向量空间 $\mathbb{V}$ 和其上的两组基 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 和 $\mathbf{f}_1, \ldots, \mathbf{f}_m$。则 $n = m$。

推论 83.

假设 $\dim(\mathbb{V}) = n$，并且 $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{V}$ 是线性无关的，则：$m \leqslant n$.

推论 84.

假设 $\dim(\mathbb{V}) = n$，并且 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{V}$ 是线性无关的，则 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是 $\mathbb{V}$ 的一个基。

现在我们来证明 Steinitz 交换引理。

Steinitz 交换引理的证明. 我们对 m 使用归纳法。

$m = 1$ 的情况:

由于 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 是 $\mathbb{V}$ 的一个基，从而存在 $c_1, \ldots, c_n$ 使得:

$$\mathbf{v}_1 = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n$$

显然 $\mathbf{v}_1 \neq \mathbf{0}$，从而存在 $c_i \neq 0$，因此我们有:

$$\mathbf{e}_i = \frac{1}{c_i}\mathbf{v}_1 - \frac{c_1}{c_i}\mathbf{e}_1 - \cdots - \frac{c_{i-1}}{c_i}\mathbf{e}_{i-1} - \frac{c_{i+1}}{c_i}\mathbf{e}_{i+1} - \cdots - \frac{c_n}{c_i}\mathbf{e}_n$$

即: $\mathbf{v}_1, \mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \mathbf{e}_{i+1}, \ldots, \mathbf{e}_n$ 是 $\mathbb{V}$ 的一组基。

Steinitz 交换引理的证明 (续). 我们还需要证明:

$$\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{n-m}}$$

是线性无关的。考察如下的线性组合:

$$c_1\mathbf{e}_1 + \cdots + c_{i-1}\mathbf{e}_{i-1} + c\mathbf{v}_1 + \cdots + c_n\mathbf{e}_n = \mathbf{0}$$

- 如果 $c = 0$, 则由于 $\mathbf{e}_i$ 是线性无关的, 从而 $c_1 = \cdots = c_{i-1} = c_{i+1} = \cdots = c_n = 0$。
- 如果 $c \neq 0$, 由于 $\mathbf{v}_1 \neq \mathbf{0}$, 从而 $\mathbf{v}_1$ 可以由 $\mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \mathbf{e}_{i+1}, \ldots, \mathbf{e}_n$ 的线性组合表示, 从而 $\mathbf{e}_i$ 可以由 $\mathbf{e}_1, \ldots, \mathbf{e}_{i-1}, \mathbf{e}_{i+1}, \ldots, \mathbf{e}_n$ 的线性组合表示, 矛盾。

Steinitz 交换引理的证明 (续).

归纳步骤:

假设命题对于 $\leqslant m-1$ 的情况成立，对于 $=m$ 的情况，令 $\mathbf{v}_1,\dots,\mathbf{v}_m$ 是线性无关的，注意到 $\mathbf{v}_1,\dots,\mathbf{v}_{m-1}$ 也是线性无关的，从而由归纳假设，存在 $1\leqslant i_1<\cdots<i_{n-m+1}\leqslant n$，使得:

$$\mathbf{v}_1,\dots,\mathbf{v}_{m-1},\mathbf{e}_{i_1},\dots,\mathbf{e}_{i_{n-m+1}}$$

是 $\mathbb{V}$ 中的一组基。从而存在不全为 0 的 $c_1,\dots,c_{m-1},d_1,d_2,\dots,d_{n-m+1}$ 使得:

$$\mathbf{v}_m=c_1\mathbf{v}_1+\cdots+c_{m-1}\mathbf{v}_{m-1}+d_1\mathbf{e}_{i_1}+\cdots+d_{m-1}\mathbf{e}_{i_{n-m+1}}$$

Steinitz 交换引理的证明 (续).

注意到存在 $l \in [n-m+1]$ 使得 $d_l \neq 0$，从而：

$$\mathbf{e}_{i_l} = \frac{1}{d_l}\mathbf{v}_m - \frac{c_1}{d_l}\mathbf{v}_1 - \cdots - \frac{c_{m-1}}{d_l}\mathbf{v}_{m-1} - \frac{d_1}{d_l}\mathbf{e}_{i_1} - \cdots - \frac{d_{l-1}}{d_l}\mathbf{e}_{i_{l-1}} - \frac{d_{l+1}}{d_l}\mathbf{e}_{i_{l+1}} - \cdots - \frac{d_{n-m+1}}{d_l}\mathbf{e}_{i_{n-m+1}}$$

即：$\mathbf{e}_{i_l}$ 可以由 $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{l-1}}, \mathbf{e}_{i_{l+1}}, \ldots, \mathbf{e}_{i_{n-m+1}}$ 表示。

进一步可以验证：

$$\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{l-1}}, \mathbf{e}_{i_{l+1}}, \ldots, \mathbf{e}_{i_{n-m+1}}$$

是 $\mathbb{V}$ 的一组基，即归纳步骤成立，引理得证。 □

向量空间 $\mathbb{Z}_0$ 的维度 (I)

回顾向量空间 $\mathbb{Z}_0$:

$$\dim(\mathbb{Z}_0) = \{\mathbf{0}\}$$

定理 85.

$\dim(\mathbb{Z}_0) = 0.$

其中的关键在于:

$$\sum_{\mathbf{v} \in \emptyset} \mathbf{v} = \mathbf{0}$$

$\sum_{v\in\emptyset} v = 0$ 的原因

关键在于加法是可交换的，令 T 是一个有限的向量集，考察 T 的一个划分：

$$T = T_1 \cup T_2,\ \text{其中}\ T_1, T_2\ \text{满足：}\ T_1 \cap T_2 = \emptyset,\ T_1 \cup T_2 = T$$

则我们有：

$$\sum_{v\in T} v = \sum_{v\in T_1} v + \sum_{v\in T_2} v$$

显然有：

$$\sum_{v\in T} v = \sum_{v\in\emptyset} v + \sum_{v\in T} v$$

从而：

$$\sum_{v\in\emptyset} v = \sum_{v\in T} v - \sum_{v\in T} v = 0$$

无限维的向量空间 (I)

回顾之前函数构成的向量空间:

$$\mathbb{F} = \mathbb{F}(\mathbb{N} \to \mathbf{R}) = \{f \mid f : \mathbb{N} \to \mathbb{R}\}$$

其中:

- 给定 $f_1, f_2 \in \mathbb{F}$，定义函数 $f_1 + f_2 : \mathbb{F} \to \mathbb{R}$ 为:

$$(f_1 + f_2)(x) = f_1(x) + f_2(x)$$

- 对于任意的 $f \in \mathbb{F}$ 和 $c \in \mathbb{R}$，定义函数 $cf : \mathbb{N} \to \mathbb{R}$ 为:

$$(cf)(x) = cf(x)$$

$\mathbb{F}$ 的维度是多少?

定理 86.

$\mathbb{F}$ 是无限维的。

事实上，对于任意的 $i \in \mathbb{N}$，定义：

$$f_i : \mathbb{N} \to \mathbb{R},\ f_i(x) = \begin{cases} 1, & \text{如果 } x = i \\ 0, & \text{其他情况} \end{cases}$$

则对于任意的 $n \in \mathbb{N}$，我们有：

$$f_0, f_1, \ldots, f_n$$

是线性无关的。

有限维的向量空间

如果一个向量空间存在一个有限的基，则称这个向量空间是有限维的。

现在我们来看一下子空间的维度:

- 假设 $\mathbb{W}$ 是 $\mathbb{V}$ 的一个子空间，那么 $\mathbb{W}$ 的维度和 $\mathbb{V}$ 的维度有什么关系?

定理 87.

给定一个向量空间 $\mathbb{V}$ 和其子空间 $\mathbb{W}$，如果 $\mathbb{V}$ 是有限的，则 $\mathbb{W}$ 也是有限的，并且:

$$\dim(\mathbb{W}) \leqslant \dim(\mathbb{V})$$

令量空间 $\mathbb{V}$ 的维度 $\dim(\mathbb{V}) = n$，并且其一组基为：

$$\mathbf{w}_1, \ldots, \mathbf{w}_n$$

我们希望从 $\mathbb{V}$ 中慢慢的扩展出一组基

$$\mathbf{v}_1, \ldots, \mathbf{v}_k$$

使得：

$$\mathbb{W} = \mathrm{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$$

定理87的证明. 我们沿用上一页的记号，对 k 进行构造。初始化 $k=0$，如果:

$$\mathbb{W} \neq \mathrm{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}).$$

则存在 $\mathbf{v}_{k+1} \in \mathbb{W} \setminus \mathrm{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$ 使得:

$$\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}$$

是线性无关的。注意到 $\dim(\mathbb{V}) = n$，从而由 Steinitz 交换引理，必然有:

$$k+1 \leqslant n$$

从而 $\dim(\mathbb{W}) \leqslant \dim(\mathbb{V})$。 $\square$

阶段总结

- 线性相关和线性无关的概念。
- 向量空间的基，Steinitz 交换引理。
- 向量空间的维度, 子空间的维度。

矩阵的列秩与行秩 (Column Rank and Row Rank of Matrices)

矩阵的列秩与行秩 (Column Rank and Row Rank of Matrices)

列空间的秩与行空间的秩

令 A 是一个 $m \times n$ 的矩阵，其列空间定义为:

$$\mathbf{C}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

$\mathbf{C}(A)$ 显然是 $\mathbb{R}^m$ 的一个子空间，所以我们有:

定理 88.

$\mathbf{C}(A)$ 是 $\mathbb{R}^m$ 的一个子空间, 且

$$\dim(\mathbf{C}(A)) \leqslant \dim(\mathbb{R}^m) = m.$$

事实上，注意到 A 是一个 $m \times n$ 的矩阵，我们还有：

定理 89.

$$\dim(\mathbf{C}(A)) \leqslant n$$

直观理解

直观上理解，n 个 $\mathbb{R}^m$ 的列向量的线性组合组成的空间的维数不会超过 n。

$\dim(\mathbf{C}(A)) \leqslant n$ 的证明

证明. 令 A 的列向量为 $\mathbf{a}_1, \ldots, \mathbf{a}_n$，则可以在其中选择一个子序列:

$$\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}$$

满足:

- $i_1 < i_2 < \cdots < i_k$.
- $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}$ 是线性无关的。
- k 是最大的。

则我们有:

$$\mathbf{C}(A) = \operatorname{span}(\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\})$$

即:

$$\dim(\mathbf{C}(A)) = k \leqslant n$$

$\square$

定义 90 **[列秩 (Column Rank)].**

给定一个 $m \times n$ 的矩阵 A，其列秩 (Column Rank) 定义为：

$$\text{column—rank}(A) = \dim(\mathbf{C}(A))$$

例 91.

矩阵 $A = \begin{bmatrix} 1 & 2 & 3 & 6 \\ 4 & 5 & 9 & 18 \\ 7 & 8 & 15 & 30 \end{bmatrix}$ 的列秩为 2.

定理 92.

给定一个 $m \times n$ 的矩阵 A，则其列秩为 m 当且仅当存在一个 $n \times m$ 的矩阵 B 使得:

$$AB = I$$

推论 93.

如果 $n \times n$ 的矩阵 A 是可逆的，则其列秩为 n。

问题 94.

假设一个 $n \times n$ 的矩阵 A 的列秩为 n，则 A 是否一定是可逆的?

- 如果 A 的列秩为 n，则对于任意的 $\mathbf{b} \in \mathbb{R}^n$，方程

$$A\mathbf{x} = \mathbf{b}$$

都有唯一解。

- 所以上述问题问的是我们是否可以写成:

$$\mathbf{x} = A^{-1}\mathbf{b}$$

一个简单的回顾

目前为止，我们已经展示了，对于一个 $n \times n$ 的矩阵 A 来说：

1. 如果存在 B, C 使得 $AB = CA = I$，则 A 是可逆的，并且有 $A^{-1} = B = C$。
2. $\text{column-rank}(A) = n$ 当且仅当存在一个 $n \times n$ 的矩阵 B 使得 $AB = I$。

接下来的目标

$$\text{column-rank}(A) = n \text{ 当且仅当存在一个 } n \times n \text{ 的矩阵 } B \text{ 使得 } BA = I\text{。}$$

令 A 是一个 $m \times n$ 的矩阵，并且我们将其写成行向量的形式 $A = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$，回顾我们曾定义过的对于矩阵 A 的行空间 $\mathbf{C}(A^\top)$:

$$\mathbf{C}(A^\top) = \mathrm{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}) \Rightarrow \dim(\mathbf{C}(A^\top)) = \dim(\mathrm{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}))$$

定义 95 **[行秩 (Row Rank)].**

矩阵 A 的行秩 (Row Rank) 定义为:

$$\text{row—rank}(A) = \dim(\mathbf{C}(A^\top))$$

引理 96.

下面的叙述是等价的:

1. A 的行秩是 n.
2. $A^\top$ 的列秩是 n.
3. 存在一个矩阵 B，使得 $A^\top B = I$。
4. 存在一个矩阵 B，使得 $BA = I$

还欠缺什么?

定理 97.

给定一个 $m \times n$ 的矩阵，我们有 $\text{row–rank}(A) = \text{column–rank}(A)$.

证明. 我们只需证明对任意的矩阵 A，其有 row—rank$(A) \leqslant$ column—rank(A) 即可。

设 row—rank$(A) = r$，column—rank$(A) = c$。则存在 A 的 c 个列向量 $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_c}$ 是线性无关的，并且对于 A 中的每个列向量 $\mathbf{a}_j (j \leqslant n)$，存在 $p_{j1}, \ldots, p_{jc}$ 使得:

$$\mathbf{a}_j = p_{j1}\mathbf{a}_{i_1} + \cdots + \mathbf{p}_{jc}\mathbf{a}_{i_c}$$

从而我们有:

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{i_1} & \cdots & \mathbf{a}_{i_c} \end{bmatrix} \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{c1} & \cdots & p_{cn} \end{bmatrix}$$

记等式最右面的矩阵为 P,P 是一个 $c \times n$ 的矩阵，则有:

$$\text{row—rank}(A) \leqslant \text{row—rank}(P) \leqslant c = \text{column—rank}(A)$$

$\square$

阶段总结

- 矩阵的列秩和行秩。
- row—rank(A) $=$ column—rank(A) 的一个简单但不深刻的证明
- 方阵的秩与方阵的逆的关系。

矩阵的秩跟解方程有什么关系?

令 A 是一个 $m \times n$ 的矩阵，并且令其写为如下的形式：

$$A = \begin{bmatrix} \mathbf{a}_1, \ldots \mathbf{a}_n \end{bmatrix}$$

其中每个 $\mathbf{a}_i \in \mathbb{R}^m$.

引理 98.

- 对于任意的 $\mathbf{b} \in \mathbb{R}^m$：

$$A\mathbf{x} = \mathbf{b}\text{有解} \iff \mathbf{b} \in \mathbf{C}(A) = \operatorname{span}(\{\mathbf{a}_1, \ldots \mathbf{a}_n\})$$

  即方程 $A\mathbf{x} = \mathbf{b}$ 对任意 $\mathbf{b} \in \mathbb{R}^m$ 都有解当且仅当其列秩为 m，即 $\dim(\mathbf{C}(A)) = m$。
- 方程 $A\mathbf{x} = \mathbf{b}$ 对某些 $\mathbf{b} \in \mathbb{R}^m$ 有两个不同的解当且仅当 $\mathbf{a}_1, \ldots \mathbf{a}_n$ 是线性相关的。
- 方程 $A\mathbf{x} = \mathbf{b}$ 对任意 $\mathbf{b} \in \mathbb{R}^m$ 都有唯一解当且仅当：

$$\mathbf{a}_1, \ldots \mathbf{a}_n\text{是}\mathbb{R}^m\text{的一组基}$$

  即：$m = n$.

引理 99.

令 A 是一个 $n \times n$ 的矩阵, 下面的叙述是等价的:

1. A 是可逆的。
2. 方程 $Ax = b$ 对任意 $b \in \mathbb{R}^m$ 都有唯一解.
3. $\dim(\mathbf{C}(A)) = n$.
4. 存在矩阵 B 使得 $AB = I$。
5. 存在矩阵 C 使得 $CA = I$。

从而我们很自然的得到 A 的逆 $A^{-1} = B = C$ 满足:

$$AA^{-1} = A^{-1}A = I$$

怎么求 A^{-1}?

我们将使用 Gauss–Jordan 消元法，也就是 Gauss 消元法的一种拓展。

通过 Gauss–Jordan 消元法来求 A^{-1}

回顾-高斯消元法

上海师范大学 Shanghai Normal University

来考虑下面的方程组:

$$\begin{aligned} x - 2y &= 1 \\ 3x + 2y &= 11 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x - 2y &= 1 (\text{将第一个式子乘以 } 3) \\ 8y &= 8 (\text{减去上面的式子消去 } x) \end{aligned}$$

新的方程是上三角的 (upper triangular),类似的,有:

$$\begin{aligned} 4x - 8y &= 4 \\ 3x + 2y &= 11 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} {\color{red}4}x - 8y &= 4 (\text{将第一个式子乘以 } \tfrac{3}{4}) \\ {\color{red}8}y &= 8 (\text{减去上面的式子消去 } x) \end{aligned}$$

- 首元 (pivot):行中第一个做其他行消去的非零元素。在完成消去后,首元在对角线上。
- 倍数 (multiplier):用来消去其他行的倍数。

例 100.

$$\begin{aligned} x - 2y &= 1 \\ 3x - 6y &= 11 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x - 2y &= 1 \\ 0y &= 8 \end{aligned}$$

此时方程没有解。

例 101.

$$\begin{aligned} x - 2y &= 1 \\ 3x - 6y &= 3 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x - 2y &= 1 \\ 0y &= 0 \end{aligned}$$

此时方程有无数的解。

当 n 个方程没有 n 个首元的时候，我们就会遇到上面的情况。

例 102.

$$\begin{aligned} -2y &= 1 \\ 3x - 6y &= 11 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} 3x - 6y &= 1 \\ -2y &= 8 \end{aligned}$$

但我们只需要进行一次行变换就可以使得方程继续消元下去。

总结

- 前两个方程组并没有 2 个首元，我们称其是奇异的 (singular)。
- 第三个方程组有 2 个首元，我们称其是非奇异的 (nonsingular)

$$\begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ 2 & -2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 3 \end{bmatrix} \implies \begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 8 \end{bmatrix}$$

Ax=b　　　　Ux=c

1. 通过第一个方程上的第一个首元将所有后续方程对应位置上的系数全部变成 0。
2. 通过新的第二个方程上的第二个首元将所有后续方程对应位置上的系数全部变成 0。

3-n 重复以上操作找到 n 个首元，最终得到了一个上三角形的矩阵 U。

消元矩阵 $E_{ij}(-k)$

$$
\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 1 & \cdots & -k & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix}
\begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \\ a_n \end{bmatrix}
=
\begin{bmatrix} a_1 \\ \vdots \\ a_i - ka_j \\ \vdots \\ a_j \\ \vdots \\ a_n \end{bmatrix}
$$

引理 103.

$E_{ij}^{-1}(-k) = E_{ij}(k)$

当不关心 $-k$ 的具体值得时候，我们后面会省略 $-k$ 来表示这样一类矩阵。

回顾：消元的矩阵形式 (II)

Shanghai Normal University

置换矩阵 P_{ij}

$$\begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_j \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix}$$

引理 104.

$P_{ij}^{-1} = P_{ij}$

小插曲：$A = LU$(I)

假设矩阵 A 是 3×3 的，且消元过程中不需要进行行交换，则我们有：

$$E_{32}E_{31}E_{21}A = U$$

一般的，对于 $n \times n$ 的情形，且消元过程中不需要进行行交换，则我们有：

$$E_{n(n-1)}E_{n(n-2)}E_{(n-1)(n-2)}\cdots E_{31}E_{21}A = U$$

从而我们有：

$$A = E_{21}E_{31}\cdots E_{(n-1)(n-2)}E_{n(n-2)}E_{n(n-1)}U = LU$$

这里 L 是一个下三角形矩阵。

小插曲：$A = LU$(II)

关于 E_{ij} 的乘积还有一个有意思的发现：

$$E_{32}E_{31}E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -3 & 2 & 1 \end{bmatrix}$$

$$E_{21}E_{31}E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -3 & -2 & 1 \end{bmatrix}$$

如果需要进行行变换，A 和 U 的形式又是怎么样的？

- $PA = LU$.

上海师范大学 Shanghai Normal University

高斯消元法告诉我们:

$$\begin{aligned} x-2y-6z &= 1 \\ 2z &= 4 \\ 3x+2y &= 11 \end{aligned} \Longrightarrow \begin{aligned} x-2y-6z &= 1 \\ 2z &= 4 \\ 8y+18z &= 8 \end{aligned} \Longrightarrow \begin{aligned} x-2y-6z &= 1 \\ 8y+18z &= 8 \\ 2z &= 4 \end{aligned}$$

也就是:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{P_{23}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}}_{E_{31}} \underbrace{\begin{bmatrix} 1 & -2 & -6 & 1 \\ 0 & 0 & 2 & 4 \\ 3 & 2 & 8 & 11 \end{bmatrix}}_{\begin{bmatrix} A & b \end{bmatrix}} = \begin{bmatrix} 1 & -2 & -6 & 1 \\ 0 & 8 & 18 & 8 \\ 0 & 0 & 2 & 4 \end{bmatrix}$$

通过回代，我们就可以得到:

$$z=2, \qquad y=-3.5, \qquad x=6$$

$$\begin{aligned} x-2y-6z &= 1 \\ 2z &= 4 \\ 3x+2y &= 11 \end{aligned} \Longrightarrow \begin{aligned} x-2y-6z &= 1 \\ 2z &= 4 \\ 8y+18z &= 8 \end{aligned} \Longrightarrow \begin{aligned} x-2y-6z &= 1 \\ 8y+18z &= 8 \\ 2z &= 4 \end{aligned}$$

Jordan 继续上述的操作:

$$\Longrightarrow \begin{aligned} x-2y-6z &= 1 \\ 8y &= -28 \\ 2z &= 4 \end{aligned} \Longrightarrow \begin{aligned} x-2y &= 13 \\ 8y &= -28 \\ 2z &= 4 \end{aligned} \Longrightarrow \begin{aligned} x &= 6 \\ 8y &= -28 \\ 2z &= 4 \end{aligned}$$

高斯若尔当消元的矩阵形式

换言之，将
$$\begin{aligned} x-2y-6z &= 1\\ 8y+18z &= 8\\ 2z &= 4 \end{aligned}$$
进行变换的下述过程：

$$\Longrightarrow \begin{aligned} x-2y-6z &= 1\\ 8y &= -28\\ 2z &= 4 \end{aligned} \Longrightarrow \begin{aligned} x-2y &= 13\\ 8y &= -28\\ 2z &= 4 \end{aligned} \Longrightarrow \begin{aligned} x &= 6\\ 8y &= -28\\ 2z &= 4 \end{aligned}$$

也可以看成：

$$\underbrace{\begin{bmatrix} 1 & \frac{1}{4} & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}}_{E_{12}} \underbrace{\begin{bmatrix} 1 & 0 & 3\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}}_{E_{13}} \underbrace{\begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & -9\\ 0 & 0 & 1 \end{bmatrix}}_{E_{23}} P_{23}E_{31}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 6\\ 0 & 8 & 0 & -28\\ 0 & 0 & 2 & 4 \end{bmatrix}$$

高斯若尔当消元的最后一个步骤

我们最后还有一个步骤:

$$\begin{aligned} x &= 6 \\ 8y &= -28 \\ 2z &= 4 \end{aligned} \implies \begin{aligned} x &= 6 \\ y &= -3.5 \\ z &= 2 \end{aligned}$$

显然可以用一个对角矩阵来刻画:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}}_{D} E_{12}E_{13}E_{23}P_{23}E_{31}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -3.5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

高斯若尔当消元的总结

高斯若尔当消元法的过程:

$$\begin{aligned} x-2y-6z &= 1 \\ 2z &= 4 \\ 3x+2y &= 11 \end{aligned} \Longrightarrow \begin{aligned} x &= 6 \\ y &= -3.5 \\ z &= 2 \end{aligned}$$

也就是:

$$\begin{bmatrix} 1 & -2 & -6 \\ 0 & 0 & 2 \\ 3 & 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 11 \end{bmatrix} \Longrightarrow \cdots \Longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ -3.5 \\ 2 \end{bmatrix}$$

可以被表示为:

$$DE_{12}E_{13}E_{23}P_{23}E_{31}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -3.5 \\ 0 & 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} I & \mathbf{x} \end{bmatrix}$$

Shanghai Normal University 上海师范大学

$$DE_{12}E_{13}E_{23}P_{23}E_{31}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -3.5 \\ 0 & 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} I & \mathbf{x} \end{bmatrix}$$

中，一个关键的点在于

- $DE_{12}E_{13}E_{23}P_{23}E_{31}$ 的选择与 $\mathbf{b}$ 无关。

从而对于任意 $\mathbf{b} \in \mathbb{R}^n$ 来说，我们都可以通过如下的方式：

$$DE_{12}E_{13}E_{23}P_{23}E_{31}\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} I & \mathbf{x} \end{bmatrix}$$

求解方程 $A\mathbf{x} = \mathbf{b}$

重新考察矩阵 $A = \begin{bmatrix} 1 & -2 & -6 \\ 0 & 0 & 2 \\ 3 & 2 & 0 \end{bmatrix}$，如果其是可逆的，则我们可以写出：

$$A^{-1} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}$$

使得：

$$A^{-1}A = A\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 & A\mathbf{x}_3 \end{bmatrix} = I$$

也就是我们得到了三个方程组：

$$A\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{e}_1, \quad A\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \mathbf{e}_2, \quad A\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{e}_3,$$

通过 Gauss–Jordan 消元我们就可以解出上述方程。

用高斯若尔当来计算 A^{-1}(II)

对于每个方程组:

$$A\mathbf{x}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix} = \mathbf{e}_1, \quad A\mathbf{x}_1 = \begin{bmatrix}0\\1\\0\end{bmatrix} = \mathbf{e}_2, \quad A\mathbf{x}_1 = \begin{bmatrix}0\\0\\1\end{bmatrix} = \mathbf{e}_3,$$

我们都有:

$$\mathrm{D}\mathrm{E}_{12}\mathrm{E}_{13}\mathrm{E}_{23}\mathrm{P}_{23}\mathrm{E}_{31}\begin{bmatrix}A & \mathbf{e}_i\end{bmatrix} = \begin{bmatrix}\mathrm{I} & \mathbf{x}_i\end{bmatrix}$$

也就是:

$$\mathrm{D}\mathrm{E}_{12}\mathrm{E}_{13}\mathrm{E}_{23}\mathrm{P}_{23}\mathrm{E}_{31}\begin{bmatrix}A & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3\end{bmatrix} = \begin{bmatrix}\mathrm{I} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{bmatrix} \quad \Rightarrow \quad \mathrm{D}\mathrm{E}_{12}\mathrm{E}_{13}\mathrm{E}_{23}\mathrm{P}_{23}\mathrm{E}_{31}A = \mathrm{I}$$

从而:

$$A^{-1} = \mathrm{I}A^{-1} = \mathrm{D}\mathrm{E}_{12}\mathrm{E}_{13}\mathrm{E}_{23}\mathrm{P}_{23}\mathrm{E}_{31}AA^{-1} = \mathrm{D}\mathrm{E}_{12}\mathrm{E}_{13}\mathrm{E}_{23}\mathrm{P}_{23}\mathrm{E}_{31} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{bmatrix}$$

高斯若尔当消元法的一个总结

1. 考察一个矩阵：

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

2. 注意到对于任意 $\mathbf{b} \in \mathbb{R}^3$，方程 $A\mathbf{x} = \mathbf{b}$ 都有解，这是因为通过高斯消元法可以发现 A 有3 个首元。从而我们通过高斯-若尔当消元法可以解出：

$$DE_{12}E_{13}E_{23}P_{23}E_{31}\begin{bmatrix} A & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} = \begin{bmatrix} I & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}$$

   即存在 $B = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}$ 使得 $AB = I$.

3. 另一方面，如果令 $C = DE_{12}E_{13}E_{23}P_{23}E_{31}$，我们也有 $CA = I$.
4. 最终我们可以发现 A 是可逆的，并且：

$$A^{-1} = B = C = DE_{12}E_{13}E_{23}P_{23}E_{31}$$

上海师范大学 Shanghai Normal University

定理 105.

令 A 是一个 $n \times n$ 的矩阵，则 A^{-1} 存在当且仅当 A 有 n 个首元。

证明. 假设 A 有 n 个首元:

- 通过 Gauss–Jordan 消元法，我们可以解出所有形如 $A\mathbf{x} = \mathbf{e}_i$ 的方程，即:

$$D \cdots E \cdots P \cdots E \begin{bmatrix} A & \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} I & \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}$$

  也就是说，令 $B = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}$，我们有 $AB = I$.
- 另一方面，令 $C = D \cdots E \cdots P \cdots E$，我们有 $CA = I$.
- 从而 A 是可逆的，并且

$$A^{-1} = B = C = D \cdots E \cdots P \cdots E$$

A^{-1} 的存在性 (II)

另一方面，我们证明，如果存在一个 B 使得 $AB = I$，则 A 有 n 个首元。反设 A 没有 n 个首元，这意味着高斯消元法会得到某一行全 0 的情况，即：

$$E \cdots P \cdots EA = \begin{bmatrix} \vdots \\ \mathbf{0} \\ \vdots \end{bmatrix}$$

令 $M = E \cdots P \cdots E$，注意到 M 是可逆的，但是：

$$M = MI = MAB = \begin{bmatrix} \vdots \\ \mathbf{0} \\ \vdots \end{bmatrix} B$$

这意味着 M 存在一行全是 0，与 M 是可逆的矛盾。

引理 106.

令 A 是一个 $n \times n$ 的矩阵, 下面的叙述是等价的:

1. A 是可逆的。
2. 方程 $A\mathbf{x} = \mathbf{b}$ 对任意 $\mathbf{b} \in \mathbb{R}^m$ 都有唯一解.
3. $\dim(\mathbf{C}(A)) = n$.
4. $\dim(\mathbf{C}(A^T)) = n$.
5. 存在矩阵 B 使得 $AB = I$。
6. 存在矩阵 C 使得 $CA = I$。
7. A 有 n 个首元。

- 通过 Gauss–Jordan 消元法求 A^{-1}.

解线性方程组 (II)(Solving Linear Equations(II))

解线性方程组 (II)(Solving Linear Equations(II))

矩阵的秩

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned} \quad \Leftrightarrow \quad \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

对于其中每个 $i \in [m]$，我们定义：

$$j_i = \begin{cases} +\infty & \text{如果第 } i \text{ 行是零行} \\ \min\{j \in [n] : a_{ij} \neq 0\} & \text{o.w.} \end{cases}$$

即 a_{j_i} 是第 j 行中最左边的不为 0 的系数。

行阶梯形 (Row Echelon Form)

这个方程组，或者说对应的系数矩阵，是行阶梯形的，如果存在 $0 \leqslant r \leqslant m$ 使得：

$$j_1 < j_2 < \cdots < j_r, \quad j_{r+1} = \cdots = j_m = +\infty$$

这也意味着该方程组具有 r 个首元 $a_{1j_1}, \ldots, a_{rj_r}$。

行阶梯形的例子

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

$$\begin{aligned} 4x_1 - 13x_2 + 2x_4 &= 0 \\ 11x_2 + x_4 &= 6 \end{aligned} \quad \Leftrightarrow \quad \begin{bmatrix} 4 & -13 & 0 & 2 \\ 0 & 11 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

$$\begin{aligned} 4x_1 - 13x_2 + 2x_4 &= 0 \\ 11x_2 + x_4 &= 6 \\ 0 &= 7 \end{aligned} \quad \Leftrightarrow \quad \begin{bmatrix} 4 & -13 & 0 & 2 \\ 0 & 11 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 7 \end{bmatrix}$$

高斯消元法的精确描述 (I)

我们维护一个值 $r \leqslant m$ 使得:

R1 第一行到第 r 行已经满足行阶梯形的形式 (Row Echelon Form)，即:

$$j_1 < j_2 < \cdots < j_r \leqslant n$$

R2 对于矩阵中第 $r+1$ 行到第 m 行第 1 列到第 j_r 列是一个全零矩阵，即:

$$a_{ij} = 0, \quad \forall i \in [r+1, m], \forall j \in [1, j_r]$$

上海师范大学 Shanghai Normal University

第 1 步

1. 选择最小的 $j \in [n]$ 使得第 j 列有非零的元素，如果这样的 j 不存在，意味着 $A = O$ 并且当 $r = 0$ 时 R1，R2 已经满足，算法结束。
2. 选择最小的 $i \in [m]$ 使得第 $a_{ij} \neq 0$。
3. 将第 1 行与第 i 行进行交换，交换后我们得到在新的第一行中 $a_{1j} \neq 0$, 并且对于所有的 $t \in [1, j-1]$ 都有 $a_{1t} = 0$，此时我们有：$j_1 = j$.
4. 对每个 $i' > 1$，我们将第 i' 行替换成：

$$(\text{row } i') - \frac{a_{i'j}}{a_{1j}}(\text{row } 1)$$

替换之后我们有对于所有的 $i' > 1$，$a_{i'j} = 0$。

显然此时对于 $r = 1$，R1 和 R2 都是满足的。

第 $i+1$ 步

此时矩阵对于 $r=i$，R1 和 R2 都是满足的，我们继续进行下一步：

1. 选择最小的 $j \in [j_i+1, n]$ 使得存在 $i' \in [i+1, m]$ 使得 $a_{i'j} \neq 0$，如果这样的 j 不存在，算法结束。
2. 选择最小的 $i' \in [i+1, m]$ 使得第 $a_{i'j} \neq 0$。
3. 将第 $i+1$ 行与第 i' 行进行交换，交换后我们得到在新的第 $i+1$ 行中 $a_{(i+1)j} \neq 0$，并且对于所有的 $t \in [1, j-1]$ 都有 $a_{(i+1)t} = 0$，此时我们有：$j_{i+1} = j$.
4. 对每个 $i' > i+1$，我们将第 i' 行替换成：

$$(\text{row } i') - \frac{a_{i'j}}{a_{(i+1)j}} (\text{row } (i+1))$$

替换之后我们有对于所有的 $i' > i+1$，$a_{i'j} = 0$。

显然此时对于 $r = i+1$，R1 和 R2 都是满足的。

一个例子 (I)

考察这样一个矩阵:

$$A=\begin{bmatrix}0&0&6&4&3\\0&0&0&0&-3\\0&2&4&2&0\\0&-2&-1&0&2\end{bmatrix}$$

第一步的具体过程

$$\Rightarrow\begin{bmatrix}0&0&6&4&3\\0&0&0&0&-3\\0&2&4&2&0\\0&-2&-1&0&2\end{bmatrix}\Rightarrow\begin{bmatrix}0&2&4&2&0\\0&0&0&0&-3\\0&0&6&4&3\\0&-2&-1&0&2\end{bmatrix}\Rightarrow\begin{bmatrix}0&2&4&2&0\\0&0&0&0&-3\\0&0&6&4&3\\0&-2&-1&0&2\end{bmatrix}$$

$$\Rightarrow\begin{bmatrix}0&2&4&2&0\\0&0&0&0&-3\\0&0&6&4&3\\0&-2&-1&0&2\end{bmatrix}\Rightarrow\begin{bmatrix}0&2&4&2&0\\0&0&0&0&-3\\0&0&6&4&3\\0&0&3&2&2\end{bmatrix}\Rightarrow\begin{bmatrix}0&2&4&2&0\\0&0&0&0&-3\\0&0&6&4&3\\0&0&3&2&2\end{bmatrix}$$

一个例子 (II)

第二步的具体过程

$$\begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 3 & 2 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 3 & 2 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 3 & 2 & 2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 3 & 2 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

第三步的具体过程

$$\begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 6 & 4 & 3 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

此时我们有:

$$j_1 = 2, \quad j_2 = 3, \quad j_3 = 5$$

引理 107.

给定一个 $m \times n$ 的矩阵 A，高斯消元法将 A 变成一个有 r 个首元的行阶梯形矩阵 U。

我们利用首元的个数来定义矩阵的秩。

定义 108 [矩阵的秩 (Rank)].

给定一个 $m \times n$ 的矩阵 A，其秩 (rank) 定义为：

$$\text{rank}(A) = \text{矩阵 } A \text{ 的首元的个数} = r$$

定理 109.

$$\text{rank}(A) = \text{row—rank}(A) = \text{column—rank}(A)$$

上海师范大学 Shanghai Normal University

为了证明定理109，我们考虑行最简形矩阵。

行最简形 (Reduced Row Echelon Form)

回顾一个 $m \times n$ 矩阵 A，定义：

$$j_i = \begin{cases} +\infty & \text{如果第 } i \text{ 行是零行} \\ \min\{j \in [n] : a_{ij} \neq 0\} & \text{o.w.} \end{cases}$$

则称 A 是行最简形的 (Reduced Row Echelon Form)，如果：

1. 其是行阶梯形的，即存在 $0 \leqslant r \leqslant m$ 使得：

$$j_1 < j_2 < \cdots < j_r, \quad j_{r+1} = \cdots = j_m = +\infty$$

2. $a_{1j_1} = \cdots = a_{rj_r} = 1$，即所有的首元都是 1。
3. 对于所有的 $l \in [1, r], i \in [1, l-1] \cup [l+1, m]$，我们都有 $a_{ij_l} = 0$，即在首元的那一列中，除了首元之外的所有元素都是 0。

行最简形的例子

上海师范大学 Shanghai Normal University

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 0 & 3 & 4 & 0 \\ 0 & 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

定理 110.

高斯若尔当消元法会将矩阵 A 变成一个行最简形矩阵 R。

定理 111.

我们有:

1. $\text{rank}(A) = \text{rank}(R)$.
2. $\text{rank}(R) = \text{column–rank}(R) = \text{row–rank}(R)$.

行变换

我们知道高斯若尔当消元法是通过一系列的行变换来实现的，其中一共有三种行变换:

1. 行加法（Row Addition）。
2. 行交换（Row Exchange）。
3. 行乘法（Row Multiplication）。

我们称其为**初等行变换 (Row Elementary Operations)**。

回顾：行加法

Shanghai Normal University 上海师范大学

$$E_{ij}(-k)A = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 1 & \cdots & -k & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i - k\mathbf{a}_j \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

引理 112.

$E_{ij}^{-1}(-k) = E_{ij}(k)$

$$P_{ij}A = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

引理 113.

$P_{ij}^{-1} = P_{ij}$

$$D_i(k)A = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & k & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ k\mathbf{a}_i \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

引理 114.

$D_i(k)^{-1} = D_i(\frac{1}{k})$

定理 115.

初等行变换不改变矩阵的行秩和列秩。

证明. 我们先来证明行变换不改变矩阵的行秩。事实上，令 $A = \begin{bmatrix} \mathbf{a}_1^T & \mathbf{a}_i^T & \cdots & \mathbf{a}_m^T \end{bmatrix}^T$，即令其写成行向量的形式，这里 $\mathbf{a}_i$ 是 $1 \times n$ 的行向量。则经过一次初等行变换后，矩阵会变成如下的形式：

$$A' = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_i' \\ \vdots \\ \mathbf{a}_m' \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ k\mathbf{a}_i \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \text{或者} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i - k\mathbf{a}_j \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \text{或者} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$$

行变换的性质 (II)

$$A' = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_i' \\ \vdots \\ \mathbf{a}_m' \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ k\mathbf{a}_i \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \text{或者} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i - k\mathbf{a}_j \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \text{或者} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_i \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$$

我们有:

$$\begin{aligned} \mathbf{C}(A^\top) &= \text{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_m\}) \\ &= \text{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_m\}) \\ &= \text{span}(\{\mathbf{a}_1, \ldots, k\mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_m\}) \\ &= \text{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_i - k\mathbf{a}_j, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_m\}) \end{aligned}$$

即:

$$\text{row—rank}(A) = \dim(\mathbf{C}(A^\top)) = \dim(\mathbf{C}(A'^\top)) = \text{row—rank}(A')$$

我们现在来考虑列秩。

- 事实上，由于我们已经证明列秩和行秩是相等的，所以我们已经可以得到列秩不改变的结论。
- 但我们依旧想直接证明一下。我们只在这里考虑列交换的情况，其他的情况大家可以自行练习。

我们将 A 写成列向量的形式：

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_i & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n \end{bmatrix}$$

行变换的性质 (IV)

交换第 i 行和第 j 行后:

$$A' = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1' & \cdots & \mathbf{a}_i' & \cdots & \mathbf{a}_j' & \cdots & \mathbf{a}_n' \end{bmatrix}$$

- 我们比较一下 $\mathbf{a}_i$ 和 $\mathbf{a}_j$:

$$\mathbf{a}_i = \begin{bmatrix} a_{i1} & \ldots & a_{ik} & \ldots & a_{jk} & \ldots & a_{in} \end{bmatrix}^\top$$
$$\mathbf{a}_i' = \begin{bmatrix} a_{i1} & \ldots & a_{jk} & \ldots & a_{ik} & \ldots & a_{in} \end{bmatrix}^\top$$

行变换的性质 (V)

我们现在需要证明:

$$\dim(\text{span}(\{\mathbf{a}_1,\ldots,\mathbf{a}_n\})) = \dim(\text{span}(\{\mathbf{a}'_1,\ldots,\mathbf{a}'_n\}))$$

我们可以发现，对于任意的 $t \in [n]$, $1 \leqslant k_1 \leqslant k_2 \leqslant \cdots \leqslant k_t \leqslant n$ 和任意的 $c_1,\ldots,c_t \in \mathbb{R}$, 我们有:

$$\sum_{l\in[t]} c_l \mathbf{a}_{k_l} = \mathbf{0} \quad \Leftrightarrow \quad \sum_{l\in[t]} c_l \mathbf{a}'_{k_l} = \mathbf{0}$$

也就是说:

- $\mathbf{a}_{k_1},\ldots\mathbf{a}_{k_t}$ 是线性相关的当且仅当 $\mathbf{a}'_{k_1},\ldots\mathbf{a}'_{k_t}$ 是线性相关的。
- $\mathbf{a}_{k_1},\ldots\mathbf{a}_{k_t}$ 是 $\text{span}(\mathbf{a}_1,\ldots,\mathbf{a}_n)$ 的一组基当且仅当 $\mathbf{a}'_{k_1},\ldots\mathbf{a}'_{k_t}$ 是 $\text{span}(\mathbf{a}'_1,\ldots,\mathbf{a}'_n)$ 的一组基。

从而:

$$\dim(\mathbf{C}(A)) = \dim(\text{span}(\{\mathbf{a}_1,\ldots,\mathbf{a}_n\})) = \dim(\text{span}(\{\mathbf{a}'_1,\ldots,\mathbf{a}'_n\})) = \dim(\mathbf{C}(A'))$$

$\square$

定理109的证明

让我们回到定理109。

定理 109.

$$\text{rank}(A) = \text{row}-\text{rank}(A) = \text{column}-\text{rank}(A)$$

证明.

1. 我们将 A 变成行阶梯形矩阵 U，并且假设其有 r 个首元，则 $\text{rank}(A) = r$。
2. 进一步我们将 U 变换成行最简形，得到 R，其还是有 r 个首元。
3. $\text{rank}(R) = \text{column}-\text{rank}(R) = \text{row}-\text{rank}(R) = r = \text{rank}(A)$.
4. 注意到我们只使用了初等行变换，从而：

$$\text{column}-\text{rank}(A) = \text{column}-\text{rank}(R), \quad \text{row}-\text{rank}(A) = \text{row}-\text{rank}(R)$$

即：

$$\text{rank}(A) = \text{row}-\text{rank}(A) = \text{column}-\text{rank}(A)$$

- 矩阵的秩 rank(A) = 矩阵的首元个数。
- rank(A) = row−rank(A) = column−rank(A).

引理 116.

令 A 是一个 $n \times n$ 的矩阵, 下面的叙述是等价的:

1. A 是可逆的。
2. 方程 $A\mathbf{x} = \mathbf{b}$ 对任意 $\mathbf{b} \in \mathbb{R}^m$ 都有唯一解.
3. rank$(A) = n$.
4. column−rank$(A) = n$.
5. row−rank$(A) = n$.
6. 存在矩阵 B 使得 $AB = I$。
7. 存在矩阵 C 使得 $CA = I$。

当 A 是可逆矩阵的时候，我们已经足够清楚 $A\mathbf{x} = \mathbf{b}$ 的解了。

问题 117.

那对于任意的 A，比如 A 不是可逆的，或者说 A 不是方阵的情况那?

解线性方程组 (II)(Solving Linear Equations(II))

$A\mathbf{x} = \mathbf{0}$ 的解

回顾：矩阵的零空间

定义 118 **[零空间 (Null Space)].**

给定一个 $m \times n$ 的矩阵 A，定义其零空间 $\mathsf{N}(A)$ 为：

$$\mathsf{N}(A) = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$$

即 $\mathsf{N}(A)$ 是所有满足 $A\mathbf{x} = \mathbf{0}$ 的 $\mathbf{x}$ 的集合。

定理 119.

零空间 $\mathsf{N}(A)$ 是 $\mathbb{R}^n$ 的一个子空间。

引理 120.

$\mathbf{N}(A) \subseteq \mathbb{R}^n$，从而 $\dim(\mathbf{N}(A)) \leqslant n$.

引理 121.

$\dim(\mathbf{C}(A)) = n$ 当且仅当 $\dim(\mathbf{N}(A)) = 0$。

说明

事实上，这是线性代数基本定理的一个特殊情况：

$$\dim(\mathbf{C}(A)) + \dim(\mathbf{N}(A)) = n$$

证明. 令

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$$

则我们有:

$$\begin{aligned} \dim(\mathbf{C}(A)) = \dim(\mathrm{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_n\})) = n \\ &\Leftrightarrow \mathbf{a}_1, \ldots, \mathbf{a}_n \text{是 } \mathbf{C}(A) \text{ 的一组基} \\ &\Leftrightarrow \mathbf{a}_1, \ldots, \mathbf{a}_n \text{线性无关} \\ &\Leftrightarrow \mathbf{N}(A) = \mathbb{Z}_0 \\ &\Leftrightarrow \dim(\mathbf{N}(A)) = 0 \end{aligned}$$

$\square$

问题 122.

一般情况下怎么去计算 $\mathbf{N}(A)$ 的一组基。特别的 $\dim(\mathbf{N}(A))$?

一些例子 (I)

$$\begin{aligned} x + 2y &= 0 \\ 2x + 4y &= 0 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x + 2y &= 0 \\ 0 &= 0 \end{aligned}$$

我们称 y 是自由的 (free)。所以:

$$\mathbf{N}(A) = \{(-2y, y) \mid y \in \mathbb{R}\}$$

我们称下列的解是特解 (special solution):

$$\mathbf{s} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

- $\mathbf{s}$ 是将 y 设为 1 得到的解。
- $\mathbf{N}(A)$ 由所有 $\mathbf{s}$ 的线性组合构成，即:

$$\mathbf{N}(A) = \operatorname{span}(\{\mathbf{s}\})$$

$$x + 2y + 3z = 0, \quad \text{i.e.} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

在这个例子中 y 和 z 都是自由变量。所以我们可以选择两个特解:

$$\mathbf{s}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{s}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

同样我们有:

$$\mathsf{N}(A) = \mathrm{span}(\{\mathbf{s}_1, \mathbf{s}_2\})$$

求零空间的思路

1. 将矩阵 A 变成行阶梯形矩阵 U(Row echelon form)。
2. 寻找到 U 的特解。

例 123.

让我们从这几个例子再思考一下。

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} A \\ 2A \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 4 \\ 6 & 16 \end{bmatrix}, \quad C = \begin{bmatrix} A & 2A \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix}$$

矩阵 A 和矩阵 B

矩阵 A

$$\mathsf{A}\mathbf{x} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

从而我们有:

$$\mathsf{N}(A) = \mathbb{Z}_0 = \{\mathbf{0}\}$$

矩阵 B

$$\mathrm{B}\mathbf{x} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 4 \\ 6 & 16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 2 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \implies \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

从而我们有:

$$\mathsf{N}(B) = \mathbb{Z}_0 = \{\mathbf{0}\}$$

矩阵 C

$$\mathbf{C}\mathbf{x} = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \Longrightarrow \quad \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

这里存在两个自由变量:

$$\{x_3, x_4\}$$

我们从而可以选择两个特解:

$$\mathbf{s}_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{s}_2 = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

使得:

$$\mathbf{N}(\mathbf{C}) = \text{span}(\{\mathbf{s}_1, \mathbf{s}_2\})$$

首元列和自由列

$$C\mathbf{x} = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \Longrightarrow \quad \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

在其行阶梯形中:

$$\begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix}$$

- 我们称首元 (pivot) 所在的列为首元列 (pivot column)，对应的变量 x_1, x_2 称为主变量 (pivot variables)
- 剩余的列则称为自由列 (free columns)，对应的变量 x_3, x_4 称为自由变量 (free variables)

行最简形下的首元列视角 (I)

$$\begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 2 & 0 & 4 \end{bmatrix} \Rightarrow \begin{bmatrix} {\color{red}1} & {\color{red}0} & 2 & 0 \\ {\color{red}0} & {\color{red}1} & 0 & 2 \end{bmatrix}$$

在行最简形 (Reduced Row Echelon Form) 下，所有的主元列构成了一个 $r \times r$ 的单位矩阵，其中 $r = \text{rank}(A)$.

问题 124.

通过不同的初等行变换，我们是否可以得到不同的首元列和自由列?

行最简形下的首元列视角 (II)

$$A = \begin{bmatrix} \color{red}{p} & \color{red}{p} & f & \color{red}{p} & f \\ | & | & | & | & | \\ | & | & | & | & | \\ | & | & | & | & | \end{bmatrix} \quad \Longrightarrow \quad R = \begin{bmatrix} 1 & 0 & a & 0 & c \\ 0 & 1 & b & 0 & d \\ 0 & 0 & 0 & 1 & e \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

3 个首元列 p

2 个自由列 f

$\operatorname{rank} r = 3$

从而 $A\mathbf{x} = \mathbf{0}$ 等价于:

$$\begin{bmatrix} 1 & 0 & a & 0 & c \\ 0 & 1 & b & 0 & d \\ 0 & 0 & 0 & 1 & e \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \mathbf{0}, \quad \text{i.e.} \quad \begin{aligned} x_1 + ax_3 + cx_5 &= 0 \\ x_2 + bx_3 + dx_5 &= 0 \\ x_4 + ex_5 &= 0 \end{aligned}$$

主变量和自由变量

$$\begin{bmatrix} 1 & 0 & a & 0 & c \\ 0 & 1 & b & 0 & d \\ 0 & 0 & 0 & 1 & e \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \mathbf{0}, \quad \text{i.e.} \quad \begin{aligned} x_1 + ax_3 + cx_5 &= 0 \\ x_2 + bx_3 + dx_5 &= 0 \\ x_4 + ex_5 &= 0 \end{aligned}$$

我们有:

首元	自由变量
x_1, x_2, x_4	x_3, x_5

也即，x_3, x_5 是可以任意选择的，而 x_1, x_2, x_4 则由 x_3, x_5 决定:

$$\begin{aligned} x_1 &= -ax_3 - cx_5, \\ x_2 &= -bx_3 - dx_5, \\ x_4 &= -ex_5 \end{aligned}$$

特殊解

$$x_1 = -ax_3 - cx_5,$$

$$x_2 = -bx_3 - dx_5,$$

$$x_4 = -ex_5$$

通过选择自由解 $(x_3, x_5) \in \{(1,0),(0,1)\}$ 我们得到了两个特殊解:

$$\mathbf{s}_1 = \begin{bmatrix} -a \\ -b \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{s}_2 = \begin{bmatrix} -c \\ -d \\ 0 \\ -e \\ 1 \end{bmatrix}$$

并且所有 $A\mathbf{x} = \mathbf{0}$ 的解都可以表示成 $\mathbf{s}_1, \mathbf{s}_2$ 的线性组合（为什么？）, 即:

$$\mathbf{N}(A) = \mathbf{N}(R) = \text{span}(\{\mathbf{s}_1, \mathbf{s}_2\})$$

一般来说，令 A 是 $m \times n$ 的矩阵，我们考虑 $Ax = \mathbf{0}$ 的解。我们先使用 Gauss—Jordan 消元法将 A 转化成行最简形 R，即：

$$Ax = \mathbf{0} \quad \longleftrightarrow \quad \begin{bmatrix} 0 & \cdots & b_{1j_1} & \cdots & b_{1n} \\ 0 & \cdots & 0 & \cdots & b_{2n} \\ 0 & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{j1} \\ \vdots \\ x_n \end{bmatrix} = \mathbf{0}$$

其 $\text{rank}(A) = r$ 意味着存在 r 个首元：

$$b_{1j_1} = b_{2j_2} = \cdots = b_{rj_r} = 1$$

也就是

首元	自由变量
$x_{j_1}, x_{j_2}, \cdots, x_{j_r}$	$x_1, \cdots, x_{j_1-1}, x_{j_1+1}, \cdots, x_{j_2-1}, \cdots, x_{j_r+1}, \cdots, x_n$

$\mathbf{R}\mathbf{x} = \mathbf{0}$ 对应的方程组为:

$$\begin{aligned} x_{j_1} + b_{1,j_1+1}x_{j_1+1} + \cdots + b_{1,j_2-1}x_{j_2-1} + b_{1,j_2+1}x_{j_2+1} + \cdots + b_{1n}x_n &= 0 \\ x_{j_2} + b_{2,j_2+1}x_{j_2+1} + \cdots + b_{2n}x_n &= 0 \\ &\vdots \\ x_{j_r} + \cdots + b_{rn}x_n &= 0 \end{aligned}$$

从而我们可以构造出$n - r$ 个特殊解:

$$\mathbf{s}_1 = \begin{bmatrix} 1 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \mathbf{s}_{j_1-1} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \mathbf{s}_{j_1+1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -b_{1,j_1+1} \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots,$$

这意味着:

$$\mathbf{N}(A)=\mathbf{N}(R)=\operatorname{span}(\{\mathbf{s}_1,\cdots,\mathbf{s}_{j_1-1},\mathbf{s}_{j_1+1},\cdots,\mathbf{s}_{j_2-1},\cdots,\mathbf{s}_{j_r+1},\cdots,\mathbf{s}_n\})$$

定理 125.

对于任意的 $m\times n$ 的矩阵 A，我们有:

$$\operatorname{rank}(A)+\dim(\mathbf{N}(A))=n$$

定理 126 [Fundamental Theorem of Linear Algebra, Part I].

令 A 是一个 $m\times n$ 的矩阵并且 $\operatorname{rank}(A)=r$，则:

1. $\dim(\mathbf{C}(A))=\dim(\mathbf{C}(A^\top))=r$。
2. $\dim(\mathbf{N}(A))=n-r$，$\dim(\mathbf{N}(A^\top))=m-r$。

- 零空间的基的计算。
- 线性代数的基本定理，第一部分

所以我们目前对于 $A\mathbf{x}=\mathbf{0}$ 的解已经有了一个比较清晰的认识。

问题 127.

那对于 $A\mathbf{x}=\mathbf{b}$ 的解呢?

解线性方程组 (II)(Solving Linear Equations(II))

$A\mathbf{x} = \mathbf{b}$ 的解

三个问题

1. 什么时候 $A\mathbf{x} = \mathbf{b}$ 有解?
2. 如果 $A\mathbf{x} = \mathbf{b}$ 有解，其解的结构是什么?
3. 怎么计算 $A\mathbf{x} = \mathbf{b}$ 的解?

回顾

当 $\mathbf{b} = 0$ 的时候，我们已经有了一个比较清晰的认识。其解的结构就是 $\mathbf{N}(A)$，一个维度为 $n - \text{rank}(A)$ 的 $\mathbb{R}^n$ 的子空间。

定理 128.

$A\mathbf{x}=\mathbf{b}$ 有解当且仅当 $\mathbf{b}\in\mathbf{C}(A)$

定理 129.

$A\mathbf{x}=\mathbf{b}$ 有解当且仅当

$$\operatorname{rank}(A)=\operatorname{rank}(\begin{bmatrix}A & \mathbf{b}\end{bmatrix})$$

证明. 将 A 写成列向量的形式，即:

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$$

则我们有:

$$\begin{aligned}
& A\mathbf{x} = \mathbf{b} \text{ 有解} \\
& \qquad \iff \mathbf{b} \in \mathbf{C}(A) = \operatorname{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}) \\
& \qquad \iff \operatorname{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}) = \operatorname{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n, \mathbf{b}\}) \\
& \qquad \iff \dim(\operatorname{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\})) = \dim(\operatorname{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n, \mathbf{b}\})) \\
& \qquad \iff \text{column–rank}(A) = \text{column–rank}(\begin{bmatrix} A & \mathbf{b} \end{bmatrix}) \\
& \qquad \iff \operatorname{rank}(A) = \operatorname{rank}(\begin{bmatrix} A & \mathbf{b} \end{bmatrix})
\end{aligned}$$

$\square$

现在我们来考察 $A\mathbf{x}=\mathbf{b}$ 的解的结构，假设 $\mathbf{x}_p$ 是其一个解，即 $A\mathbf{x}_p=\mathbf{b}$，我们也称其为特解 (particular solution)。

定理 130.

$$A\mathbf{x}=\mathbf{b} \qquad \Longleftrightarrow \qquad \mathbf{x}-\mathbf{x}_p \in \mathbf{N}(A)$$

证明. 只需注意到：

$$A(\mathbf{x}-\mathbf{x}_p)=A\mathbf{x}-A\mathbf{x}_p=\mathbf{b}-\mathbf{b}=\mathbf{0}$$

□

$A\mathbf{x} = \mathbf{b}$ 的通解 (I)

事实上，令 $\mathbf{x}_p$ 为 $A\mathbf{x} = \mathbf{b}$ 的任一特解，令:

$$\mathbf{s}_1, \dots, \mathbf{s}_l$$

是 $A\mathbf{x} = \mathbf{0}$ 的一组特殊解，即其零空间 $\mathbf{N}(A)$ 的一组基，从而 $l = n - \mathrm{rank}(A)$。则任何一个 $A\mathbf{x} = \mathbf{b}$ 的解都可以表示为:

$$\mathbf{x} = \mathbf{x}_p + c_1\mathbf{s}_1 + \cdots + c_l\mathbf{s}_l$$

即一个特解 + 一个齐次解 ($A\mathbf{x} = \mathbf{0}$) 的形式。

注意到 A 的秩 r 应当满足:

$$r = \text{rank}(A) = \text{column--rank}(A) \leqslant \min\{m, n\}$$

从而我们有:

m	n	$\dim(\mathbf{N}(A))$	$Ax = b$ 的解的个数
$= r$	$= r$	0	1
$= r$	$> r$	$\geqslant 1$	∞
$> r$	$= r$	0	0 or 1
$> r$	$> r$	$\geqslant 1$	0 or ∞

怎么计算 $A\mathbf{x} = \mathbf{b}$ 的解

对其增广矩阵使用 Gauss–Jordan:

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$$

正交和投影 (Orthogonality and Projection)

正交和投影 (Orthogonality and Projection)

正交性

$Ax = 0$ 的解与行空间 A

我们来从几何的角度来看 $A\mathbf{x} = \mathbf{0}$ 的解。记矩阵 A 的形式如下：

$$A = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}$$

则每个 $\mathbf{a}_i$ 可以视作一个 $n \times 1$ 的矩阵，即：

$$\mathbf{a}_i = \begin{bmatrix} a_{i1} \\ \vdots \\ a_{in} \end{bmatrix}$$

则对于任意 $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^n$ 有：

$$A\mathbf{x} = \mathbf{0} \iff a_{i1}x_1 + \cdots + a_{in}x_n = 0 \text{ 对于任一 } i \in [m]$$

$\iff$ 对于任一 $i \in [m]$ $\mathbf{a}_i \cdot \mathbf{x} = 0$，即 $\mathbf{x}$ 与 $\mathbf{a}_i$ 都是垂直 (正交)的。

$$\iff \text{对于任一 } i \in [m] \ \mathbf{a}_i^\top \mathbf{x} = 0$$

定理 131.

给定一个矩阵 A，其行空间 $\mathbf{C}(A^{\mathsf{T}})$ 和零空间 $\mathbf{N}(A)$ 是正交的 (orthogonal)，即对于任意的 $\mathbf{u} \in \mathbf{C}(A^{\mathsf{T}})$ 和 $\mathbf{v} \in \mathbf{N}(A)$，我们都有：

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^{\mathsf{T}}\mathbf{v} = 0$$

特别的，其逆命题也是成立的，即如果存在 $\mathbf{v} \in \mathbb{R}^n$ 满足 $\mathbf{v}$ 与 $\mathbf{C}(A^{\mathsf{T}})$ 中的任何一个 $\mathbf{u}$ 都是垂直的，则：

$$A\mathbf{v} = \mathbf{0}, \text{ 即：} \mathbf{v} \in \mathbf{N}(A)$$

定理131的证明. 记 A 是之前的形式:

$$A = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}$$

则 $\mathbf{u} \in \mathbf{C}(A^\top)$ 等价于存在 $\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \in \mathbb{R}^m$ 使得:

$$\mathbf{u} = c_1\mathbf{a}_1 + \cdots + c_m\mathbf{a}_m = A^\top\mathbf{c}$$

从而对于任意 $\mathbf{v} \in \mathbf{N}(A)$ 有:

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^\top\mathbf{v} = (A^\top\mathbf{c})^\top\mathbf{c} = \mathbf{c}^\top A\mathbf{v} = \mathbf{c}^\top\mathbf{0} = 0$$

$\square$

定义 132 **[Orthogonal Subspaces].**

令 $n \geqslant 0$，$\mathbb{V}$ 和 $\mathbb{W}$ 是 $\mathbb{R}^n$ 的两个子空间，我们称 $\mathbb{V}$ 和 $\mathbb{W}$ 是正交的 (orthogonal)，记作：

$$\mathbb{V} \perp \mathbb{W}$$

如果每个 $\mathbb{V}$ 中的向量 $\mathbf{v}$ 和 W 中的任何一个向量 $\mathbf{w}$ 都是垂直的 (perpendicular)，即：

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^{\mathsf{T}}\mathbf{w} = 0$$

我们同样用 $\mathbf{v} \perp \mathbf{w}$ 来表示 $\mathbf{v} \cdot \mathbf{w} = 0$

例 133.

- $\{(x, 0) \mid x \in \mathbb{R}\}$ 和 $\{(0, y) \mid y \in \mathbb{R}\}$ 是正交的。
- 任何一个向量空间 $\mathbb{V}$ 和 $\mathsf{Z} = \{\mathbf{0}\}$ 都是正交的。
- $\{(x, 0, 0) \mid x, z \in \mathbb{R}\}$ 和 $\{(0, y, z) \mid y, z \in \mathbb{R}\}$ 是正交的。

基与正交的关系 (I)

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是 $\mathbb{R}^n$ 的两个子空间:

- $\mathbb{V}$ 的一组基为 $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$
- $\mathbb{W}$ 的一组基为 $\{\mathbf{w}_1, \cdots, \mathbf{w}_l\}$。

如果 $\mathbb{V}$ 和 $\mathbb{W}$ 是正交的，显然这两组向量是互相正交的，那么问题反过来呢?

定理 134.

$\mathbb{V} \perp \mathbb{W}$ 当且仅当对任意的 $i \in [k]$, $j \in [l]$ 我们有: $\mathbf{v}_i \perp \mathbf{w}_j$.

定理134的证明. 我们只需证明 $\Leftarrow$ 的方向，另一边直接由定义可得。

假设对于任意的 $\mathbf{v}_i$ 和 $\mathbf{w}_j$，我们有：$\mathbf{v}_i \perp \mathbf{w}_j$，则对于任意的 $\mathbf{v} \in \mathbb{V}$ 和 $\mathbf{w} \in W$，存在 $\mathbf{a} \in \mathbb{R}^k$ 和 $\mathbf{b} \in \mathbb{R}^l$ 满足:

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{bmatrix} \mathbf{a}, \quad \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_l \end{bmatrix} \mathbf{b}$$

从而:

$$\begin{aligned}
\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^\top \mathbf{w} = \mathbf{a}^\top \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_k^\top \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_l \end{bmatrix} \mathbf{b} \\
= \mathbf{a}^\top \begin{bmatrix} \mathbf{v}_1^\top \mathbf{w}_1 & \cdots & \mathbf{v}_1^\top \mathbf{w}_l \\ \vdots & \ddots & \vdots \\ \mathbf{v}_k^\top \mathbf{w}_1 & \cdots & \mathbf{v}_k^\top \mathbf{w}_l \end{bmatrix} \mathbf{b} \\
= \mathbf{a}^\top \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \mathbf{b} = 0
\end{aligned}$$

$\square$

我们利用定理134来给出 $\mathsf{C}(A^\top)\perp\mathsf{N}(A)$ 的另一个证明。

1. 记 $A^\top$ 的列向量为 $\mathbf{a}_1,\cdots,\mathbf{a}_m$，则可以从中选出 $\mathsf{C}(A^\top)$ 的一组基:

$$\{\mathbf{a}_{i_1},\cdots,\mathbf{a}_{i_r}\}$$

   其中 $r=\operatorname{rank}(A)$.

2. 类似的选出 $\mathsf{N}(A)$ 的一组基:

$$\{\mathbf{x}_1,\cdots,\mathbf{x}_{n-r}\}$$

3. 对任意的 $k\in[r]$ 和 $j\in[n-r]$ 我们有: $\mathbf{a}_{i_k}\perp\mathbf{x}_j$.

$\square$

直观理解

$\mathsf{C}(A)$ 和 $\mathsf{N}(A)$ 可以看成将 $\mathbb{R}^n$ 分解成了两个正交的子空间。

定义 135 **[Orthogonal Complements].**

令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的一个子空间，我们称 $\mathbb{V}$ 的正交补 (orthogonal complement) 为：

$$\mathbb{V}^{\perp}=\{\mathbf{v}\in\mathbb{R}^n \mid \mathbf{v}\perp\mathbf{u},\ \text{对于任意的 } \mathbf{u}\in\mathbb{V}\}$$

例 136.

- 考察 $\mathbb{R}^2$ 的子空间 $\{(c,0)\mid c\in\mathbb{R}\}$，其正交补为：$\{(0,c)\mid c\in\mathbb{R}\}$
- 考察 $\mathbb{R}^2$ 的子空间 $\{(c,2c)\mid c\in\mathbb{R}\}$，其正交补为：$\{(-2c,c)\mid c\in\mathbb{R}\}$
- 考察 $\mathbb{R}^3$ 的子空间 $\{(x,y,z)\mid x+y+z=0\}$，其正交补为：$\{(c,c,c)\mid c\in\mathbb{R}\}$

引理 137.

令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的一个子空间，则：

1. $\mathbb{V}^{\perp}$ 是一个子空间。
2. $\mathbb{V} \perp \mathbb{V}^{\perp}$。
3. 令 $\mathbb{W}$ 是 $\mathbb{R}^n$ 的子空间，如果 $\mathbb{W} \perp \mathbb{V}$，则 $\mathbb{W} \subseteq \mathbb{V}^{\perp}$，即 $\mathbb{V}^{\perp}$ 是最大的与 $\mathbb{V}$ 正交的子空间。
4. $(\mathbb{V}^{\perp})^{\perp} = \mathbb{V}$。

定理 138 [Fundamental Theorem of Linear Algebra, Part II].

令 A 是一个 $m \times n$ 的矩阵，则其零空间 $\mathsf{N}(A)$ 是行空间 $\mathsf{C}(A^\top)$ 的正交补，即：

$$\mathsf{N}(A) = (\mathsf{C}(A^\top))^\perp$$

我们再来从几何的角度理解一下矩阵 A。

我们已经介绍了矩阵 A 的四个空间:

1. $\mathbf{C}(A)$: A 的列空间，即所有的 $A\mathbf{x}$ 的集合。
2. $\mathbf{N}(A)$: A 的零空间，即 $A\mathbf{x}=\mathbf{0}$ 的解的集合。
3. $\mathbf{C}(A^{\mathsf{T}})$: A 的行空间，即所有的 $A^{\mathsf{T}}\mathbf{y}$ 的集合。
4. $\mathbf{N}(A^{\mathsf{T}})$: A^{T} 的零空间，即 $A^{\mathsf{T}}\mathbf{x}=\mathbf{0}$ 的解的集合。

我们同样引入 A^{T} 的零空间 $\mathbf{N}(A^{\mathsf{T}})$，其是 $A^{\mathsf{T}}\mathbf{y}=\mathbf{0}$ 的解的集合，即:

$$\mathbf{y}^{\mathsf{T}}A=\mathbf{0}$$

的解的集合，我们称其为 A 的左零空间 (Left Nullspace)。

我们再来回顾一下线性代数基本定理:

定理 138 [Fundamental Theorem of Linear Algebra, Part I].

令 A 是一个 $m \times n$ 的矩阵并且 $\text{rank}(A) = r$，则:

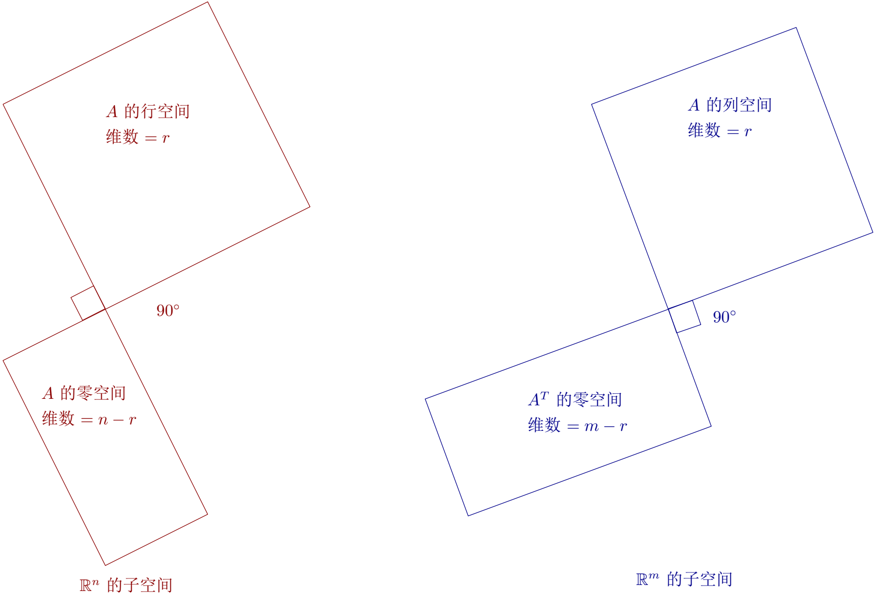
1. $\dim(\mathbf{C}(A)) = \dim(\mathbf{C}(A^\top)) = r$。
2. $\dim(\mathbf{N}(A)) = n - r$, $\dim(\mathbf{N}(A^\top)) = m - r$。

定理 138 [Fundamental Theorem of Linear Algebra, Part II].

令 A 是一个 $m \times n$ 的矩阵，则:

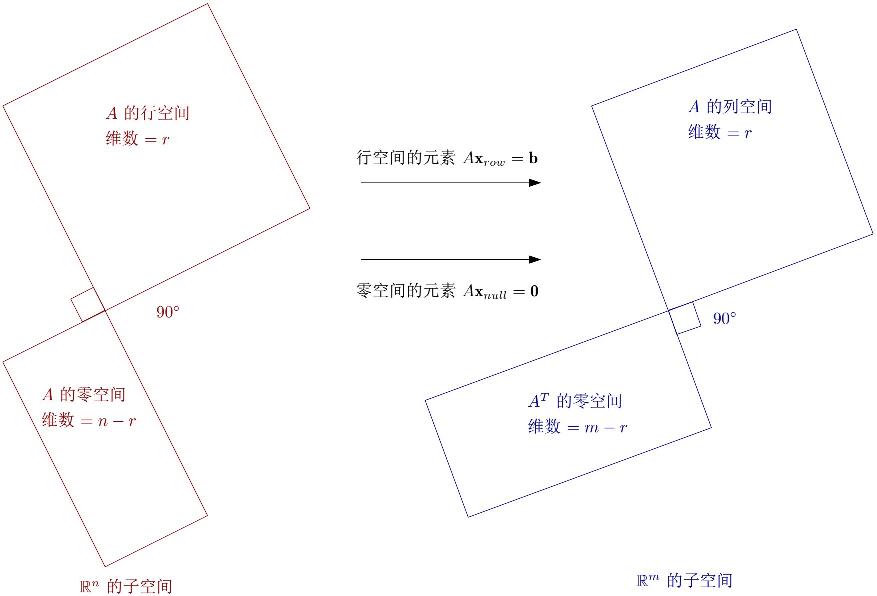
1. $\mathbf{N}(A) = (\mathbf{C}(A^\top))^\perp$
2. $\mathbf{N}(A^\top) = (\mathbf{C}(A))^\perp$

$m \times n$ 的矩阵 A 的四个空间



矩阵 A 的空间理解 (II)

$m \times n$ 的矩阵 A 的四个空间



我们考虑 $\mathbb{R}^2$ 的子集:

$$\mathbb{V} = \{(x, 0) \mid x \in \mathbb{R}\}$$

其正交补为:

$$\mathbb{V}^{\perp} = \{(0, y) \mid y \in \mathbb{R}\}$$

注意到: $\mathbb{R} \neq \mathbb{V} \cup \mathbb{V}^{\perp}$, 但每个 $(x, y) \in \mathbb{R}$ 都可以表示为:

$$(x, y) = (x, 0) + (0, y)$$

引理 139.

令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的一个子空间, 则对于任一 $\mathbf{x} \in \mathbb{R}^n$, 我们都存在唯一的 $\mathbf{v} \in \mathbb{V}$ 和 $\mathbf{v}^{\perp} \in \mathbb{V}^{\perp}$ 使得:

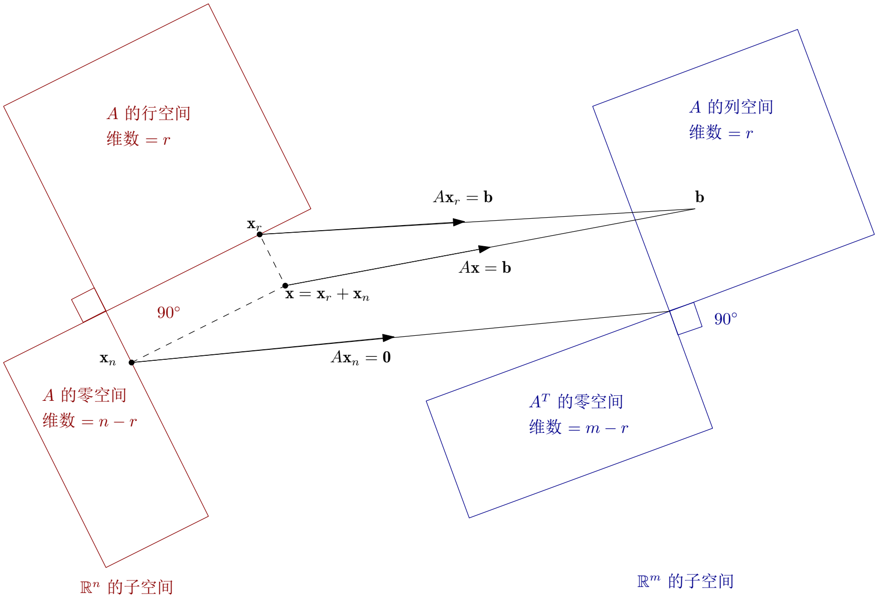
$$\mathbf{x} = \mathbf{v} + \mathbf{v}^{\perp}$$

换句话说,

$$\mathbb{R}^n = \mathbb{V} + \mathbb{V}^{\perp} = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in \mathbb{V} \text{ and } \mathbf{v} \in \mathbb{V}^{\perp}\}$$

矩阵 A 的空间理解 (III)

$m \times n$ 的矩阵 A 的四个空间



引理 140.

令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的一个子空间，则对于任一 $\mathbf{x} \in \mathbb{R}^n$，我们都存在唯一的 $\mathbf{v} \in \mathbb{V}$ 和 $\mathbf{v}^\perp \in \mathbb{V}^\perp$ 使得:

$$\mathbf{x} = \mathbf{v} + \mathbf{v}^\perp$$

换句话说，

$$\mathbb{R}^n = \mathbb{V} + \mathbb{V}^\perp = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in \mathbb{V} \text{ and } \mathbf{v} \in \mathbb{V}^\perp\}$$

说明

1. 我们需要一些额外的手段(投影，Projection)来证明上述结论，也就是我们接下来要讨论的内容。
2. 作为一个作业，你们被要求先来尝试证明其唯一性。

阶段总结

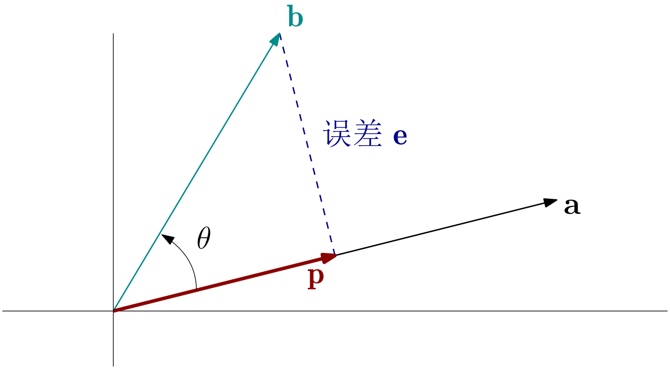
- 正交的概念。子空间正交。
- 正交补的概念。线性代数基本定理的第二部分。
- 矩阵的四个空间的几何直观。
- 正交补的性质，待证明的引理140。

投影

正交和投影 (Orthogonality and Projection)

投影到一条直线

假设一条线的方向是 a $= (a_1, \cdots, a_m)$。考虑任一个向量 b $= (b_1, \cdots, b_m)$，我们希望在这条直线上找到 p，使得 p 到 b 的距离最小。



寻找最小的 e

关键在于发现 b 和 p 的最小误差是与 a(p) 垂直的。我们称 p 是 b 在 a 上的投影。

投影的计算 (I)

假设:

$$\mathbf{p} = \hat{x}\mathbf{a}$$

则 $\mathbf{e} = \mathbf{b} - \mathbf{p}$，注意到 $\mathbf{e} \perp \mathbf{a}$，则我们有:

$$0 = \mathbf{a} \cdot \mathbf{e} = \mathbf{a}^{\mathsf{T}}(\mathbf{b} - \mathbf{p}) = \mathbf{a}^{\mathsf{T}}(\mathbf{b} - \hat{x}\mathbf{a}) = \mathbf{a}^{\mathsf{T}}\mathbf{b} - \hat{x}\mathbf{a}^{\mathsf{T}}\mathbf{a}$$

从而我们有:

$$\hat{x} = \frac{\mathbf{a}^{\mathsf{T}}\mathbf{b}}{\mathbf{a}^{\mathsf{T}}\mathbf{a}}$$

即我们所需要求的投影 $\mathbf{p}$ 为:

$$\mathbf{p} = \frac{\mathbf{a}^{\mathsf{T}}\mathbf{b}}{\mathbf{a}^{\mathsf{T}}\mathbf{a}}\mathbf{a}$$

另一个算法

注意到: $\mathbf{p} = \frac{\|\mathbf{p}\|}{\|\mathbf{a}\|}\mathbf{a}$，$\|\mathbf{p}\| = \|\mathbf{b}\|\cos\theta$ 以及 $\cos\theta = \frac{\mathbf{a}^{\mathsf{T}}\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$，我们有:

$$\hat{x} = \frac{\|\mathbf{b}\|\cos\theta}{\|\mathbf{a}\|} = \frac{\|\mathbf{b}\|}{\|\mathbf{a}\|}\frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|} = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|^2} = \frac{\mathbf{a}^{\mathsf{T}}\mathbf{b}}{\mathbf{a}^{\mathsf{T}}\mathbf{a}}$$

我们来证明，当误差最小的时候恰好为 $\mathbf{e}$ 与 $\mathbf{p}$ 垂直的时候：

$$
\begin{aligned}
\|\mathbf{b}-x\mathbf{a}\|^2 &= \|\mathbf{b}-\mathbf{p}+\mathbf{p}-x\mathbf{a}\|^2 \\
&= \|\mathbf{b}-\mathbf{p}\|^2 + \|\mathbf{p}-x\mathbf{a}\|^2 + 2(\mathbf{b}-\mathbf{p})\cdot(\mathbf{p}-x\mathbf{a}) \\
&= \|\mathbf{b}-\mathbf{p}\|^2 + \|\hat{x}\mathbf{a}-x\mathbf{a}\|^2 + 2(\mathbf{b}-\mathbf{p})\cdot(\hat{x}\mathbf{a}-x\mathbf{a}) \\
&= \|\mathbf{b}-\mathbf{p}\|^2 + (\hat{x}-x)^2\|\mathbf{a}\|^2 + 2(\hat{x}-x)(\mathbf{b}-\mathbf{p})\cdot\mathbf{a} \\
&= \|\mathbf{b}-\mathbf{p}\|^2 + (\hat{x}-x)^2\|\mathbf{a}\|^2 \\
&\geqslant \|\mathbf{b}-\mathbf{p}\|^2
\end{aligned}
$$

最后一个不等式等号成立当且仅当 $x=\hat{x}$，所以我们得到 $\mathbf{p}$ 是 $\mathbf{a}$ 方向这条线上唯一的一个点使得其与 $\mathbf{b}$ 的距离是最近的。

例 141.

1. 对于 $\mathbf{b}=\mathbf{a}=\begin{bmatrix}2\\3\end{bmatrix}$ 来说，其投影 $\mathbf{p}=\begin{bmatrix}2\\3\end{bmatrix}$
2. 对于 $\mathbf{b}=\begin{bmatrix}3\\-2\end{bmatrix}$，$\mathbf{a}=\begin{bmatrix}2\\3\end{bmatrix}$ 来说，其投影 $\mathbf{p}=\begin{bmatrix}0\\0\end{bmatrix}$
3. 对于 $\mathbf{b}=\begin{bmatrix}1\\1\\1\end{bmatrix}$ 和 $\mathbf{a}=\begin{bmatrix}1\\2\\2\end{bmatrix}$ 来说，$\mathbf{a}^\top\mathbf{b}=5$，$\|\mathbf{a}\|^2=9$，从而其投影 $\mathbf{p}$ 为：

$$\mathbf{p}=\hat{x}\mathbf{a}=\frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a}=\frac{5}{9}\mathbf{a}=\begin{bmatrix}\frac{5}{9}\\\frac{10}{9}\\\frac{10}{9}\end{bmatrix}$$

给定 $\mathbf{a} \in \mathbb{R}^m$ 和 $\mathbf{b} \in \mathbb{R}^m$，前面我们已经给出了 $\mathbf{b}$ 在 $\mathbf{a}$ 上的投影 $\mathbf{p}$，是否可以找到一个矩阵 P，使得我们有：

$$P\mathbf{b} = \mathbf{p}$$

解 142.

$$P = \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top\mathbf{a}}$$

这里 P 是一个 $m \times m$ 的矩阵。

证明.

$$P\mathbf{b} = \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top\mathbf{a}}\mathbf{b} = \frac{\mathbf{a}\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}} = \frac{\mathbf{a}^\top\mathbf{b}\mathbf{a}}{\mathbf{a}^\top\mathbf{a}} = \frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a} = \mathbf{p}$$

□

说明

注意 $\mathbf{a}^\top\mathbf{b}$ 既可以当成 1×1 的矩阵，也可以当成是一个 $\mathbb{R}$ 中的数。

回顾投影的误差是：

$$\mathbf{e} = \mathbf{b} - \mathbf{p}$$

从而当 P 是投影矩阵的时候，我们有：

$$(I - P)\mathbf{b} = I\mathbf{b} - P\mathbf{b} = \mathbf{b} - \mathbf{p}$$

注意到 $\mathbf{e}$ 是与 $\mathbf{p}$ 垂直的，从而 $I - P$ 是一个将 $\mathbf{b}$ 投影到与 $\mathbf{a}$ 正交的子空间的投影矩阵。

正交和投影 (Orthogonality and Projection)

投影到一个子空间

投影到一个子空间

我们现在来考虑对一个子空间的投影。令 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m$ 是线性无关的，即他们是下列子空间的一组基：

$$\mathbb{V} = \mathrm{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\})$$

与到一条线的投影相同，$\mathbf{b}$ 到 $\mathbb{V}$ 的投影应该是：

$$\mathbb{V} \text{ 中离 } \mathbf{b} \text{ 最近的元素 (可能是唯一的？)}$$

也就是说，我们需要寻找到 $\mathbb{V}$ 中的一个向量 $\mathbf{p}$：

$$\mathbf{p} = \hat{x}_1\mathbf{a}_1 + \cdots + \hat{x}_n\mathbf{a}_n$$

使得 $\|\mathbf{b} - \mathbf{p}\|$ 最小

记号

记 $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)$ 和 $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$，则我们有：

$$\mathbf{p} = A\hat{\mathbf{x}}$$

这里 A 是一个 $m \times n$ 的矩阵。

令

$$\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - A\hat{\mathbf{x}}$$

我们首先证明:

$$\mathbf{e} \perp \mathbb{V}$$

证明. 对于任意的 $\mathbf{v} \in \mathbb{V}$，我们有:

$$\mathbf{v} = A\mathbf{y}$$

从而:

$$\begin{aligned}
\|\mathbf{b} - \mathbf{v}\|^2 &= \|\mathbf{b} - A\mathbf{y}\|^2 \\
&= \|\mathbf{b} - A\hat{\mathbf{x}} + A\hat{\mathbf{x}} - A\mathbf{y}\|^2 \\
&= \|\mathbf{e}\|^2 + \|A(\hat{\mathbf{x}} - \mathbf{y})\|^2 + 2\mathbf{e} \cdot A(\hat{\mathbf{x}} - \mathbf{y}) \\
&= \|\mathbf{e}\|^2 + \|A(\hat{\mathbf{x}} - \mathbf{y})\|^2 \\
&\geqslant \|\mathbf{e}\|^2
\end{aligned}$$

这也意味着:

$$\mathbf{e} \perp \mathbf{a}_1, \cdots, \mathbf{e} \perp \mathbf{a}_n$$

从而我们有:

$$\begin{cases} \mathbf{a}_1^\top (b - A\hat{\mathbf{x}}) = 0 \\ \vdots \\ \mathbf{a}_n^\top (b - A\hat{\mathbf{x}}) = 0 \end{cases}$$

即:

$$A^\top (b - A\hat{\mathbf{x}}) = \mathbf{0}$$

p 的计算-投影矩阵 P

我们可以看到:

$$A^\top(b - A\hat{\mathbf{x}}) = \mathbf{0} \iff A^\top A\hat{\mathbf{x}} = A^\top \mathbf{b}$$

注意到 A 是一个 $m \times n$ 的矩阵:

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$$

并且其列向量是线性无关的，则 $A^\top A$ 是 $n \times n$ 的矩阵，并且如果我们可以证明 $A^\top A$ 是可逆的，则我们有:

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

并且我们可以得到 $\mathbf{b}$ 到 $\mathbb{V}(= \mathrm{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_n\})$ 的投影为:

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

对应的投影矩阵为:

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

考虑 $\mathbb{R}^3$，考虑矩阵 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ 的列空间和 $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$，我们来计算其投影和对应的投影矩阵。

1. $A^\top A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$, $A^\top \mathbf{b} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$
2. 解方程：$A^\top A\hat{\mathbf{x}} = A^\top \mathbf{b}$:
$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$
可得：$\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2) = (5, -3)$
3. 其投影 $\mathbf{p} = A\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$，误差为 $\mathbf{e} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$，投影矩阵 $P = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$

现在我们来证明 $A^\top A$ 的可逆性，注意到：

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$$

并且其列向量是线性无关的，所以其是列满秩的。

定理 143.

令 A 是一个 $m \times n$ 的矩阵，并且 $\text{rank}(A) = n$，则 $A^\top A$ 是可逆的。

上海师范大学 Shanghai Normal University

定理143的证明. 我们证明:column−rank$(A^\top A) = n$，即等价的:

$$A^\top A\mathbf{x} = \mathbf{0} \text{ 只有 } \mathbf{0} \text{ 一个解。}$$

事实上，我们有:

$$\begin{aligned}
A^\top A\mathbf{x} = \mathbf{0} \quad &\Longrightarrow \quad \mathbf{x}^\top A^\top A\mathbf{x} = 0 \\
&\Longleftrightarrow \quad (A\mathbf{x})^\top A\mathbf{x} = 0 \\
&\Longleftrightarrow \quad A\mathbf{x} \cdot A\mathbf{x} = 0 \\
&\Longleftrightarrow \quad \|A\mathbf{x}\| = 0 \\
&\Longleftrightarrow \quad A\mathbf{x} = \mathbf{0} \\
&\Longleftrightarrow \quad \mathbf{x} = \mathbf{0} \quad (\text{这是因为 rank}(A) = n)
\end{aligned}$$

$\square$

1. 我们的目标是计算 $\mathbf{b}$ 到下列空间:

$$\mathbb{V} = \mathrm{span}(\{\mathbf{a}_1, \ldots, \mathbf{a}_n\})$$

的投影 $\mathbf{p}$，其中 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 是线性无关的，$\mathbf{p} \in \mathbb{V}$.

2. 我们令 $\mathbf{p} \in \mathbb{V}$ 是满足其误差 $\mathbf{e} = \mathbf{b} - \mathbf{p}$ 与 $\mathbb{V}$ 垂直的向量。我们证明了，对于任意的 $\mathbf{v} \in \mathbb{V}$:

$$\|\mathbf{b} - \mathbf{v}\| = \min_{\mathbf{u} \in \mathbb{V}} \|\mathbf{b} - \mathbf{u}\| \quad \Longleftrightarrow \quad \mathbf{v} = \mathbf{p}$$

3. 我们得到了相应的投影矩阵 $P = A(A^\top A)^{-1}A^\top$，即:

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

并且我们证明了当 $\mathrm{rank}(A) = n$ 时 $(A^\top A)^{-1}$ 是存在的，这也说明了 $\mathbf{p}$ 的唯一性。

引理 140.

令 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的一个子空间，则对于任一 $\mathbf{x} \in \mathbb{R}^n$，我们都存在唯一的 $\mathbf{v} \in \mathbb{V}$ 和 $\mathbf{v}^{\perp} \in \mathbb{V}^{\perp}$ 使得:

$$\mathbf{x} = \mathbf{v} + \mathbf{v}^{\perp}$$

换句话说，

$$\mathbb{R}^n = \mathbb{V} + \mathbb{V}^{\perp} = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in \mathbb{V} \text{ and } \mathbf{v} \in \mathbb{V}^{\perp}\}$$

证明. 令 $\mathbf{a}_1, \cdots, \mathbf{a}_k$ 表示 $\mathbb{V}$ 的一组基，并且:

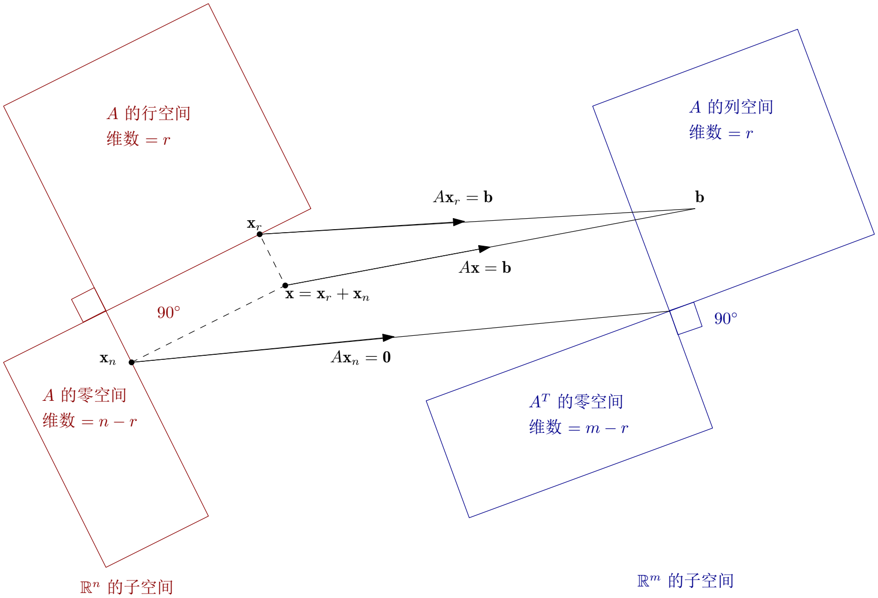
$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_k \end{bmatrix}$$

则对于任意的 $\mathbf{x} \in \mathbb{R}^n$，令 $\mathbf{u} = A(A^{\top}A)^{-1}A^{\top}\mathbf{x}$，则我们有:

$$\mathbf{x} = \mathbf{u} + (\mathbf{x} - \mathbf{u}), \quad \mathbf{u} \in \mathbb{V}, \ \mathbf{x} - \mathbf{u} \in \mathbb{V}^{\perp}$$

□

$m \times n$ 的矩阵 A 的四个空间



阶段总结

Shanghai Normal University 上海师范大学

- 投影到一条直线:

$$\mathbf{p}=\frac{\mathbf{a}^\top\mathbf{b}}{\mathbf{a}^\top\mathbf{a}}\mathbf{a},\quad P=\frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top\mathbf{a}}$$

- 投影到一个子空间:

$$\mathbf{p}=A\hat{x}=A(A^\top A)^{-1}A^\top\mathbf{b},\quad P=A(A^\top A)^{-1}A^\top$$

最后让我们回到方程 $A\mathbf{x} = \mathbf{b}$。

问题 141.

如果其没有解，我们如何找出一个 $\hat{\mathbf{x}}$ 使其是最为接近的一组解？

- 投影-最小二乘法 (Least Squares Approximation)!

最小二乘，标准正交基 (Least Square Approximations, Orthonormal Bases)

最小二乘，标准正交基 (Least Square Approximations, Orthonormal Bases)

最小二乘法

很多时候，$Ax = b$ 并不一定有解，但我们依旧想找到合适的 $\hat{x}$ 去表示。

例 142.

我们想研究人群受全日制教育的年数与收入是否存在关系。通过调查，我们得到了 n 个人的数据 $(t_1, y_1), \cdots, (t_n, y_n)$。这里 t_i 表示第 i 个人受全日制教育的年数，y_i 表示第 i 个人 35 岁的年收入。

假设其满足线性关系，即 $y = f(t) = kt + b$，则我们需要求解下列方程:

$$\begin{cases} kt_1 + b = y_1 \\ \cdots \\ kt_n + b = y_n \end{cases}, \quad \text{即} \quad \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

上海师范大学 Shanghai Normal University

我们也可以假设其满足二次关系，即 $y = f(t) = at^2 + bt + c$，则我们需要求解下列方程：

$$\begin{cases} at_1^2 + bt_1 + c = y_1 \\ \cdots \\ at_n^2 + bt_n + c = y_n \end{cases}, \quad \text{即} \quad \begin{bmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_n^2 & t_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

这类方程往往没有解 (方程数目远多于未知数)，但我们依旧希望能找到一个 x 使得 Ax 与 b 尽可能的接近。

$\min \|b - Ax\|$ 即 Ax是 $C(A)$ 上的投影。

$\min \|\mathbf{b} - A\mathbf{x}\|$ 的步骤 (I)

假设 A 是一个 $m \times n$ 的矩阵。

1. 我们知道如果 $\hat{\mathbf{x}} \in \mathbb{R}$ 满足:

$$A^{\mathsf{T}}(A\hat{x} - \mathbf{b}) = \mathbf{0} \quad 即 \quad A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$$

   则 $A\hat{\mathbf{x}} - \mathbf{b}$ 与 $\mathbf{C}(A)$ 正交。

2. 我们可以得到 $\hat{\mathbf{x}}$ 的表达式:

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

   并且我们知道 $\hat{x}$ 是唯一的。

3. 我们称 $\hat{x}$ 就是最小二乘解（least square solution），因为其误差的长度 $\|\mathbf{e}\|$

$$\mathbf{e} = \mathbf{b} - A\hat{x}$$

   是所有 $\mathbf{b} - A\mathbf{x}$ 中最小的。

问题 143.

上述步骤存在什么问题?

我们假设了:

$$\operatorname{rank}(A) = n$$

当 $\operatorname{rank}(A) < n$ 的时候，$A^\top A$ 不一定是可逆的，也就是 $A^\top A$ 不存在。

例 144.

令 $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$，则 $A^\top A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$，则 $\operatorname{rank}(A) = \operatorname{rank}(A^\top A) = 1$，$A^\top A$ 不可逆。

我们再来审视上面的步骤:

1. 我们知道如果 $\hat{\mathbf{x}} \in \mathbb{R}$ 满足:

$$A^{\mathsf{T}}(A\hat{\mathbf{x}} - \mathbf{b}) = \mathbf{0} \quad 即 \quad A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$$

   则 $A\hat{\mathbf{x}} - \mathbf{b}$ 与 $\mathbf{C}(A)$ 正交。

2. 我们可以得到 $\hat{\mathbf{x}}$ 的表达式:

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

   并且我们知道 $\hat{\mathbf{x}}$ 是唯一的。

3. $\hat{\mathbf{x}}$ 满足其误差 $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ 的长度 $\|\mathbf{e}\|$ 是所有 $\mathbf{b} - A\mathbf{x}$ 中最小的, 即:

$$\|\mathbf{e}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\| = \min\{\|\mathbf{b} - \mathbf{v}\| \mid \mathbf{v} \in \mathbf{C}(A)\}$$

上海师范大学 Shanghai Normal University

我们希望找到 $\mathbf{v} \in \mathbf{C}(A)$ 使得 $\|\mathbf{b} - \mathbf{v}\|$ 最小。

1. 选择 $\mathbf{C}(A)$ 的一组基 $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_r}\}$，其中 $r = \text{rank}(A)$。
2. 定义 $m \times r$ 的矩阵：
$$A' = \begin{bmatrix} \mathbf{a}_{i_1} & \cdots & \mathbf{a}_{i_r} \end{bmatrix}$$
显然我们有：$\mathbf{C}(A) = \mathbf{C}(A')$
3. rank(A') 是列满秩的，所以我们可以利用前面的方法来找到 $A'\mathbf{x}' = \mathbf{b}$ 的最优近似解，即：
$$\hat{\mathbf{x}}' = (A'^{\top}A')^{-1}A'^{\top}\mathbf{b} \in \mathbb{R}^n$$
显然其误差 $\mathbf{e} = \mathbf{b} - A'\hat{\mathbf{x}}'$ 是所有 $\mathbf{b} - A\mathbf{x}$ 中长度最小的，即：
$$\|\mathbf{e}\| = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\| = \min\{\|\mathbf{b} - \mathbf{v}\| \mid \mathbf{v} \in \mathbf{C}(A')\} == \min\{\|\mathbf{b} - \mathbf{v}\| \mid \mathbf{v} \in \mathbf{C}(A)\}$$
4. 我们需要的 $\hat{\mathbf{x}} \in \mathbb{R}^n$ 只要满足：
$$A\hat{\mathbf{x}} = A'\hat{\mathbf{x}}'$$

我们有:

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

如果 $A^\top A = I$,则上述公式可以简化为:

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

定理 145.

令 $A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}$ 是一个 $m \times n$ 的矩阵,则 $A^\top A = I$ 当且仅当下列的两个条件满足:

1. 对于任意的 $i \neq j$ 有 $\mathbf{a}_i \perp \mathbf{a}_j$。
2. 对于任意的 $i \in [n]$,$\mathbf{a}_i$ 是单位向量,即 $\|\mathbf{a}_i\| = 1$。

上海师范大学 Shanghai Normal University

定理146的证明. 只要注意到:

$$A^\top A = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{a}_1 & \cdots & \mathbf{a}_1^\top \mathbf{a}_n \\ \vdots & \ddots & \vdots \\ \mathbf{a}_n^\top \mathbf{a}_1 & \cdots & \mathbf{a}_n^\top \mathbf{a}_n \end{bmatrix}$$

从而:

$$A^\top A = I \iff \mathbf{a}_i^\top \mathbf{a}_j = \begin{cases} 1 & \text{当 } i = j \\ 0 & \text{当 } i \neq j \end{cases}$$

$\square$

我们再来看一下 $\hat{x} = A^{\mathsf{T}}\mathbf{b}$ 的情况。注意到此时 A 中每个列向量 $\mathbf{a}_i$ 都满足 $\|\mathbf{a}_i\| = 1$，从而对于所有的 $i \neq [n]$，我们有：

$$\hat{x}_i = \mathbf{a}_i^{\mathsf{T}}\mathbf{b} = \mathbf{a}_i \cdot \mathbf{b} = \frac{\mathbf{a}_i \cdot \mathbf{b}}{\|\mathbf{a}_i\|\|\mathbf{b}\|}\|\mathbf{b}\| = \|\mathbf{b}\| \cos\theta_i$$

这里 θ_i 是 $\mathbf{b}$ 和 $\mathbf{a}_i$ 的夹角。从而 $\mathbf{b}$ 到 $\mathbf{C}(A)$ 的投影 $\mathbf{p}$ 可以表示为：

$$\mathbf{p} = \sum_{i=1}^{n} \hat{x}_i \mathbf{a}_i = \sum_{i=1}^{n} (\|\mathbf{b}\| \cos\theta_i)\mathbf{a}_i$$

也就是：

p 是 b 分别到每条线 span($\{\mathbf{a}_i\}$) 上的投影的和。

说明

上述的性质并不是总对任意的 A 都成立。

上海师范大学 Shanghai Normal University

回顾 $A^\top A = I$ 成立的要求:

定理 146.

令 $A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}$ 是一个 $m \times n$ 的矩阵，则 $A^\top A = I$ 当且仅当下列的两个条件满足:

1. 对于任意的 $i \neq j$ 有 $\mathbf{a}_i \perp \mathbf{a}_j$。
2. 对于任意的 $i \in [n]$，$\mathbf{a}_i$ 是单位向量，即 $\|\mathbf{a}_i\| = 1$。

这其中核心是:

A 的列向量是 $\mathbf{C}(A)$ 的一组标准正交基 (orthonormal basis)。

最小二乘，标准正交基（Least Square Approximations, Orthonormal Bases）

标准正交基和 Gram-Schmidt 正交化

定义 147 **[Orthonormal Vectors].**

$\mathbf{q}_1, \cdots, \mathbf{q}_n \in \mathbb{R}^m$ 是标准正交的 (orthonormal)，如果：

$$\mathbf{q}_i^\top \mathbf{q}_j = \begin{cases} 0 & \text{当 } i \neq j\text{，即 } \mathbf{q}_i \text{ 和 } \mathbf{q}_j \text{ 是正交的} \\ 1 & \text{当 } i = j\text{，即 } \mathbf{q}_i \text{ 是单位向量} \end{cases}$$

我们将列向量是标准正交的矩阵记为 Q，显然我们有：

$$Q^\top Q = I$$

问题 148.

$QQ^\top = I$ 是否成立?

定义 149 **[正交矩阵 (Orthogonal Matrix)].**

称一个 $n \times n$ 的矩阵 Q 是正交矩阵 (Orthogonal Matrix)，如果:

$$Q^{\top}Q = I$$

或者等价的说，其列向量 $\mathbf{q}_1, \cdots, \mathbf{q}_n \in \mathbb{R}^n$ 也是标准正交的。

定理 150.

令 Q 是一个 $n \times n$ 的矩阵，则:

$$Q\text{是正交矩阵} \quad \Longleftrightarrow \quad Q\text{是可逆的并且 } Q^{-1} = Q^{\top}$$

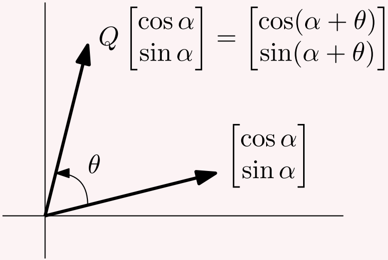
推论 151.

如果 Q 是一个正交矩阵，则其行向量也是标准正交的。

例 152.

我们第一个例子是 $\mathbb{R}^2$ 上的旋转矩阵:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad Q^{-1} = Q^{\top} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$



高维的情况

旋转也可以推广到高维的情况，相应的旋转被称作Givens 变换。

例 153.

我们第二个例子是 $\mathbb{R}^3$ 上的置换矩阵:

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q^{-1} = Q^{\mathsf{T}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

定理 154.

任意 $n \times n$ 的置换矩阵都是正交矩阵。

上海师范大学 Shanghai Normal University

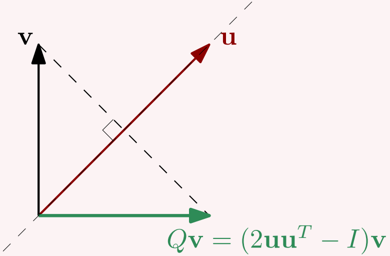
例 155.

我们第三个例子是 $\mathbb{R}^n$ 上的反射矩阵，令 $\mathbf{u} \in \mathbb{R}^n$ 是一个单位向量，定义矩阵 Q:

$$Q = 2\mathbf{u}\mathbf{u}^\top - I$$

则我们有:

$$Q^\top Q = (2\mathbf{u}\mathbf{u}^\top - I)^\top(2\mathbf{u}\mathbf{u}^\top - I) = 4\mathbf{u}\mathbf{u}^\top - 4\mathbf{u}\mathbf{u}^\top + I = I, \quad \text{即 } Q^{-1} = Q^\top = Q$$



理解反射矩阵的关键是要意识到当 $\mathbf{u}$ 是单位向量的时候，$\mathbf{u}\mathbf{u}^\top$ 是投影到 $\mathbf{u}$ 上的投影矩阵。该类矩阵也被称作Householder 矩阵。

定理 156.

令 Q 是一个 $n \times n$ 的正交矩阵，则对于任意的 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，我们有：

1. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$
2. $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$

证明.

- $\|Q\mathbf{x}\|^2 = Q\mathbf{x} \cdot Q\mathbf{x} = \mathbf{x}^\top Q^\top Q\mathbf{x} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|^2.$
- $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x}^\top Q^\top Q\mathbf{y} = \mathbf{x}^\top \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$

$\square$

令 $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ 是标准正交的，并且：

$$Q = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix}$$

显然 $\mathrm{rank}(Q) = n$。从而 $Q\mathbf{x} = \mathbf{b}$ 的最小二乘解为：$\hat{\mathbf{x}} = Q^\top \mathbf{b}$，对应的投影矩阵为 $QQ^\top$，从而 $\mathbf{b}$ 到 $\mathbf{C}(Q)$ 的投影 $\mathbf{p}$ 为：

$$\mathbf{p} = Q\hat{\mathbf{x}} = QQ^\top \mathbf{b} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^\top \mathbf{b} \\ \vdots \\ \mathbf{q}_n^\top \mathbf{b} \end{bmatrix} = \mathbf{q}_1 \mathbf{q}_1^\top \mathbf{b} + \cdots + \mathbf{q}_n \mathbf{q}_n^\top \mathbf{b}$$

也就是：

$\mathbf{p}$ 是 $\mathbf{b}$ 分别到每条线 $\mathrm{span}(\{\mathbf{q}_i\})$ 上的投影的和。

说明

上述表达也等价于 Q 是一个正交矩阵。

回顾 $A^\top A = I$

前面我们已经提到过，当 $A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}$ 满足下列条件：

$$A^\top A = I$$

即其列向量是标准正交的，则 $\mathbf{b}$ 到 $\mathbf{C}(A)$ 的投影 $\mathbf{p}$ 可以表示为：

$$\mathbf{p} = \mathbf{a}_1\mathbf{a}_1^\top\mathbf{b} + \cdots + \mathbf{a}_n\mathbf{a}_n^\top\mathbf{b}$$

现在我们假设 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ 只是两两正交的，则:

$$\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}, \cdots, \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$$

是标准正交的，并且:

$$\operatorname{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}) = \operatorname{span}(\{\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}, \cdots, \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}\})$$

从而 $\mathbf{b}$ 到 $\mathbf{C}(A)$ 上的投影 $\mathbf{p}$ 可以表示为:

$$\begin{aligned}
\mathbf{p} &= \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}\left(\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}\right)^\top \mathbf{b} + \cdots + \frac{\mathbf{a}_n}{\|\mathbf{a}_n\|}\left(\frac{\mathbf{a}_n}{\|\mathbf{a}_n\|}\right)^\top \mathbf{b} \\
&= \sum_{i \in [n]} \frac{\mathbf{a}_i \mathbf{a}_i^\top}{\mathbf{a}_i^\top \mathbf{a}_i} \mathbf{b} = \sum_{i \in [n]} \frac{\mathbf{a}_i^\top \mathbf{b}}{\mathbf{a}_i^\top \mathbf{a}_i} \mathbf{a}_i = \sum_{i \in [n]} \frac{\|\mathbf{b}\| \cos\theta_i}{\|\mathbf{a}_i\|} \mathbf{a}_i
\end{aligned}$$

这里 θ_i 是 $\mathbf{b}$ 和 $\mathbf{a}_i$ 的夹角。

我们再考虑一个一般的情况。如果 $A^{\mathsf{T}}A \neq I$,

$$\text{我们是否可以找到一个 } Q \text{ 使得 } Q^{\mathsf{T}}Q = I \text{ 并且 } \mathbf{C}(Q) = \mathbf{C}(A)?$$

等价地说，找到一个 Q 使得:

$$Q \text{ 的列向量组成了 } \mathbf{C}(A) \text{ 的一组标准正交基。}$$

定理 157.

令 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m$ 是线性无关的。则存在一组向量 $\mathbf{q}_1, \cdots, \mathbf{q}_n \in \mathbb{R}^m$，满足:

- 对于任意的 $i \neq j \in [n]$，$\mathbf{q}_i \perp \mathbf{q}_j$。
- $\mathrm{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}) = \mathrm{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_n\})$

推论 158.

一组 $\mathrm{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_n\})$ 的标准正交基为:

$$\frac{\mathbf{q}_1}{\|\mathbf{q}_1\|}, \cdots, \frac{\mathbf{q}_n}{\|\mathbf{q}_n\|}$$

我们将给出该定理的一个构造性证明，该过程也被称作Gram-Schmidt 正交化。

两个的情况

假设 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$，则 $\mathbf{x}$ 到 $\operatorname{span}(\{\mathbf{y}\})$ 上的投影为：

$$\mathbf{p} = \frac{\mathbf{y}^\mathsf{T}\mathbf{x}}{\mathbf{y}^\mathsf{T}\mathbf{y}}\mathbf{y}$$

其误差 $\mathbf{e} = \mathbf{x} - \mathbf{p}$ 满足 $\mathbf{e} \perp \mathbf{y}$。从而我们有：

$$\operatorname{span}(\{\mathbf{x}, \mathbf{y}\}) = \operatorname{span}(\{\mathbf{e}, \mathbf{y}\})$$

证明. 我们通过归纳的方式来获取 $\mathbf{q}_1, \cdots, \mathbf{q}_n$。
令 $\mathbf{q}_1 = \mathbf{a}_1$，假设对于 $k < n$，我们已经找到了 $\mathbf{q}_1, \cdots, \mathbf{q}_k$，满足:

- R1 对于任意的 $i \neq j \in [k]$，$\mathbf{q}_i \perp \mathbf{q}_j$。
- R2 $\mathrm{span}(\{\mathbf{a}_1, \cdots, \mathbf{a}_k\}) = \mathrm{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_k\})$

由于 $\mathbf{a}_1, \cdots, \mathbf{a}_k$ 是线性无关的，从而 $\mathbf{q}_1, \cdots, \mathbf{q}_k$ 也是线性无关的，并且 $\mathbf{q}_i \neq \mathbf{0}$.

证明. 我们定义:

$$\mathbf{q}_{k+1} = \mathbf{a}_{k+1} - \sum_{i=1}^{k} \frac{\mathbf{q}_i^\top \mathbf{a}_{k+1}}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{q}_i$$

注意到，事实上 $\sum_{i=1}^{k} \frac{\mathbf{q}_i^\top \mathbf{a}_{k+1}}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{q}_i$ 就是 $\mathbf{a}_{k+1}$ 到 $\mathrm{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_k\})$ 的投影 $\mathbf{p}$:

$$\mathbf{p} = \sum_{i=1}^{k} \frac{\mathbf{q}_i \mathbf{q}_i^\top}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{a}_{k+1} = \sum_{i=1}^{k} \frac{\mathbf{q}_i^\top \mathbf{a}_{k+1}}{\mathbf{q}_i^\top \mathbf{q}_i} \mathbf{q}_i = \sum_{i \in [k]} \frac{\|\mathbf{a}_{k+1}\| \cos \theta_i}{\|\mathbf{q}_i\|} \mathbf{q}_i$$

这里 θ_i 是 $\mathbf{a}_{k+1}$ 与 θ_i 的夹角。从而:

- $\mathbf{q}_{k+1} = \mathbf{a}_{k+1} - \mathbf{p}$ 与每个向量 $\mathbf{q}_i (i \in [k])$ 正交。
- $\mathrm{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_k, \mathbf{a}_{k+1}\}) = \mathrm{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_k, \mathbf{a}_{k+1} - \mathbf{p}\}) = \mathrm{span}(\{\mathbf{q}_1, \cdots, \mathbf{q}_k, \mathbf{q}_{k+1}\})$

□

一个例子

上海师范大学 Shanghai Normal University

考察如下三个 $\mathbb{R}^3$ 中的向量:

$$\mathbf{a}=\begin{bmatrix}1\\-1\\0\end{bmatrix},\ \mathbf{b}=\begin{bmatrix}2\\0\\-2\end{bmatrix},\ \mathbf{c}=\begin{bmatrix}3\\-3\\3\end{bmatrix}$$

我们可以通过 Gram-Schmidt 正交化来找到一组正交的向量组:

1. $\mathbf{q}_1=\mathbf{a}=\begin{bmatrix}1\\-1\\0\end{bmatrix}$
2. $\mathbf{q}_2=\mathbf{b}-\frac{\mathbf{q}_1^\top\mathbf{b}}{\mathbf{q}_1^\top\mathbf{q}_1}\mathbf{q}_1=\begin{bmatrix}1\\1\\-2\end{bmatrix}$
3. $\mathbf{q}_3=\mathbf{c}-\frac{\mathbf{q}_1^\top\mathbf{c}}{\mathbf{q}_1^\top\mathbf{q}_1}\mathbf{q}_1-\frac{\mathbf{q}_2^\top\mathbf{c}}{\mathbf{q}_2^\top\mathbf{q}_2}\mathbf{q}_2=\begin{bmatrix}1\\1\\1\end{bmatrix}$

单位化后即可得到一组标准正交基: $\{\frac{1}{\sqrt{2}}\mathbf{q}_1,\frac{1}{\sqrt{6}}\mathbf{q}_2,\frac{1}{\sqrt{3}}\mathbf{q}_3\}$.

再次回顾整个过程，令 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m$ 是线性无关的一组向量，我们利用 Gram-Schmidt 正交化得到一组正交的向量组 $\mathbf{q}_1, \cdots, \mathbf{q}_n$：

$$
\begin{aligned}
\mathbf{q}_1 &= \mathbf{a}_1 \\
\mathbf{q}_2 &= \mathbf{a}_2 - \frac{\mathbf{q}_1^\top \mathbf{a}_2}{\mathbf{q}_1^\top \mathbf{q}_1}\mathbf{q}_1 \\
\mathbf{q}_3 &= \mathbf{a}_3 - \frac{\mathbf{q}_1^\top \mathbf{a}_3}{\mathbf{q}_1^\top \mathbf{q}_1}\mathbf{q}_1 - \frac{\mathbf{q}_2^\top \mathbf{a}_3}{\mathbf{q}_2^\top \mathbf{q}_2}\mathbf{q}_2 \\
&\vdots \\
\mathbf{q}_n &= \mathbf{a}_n - \frac{\mathbf{q}_1^\top \mathbf{a}_n}{\mathbf{q}_1^\top \mathbf{q}_1}\mathbf{q}_1 - \frac{\mathbf{q}_2^\top \mathbf{a}_n}{\mathbf{q}_2^\top \mathbf{q}_2}\mathbf{q}_2 - \cdots - \frac{\mathbf{q}_{n-1}^\top \mathbf{a}_n}{\mathbf{q}_{n-1}^\top \mathbf{q}_{n-1}}\mathbf{q}_{n-1}
\end{aligned}
$$

可逆矩阵的 QR 分解 (II)

定义下列矩阵:

$$A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}, \quad \hat{Q} = \begin{bmatrix} \mathbf{q}_1, \cdots, \mathbf{q}_n \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{\mathbf{q}_1}{\|\mathbf{q}_1\|}, \cdots, \frac{\mathbf{q}_n}{\|\mathbf{q}_n\|} \end{bmatrix}$$

则我们有:

$$A = \hat{Q}R = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} 1 & \frac{\mathbf{q}_1^\top \mathbf{a}_2}{\mathbf{q}_1^\top \mathbf{q}_1} & \cdots & \frac{\mathbf{q}_1^\top \mathbf{a}_n}{\mathbf{q}_1^\top \mathbf{q}_1} \\ 0 & 1 & \cdots & \frac{\mathbf{q}_2^\top \mathbf{a}_n}{\mathbf{q}_2^\top \mathbf{q}_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$A = QR = \begin{bmatrix} \frac{\mathbf{q}_1}{\|\mathbf{q}_1\|} & \cdots & \frac{\mathbf{q}_n}{\|\mathbf{q}_n\|} \end{bmatrix} \begin{bmatrix} \|\mathbf{q}_1\| & \frac{\mathbf{q}_1^\top \mathbf{a}_2}{\|\mathbf{q}_1\|} & \cdots & \frac{\mathbf{q}_1^\top \mathbf{a}_n}{\|\mathbf{q}_1\|} \\ 0 & \|\mathbf{q}_2\| & \cdots & \frac{\mathbf{q}_2^\top \mathbf{a}_n}{\|\mathbf{q}_2\|} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\mathbf{q}_n\| \end{bmatrix}$$

这意味着任何一个可逆矩阵 A 都可以分解为一个正交矩阵 Q 和一个主对角线上是正数的上三角矩阵 R 的乘积，即 QR 分解。

定理 159.

令 A 是一个 $n \times n$ 的可逆矩阵，则存在一个正交矩阵 Q 和一个主对角线上是正数的上三角矩阵 R，使得

$$A = QR$$

特别的，该分解是唯一的。（唯一性留作练习）

再次回顾投影的计算

注意到:

$$A^{\top}A = R^{\top}Q^{\top}QR = R^{\top}R$$

从而:

$$A^{\top}A\hat{\mathbf{x}} = A^{\top}\mathbf{b} \quad \Longrightarrow \quad R^{\top}RA\hat{\mathbf{x}} = R^{\top}Q^{\top}\mathbf{b} \quad \Longrightarrow \quad RA\hat{\mathbf{x}} = Q^{\top}\mathbf{b}$$

从而我们有:

$$\hat{\mathbf{x}} = R^{-1}Q^{\top}\mathbf{b}$$

这里 R^{-1} 是一个上三角矩阵，从而 $\hat{\mathbf{x}}$ 可以通过回代法求解。

阶段总结

- $A\mathbf{x} = \mathbf{b}$ 无解的情况，通过最小二乘法（投影的方式）找到了最优的近似解 $\hat{\mathbf{x}}$。
- 正交矩阵和标准正交基。
- 找到一组正交基的方法：Gram-Schmidt 正交化。
- 矩阵的 QR 分解。

行列式 (Determinants)

行列式 (Determinants)

什么是行列式

什么是行列式？

第一次学行列式的时候，碰到的两个问题：

- 什么是行列式？
- 行列式有什么用？

目　录

学到的知识

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

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

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}}_{(A)} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$A_i = \begin{bmatrix} a_{11} & b_1 & \cdots & a_{1n} \\ a_{21} & b_2 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & b_n & \cdots & a_{nn} \end{bmatrix}$$

$$x_1 = \frac{detA_1}{detA}$$

$$x_2 = \frac{detA_2}{detA}$$

$$\vdots$$

$$x_n = \frac{detA_n}{detA}$$

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = (-1)^2 a_{11}A_{11} + \cdots + (-1)^{n+1} a_{1n}A_{1n}$$

$$= (-1)^2 a_{11} \begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{vmatrix} + \cdots + (-1)^{n+1} a_{1n} \begin{vmatrix} a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

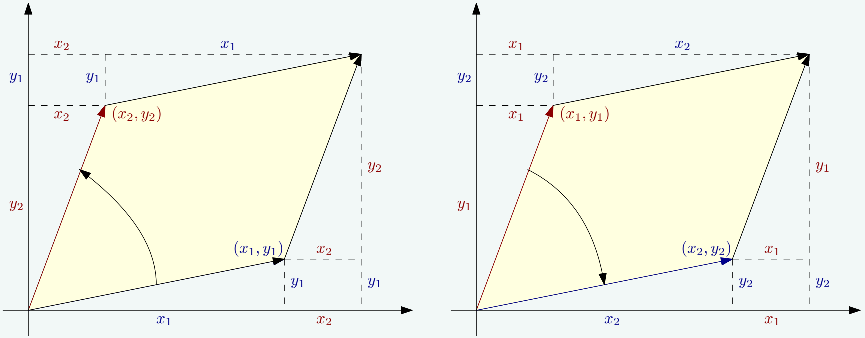
但是

为什么?

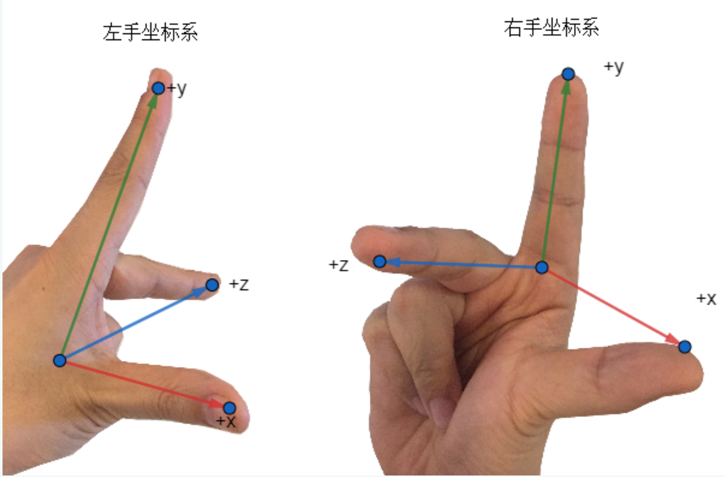
我们以 2×2 的行列式为例：

$$\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = x_1 y_2 - x_2 y_1$$

向量 (x_1, y_1) 和 (x_2, y_2) 围成的平行四边形的有向面积



左手系 VS 右手系



n 维空间的有向体积

对于 n 维空间 n 个向量构成的多面体:

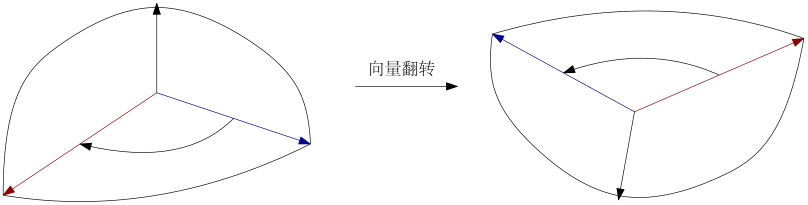
$$a_1, \cdots, a_i, \cdots, a_j, \cdots a_n, \quad \text{记其体积为 } D(a_1, \cdots, a_i, \cdots, a_j, \cdots a_n)$$

我们将其两个向量对换位置:

$$a_1, \cdots, a_j, \cdots, a_i, \cdots a_n, \quad \text{记其体积为 } D(a_1, \cdots, a_j, \cdots, a_i, \cdots a_n)$$

我们应当有:

$$D(a_1, \cdots, a_i, \cdots, a_j, \cdots a_n) = -D(a_1, \cdots, a_j, \cdots, a_i, \cdots a_n)$$



行列式的一个前瞻

行列式想求取的是：

在 n 维空间的 n 个向量张成的平行六面体 (n-dimensional parallelepiped) 的有向体积。

但行列式的值，存储了对应方阵的很多信息。

- (Pivot Formula)，我们将证明 $\det A$ 和 A 中所有首元的乘积的绝对值相同。
- (Big Formula), 我们将说明 $\det A$ 是 $n!$ 个不同元素组合之和。
- (Cofactor Expansion)，我们将证明 $\det A$ 是 A 中一些子行列式的线性组合。

从而我们可以得到很多矩阵的性质。

1. (Invertibility)，A 可逆当且仅当 $\det A \neq 0$。
2. (rank)，A 的秩等于 A 的行列式的非零子式的个数。
3. (Cramer Rules)，我们可以用行列式来求解线性方程组。
4. (Eigenvalues)，这是我们接下来要讨论的内容。
5. $\cdots$

令 $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ 是一个 $n \times n$ 的矩阵，方阵 A 的行列式 $\det(A)$是由 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 在 n 维空间长成的平行六面体 (n-dimensional parallelepiped) 的有向体积。

我们希望通过其的基本性质最终给出行列式对应的值，即最终一步一步得到 $\det(A)$ 相当于关于 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 的函数:

$$\det(A) = D(\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

行列式的性质

对于 $\mathbb{R}^n$ 的如下 n 个向量:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

则我们有:

$$D(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1$$

这也就是

- $\det(I) = 1$。

对于 n 个 $\mathbb{R}^n$ 的向量

$$\mathbf{a}_1, \cdots, \mathbf{a}_n$$

我们有:

$$D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots \mathbf{a}_n)$$

这也就是:

- 交换两列改变行列式的符号:

$$\det(\begin{bmatrix}\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n\end{bmatrix}) = -\det(\begin{bmatrix}\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n\end{bmatrix})$$

上海师范大学 Shanghai Normal University

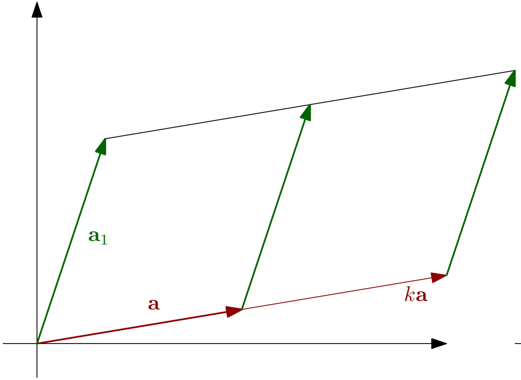
行列式的值应当对于每列都是线性的:

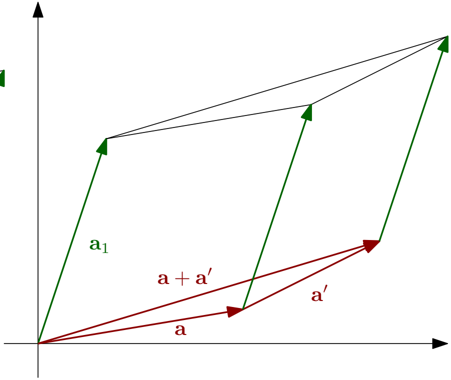
- $D(\mathsf{a}_1, \cdots, c\mathsf{a}_i, \cdots, \mathsf{a}_n) = cD(\mathsf{a}_1, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_n)$。
- $D(\mathsf{a}_1, \cdots, \mathsf{a}_i + \mathsf{a}_i', \cdots, \mathsf{a}_n) = D(\mathsf{a}_1, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_n) + D(\mathsf{a}_1, \cdots, \mathsf{a}_i', \cdots, \mathsf{a}_n)$。

这也就是:

- 行列式的值对于每一列都是满足线性的:

$$\det\left(\begin{bmatrix} \mathsf{a}_1 & \cdots & (c\mathsf{a}_i + d\mathsf{a}_i') & \cdots & \mathsf{a}_n \end{bmatrix}\right) = c\det\left(\begin{bmatrix} \mathsf{a}_1 & \cdots & \mathsf{a}_i & \cdots & \mathsf{a}_n \end{bmatrix}\right) + d\det\left(\begin{bmatrix} \mathsf{a}_1 & \cdots & \mathsf{a}_i' & \cdots & \mathsf{a}_n \end{bmatrix}\right)$$





记 S_1 是 a, a_1 张成的平行四边形的有向面积，S_2 是 $\mathsf{ka}, \mathsf{a}_1$ 张成的平行四边形的有向面积,S_3 是 $\mathsf{a}', \mathsf{a}_1$ 张成的平行四边形的有向面积,S_4 是 $\mathsf{a}+\mathsf{a}', \mathsf{a}_1$ 张成的平行四边形的有向面积，则:

$$S_2 = kS_1$$

$$S_4 = S_1 + S_3$$

回顾我们目前所做的，对于 n 维空间的 n 个向量 $\mathbf{a}_1, \cdots, \mathbf{a}_n$，我们希望定义一个函数 D, 满足下列性质:

- $D(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1$
- $D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots \mathbf{a}_n)$
- $D(\mathbf{a}_1, \cdots, c\mathbf{a}_i + d\mathbf{a}_i', \cdots, \mathbf{a}_n) = cD(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n) + dD(\mathbf{a}_1, \cdots, \mathbf{a}_i', \cdots, \mathbf{a}_n)$

我们希望说明:

D是存在的且唯一的

而这就是我们想要的行列式，即:

$$\det(A) = \det\left(\begin{bmatrix} \mathbf{a}_1 \cdots \mathbf{a}_n \end{bmatrix}\right) = D(\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

我们也会将其记作:

$$|A| \text{ 或者 } \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

在接下来的内容中，我们将分不加区分的使用 $n \times n$ 的矩阵 A 和其 n 个对应的列向量 $\mathbf{a}_1, \cdots, \mathbf{a}_n$。

定理 160.

- 如果 A 存在一个列向量是 $\mathbf{0}$，则 $\det(A) = 0$。
- 如果 A 存在两列向量相同，则 $\det(A) = 0$。
- 如果 A 存在一列是其他列的倍数，则 $\det(A) = 0$。

证明. 不妨假设 A 的第一列是 $\mathbf{a}_1 = \mathbf{0}$，则我们有:

$$D(\mathbf{a}_1', \cdots, \mathbf{a}_n) = D(\mathbf{a}_1 + \mathbf{a}_1', \cdots, \mathbf{a}_n) = D(\mathbf{a}_1, \cdots, \mathbf{a}_n) + D(\mathbf{a}_1', \cdots, \mathbf{a}_n)$$

从而:

$$\det(A) = D(\mathbf{a}_1, \cdots, \mathbf{a}_n) = D(\mathbf{0}, \cdots, \mathbf{a}_n) = 0$$

证明.

- 如果 A 存在两列向量相同，不妨记为 $\mathbf{a}_i = \mathbf{a}_j$，则我们有：

$$D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n)$$

  从而：

$$\det(A) = D(\mathbf{a}_1, \cdots, \mathbf{a}_n) = 0$$

- 如果 A 存在一列是其他列的倍数，不妨记为 $\mathbf{a}_i = c\mathbf{a}_j$，则我们有：

$$\det(A) = D(\mathbf{a}_1, \cdots, c\mathbf{a}_j, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = cD(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n) = 0$$

$\square$

定理 161.

如果 rank$(A) < n$, 则 $\det(A) = 0$。

证明. 由 rank$(A) < n$ 可知 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 是线性相关的，即存在 $1 \leqslant i_0 < i_1 < \cdots < i_k \leqslant n$ 和 $c_1, \ldots, c_k$ 使得:

$$\mathbf{a}_{i_0} = c_1\mathbf{a}_{i_1} + \cdots c_k\mathbf{a}_{i_k}$$

从而我们有:

$$\begin{aligned}\det(A) =& D(\mathbf{a}_1, \cdots, \mathbf{a}_{i_0}, \cdots, \mathbf{a}_n) \\ =& D(\mathbf{a}_1, \cdots, c_1\mathbf{a}_{i_1} + \cdots c_k\mathbf{a}_{i_k}, \cdots, \mathbf{a}_n) \\ =& \sum_{j=1}^{k} c_j D(\mathbf{a}_1, \cdots, \mathbf{a}_{i_j}, \cdots, \mathbf{a}_{i_j}, \cdots, \mathbf{a}_n) \\ =& 0\end{aligned}$$

□

定理 162.

将矩阵的某些列的线性组合加到另一列上 (其他列不改变)，行列式的值保持不变。

证明. 利用:

- $D(\mathbf{a}_1, \cdots, c\mathbf{a}_j, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = cD(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n) = 0$

可知:

$$D(\mathbf{a}_1, \cdots, \mathbf{a}_i + \sum_{i=1}^{n} c_i \mathbf{a}_i, \cdots, \mathbf{a}_n) = D(\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

□

定理 163.

令 A 是一个 $n \times n$ 的对角矩阵:

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}$$

则我们有:

$$\det(A) = a_1 a_2 \cdots a_n$$

定理 164.

令 A 是一个 $n \times n$ 的三角矩阵 (triangluar matrix):

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \quad \text{或者} \quad A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

则我们有:

$$\det(A) = a_{11}a_{22}\cdots a_{nn}$$

阶段总结

- 行列式的几何意义。
- 行列式需要满足的基本性质。
- 由这些性质衍生的一些性质。

我们的目标

考虑一个 $\mathbb{R}^{n\times n}\to\mathbb{R}$ 的函数 D，其满足下列三个性质：

- $D(\mathbf{e}_1,\cdots,\mathbf{e}_n)=1$
- $D(\mathbf{a}_1,\cdots,\mathbf{a}_i,\cdots,\mathbf{a}_j,\cdots\mathbf{a}_n)=-D(\mathbf{a}_1,\cdots,\mathbf{a}_j,\cdots,\mathbf{a}_i,\cdots\mathbf{a}_n)$
- $D(\mathbf{a}_1,\cdots,c\mathbf{a}_i+d\mathbf{a}_i',\cdots,\mathbf{a}_n)=cD(\mathbf{a}_1,\cdots,\mathbf{a}_i,\cdots,\mathbf{a}_n)+dD(\mathbf{a}_1,\cdots,\mathbf{a}_i',\cdots,\mathbf{a}_n)$

我们希望说明，这样的函数D 存在而且唯一，从而其就是对应的行列式 $\det(A)$。

我们的思路

为了证明这样的 D 存在且唯一，我们希望：

- 通过 D 的性质，我们尝试计算出对于任意的 $n \times n$ 的矩阵 A，其对应的函数值。
- 如果对于每个 $n \times n$ 的矩阵 A，我们都能计算出一个唯一的函数值，那么我们就可以说明 D 是存在且唯一的。

行列式的计算

列加法

令 $i, j \in [n]$，我们将第 j 列的 l 倍加到第 i 列上，我们有:

$$A' = \begin{bmatrix} a_{11} & \cdots & a_{1i} + l a_{1j} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ni} + l a_{nj} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}$$

则我们有:

$$\det(A') = \det(A)$$

列交换

令 $i, j \in [n]$，我们将第 j 列和第 i 列互换，我们有:

$$A' = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1i} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{ni} & \cdots & a_{nn} \end{bmatrix}$$

则我们有:

$$\det(A') = -\det(A)$$

令 $i \in [n]$ 和 $l \in \mathbb{R}$，我们将第 i 列乘以 l 倍，我们有:

$$A' = \begin{bmatrix} a_{11} & \cdots & la_{1i} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & la_{ni} & \cdots & a_{nn} \end{bmatrix}$$

则我们有:

$$\det(A') = l\det(A)$$

计算行列式 A 的值

1. 我们对 A 进行初等列变换，或者等价地说，对 $A^{\top}$ 进行初等行变换。最终我们可以得到一个 $A^{\top}$ 的行阶梯形 R。从而 $R^{\top}$ 是一个下三角矩阵。令其对角线的元素为 $d_1, \cdots, d_n$，则我们有:
$$\det(R^{\top}) = d_1 \cdots d_n$$
2. 我们根据 A 变成 $R^{\top}$ 的过程，注意到这个过程中得每一步都是列变换，从而根据前面得结论一步步反推得到 $\det(A)$ 的值。

定理 165.

$\operatorname{rank}(A) = n$ 当且仅当 $\det(A) \neq 0$。

考察 2×2 的矩阵:

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

我们有:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \Longrightarrow \begin{bmatrix} a & 0 \\ c & d-\frac{cb}{a} \end{bmatrix}$$

从而:

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d-\frac{cb}{a} \end{vmatrix} = a\cdot(d-\frac{cb}{a}) = ad-bc$$

一些例子 (II)

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

$$\Longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & -\frac{3}{2} \\ -\frac{1}{2} & -2 & \frac{7}{2} \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & -2 & \frac{1}{2} \end{bmatrix}$$

则我们有:

$$\det\left(\begin{bmatrix} 2 & 0 & 1 \\ 1 & -4 & -1 \\ -1 & 8 & 3 \end{bmatrix}\right) = \begin{vmatrix} 2 & 0 & 1 \\ 1 & -4 & -1 \\ -1 & 8 & 3 \end{vmatrix} = 2 \cdot (-4) \cdot \begin{vmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & -2 & \frac{1}{2} \end{vmatrix} = 2 \cdot (-4) \cdot \frac{1}{2} = -4$$

$$\begin{bmatrix} 2 & 1 & 4 & 1 \\ 3 & -1 & 2 & 1 \\ 1 & 2 & 3 & 2 \\ 5 & 0 & 6 & 2 \end{bmatrix} \Longrightarrow \begin{bmatrix} 2 & 0 & 0 & 0 \\ 3 & -\frac{5}{2} & -4 & -\frac{1}{2} \\ 1 & \frac{3}{2} & 1 & \frac{3}{2} \\ 5 & -\frac{5}{2} & -4 & -\frac{1}{2} \end{bmatrix} \Longrightarrow \begin{bmatrix} 2 & 0 & 0 & 0 \\ 3 & -\frac{5}{2} & 0 & 0 \\ 1 & \frac{3}{2} & -\frac{7}{5} & -\frac{9}{10} \\ 5 & -\frac{5}{2} & 0 & 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} 2 & 0 & 0 & 0 \\ 3 & -\frac{5}{2} & 0 & 0 \\ 1 & \frac{3}{2} & -\frac{7}{5} & 0 \\ 5 & -\frac{5}{2} & 0 & 0 \end{bmatrix}$$

从而:

$$\det(\begin{bmatrix} 2 & 1 & 4 & 1 \\ 3 & -1 & 2 & 1 \\ 1 & 2 & 3 & 2 \\ 5 & 0 & 6 & 2 \end{bmatrix}) = \begin{vmatrix} 2 & 1 & 4 & 1 \\ 3 & -1 & 2 & 1 \\ 1 & 2 & 3 & 2 \\ 5 & 0 & 6 & 2 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 0 & 0 \\ 3 & -\frac{5}{2} & 0 & 0 \\ 1 & \frac{3}{2} & -\frac{7}{5} & 0 \\ 5 & -\frac{5}{2} & 0 & 0 \end{vmatrix} = 2 \cdot (-\frac{5}{2}) \cdot (-\frac{7}{5}) \cdot 0 = 0$$

$$\begin{bmatrix} a & 0 & \cdots & 1 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & a \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 0 & \cdots & 1 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a} & 0 & \cdots & a \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a} & 0 & \cdots & a-\frac{1}{a} \end{bmatrix}$$

从而:

$$\det\left(\begin{bmatrix} a & 0 & \cdots & 1 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & a \end{bmatrix}\right) = \begin{vmatrix} a & 0 & \cdots & 1 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & a \end{vmatrix} = a \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a} & 0 & \cdots & a-\frac{1}{a} \end{vmatrix} = a^{n-1}(a-\frac{1}{a}) = a^n - a^{n-2}$$

D 的存在性

回顾一下我们对 $\det A$ 的计算:

1. 根据我们的理解，$\det A$ 如果存在，一定要满足三条基本性质。
2. 我们推得了关于 $\det A$ 的一些其他性质。
3. 通过上述的讨论，我们知道了初等列变换对 $\det A$ 的影响。
4. 而通过初等列变换，我们可以将 A 转变成一个下三角矩阵 L.
5. 我们知道 $\det(L)$ 是多少，从而根据 3 和 4，我们可以从 $\det L$ 得到 $\det(A)$ 的值。

上述的过程说明了 $\det A$ 的存在性

- 我们描绘了行列式想计算的内容

  在 n 维空间的 n 个向量张成的平行六面体 (n-dimensional parallelepiped) 的有向体积
- 由此发生，我们假定了 $\det(A)$ 该满足的三条性质
  1. $D(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1$
  2. $D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots \mathbf{a}_n)$
  3. $D(\mathbf{a}_1, \cdots, c\mathbf{a}_i + d\mathbf{a}_i', \cdots, \mathbf{a}_n) = cD(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n) + dD(\mathbf{a}_1, \cdots, \mathbf{a}_i', \cdots, \mathbf{a}_n)$
- 我们通过这三条性质推导了一些其他性质。
- 我们通过初等列变换，将 A 转变成一个下三角矩阵 L，从而得到了 $\det(A)$ 的值。这一构造过程，说明了 $\det(A)$ 的存在性。

行列式更多的性质

我们将证明:

定理 166 **[Transpose].**

令 A 是一个 $n \times n$ 的矩阵，我们有:

$$\det(A) = \det(A^{\top})$$

定理 167 **[Product of Determinants].**

令 A 是一个 $n \times n$ 的矩阵，B 是一个 $n \times n$ 的矩阵，我们有:

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

这也说明，对矩阵的初等行变换对行列式的影响跟初等列变换是完全相同的

推论 168.

给定一个正交矩阵 Q，我们有：

$$\det(Q) = \pm 1$$

这说明，在 $\mathbb{R}^n$ 中由 n 个标准正交 (orthonormal) 基向量构成的平行六面体的体积是 1。

证明. 注意到：

$$Q^\top Q = I$$

从而我们有：

$$1 = \det(I) = \det(Q)\det(Q^\top) = \det(Q)\det(Q) = (\det(Q))^2 \quad \Longrightarrow \quad \det(Q) = \pm 1$$

$\square$

引理 169.

令 A 是一个 $n \times n$ 的矩阵，E 是一个 $n \times n$ 的初等矩阵，则我们有:

$$\det(AE) = \det(A)\det(E)$$

推论 170.

令 A 是一个 $n \times n$ 的矩阵，$E_1, \cdots, E_k$ 是 $n \times n$ 的初等矩阵，则我们有:

$$\det(AE_1E_2\cdots E_k) = \det(A)\det(E_1)\det(E_2)\cdots\det(E_k)$$

特别的:

$$\det(E_1E_2\cdots E_k) = \det(E_1)\det(E_2)\cdots\det(E_k)$$

注意到 E 是下列三种矩阵的一种：

$$
\underbrace{\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 1 & \cdots & k & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix}}_{E_{ij}(k)},
\underbrace{\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix}}_{P_{ij}},
\underbrace{\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & k & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix}}_{D_i(k)}
$$

注意到 AE 是对 A 进行相应的列变换操作，所以我们只需要检查每种初等矩阵是否满足 $\det(AE) = \det(A)\det(E)$ 即可。

令 $A = \begin{bmatrix} \mathsf{a}_1, \cdots, \mathsf{a}_n \end{bmatrix}$，注意到:

$$AE_{ij}(k) = \begin{bmatrix} \mathsf{a}_1, \cdots, \mathsf{a}_i + k\mathsf{a}_j, \cdots, \mathsf{a}_j, \cdots, \mathsf{a}_n \end{bmatrix}$$

从而我们有:

$$\begin{aligned} \det(E_{ij}(k)) &= \det(I) = 1 \\ \det(AE_{ij}(k)) &= D(\mathsf{a}_1, \cdots, \mathsf{a}_i + k\mathsf{a}_j, \cdots, \mathsf{a}_j, \cdots, \mathsf{a}_n) \\ &= D(\mathsf{a}_1, \cdots, \mathsf{a}_i, \cdots, \mathsf{a}_j, \cdots, \mathsf{a}_n) + kD(\mathsf{a}_1, \cdots, \mathsf{a}_j, \cdots, \mathsf{a}_j, \cdots, \mathsf{a}_n) \\ &= \det(A) \end{aligned}$$

即:

$$\det(AE_{ij}(k)) = \det(A)\det(E_{ij}(k))$$

令 $A = \left[\mathbf{a}_1, \cdots, \mathbf{a}_n\right]$，注意到:

$$AP_{ij} = \left[\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n\right]$$

从而我们有:

$$\begin{aligned}\det(P_{ij}) &= -\det(I) = -1 \\ \det(AP_{ij}) &= D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n) \\ &= -D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n) \\ &= -\det(A)\end{aligned}$$

即:

$$\det(AP_{ij}) = \det(A)\det(P_{ij})$$

令 $A = \begin{bmatrix} \mathbf{a}_1, \cdots, \mathbf{a}_n \end{bmatrix}$，注意到:

$$AD_i(k) = \begin{bmatrix} \mathbf{a}_1, \cdots, k\mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n \end{bmatrix}$$

从而我们有:

$$\begin{aligned} \det(D_i(k)) &= k\det(I) = k \\ \det(AD_i(k)) &= D(\mathbf{a}_1, \cdots, k\mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n) \\ &= kD(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_n) \\ &= k\det(A) \end{aligned}$$

即:

$$\det(AD_i(k)) = \det(A)\det(D_i(k))$$

回顾 Gauss–Jordan 消元法:

$$D\cdots E\cdots P\cdots E\begin{bmatrix}A & I\end{bmatrix}=\begin{bmatrix}I & A^{-1}\end{bmatrix}$$

我们有:

定理 171.

每个可逆矩阵都可以由一系列的初等矩阵的乘积得到。

定理 166 **[Transpose].**

令 A 是一个 $n \times n$ 的矩阵，我们有：

$$\det(A) = \det(A^\top)$$

证明.

1. 对于每个初等矩阵 E，我们有 $\det(E) = \det(E^\top)$。
2. 如果 A 不是可逆的，则 $\mathrm{rank}(A) = \mathrm{rank}(A^\top) < n$，从而

$$\det(A) = \det(A^\top) = 0$$

3. 如果 A 是可逆的，则存在一系列的初等矩阵 $E_1, \cdots E_l$ 使得：

$$A = E_1 \cdots E_l$$

从而 $A^\top = E_l^\top \cdots E_1^\top$，因此：

$$\begin{aligned}\det(A) &= \det(E_1) \cdots \det(E_l) \\ &= \det(E_1^\top) \cdots \det(E_l^\top) = \det(E_l^\top) \cdots \det(E_1^\top) = \det(E_l^\top \cdots E_1^\top) = \det(A^\top)\end{aligned}$$

我们现在来证明:

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

证明. 如果 B 是可逆的，则存在一系列的初等矩阵 $E_1,\cdots E_k$ 使得:

$$B=E_1\cdots E_k$$

从而:

$$\begin{aligned}\det(AB)=\det(AE_1\cdots E_k)&=\det(A)\det(E_1)\cdots\det(E_k)\\&=\det(A)\det(E_1\cdots E_k)=\det(A)\det(B)\end{aligned}$$

如果 B 不是可逆的，则 AB 也不可逆，否则:

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

所以此时我们有:

$$\det(AB)=\det(A)\det(B)=0$$

$\square$

一个运用-Cramer 法则 (I)

我们来展示 $\det(AB) = \det(A)\det(B)$ 的一个运用。考察下列的方程组:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= b_n \end{aligned} \quad \text{即} \quad \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \text{(记作 } A\mathbf{x} = \mathbf{b}\text{)}$$

注意到:

$$A \begin{bmatrix} x_1 & 0 & 0 & \cdots & 0 \\ x_2 & 1 & 0 & \cdots & 0 \\ x_3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} A\mathbf{x} & A\mathbf{e}_2 & A\mathbf{e}_3 & \cdots & A\mathbf{e}_n \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

记下列矩阵为 A_1:

$$\begin{bmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

注意到:

$$\begin{vmatrix} x_1 & 0 & 0 & \cdots & 0 \\ x_2 & 1 & 0 & \cdots & 0 \\ x_3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n & 0 & 0 & \cdots & 1 \end{vmatrix} = \begin{vmatrix} x_1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix} = x_1$$

则我们有:

$$\det(A)x_1 = \det(A_1), \quad \text{即:} \quad x_1 = \frac{\det(A_1)}{\det(A)}$$

一个运用-Cramer 法则 (III)

对于任意的 $i \in [n]$，令 A_i 是将 A 的第 i 列替换成 $\mathbf{b}$ 后的矩阵，则我们有：

$$A\begin{bmatrix} 1 & \cdots & x_1 & \cdots & 0 \\ 0 & \cdots & x_2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & 0 \\ 0 & \cdots & x_n & \cdots & 1 \end{bmatrix} = \left[A\mathbf{e}_1 \cdots A\mathbf{x} \cdots A\mathbf{e}_n\right] = A_i \quad \Longrightarrow \det(A)x_i = \det(A_i)$$

从而方程组 $A\mathbf{x} = \mathbf{b}$ 的解可以表示为：

$$\mathbf{x} = \begin{bmatrix} \frac{\det(A_1)}{\det(A)} \\ \vdots \\ \frac{\det(A_n)}{\det(A)} \end{bmatrix}$$

这就是克拉默法则 (Cramer's Rule)

另一个应用-$\det(A)$ 跟首元的关系

上海师范大学 Shanghai Normal University

令可逆矩阵 A 的首元为 $p_1, \cdots, p_n$，通过 Gauss 消元法我们可以将矩阵 A 变为如下的行阶梯形矩阵:

$$U = \begin{bmatrix} p_1 & * & \cdots & * \\ 0 & p_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_n \end{bmatrix}$$

由于上述过程只使用了行加法和行交换操作 (列加法或者列交换)，从而我们有:

$$\det(A) = \det(U) = p_1 p_2 \cdots p_n \text{ 或者 } \det(A) = -\det(U) = -p_1 p_2 \cdots p_n$$

即:

$$|\det(A)| = |p_1 p_2 \cdots p_n|$$

行列式的正式定义

考虑一个 $n \times n$ 的矩阵:

$$A = \begin{bmatrix} a_{11} & \cdots & \mathbf{a}_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & \mathbf{a}_{nn} \end{bmatrix}$$

注意到其每个列向量:

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix} = a_{1j} \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + a_{nj} \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} = \sum_{i=1}^{n} \mathbf{a}_{ij} \mathbf{e}_i$$

从而由行列式的线性性，我们有:

$$D(\mathbf{a}_1, \cdots, \mathbf{a}_n) = D(\sum_{i=1}^{n} \mathbf{a}_{ij} \mathbf{e}_i, \cdots, \mathbf{a}_n) = \sum_{i=1}^{n} a_{i1} D(\mathbf{e}_1, \cdots, \mathbf{a}_n)$$

如果我们将矩阵的每一列都按上述方式展开，则：

$$D(\mathbf{a}_1,\cdots,\mathbf{a}_n)=\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n}^{n}a_{i_1 1}a_{i_2 2}\cdots a_{i_n n}D(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\cdots,\mathbf{e}_{i_n})$$

上述等式一共有n^n项。由行列式的性质，如果存在 $i_k=i_j$，则我们有：

$$D(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\cdots,\mathbf{e}_{i_k},\cdots,\mathbf{e}_{i_j},\cdots\mathbf{e}_{i_n})=0$$

从而上述等式可以转化为：

$$D(\mathbf{a}_1,\cdots,\mathbf{a}_n)=\sum_{i_1\in[n]}\sum_{\substack{i_2\in[n]\\ i_2\neq i_1}}\cdots\sum_{\substack{i_n\in[n]\\ i_n\neq i_1,\cdots,i_n\neq i_{n-1}}}a_{i_1 1}a_{i_2 2}\cdots a_{i_n n}D(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\cdots,\mathbf{e}_{i_n})$$

我们来观察一下这个式子:

$$D(\mathbf{a}_1,\cdots,\mathbf{a}_n)=\sum_{i_1\in[n]}\sum_{\substack{i_2\in[n]\\ i_2\neq i_1}}\cdots\sum_{\substack{i_n\in[n]\\ i_n\neq i_1,\cdots,i_n\neq i_{n-1}}} a_{i_1 1}a_{i_2 2}\cdots a_{i_n n}D(\mathbf{e}_{i_1},\mathbf{e}_{i_2},\cdots,\mathbf{e}_{i_n})$$

事实上, $i_1,\cdots,i_n$是 $1,\cdots,n$ 的一个置换。

例 167.

- $1,3,2,4$ 就是 $1,2,3,4$ 的一个置换。
- c,d,e,a,b 是 a,b,c,d,e 的一个置换。

定义 168 **[Permutation].**

固定一个 $n \in \mathbb{N}$，一个 $[n]$ 的置换 (Permutation) 是一个 $[n] \to [n]$ 的双射函数 (bijective function)σ，并且我们定义：

$$\mathrm{Perm}(n) = \{\sigma \mid \sigma\text{是 } [n] \text{ 的一个置换。}\}$$

从而我们可以将上述表达式改写为：

$$\begin{aligned} D(\mathbf{a}_1, \cdots, \mathbf{a}_n) &= \sum_{i_1 \in [n]} \sum_{\substack{i_2 \in [n] \\ i_2 \neq i_1}} \cdots \sum_{\substack{i_n \in [n] \\ i_n \neq i_1, \cdots, i_n \neq i_{n-1}}} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} D(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \cdots, \mathbf{e}_{i_n}) \\ &= \sum_{\sigma \in \mathrm{Perm}(n)} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} D(\mathbf{e}_{\sigma}(1), \cdots, \mathbf{e}_{\sigma(n)}) \end{aligned}$$

我们需要搞清楚 $D(\mathbf{e}_\sigma(1), \cdots, \mathbf{e}_{\sigma(n)})$ 是什么，注意到:

$$P = \left[\mathbf{e}_{\sigma(1)}, \cdots, \mathbf{e}_{\sigma(n)}\right]$$

是一个置换矩阵。

1. P 可以通过若干次列交换操作变成单位矩阵 I。
2. 每次列交换操作改变其行列式的符号。
3. 从而 $\det(P) = (-1)^k$，这里 k 是列交换的次数。

注意!

这里的 k 是不唯一的!

1. 我们希望定义 $\det A$ 来表示：由 $\mathbf{a}_1, \cdots, \mathbf{a}_n$ 在 n 维空间长成的平行六面体 (n-dimensional parallelepiped) 的有向体积。
2. 我们发现 $\det A$ 需要满足三个基本的性质：
   - $D(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1$
   - $D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots \mathbf{a}_n)$
   - $D(\mathbf{a}_1, \cdots, c\mathbf{a}_i + d\mathbf{a}_i', \cdots, \mathbf{a}_n) = cD(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n) + dD(\mathbf{a}_1, \cdots, \mathbf{a}_i', \cdots, \mathbf{a}_n)$
3. 通过这三个基本的性质，我们计算出了行列式的表达：

$$\begin{aligned}\det(A) &= \sum_{\sigma \in Perm(n)} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} D(\mathbf{e}_\sigma(1), \cdots, \mathbf{e}_{\sigma(n)}) \\ &= \sum_{\sigma \in Perm(n)} (-1)^{k_\sigma} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}\end{aligned}$$

这里 k_σ 是置换矩阵 $\left[\mathbf{e}_{\sigma(1)}, \cdots, \mathbf{e}_{\sigma(n)}\right]$ 变成单位矩阵 I 所需要的列交换的次数。但我们还面临一个问题：

k_σ 并不唯一，所以我们不能用上述式子作为 $\det(A)$ 的定义。

令 $\sigma : [n] \to [n]$ 是一个置换，我们用如下的方式进行简写：

$$\sigma(1)\sigma(2)\cdots\sigma(n)$$

比如，我们称：

$$4213\text{就是 } 1234 \text{ 的一个置换。}$$

观察这些置换的特点：

- 其中存在在前面但是更大的数，比如在 4213 中，4 在 2 前面，但是 $4 > 2$.

我们称这样的一堆数为逆序对，而一个置换中的逆序对的个数为这个置换的逆序数 (Inversion Number)

定义 169 **[Inversion numbers].**

给定一个置换 $\sigma = \sigma(1)\sigma(2)\cdots\sigma(n)$，其逆序数 (Inversion number)$\tau(\sigma)$定义为：

$$\tau(\sigma) = |\{(p, q) \mid 1 \leqslant p < q \leqslant n \text{ 并且 } \sigma(p) > \sigma(q)\}|$$

例 170.

- $\tau(4213) = 4.$
- $\tau(1324) = 1$
- $\tau(3241) = 3.$

观察到:

$$\tau(4213)=4 \implies 4213 \to 2413 \to 2143 \to 2134 \to 1234$$

$$\tau(1324)=1 \implies 1324 \to 1234$$

$$\tau(3241)=3 \implies 3214 \to 2314 \to 2134 \to 1234$$

逆序数恰好可以作为一个将置换 $\sigma=\sigma(1)\sigma(2)\cdots\sigma(n)$ 变回 $123\cdots n$ 的个数!

定理 171.

给定置换矩阵

$$P=\left[\mathbf{e}_{\sigma(1)},\cdots,\mathbf{e}_{\sigma(n)}\right]$$

我们可以通过 $\tau(\sigma)$ 次列交换将其变回单位矩阵 I。

这意味着我们可以用 $\tau(\sigma)$ 来定义 k_σ!

定理171的证明. 注意到对于一个置换 σ，我们有:

$$\tau(\sigma)=\sum_{i=1}^{n}|\{t \mid t>i \text{ 并且 } \sigma(i)>\sigma(t)\}|$$

从而对于置换矩阵我们可以定义如下的列交换操作, 初始令 $k=n$:

1. 不妨令 $k=\sigma(t)$，令 $t_1<t_2<\cdots<t_j$ 是满足 $k>\sigma(t_i)$ 的位置。依次将第 k 列与第 t_i 列进行交换，共交换 j 次。
2. 令 $k=k-1$，重复上述过程。

注意到上述列交换的总数恰好是置换 σ 的逆序数 (为什么？)，从而定理得证。 □

现在我们就可以得到行列式的正式定义了。

定义 172.

令 A 是一个 $n \times n$ 的矩阵，我们定义：

$$\det(A) = \sum_{\sigma \in \mathrm{Perm}(n)} (-1)^{\tau(\sigma)} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}$$

定理 173.

$\det(A)$ 满足下列性质：

- $D(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1$
- $D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots \mathbf{a}_n)$
- $D(\mathbf{a}_1, \cdots, c\mathbf{a}_i + d\mathbf{a}_i', \cdots, \mathbf{a}_n) = cD(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n) + dD(\mathbf{a}_1, \cdots, \mathbf{a}_i', \cdots, \mathbf{a}_n)$

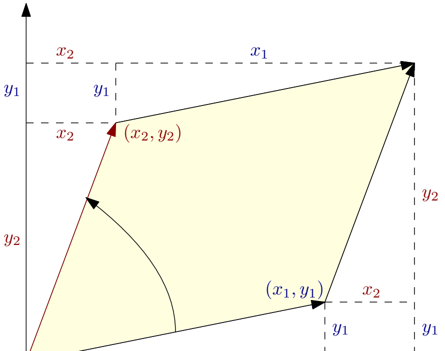
例子-2×2 的情况

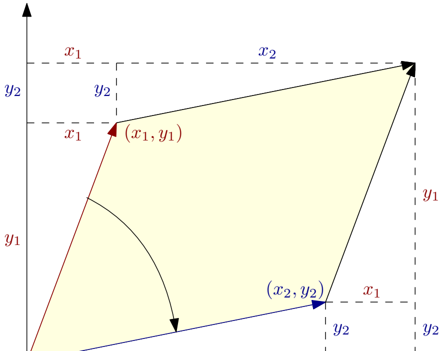
对于一个 2×2 的矩阵:

$$A=\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}$$

一共有 $12, 21$ 两种不同的置换，从而我们有:

$$\det(A)=(-1)^{\tau(12)}x_1y_2+(-1)^{\tau(21)}x_2y_1=x_1y_2-x_2y_1$$





对于一个 3×3 的矩阵:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

其一共有 $123, 132, 213, 231, 312, 321$ 六种不同的置换，并且:

$$\tau(123) = 0,\ \tau(132) = 1,\ \tau(213) = 1,\ \tau(231) = 2,\ \tau(312) = 2,\ \tau(321) = 3$$

从而我们有:

$$\begin{aligned}\det(A) =& (-1)^{\tau(123)} a_{11}a_{22}a_{33} + (-1)^{\tau(132)} a_{11}a_{23}a_{32} + (-1)^{\tau(213)} a_{12}a_{21}a_{33} \\ &+ (-1)^{\tau(231)} a_{12}a_{23}a_{31} + (-1)^{\tau(312)} a_{13}a_{21}a_{32} + (-1)^{\tau(321)} a_{13}a_{22}a_{31} \\ =& a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31}\end{aligned}$$

例子-3×3 的情况 (II)

这也就是书本中常见的三阶行列式计算的记忆方式:

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

定理 174 **[Uniqueness].**

$\det(A)$ 是唯一一个 $\underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n\text{个}} \to \mathbb{R}^n$ 的函数满足下述三个性质:

- $D(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1$
- $D(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots \mathbf{a}_n) = -D(\mathbf{a}_1, \cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots \mathbf{a}_n)$
- $D(\mathbf{a}_1, \cdots, c\mathbf{a}_i + d\mathbf{a}_i', \cdots, \mathbf{a}_n) = cD(\mathbf{a}_1, \cdots, \mathbf{a}_i, \cdots, \mathbf{a}_n) + dD(\mathbf{a}_1, \cdots, \mathbf{a}_i', \cdots, \mathbf{a}_n)$

证明. 我们前面已经证明，满足这三个性质的函数具备如下的性质:

$$D(\mathbf{a}_1, \cdots, \mathbf{a}_n) = \sum_{\sigma \in \mathrm{Perm}(n)} (-1)^{k_\sigma} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}$$

根据前面的讨论，k_σ 可以用 $\tau(\sigma)$ 来替代，从而:

$$D(\mathbf{a}_1, \cdots, \mathbf{a}_n) = \sum_{\sigma \in \mathrm{Perm}(n)} (-1)^{\tau(\sigma)} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} = \det(A)$$

这就完成了唯一性的证明。 $\square$

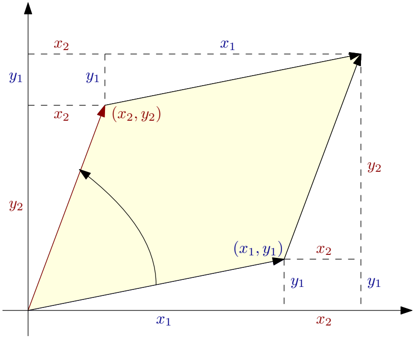
行列式的展开

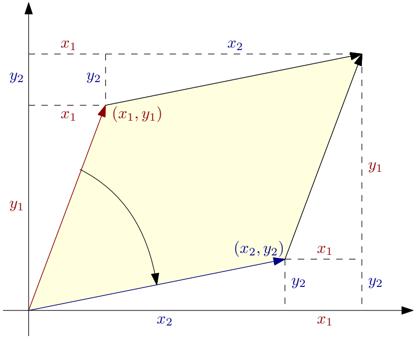
对于一个 2 × 2 的矩阵:

$$A = \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}$$

一共有 12, 21 两种不同的置换，从而我们有:

$$\det(A) = (-1)^{\tau(12)} x_1 y_2 + (-1)^{\tau(21)} x_2 y_1 = x_1 y_2 - x_2 y_1$$





这可以看成是 det(A)代数余子式展开 (Cofactor Expansion)的一个特殊情况。

令 A 是一个 $n \times n$ 的矩阵，并且 $n \geqslant 2$。对于任意的 $i, j \in [n]$，我们定义：

M_{ij} 是将 A 第 i 行第 j 列的元素删去后得到的 $n-1 \times n-1$ 的矩阵。

定义 175 **[Cofactor].**

下列数：

$$C_{ij} = (-1)^{i+j} \det(M_{ij})$$

称之为代数余子式 (Cofactor)，特别的 $\det(M_{ij})$ 称之为余子式。

定理 176.

1. $\det(\begin{bmatrix} a \end{bmatrix}) = a$.
2. 对于任意的 $n \geqslant 2$，并且 A 是一个 $n \times n$ 的矩阵:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

我们有:

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

定理176的证明 (I)

我们对其第一列展开:

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ 0 & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} 0 & \cdots & a_{1n} \\ a_{12} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{vmatrix} + \cdots + \begin{vmatrix} 0 & \cdots & a_{1n} \\ 0 & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

所以我们只需要证明，对任意的 $i \in [n]$ 有:

$$\begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = a_{i1}C_{i1} = (-1)^{i+1}a_{i1}\det(M_{i1}) = (-1)^{i-1}a_{i1}\det(M_{i1})$$

定理176的证明 (II)

对其第 i 行向上交换 $i-1$ 次可得:

$$\begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{i-1,2} & \cdots & a_{i-1,n} \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = (-1) \begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ 0 & a_{i-1,2} & \cdots & a_{i-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} \cdots = (-1)^{i-1} \begin{vmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \\ 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{i-1,2} & \cdots & a_{i-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

所以我们只要证明:

$$\begin{vmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \\ 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{i-1,2} & \cdots & a_{i-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = a_{i1} \det(M_{i1})$$

通过行列式的性质，我们注意到：

$$\begin{vmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \\ 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{i-1,2} & \cdots & a_{i-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{i1} & 0 & \cdots & 0 \\ 0 & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{i-1,2} & \cdots & a_{i-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{i1} & O_{1\times(n-1)} \\ O_{(n-1)\times 1} & M_{i1} \end{vmatrix}$$

也就是要证明：

$$\begin{vmatrix} a_{i1} & O_{1\times(n-1)} \\ O_{(n-1)\times 1} & M_{i1} \end{vmatrix} = a\det(M_{i1})$$

上海师范大学 Shanghai Normal University

引理 177.

我们有： $\det(A) = \begin{vmatrix} a & O \\ O & B \end{vmatrix} = a\det(B)$

证明.

$$\det(A) = \begin{vmatrix} a & O \\ O & B \end{vmatrix} = \sum_{\sigma\in \mathrm{Perm}(n)} (-1)^{\tau(\sigma)} a_{\sigma(1)1}\cdots a_{\sigma(n)n}$$

$$= \sum_{\substack{\sigma\in \mathrm{Perm}(n)\\ \sigma(1)=1}} (-1)^{\tau(\sigma)} a\, a_{\sigma(2)2}\cdots a_{\sigma(n)n}$$

$$= a \sum_{\substack{\sigma\in \mathrm{Perm}(n)\\ \sigma(1)=1}} (-1)^{\tau(\sigma)} b_{\sigma(2)-1,1}\cdots b_{\sigma(n)-1,n}$$

$$= a \sum_{\delta\in \mathrm{Perm}(n-1)} (-1)^{\tau(\delta)} b_{\delta(1),1}\cdots b_{\delta(n-1),n-1} = a\det(B)$$

□

定理 178.

1. $\det(\begin{bmatrix} a \end{bmatrix}) = a.$
2. 对于任意的 $n \geqslant 2$，并且 A 是一个 $n \times n$ 的矩阵:

$$\begin{bmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{11} & \cdots & a_{ni} & \cdots & a_{nn} \end{bmatrix}$$

我们有:

$$\det(A) = a_{1i}C_{1i} + \cdots + a_{1i}C_{1i}$$

思路

只需要将第 j 列做 $j-1$ 次列交换换到第 1 列即可。

定理178的证明. 通过 $i-1$ 次列交换，我们可以将第 i 列逐步换到第 1 列，即:

$$\begin{bmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{11} & \cdots & a_{ni} & \cdots & a_{nn} \end{bmatrix} \to \cdots \to \begin{bmatrix} a_{1i} & a_{11} & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ a_{ni} & a_{11} & \cdots & \cdots & a_{nn} \end{bmatrix}$$

从而:

$$\begin{vmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{11} & \cdots & a_{ni} & \cdots & a_{nn} \end{vmatrix} = (-1)^{i-1} \begin{vmatrix} a_{1i} & a_{11} & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ a_{ni} & a_{11} & \cdots & \cdots & a_{nn} \end{vmatrix}$$

$$= (-1)^{i-1}(a_{1i}C'_{11} + a_{2i}C'_{12} + \cdots + a_{ni}C'_{1n})$$

$$= (-1)^{i-1}(a_{1i}(-1)^{1+1}\det(M_{1i}) + a_{2i}(-1)^{2+1}\det(M_{2i}) +$$

$$\cdots + a_{ni}(-1)^{n+1}\det(M_{ni}))$$

$$= (-1)^{1+i}a_{1i}\det(M_{1i}) + (-1)^{2+i}a_{2i}\det(M_{2i}) + \cdots + (-1)^{n+i}a_{ni}\det(M_{ni})$$

$$= a_{1i}C_{1i} + \cdots + a_{1i}C_{1i}$$

定理 179.

1. $\det(\begin{bmatrix} a \end{bmatrix}) = a$.
2. 对于任意的 $n \geqslant 2$，并且 A 是一个 $n \times n$ 的矩阵:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

我们有:

$$\det(A) = a_{i1}C_{i1} + \cdots + a_{in}C_{in}$$

证明. 只要注意到 $\det(A) = \det(A^{\mathsf{T}})$ 即可。 □

计算下列行列式:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 0 & 0 & 4 \end{vmatrix}$$

证明. 我们根据其第一列展开:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 0 & 0 & 4 \end{vmatrix} = 1\begin{vmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{vmatrix} - 1\begin{vmatrix} 1 & 1 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{vmatrix} + 1\begin{vmatrix} 1 & 1 & 1 \\ 2 & 0 & 0 \\ 0 & 0 & 4 \end{vmatrix} - 1\begin{vmatrix} 1 & 1 & 1 \\ 2 & 0 & 0 \\ 0 & 3 & 0 \end{vmatrix}$$

$$= 24 - 12 - 8 - 6 = -2$$

□

我们也可以按第二列进行展开:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 0 & 0 & 4 \end{vmatrix} = (-1)^{2+1} 1 \begin{vmatrix} 1 & 0 & 0 \\ 1 & 3 & 0 \\ 1 & 0 & 4 \end{vmatrix} + (-1)^{2+2} 2 \begin{vmatrix} 1 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 4 \end{vmatrix} = -12 + 2 * 5 = -2$$

也可以选择按第第二行展开:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 0 & 0 & 4 \end{vmatrix} = (-1)^{1+2} 1 \begin{vmatrix} 1 & 1 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{vmatrix} + (-1)^{2+2} 2 \begin{vmatrix} 1 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 4 \end{vmatrix} = -12 + 2 * 5 = -2$$

记下列的矩阵为 D_n:

$$D_n = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 2 & -1 \\ 0 & \cdots & 0 & 0 & -1 & 2 \end{bmatrix}$$

计算 $\det(D_n)$

证明. 我们按第一行展开:

$$\det(D_n) = 2\begin{vmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdots & 0 & -1 & 2 & -1 \\ \cdots & 0 & 0 & -1 & 2 \end{vmatrix} + (-1)^{1+2}(-1)\begin{vmatrix} -1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{vmatrix}$$

$$= 2\begin{vmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdots & 0 & -1 & 2 & -1 \\ \cdots & 0 & 0 & -1 & 2 \end{vmatrix} + (-1)^{1+2}(-1)(-1)\begin{vmatrix} 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 2 \end{vmatrix}$$

$$= 2\det(D_{n-1}) - \det(D_{n-2})$$

$\square$

一些例子 (II)-续

我们有:

$$\det(D_n) = 2\det(D_{n-1}) - \det(D_{n-2})$$

注意到:

$$\det(D_1) = \begin{vmatrix} 2 \end{vmatrix} = 2, \quad \det(D_2) = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 4 - 1 = 3$$

我们可以根据上述递推式得到:

$$\det(D_n) = n + 1$$

□

下列是 Vandermonde 矩阵:

$$V_n = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}$$

证明其行列式值为:

$$\det(V_n) = \Pi_{1\leqslant j\leqslant i\leqslant n}(x_j - x_i)$$

证明. 我们通过归纳法来计算 $\det(V_n)$。

- $n = 2$ 时，我们有:

$$\det(V_2) = \begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix} = x_2 - x_1$$

考虑 n 的时候，我们的策略是先用行变换将第一列除了第一个元素全部变为 0，再将行列式按第一列展开，即：

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-2} & x_2^{n-2} & \cdots & x_n^{n-2} \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix} \to \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-2} & x_2^{n-2} & \cdots & x_n^{n-2} \\ 0 & x_2^{n-2}(x_2-x_1) & \cdots & x_2^{n-2}(x_n-x_1) \end{bmatrix}$$

$$\to \cdots \to \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2-x_1 & \cdots & x_n-x_1 \\ 0 & x_2(x_2-x_1) & \cdots & x_n(x_n-x_1) \\ 0 & \vdots & \ddots & \vdots \\ 0 & x_2^{n-2}(x_2-x_1) & \cdots & x_2^{n-2}(x_n-x_1) \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \\ 0 & x_2(x_2 - x_1) & \cdots & x_n(x_n - x_1) \\ 0 & \vdots & \ddots & \vdots \\ 0 & x_2^{n-2}(x_2 - x_1) & \cdots & x_2^{n-2}(x_n - x_1) \end{bmatrix}$$

从而我们有:

$$\det(V_n) = \begin{vmatrix} x_2 - x_1 & \cdots & x_n - x_1 \\ x_2(x_2 - x_1) & \cdots & x_n(x_n - x_1) \\ \vdots & \ddots & \vdots \\ x_2^{n-2}(x_2 - x_1) & \cdots & x_2^{n-2}(x_n - x_1) \end{vmatrix} = (x_2 - x_1) \cdots (x_n - x_1) \begin{vmatrix} 1 & \cdots & 1 \\ x_2 & \cdots & x_n \\ \vdots & \ddots & \vdots \\ x_2^{n-2} & \cdots & x_n^{n-2} \end{vmatrix}$$

由归纳假设我们有:

$$\begin{vmatrix} 1 & \cdots & 1 \\ x_2 & \cdots & x_n \\ \vdots & \ddots & \vdots \\ x_2^{n-2} & \cdots & x_n^{n-2} \end{vmatrix} = \Pi_{2\leqslant j\leqslant i\leqslant n}(x_j - x_i)$$

从而我们有:

$$\det(V_n) = (x_2 - x_1)\cdots(x_n - x_1)\Pi_{2\leqslant j\leqslant i\leqslant n}(x_j - x_i) = \Pi_{1\leqslant j\leqslant i\leqslant n}(x_j - x_i)$$

□

上海师范大学 Shanghai Normal University

我们已经阐述了克拉默法则 (Cramer's Rule)在解方程中的作用。现在让我们来用它解决一下 A^{-1}。注意到 $A^{-1} = \begin{bmatrix}\mathbf{x}_1 \cdots \mathbf{x}_n\end{bmatrix}$ 等价于如下 n 个方程组:

$$A\mathbf{x}_1 = \mathbf{e}_1, A\mathbf{x}_2 = \mathbf{e}_2, \cdots, A\mathbf{x}_n = \mathbf{e}_n$$

特别的:

$$A\mathbf{x}_1 = A\begin{bmatrix} x_{11} \\ \vdots \\ x_{n1} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}$$

从而，通过 Cramer's Rule 我们有对任意的 $i \in [n]$:

$$x_{i1} = \frac{1}{\det(A)} \begin{vmatrix} a_{11} & \cdots & a_{1,i-1} & 1 & a_{1,i+1} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2,i-1} & 0 & a_{2,i+1} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n,i-1} & 0 & a_{n,i+1} & \cdots & a_{nn} \end{vmatrix} = \frac{C_{1i}}{\det(A)}$$

更一般的来说，对于 $j \in [n]$

$$A\mathbf{x}_j = A\begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

同样通过 Cramer's Rule 我们有对任意的 $i \in [n]$:

$$x_{ij} = \frac{1}{\det(A)} \begin{vmatrix} a_{11} & \cdots & a_{1,i-1} & 0 & a_{1,i+1} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{j,i-1} & 1 & a_{j,i+1} & \cdots & a_{jn} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n,i-1} & 0 & a_{n,i+1} & \cdots & a_{nn} \end{vmatrix} = \frac{C_{ji}}{\det(A)}$$

从而我们有:

$$A^{-1}=\frac{1}{\det(A)}\begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix}$$

定义 180 **[Adjugate Matrix].**

称下列矩阵:

$$A^{*}=\begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix}$$

为 A 的伴随矩阵 (Adjugate Matrix)。

定理 181.

$AA^{*}=\det(A)I$

阶段总结-关于行列式

Shanghai Normal University

- 转置矩阵、矩阵乘法和行列式的关系。
  - $\det(A) = \det(A^{\mathsf{T}})$, $\det(AB) = \det(A)\det(B)$
  - 行列式跟首元的关系:

$$\det(A) = \det(U) = p_1 p_2 \cdots p_n \text{ 或者 } \det(A) = -\det(U) = -p_1 p_2 \cdots p_n$$

- 行列式的正式定义，The Big Formula:

$$\det(A) = \sum_{\sigma \in \mathrm{Perm}(n)} (-1)^{\tau(\sigma)} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}$$

- 代数余子式，行列式按行 (列) 展开:

$$\det(A) = a_{1i} C_{1i} + \cdots + a_{1i} C_{1i}$$

- Cramer's Rule，用行列式解方程，求矩阵的逆，伴随矩阵。

特征值与特征向量 (Eigenvalues and Eigenvectors)

特征值与特征向量 (Eigenvalues and Eigenvectors)

特征值介绍

我们知道斐波那契数列：$0, 1, 1, 2, 3, 5, 8, 13, 21, \cdots,,$ 即我们可以有如下的表示：

$$F_0 = 0,\ F_1 = 1,\ F_{k+2} = F_{k+1} + F_k$$

我们设法求这个递推式，我们可以重写这个式子：

$$F_{k+2} + aF_{k+1} = c(F_{k+1} + aF_k)$$

其中：

$$ac = 1$$

$$c - a = 1$$

解上述方程得：

$$\begin{cases} a & = \frac{-1+\sqrt{5}}{2} \\ c & = \frac{1+\sqrt{5}}{2} \end{cases} \quad \text{或} \quad \begin{cases} a & = \frac{-1-\sqrt{5}}{2} \\ c & = \frac{1-\sqrt{5}}{2} \end{cases}$$

上海师范大学 Shanghai Normal University

即我们有:

$$F_{k+2}+\frac{-1+\sqrt{5}}{2}F_{k+1}=\frac{1+\sqrt{5}}{2}(F_{k+1}+\frac{-1+\sqrt{5}}{2}F_k)$$

$$F_{k+2}+\frac{-1-\sqrt{5}}{2}F_{k+1}=\frac{1-\sqrt{5}}{2}(F_{k+1}+\frac{-1-\sqrt{5}}{2}F_k)$$

从而我们有:

$$F_{k+1}+\frac{-1+\sqrt{5}}{2}F_k=(\frac{1+\sqrt{5}}{2})^k(F_1+\frac{-1+\sqrt{5}}{2}F_0)$$

$$F_{k+1}+\frac{-1-\sqrt{5}}{2}F_k=(\frac{1-\sqrt{5}}{2})^k(F_1+\frac{-1-\sqrt{5}}{2}F_0)$$

两式相减:

$$F_k=\frac{1}{\sqrt{5}}((\frac{1+\sqrt{5}}{2})^k-(\frac{1-\sqrt{5}}{2})^k)$$

斐波那契数列的矩阵形式 (I)

现在我们用矩阵得过程来探索一下。由:

$$F_{k+1} = F_k + F_{k-1}$$

我们有:

$$\begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix}$$

从而我们有:

$$\begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right)^2 \begin{bmatrix} F_{k-1} \\ F_{k-2} \end{bmatrix} = \cdots = \left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right)^k \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}$$

另一方面:

$$F_{k+1} + \frac{-1+\sqrt{5}}{2}F_k = \frac{1+\sqrt{5}}{2}\left(F_k + \frac{-1+\sqrt{5}}{2}F_{k-1}\right)$$

$$F_{k+1} + \frac{-1-\sqrt{5}}{2}F_k = \frac{1-\sqrt{5}}{2}\left(F_k + \frac{-1-\sqrt{5}}{2}F_{k-1}\right)$$

我们有:

$$\begin{bmatrix} F_{k+1} + \frac{-1+\sqrt{5}}{2}F_k \\ F_{k+1} + \frac{-1-\sqrt{5}}{2}F_k \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} F_k + \frac{-1+\sqrt{5}}{2}F_{k-1} \\ F_k + \frac{-1-\sqrt{5}}{2}F_{k-1} \end{bmatrix}$$

从而通过类似的讨论我们可以得到:

$$\begin{aligned} \begin{bmatrix} F_{k+1} + \frac{-1+\sqrt{5}}{2}F_k \\ F_{k+1} + \frac{-1-\sqrt{5}}{2}F_k \end{bmatrix} &= \left(\begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}\right)^k \begin{bmatrix} F_1 + \frac{-1+\sqrt{5}}{2}F_0 \\ F_1 + \frac{-1-\sqrt{5}}{2}F_0 \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^k & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^k \end{bmatrix} \begin{bmatrix} F_1 + \frac{-1+\sqrt{5}}{2}F_0 \\ F_1 + \frac{-1-\sqrt{5}}{2}F_0 \end{bmatrix} \end{aligned}$$

斐波那契数列的矩阵形式 (III)

而注意到:

$$\begin{bmatrix} F_1 + \frac{-1+\sqrt{5}}{2}F_0 \\ F_1 + \frac{-1-\sqrt{5}}{2}F_0 \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}, \quad \begin{bmatrix} F_{k+1} + \frac{-1+\sqrt{5}}{2}F_k \\ F_{k+1} + \frac{-1-\sqrt{5}}{2}F_k \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$$

从而我们有:

$$\begin{bmatrix} F_{k+1} + \frac{-1+\sqrt{5}}{2}F_k \\ F_{k+1} + \frac{-1-\sqrt{5}}{2}F_k \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix}$$

$$\begin{bmatrix} F_{k+1} + \frac{-1+\sqrt{5}}{2}F_k \\ F_{k+1} + \frac{-1-\sqrt{5}}{2}F_k \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} F_k + \frac{-1+\sqrt{5}}{2}F_{k-1} \\ F_k + \frac{-1-\sqrt{5}}{2}F_{k-1} \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix}$$

我们得到了:

$$\begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}\begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix}$$

令:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & \frac{-1+\sqrt{5}}{2} \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix}$$

我们有:

$$XA = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}X, \quad \text{即:} \; A = X^{-1}\begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}X$$

注意到此时:

$$X^{-1} = \begin{bmatrix} \frac{5+\sqrt{5}}{10} & \frac{5-\sqrt{5}}{10} \\ \frac{\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} \end{bmatrix}$$

令 X^{-1} 的列向量是 $\mathbf{x}_1, \mathbf{x}_2$。我们有:

$$AX^{-1} = \begin{bmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 \end{bmatrix} = X^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} \lambda\mathbf{x}_1 & \lambda\mathbf{x}_2 \end{bmatrix}$$

也就是:

$$A\mathbf{x}_1 = \lambda_1\mathbf{x}_1, \quad A\mathbf{x}_2 = \lambda_1\mathbf{x}_2$$

我们称 λ_1, λ_2 是 A 的特征值 (Eigenvalue),而 $\mathbf{x}_1, \mathbf{x}_2$ 是相应的特征向量 (Eigenvector)

$A = X^{-1}\Lambda X$ 的另一个用处:

$$
\begin{aligned}
\begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} &= \left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\right)^k \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} \\
&= \left(X^{-1}\begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} X\right)^k \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} \\
&= \left(X^{-1}\begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^k & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^k \end{bmatrix} X\right) \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}
\end{aligned}
$$

我们来尝试给出特征值和特征向量的定义:

定义 182 **[特征值和特征向量，尝试定义].**

令 A 是一个 $n \times n$ 的矩阵，λ 是一个实数，$\mathbf{x}$ 是一个非零的 n 维向量。如果:

$$A\mathbf{x} = \lambda\mathbf{x}$$

则称 λ 是 A 的特征值 (Eigenvalue)，$\mathbf{x}$ 是 λ 对应的特征向量 (Eigenvector)。

让我们关注一下斐波那契的递推矩阵:

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

我们有:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5+\sqrt{5}}{10} \\ \frac{\sqrt{5}}{5} \end{bmatrix} = \frac{1+\sqrt{5}}{2} \begin{bmatrix} \frac{5+\sqrt{5}}{10} \\ \frac{\sqrt{5}}{5} \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5-\sqrt{5}}{10} \\ -\frac{\sqrt{5}}{5} \end{bmatrix} = \frac{1-\sqrt{5}}{2} \begin{bmatrix} \frac{5-\sqrt{5}}{10} \\ -\frac{\sqrt{5}}{5} \end{bmatrix}$$

从而对于矩阵 A:

- 其有两个特征值: $\lambda_1 = \frac{1+\sqrt{5}}{2}$, $\lambda_2 = \frac{1-\sqrt{5}}{2}$, 对应的特征向量为:
$$\mathbf{x}_1 = \begin{bmatrix} \frac{5+\sqrt{5}}{10} \\ \frac{\sqrt{5}}{5} \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} \frac{5-\sqrt{5}}{10} \\ -\frac{\sqrt{5}}{5} \end{bmatrix}.$$
- 回顾之前的计算:

$$AX^{-1} = A\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = X^{-1}\Lambda$$

令 A 是一个 $n \times n$ 的矩阵，将其看作一个函数$f_A : \mathbb{R}^n \to \mathbb{R}^n$，即：

$$f_A(\mathbf{x}) = A\mathbf{x}$$

则当 $\mathbf{x}$ 是 A 的特征向量，即 $A\mathbf{x} = \lambda\mathbf{x}$ 时我们有：

$$f_A(\mathbf{x}) = \lambda\mathbf{x}$$

这意味着f_A 不改变 $\mathbf{x}$ 的方向。

考察矩阵:

$$A=\begin{bmatrix}0.8 & 0.3\\ 0.2 & 0.7\end{bmatrix}$$

我们有:

$$A\begin{bmatrix}0.6\\ 0.4\end{bmatrix}=\begin{bmatrix}0.6\\ 0.4\end{bmatrix},\quad A\begin{bmatrix}1\\ -1\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1\\ -1\end{bmatrix}$$

从而 $1,\frac{1}{2}$ 是其两个特征值, $\begin{bmatrix}0.6\\ 0.4\end{bmatrix},\begin{bmatrix}1\\ -1\end{bmatrix}$ 是相应的特征向量。注意到这两个向量是线性无关的，意味着任何一个向量都可以表示成其线性组合，比如:

$$\begin{bmatrix}0.8\\ 0.2\end{bmatrix}=\begin{bmatrix}0.6\\ 0.4\end{bmatrix}+0.2\begin{bmatrix}1\\ -1\end{bmatrix}$$

从而我们有:

$$A^{100}\begin{bmatrix}0.8\\ 0.2\end{bmatrix}=(1)^{100}\begin{bmatrix}0.6\\ 0.4\end{bmatrix}+0.2(\frac{1}{2})^{100}\begin{bmatrix}1\\ -1\end{bmatrix}\approx\begin{bmatrix}0.6\\ 0.4\end{bmatrix}$$

$\lambda = 0$ 的情况

现在我们来讨论 $\lambda = 0$ 的情况，假设 $n \times n$ 的矩阵 A 存在特征值 0，即存在一个非零的 $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ 满足:

$$A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$$

显然这说明 $\dim(\mathrm{N}(A)) \geqslant 1$.

引理 183.

下述是等价的:

1. A 具有特征值 0。
2. $\mathrm{N}(A) \neq \mathbb{Z}_0 (= \{\mathbf{0}\})$。
3. $\mathrm{rank}(A) < n$，即 A 是奇异的 (singular).
4. $\det(A) = 0$.

对角矩阵的特征值

现在我们来考察对角矩阵:

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

显然我们有:

$$A\mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad \text{这里}\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

从而 Λ 的特征值为$\lambda_1, \cdots, \lambda_n$，每个$\lambda_i$对应的特征向量为$\mathbf{e}_i$。

现在我们再来回顾投影矩阵，令 A 是一个 $m \times n$ 且 $\mathrm{rank}(A) = n$ 的矩阵，则投影至 $C(A)$ 的投影矩阵 P 可以表示为：

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

注意到 $P\mathbf{x}$ 是 $\mathbf{x}$ 到 $C(A)$ 上的投影，所以 $P\mathbf{x} = \lambda\mathbf{x}$ 只有两种情况：

1. $\mathbf{x} \in C(A)$，此时 $P\mathbf{x} = 1\mathbf{x} = \mathbf{x}$
2. $\mathbf{x} \in C(A)^{\perp}$ 即 $\mathbf{x} \in N(A^{\top})$，此时 $P\mathbf{x} = 0\mathbf{x} = \mathbf{0}$

回顾特征值，实际上就是：

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad (A - \lambda I)\mathbf{x} = \mathbf{0} \text{ 有非零解}$$

定理 184.

令 A 是 $n \times n$ 的矩阵，λ 是 A 的特征值当且仅当：

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

证明.

$$\begin{aligned} \lambda\text{是 } A \text{ 的特征值} &\iff A\mathbf{x} = \lambda\mathbf{x} \text{ 有非零解} \\ &\iff (A - \lambda I)\mathbf{x} = \mathbf{0} \text{ 有非零解} \\ &\iff \mathrm{rank}(A - \lambda I) < n \\ &\iff \det(A - \lambda I) = 0 \end{aligned}$$

□

我们将 λ 看成一个变量，则对于 $n \times n$ 的矩阵 A:

$$f(\lambda) = \det(A - \lambda I_n)$$

是一个关于 λ 的 n 次多项式。(为什么？)

定义 185 **[特征多项式 (Characteristic Polynomial)].**

令 A 是 $n \times n$ 的矩阵，则称 $f(\lambda) = \det(A - \lambda I_n)$ 是 A 的特征多项式 (Characteristic Polynomial).

我们首先来看下斐波那契矩阵:

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

其特征多项式为:

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

该多项式的两个零点为:

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \quad \lambda_2 = \frac{1-\sqrt{5}}{2}$$

一些例子 (II)

对角矩阵:

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

的特征多项式为:

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$$

消元法会改变特征值!

我们来考察一下消元法:

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \implies U = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$$

我们有:

- $f_A(\lambda) = (1-\lambda)(6-\lambda) - 6 = \lambda(\lambda - 7) \implies \lambda_1 = 0,\ \lambda_2 = 7.$
- $f_U(\lambda) = (1-\lambda)(0-\lambda) - 0 = \lambda(\lambda - 1) \implies \lambda_1 = 0,\ \lambda_2 = 1.$

迹 (Trace)

我们定义 A 的主对角线上元素之和称为 A 的迹 (Trace),即:

$$\text{trace} = a_{11} + a_{22} + \cdots + a_{nn}$$

则我们有:

- $\lambda_1 + \lambda_2 + \cdots + \lambda_n = \text{trace} = a_{11} + a_{22} + \cdots + a_{nn}$。
- $\det(A) = \lambda_1\lambda_2\cdots\lambda_n$

最后我们来关注一下旋转矩阵：

$$Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{bmatrix}$$

其特征多项式为：

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

其两个零点为：

$$\lambda_1 = i,\ \lambda_2 = -i$$

特征值和特征向量是复数!

Q 有两个复数特征值:

$$\lambda_1 = i,\ \lambda_2 = -i$$

其对应的特征向量为:

$$Q\begin{bmatrix}1\\ i\end{bmatrix} = -i\begin{bmatrix}1\\ i\end{bmatrix}$$

$$Q\begin{bmatrix}i\\ 1\end{bmatrix} = i\begin{bmatrix}i\\ 1\end{bmatrix}$$

也是在$\mathbb{C}^2$上的向量!

定理 186 **[Fundamental Theorem of Algebra].**

对于任一的 $a_1, \cdots, a_n \in \mathbb{C}$，我们有

$$x^n + a_1 x^{n-1} + \cdots a_{n-1} x + a_n = (x - b_1)(x - b_2) \cdots (x - b_n)$$

这里 $b_1, \cdots b_n \in \mathbb{C}$，即任一单变元的 n 次复系数多项式恰好有 n 个复数根（可重复）。

定义 187 **[特征值和特征向量].**

令 A 是一个 $n \times n$ 的矩阵，$\lambda \in \mathbb{C}$，$\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$。如果：

$$A\mathbf{x} = \lambda\mathbf{x}$$

则称 λ 是 A 的特征值 (Eigenvalue)，$\mathbf{x}$ 是 λ 对应的特征向量 (Eigenvector)。

推论 188.

令 A 是一个 $n \times n$ 的矩阵，则在计算重根的意义下，A 恰好有 n 个 $\mathbb{C}$ 中的特征值。

例 189.

考察矩阵

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

其特征多项式为 $f(\lambda) = \lambda^2(\lambda^2 + 1)$，注意到 $f(\lambda) = 0$ 一共有 4 个解：

$$\lambda_1 = \lambda_2 = 0, \ \lambda_3 = i, \ \lambda_4 = -i$$

这里 0 是一个重根，在计算重复次数以后可以得到其一共有 4 个特征值。

在一般意义下:

- 矩阵可以有复数。
- 向量空间是可以通过复数扩充的，即数乘允许 $c \in \mathbb{C}$ 进行运算。

但是在我们这堂课中，为了保持简单:

- 我们只考虑实数的矩阵。
- 但是我们依旧会讨论到复数的特征值和特征向量。

我们可以如下计算特征值和特征向量:

1. 计算 A 的特征多项式，即 $A - \lambda I$ 的行列式 $\det(A - \lambda I)$.
2. 计算 $\det(A - \lambda I)$ 的根，我们一共会得到 n 个特征值（可能重复）。这使得 $A - \lambda I$ 变成一个奇异矩阵 (singular).
3. 对每个 λ，通过解方程 $(A - \lambda I)\mathbf{x} = \mathbf{0}$ 来获取 λ 对应的特征向量 $\mathbf{x}$.

特征值与特征向量 (Eigenvalues and Eigenvectors)

对角化矩阵

$$
\begin{aligned}
\begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} &= \left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right)^k \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} \\
&= \left(X' \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} (X')^{-1} \right)^k \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} \\
&= \left(X' \begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^k & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^k \end{bmatrix} (X')^{-1} \right) \begin{bmatrix} F_1 \\ F_0 \end{bmatrix}
\end{aligned}
$$

中间的对角矩阵 Λ 极大的简化了我们对 $\left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right)^k$ 的计算，我们把这过程称为对角化 (Diagonalization)。

定义 190 **[Diagonalization].**

一个方阵 A 是可对角化的 (Diagonalizable)，如果其存在一个可逆矩阵 X 和对角矩阵 Λ 使得:

$$A = X\Lambda X^{-1}, \text{ 或者等价的 } \Lambda = X^{-1}AX$$

推论 191.

若 $A = X\Lambda X^{-1}$，其中 $X = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，则对于任意的 $k \geqslant 1$ 我们有:

$$A^k = X\Lambda^k X^{-1} = X \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix} X^{-1}$$

我们也称这是矩阵的谱分解。

对角化的一个直观理解

$$AX = X\Lambda$$

1. X 的每一列都应该是 A 的特征向量。
2. X 是可逆的，所以 A 需要有 n 个线性无关的特征向量。

定理 192.

$n \times n$ 的矩阵可对角化当且仅当 A 有n 个线性无关的特征向量。

定理224的证明 (I)

假设 A 有 n 个特征值 $\lambda_1, \cdots, \lambda_n$(可以重复)，对应的特征向量 $\mathbf{x}_1, \cdots, \mathbf{x}_n$ 是线性无关的:

$$A\mathbf{x}_1 = \lambda_1\mathbf{x}_1, \cdots, A\mathbf{x}_n = \lambda_n\mathbf{x}_n$$

定义下列矩阵:

$$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n) = \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix}, \quad X = \begin{bmatrix} \mathbf{x}_1, \cdots, \mathbf{x}_n \end{bmatrix}$$

则我们有:

- $AX = A\begin{bmatrix} \mathbf{x}_1, \cdots, \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{x}_1, \cdots, A\mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{x}_1, \cdots, \lambda_n\mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1, \cdots, \mathbf{x}_n \end{bmatrix}\Lambda = X\Lambda.$
- $\mathrm{rank}(X) = n$，从而 X 可逆，即:

$$A = X\Lambda X^{-1}$$

另一方面，如果:

$$A = X\Lambda X^{-1}$$

定义下列矩阵:

$$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n) = \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix}, \quad X = \begin{bmatrix} \mathbf{x}_1, \cdots, \mathbf{x}_n \end{bmatrix}$$

我们自然容易验证:

- 对于每个 $\mathbf{x}_i$，有 $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$。
- $\mathbf{x}_1, \cdots, \mathbf{x}_n$ 是线性无关的。

□

上海师范大学 Shanghai Normal University

考察矩阵:

$$A=\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}$$

其特征多项式为 $f_A(\lambda)=\lambda^2-4\lambda+3=(\lambda-1)(\lambda-3)$，分别计算其对应的特征向量为:

- $\lambda=1$ 时解方程 $(A-I)\mathbf{x}=\mathbf{0}$ 得一个其特征向量为 $\begin{bmatrix}1\\1\end{bmatrix}$.
- $\lambda=3$ 时解方程 $(A-3I)\mathbf{x}=\mathbf{0}$ 得一个其特征向量为 $\begin{bmatrix}1\\-1\end{bmatrix}$

从而我们有:

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

$$A^k=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1^k & 0\\ 0 & 3^k\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1+3^k & 1-3^k\\ 1-3^k & 1+3^k\end{bmatrix}$$

下列矩阵：

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

不可对角化。

- A 的特征值为 $\lambda = 0$，只有一个特征向量。
- B 的特征值为 $\lambda = 0$，只有一个特征向量。
- C 的特征值为 $\lambda = 1$，只有一个特征向量。

可逆和对角化没有关系!

- 矩阵是否可逆取决于是否有零特征值 ($\lambda = 0$)。
- 矩阵是否可对角化取决于是否有 n 个特征向量。

定理 193.

令 $\mathbf{x}_1, \cdots, \mathbf{x}_k$ 是 A 的 k 个特征向量，并且其对应的特征值两两不相同。则 $\mathbf{x}_1, \cdots, \mathbf{x}_k$ 是线性无关的。

推论 194.

如果 $n \times n$ 的矩阵 A 有 n 个不同的特征值，则 A 是可对角化的。

定理的直观理解

$$A\mathbf{x} = \lambda\mathbf{x} = c_1\lambda\mathbf{x}_1 + \cdots c_k\lambda\mathbf{x}_k$$

$$A(c_1\mathbf{x}_1 + \cdots c_k\mathbf{x}_k) = c_1\lambda_1\mathbf{x}_1 + \cdots c_k\lambda_k\mathbf{x}_k$$

当 λ 互不相等的时候，上述两个式子不可能相等。

我们对 k 做归纳法，$k=1$ 的时候是显然的。

假设 $k \geqslant 2$，并且:

$$c_1\mathbf{x}_1+\cdots+c_k\mathbf{x}_k=\mathbf{0}$$

我们需要证明 $c_1=c_2=\cdots=c_k=0$，反设结论不成立，即存在 $c_i \neq 0$，即:

$$\mathbf{x}_i=\sum_{j\in[k]\setminus\{i\}}-\frac{c_j}{c_i}\mathbf{x}_j$$

从而我们有:

$$\lambda_i\mathbf{x}_i=A\mathbf{x}_i=\sum_{j\in[k]\setminus\{i\}}-\frac{c_j}{c_i}A\mathbf{x}_j=\sum_{j\in[k]\setminus\{i\}}-\frac{c_j}{c_i}\lambda_j\mathbf{x}_j$$

$$\lambda_i \mathbf{x}_i = \sum_{j\in[k]\setminus\{i\}} -\frac{c_j}{c_i}\lambda_j \mathbf{x}_j$$

注意到，由归纳假设:

$$\mathbf{x}_1, \cdots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1} \cdots \mathbf{x}_k \text{ 是线性无关的}$$

- 如果 $\lambda_i = 0$，我们有 $\frac{c_j}{c_i}\lambda_j = 0$，从而 $\frac{c_j}{c_i} = 0$，$\mathbf{x}_i = \mathbf{0}$，矛盾。
- 如果 $\lambda_i \neq 0$，我们有:

$$\mathbf{x}_i = \sum_{j\in[k]\setminus\{i\}} -\frac{c_j}{c_i}\mathbf{x}_j \text{ 和 } \mathbf{x}_i = \sum_{j\in[k]\setminus\{i\}} -\frac{c_j}{c_i}\frac{\lambda_j}{\lambda_i}\mathbf{x}_j$$

注意到 $\lambda_j \neq \lambda_i$，所以存在不全为 0 的 $d_1, \cdots, d_{i-1}, d_{i+1}, \cdots, d_k$ 满足:

$$d_1\mathbf{x}_1 + \cdots + d_{i-1}\mathbf{x}_{i-1} + d_{i-1}\mathbf{x}_{i-1} + \cdots + d_k\mathbf{x}_k = \mathbf{0}$$

与 $\mathbf{x}_1, \cdots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1} \cdots \mathbf{x}_k$ 是线性无关的矛盾。

$\square$

让我们再来考虑一下对角化的形式：

$$A = X\Lambda X^{-1}$$

当特征向量组成的矩阵 X 改变的时候，我们得到了无数个不同的 A。

- 但这些矩阵 A 的特征值都是相同的。

更一般的，对于

$$A = BCB^{-1}$$

即使 C 不可以对角化，我们也可以得到跟 C 具有相同特征值的一大类矩阵。

我们称这样一个关系叫作相似 (similar)。

定义 195 **[相似矩阵（Similar Matrix].**

给定矩阵 A, B，如果存在可逆矩阵 P，使得:

$$A = PBP^{-1}$$

则称矩阵 A 和 B 是相似的。

引理 196.

相似矩阵 A 和 B 的特征值相同。

证明. 假设 $B\mathbf{x} = \lambda\mathbf{x}$，则我们有:

$$A(P\mathbf{x}) = PBP^{-1}(P\mathbf{x}) = PB\mathbf{x} = \lambda P\mathbf{x}$$

$\square$

考察如下两个矩阵：

$$Q = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},\ Q' = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ 0 & 2 \end{bmatrix}$$

Q 和 Q' 是相似的，考虑如下两组基：

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\ \text{和} \mathbf{f}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ \mathbf{f}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

给定 $\mathbf{u} = (5,4) = 5\mathbf{e}_1 + 4\mathbf{e}_2$ 和 $\mathbf{v} = (5,9) = 5\mathbf{f}_1 + 4\mathbf{f}_2$，我们有：

- $Q\mathbf{u} = Q\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}\begin{bmatrix} 5 \\ 4 \end{bmatrix} = \begin{bmatrix} 2\mathbf{e}_1 & 3\mathbf{e}_2 \end{bmatrix}\begin{bmatrix} 5 \\ 4 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}\begin{bmatrix} 10 \\ 12 \end{bmatrix}$
- $Q'\mathbf{v} = Q'\begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 \end{bmatrix}\begin{bmatrix} 5 \\ 4 \end{bmatrix} = \begin{bmatrix} 2\mathbf{f}_1 & 3\mathbf{f}_2 \end{bmatrix}\begin{bmatrix} 5 \\ 4 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 \end{bmatrix}\begin{bmatrix} 10 \\ 12 \end{bmatrix}$

两个相似矩阵本质上是在不同基下的相同线性变换。

阶段总结

- 特征值和特征向量。
- 求特征值和特征向量的方法。
- 对角化矩阵。
- 相似矩阵的概念。

对称矩阵和正定矩阵 (Symmetric Matrices and Positive Definite Matrices)

对称矩阵和正定矩阵 (Symmetric Matrices and Positive Definite Matrices)

对称矩阵

我们称一个方阵是对称 (Symmetric)的，如果其满足：

$$A = A^{\mathsf{T}}$$

例 197.

下述矩阵都是对称的：

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 5 & 6 \\ 3 & 5 & 1 & 7 \\ 4 & 6 & 7 & 8 \end{bmatrix}$$

对称矩阵的对角化?

让我们思考一下如果一个对称矩阵 S 可以对角化会发生什么? 假设其可以对角化为:

$$S = X\Lambda X^{-1}$$

那我们有:

$$S^\top = (X\Lambda X^{-1})^\top = (X^\top)^{-1}\Lambda^\top X^\top = (X^\top)^{-1}\Lambda X^\top = S = X\Lambda X^{-1}$$

一个理想的状况是 $X^\top = X^{-1}$, 即$X^\top X = I$。事实上也正是如此, 我们将证明对于对称矩阵:

1. 特征值是实数。
2. 不同特征值的特征向量是正交的。

特征值是实数!

我们将首先证明对于对称矩阵 S，其特征值是实数。

定理 198.

所有实对称矩阵的特征值都是实数。

我们还可以证明一个更强的版本:

定理 199.

所有实对称矩阵的特征向量是实数，并且每个特征值都有一个对应的实特征向量。

令 $x \in \mathbb{C}$，则我们有存在 $a, b \in \mathbb{R}$, 使得:

$$x = a + bi$$

我们定义 x 的共轭复数 (complex conjugate) 为:

$$\bar{x} = a - bi$$

引理 200.

给定复数 $x, y \in \mathbb{C}$，我们有:

- 如果 $\bar{x}x = 0$，则 $x = 0$.
- $\overline{x + y} = \bar{x} + \bar{y}$.
- $\overline{xy} = \bar{x}\bar{y}$.

定义 201 **[共轭矩阵 (Conjugate Matrix)].**

对于一个矩阵 A，我们定义其共轭矩阵 $\bar{A}$ 为：

$$\bar{A}(i,j) = \overline{A(i,j)}$$

引理 202.

令 A 是一个 $m \times n$ 的复矩阵：

1. 对任意的 $\lambda \in \mathbb{C}$，我们有：$\overline{\lambda A} = \bar{\lambda}\bar{A}$.
2. 对任意的复向量 $\mathbf{x} \in \mathbb{C}^n$, 我们有：$\overline{A\mathbf{x}} = \bar{A}\bar{\mathbf{x}}$.

引理 203.

令 $\mathbf{x} \in \mathbb{C}^n$，如果 $\bar{\mathbf{x}}^\top \mathbf{x} = 0$, 则 $\mathbf{x} = \mathbf{0}$.

定理199的证明. 令 S 是 $n \times n$ 的实对称矩阵，$\lambda \in \mathbb{C}$ 是 S 的特征值，$\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ 是 λ 对应的特征向量。我们有:

$$
\begin{aligned}
& S\mathbf{x} = \lambda\mathbf{x} \\
\Longrightarrow & \bar{S}\bar{\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}} \\
\Longrightarrow & \bar{\mathbf{x}}^T\bar{S}^T = \bar{\lambda}\bar{\mathbf{x}}^T \\
\Longrightarrow & \bar{\mathbf{x}}^T S = \bar{\lambda}\bar{\mathbf{x}}^T \qquad (S \text{ 是实对称的，从而 } \bar{S}^T = S) \\
\Longrightarrow & \bar{\mathbf{x}}^T S\mathbf{x} = \bar{\lambda}\bar{\mathbf{x}}^T\mathbf{x} \\
\Longrightarrow & \lambda\bar{\mathbf{x}}^T\mathbf{x} = \bar{\lambda}\bar{\mathbf{x}}^T\mathbf{x} \\
\Longrightarrow & (\lambda - \bar{\lambda})\bar{\mathbf{x}}^T\mathbf{x} = 0
\end{aligned}
$$

显然由于 $\mathbf{x} \neq \mathbf{0}$，从而 $\bar{\mathbf{x}}^T\mathbf{x} \neq 0$，因此 $\lambda - \bar{\lambda} = 0$，即 λ 是实数。 □

定理199的证明续. 假设 $\mathbf{x}_j = a_j + b_j i$，即：

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 i \\ \vdots \\ a_n + b_n i \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + i \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \mathbf{a} + i\mathbf{b}$$

由于 $S\mathbf{x} = \lambda\mathbf{x}$，我们有：

$$S\mathbf{x} = S(\mathbf{a} + i\mathbf{b}) = \lambda(\mathbf{a} + i\mathbf{b})$$

由于 S 是实对称矩阵，λ 是实数，我们有：

$$S\mathbf{a} = \lambda\mathbf{a},\ S\mathbf{b} = \lambda\mathbf{b}$$

由于 $\mathbf{x} \neq \mathbf{0}$，$\mathbf{a}$ 和 $\mathbf{b}$ 至少有一个是 S 的特征向量。 $\square$

定理 204.

对于一个实对称矩阵 S，如果 λ_1 和 λ_2 是 S 的两个不同的特征值，$\mathbf{x}_1$ 和 $\mathbf{x}_2$ 是 λ_1 和 λ_2 对应的特征向量，则 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 是正交的。

证明. 由假设我们有：

$$S\mathbf{x}_1 = \lambda_1\mathbf{x}_1, \quad S\mathbf{x}_2 = \lambda_2\mathbf{x}_2$$

从而：

$$\begin{aligned}
\lambda_1(\mathbf{x}_1 \cdot \mathbf{x}_2) &= (\lambda_1\mathbf{x}_1) \cdot \mathbf{x}_2 = (S\mathbf{x}_1) \cdot \mathbf{x}_2 = (S\mathbf{x}_1)^{\mathsf{T}}\mathbf{x}_2 = \mathbf{x}_1^{\mathsf{T}}S^{\mathsf{T}}\mathbf{x}_2 \\
&= \mathbf{x}_1^{\mathsf{T}}S\mathbf{x}_2 = \mathbf{x}_1^{\mathsf{T}}(S\mathbf{x}_2) = \mathbf{x}_1^{\mathsf{T}}(\lambda_2\mathbf{x}_2) = \mathbf{x}_1 \cdot (\lambda_2\mathbf{x}_2) = \lambda_2(\mathbf{x}_1 \cdot \mathbf{x}_2)
\end{aligned}$$

由于 $\lambda_1 \neq \lambda_2$，从而 $\mathbf{x}_1 \cdot \mathbf{x}_2 = 0$，即 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 是正交的。 $\square$

由上述定理我们可以直接得到:

推论 205.

令 S 是一个 $n \times n$ 的实对称矩阵，如果 S 存在n 个两两不同的特征值，则 S 可以对角化。更精确的说，存在一个正交矩阵 Q 和对角矩阵 Λ 使得:

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = Q^{\mathsf{T}} S Q$$

这里 $\lambda_1, \cdots, \lambda_n$ 都是 S 的特征值。

而事实上，我们并不需要“n 个两两不同的特征值”这个条件。

定理 206.

令 S 是一个 $n \times n$ 的实对称矩阵，则 S 可以对角化。更精确的说，存在一个正交矩阵 Q 和对角矩阵 Λ 使得：

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = Q^{\mathsf{T}} S Q$$

这里 $\lambda_1, \cdots, \lambda_n$ 都是 S 的特征值。显然这是一个当且仅当的关系，因为另一个方向是显然成立的。

定理206的证明 (I)

令 S 是一个 $n \times n$ 的实对称矩阵，我们对 n 作归纳。

BASE：$n = 1$ 时是显然的。

INDUCTION: 现在令 $n \geqslant 2$。令 S 的一个特征值为 $\lambda_1 \in \mathbb{R}$，其对应的一个特征向量为 $\mathbf{x}_1 \in \mathbb{R}^n \setminus \{0\}$。我们不妨可以假设 $\|\mathbf{x}_1\| = 1$，否则我们可以令：

$$\mathbf{x}_1 \leftarrow \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$$

通过 Gram–Schmidt 正交化，我们可以从 $\mathbf{x}_1$ 扩展出一组标准正交基：

$$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$$

满足：

$$\mathbf{x}_i \cdot \mathbf{x}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

上海师范大学 Shanghai Normal University

定义矩阵 P:

$$P = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

不难验证 P 是正交矩阵，即 $P^{\top}P = I$。并且我们有:

$$p^{\top}\mathbf{x}_i = \mathbf{e}_i$$

从而:

$$\begin{aligned} P^{\top}SP &= P^{\top}S\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \\ &= \begin{bmatrix} P^{\top}S\mathbf{x}_1 & P^{\top}S\mathbf{x}_2 & \cdots & P^{\top}S\mathbf{x}_n \end{bmatrix} \\ &= \begin{bmatrix} P^{\top}\lambda_1\mathbf{x}_1 & P^{\top}S\mathbf{x}_2 & \cdots & P^{\top}S\mathbf{x}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 P^{\top}\mathbf{x}_1 & P^{\top}S\mathbf{x}_2 & \cdots & P^{\top}S\mathbf{x}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{e}_1 & P^{\top}S\mathbf{x}_2 & \cdots & P^{\top}S\mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{a}^{\top} \\ \mathbf{0} & B \end{bmatrix} \end{aligned}$$

这里 $\mathbf{a} \in \mathbb{R}^{n-1}$ 是一个列向量，B 是一个 $(n-1) \times (n-1)$ 的矩阵。

注意到:

$$\begin{bmatrix} \lambda_1 & \mathbf{0}^\mathsf{T} \\ \mathbf{a} & B^\mathsf{T} \end{bmatrix} = \left(\begin{bmatrix} \lambda_1 & \mathbf{a}^\mathsf{T} \\ \mathbf{0} & B \end{bmatrix} \right)^\mathsf{T} = \left(P^\mathsf{T} S P \right)^\mathsf{T} = P^\mathsf{T} S^\mathsf{T} P = P^\mathsf{T} S P = \begin{bmatrix} \lambda_1 & \mathbf{a}^\mathsf{T} \\ \mathbf{0} & B \end{bmatrix}$$

从而我们有 $\mathbf{a} = \mathbf{0}$，并且 $B = B^\mathsf{T}$. 从而 B 是一个 $(n-1) \times (n-1)$ 实对称矩阵。由归纳假设，我们可以对 B 进行对角化，即存在 $(n-1) \times (n-1)$ 的正交矩阵 P' 和对角矩阵 Λ' 使得:

$$\Lambda' = (P')^\mathsf{T} B P'$$

从而我们有:

$$P^\mathsf{T} S P = \begin{bmatrix} \lambda_1 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & B \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & (P')^\mathsf{T} B P' \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & P' \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & \Lambda' \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & (P')^\mathsf{T} \end{bmatrix}$$

令

$$M = \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & P' \end{bmatrix}$$

显然 M 是一个正交矩阵，并且我们有:

$$(PM)^\top SPM = M^\top P^\top SPM = \begin{bmatrix} \lambda_1 & \mathbf{0}^\top \\ \mathbf{0} & \Lambda' \end{bmatrix}$$

从而令 $Q = PM, \Lambda = \begin{bmatrix} \lambda_1 & \mathbf{0}^\top \\ \mathbf{0} & \Lambda' \end{bmatrix}$，我们有:

$$\Lambda = Q^\top SQ$$

并且有归纳假设，Λ 上对角线的元素都是 S 的特征值。 □

给定实对称矩阵 S，由上述定理，存在正交矩阵 Q 和对角矩阵 Λ 使得：

$$S = Q\Lambda Q^{\top}$$

令 $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$，$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$，我们有：

$$S = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^{\top} \\ \mathbf{q}_2^{\top} \\ \vdots \\ \mathbf{q}_n^{\top} \end{bmatrix} = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^{\top} + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^{\top} + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^{\top}$$

即：实对称矩阵的谱分解 (Spectral Decomposition)将其分解成了n 个秩为 1的矩阵之和。

我们讨论了 $n \times n$ 的实对称矩阵 S。

- 其所有的特征值都是实数，并且都有着对应的实特征向量。
- 不同特征值对应的特征向量都是正交的。
- 任何一个实对称矩阵都可以被对角化，并且其谱分解可以写成 n 个秩为 1 的矩阵之和，即：

$$S = Q^{\top} \Lambda Q = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^{\top} + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^{\top} + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^{\top}$$

对称矩阵和正定矩阵 (Symmetric Matrices and Positive Definite Matrices)

正定矩阵

在很多应用中，所有特征值是正的对称矩阵起到了非常大的作用。

定义 207 **[Positive Definite Matrix].**

一个对称矩阵 S 被称为是正定矩阵，如果其所有的特征值 λ 都满足 $\lambda > 0$。

例 208.

$$\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 4 & 3 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

定理 209.

令 S 是一个实对称矩阵，则 S 是正定的，当且仅当对于所有的 $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$，都有：

$$\mathbf{x}^\mathsf{T} S \mathbf{x} > 0$$

$\mathbf{x}^\mathsf{T} S \mathbf{x}$ 是什么？

令 $\mathbf{x} = (x_1, \cdots, x_n)$，$S = \begin{bmatrix} s_{ij} \end{bmatrix}_{n \times n}$，则我们有：

$$\mathbf{x}^\mathsf{T} S \mathbf{x} = s_{11} x_1^2 + \cdots + s_{nn} x_n^2 + 2 \sum_{1 \leqslant i \neq j \leqslant n} s_{ij} x_i x_j$$

即 $\mathbf{x}^\mathsf{T} A \mathbf{x}$ 是关于 $x_1, \cdots, x_n$ 的一个二次齐次函数，我们称其为二次型 (Quadratic Forms)。

一般地，对于 $x_1,\cdots,x_n$ 的一个二次齐次函数:

$$f(x_1,\cdots,x_n)=s_{11}x_1^2+\cdots+s_{nn}x_n^2+2\sum_{1\leqslant i\neq j\leqslant n}s_{ij}x_ix_j$$

我们总可以将其转换成如下的矩阵形式:

$$f(x_1,\cdots,x_n)=\mathbf{x}^\top S\mathbf{x},\text{ 其中: } \mathbf{x}=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix},\ S=\begin{bmatrix}s_{11}&s_{12}&\cdots&s_{1n}\\ s_{12}&s_{22}&\cdots&s_{2n}\\ \vdots&\vdots&\ddots&\vdots\\ s_{1n}&s_{2n}&\cdots&s_{nn}\end{bmatrix}$$

这意味着我们总可以用一个实对称矩阵来表示一个关于 $x_1,\cdots,x_n$ 的一个二次齐次函数。特别的，正定矩阵就是指的是那些恒大于 0 的二次型。

我们先来看一个简单的应用。

定理 210.
令 A 是 $m \times n$ 的矩阵，并且 $\mathrm{rank}(A) = n$. 则 $S = A^{\mathsf{T}}A$ 是对称且正定的。

证明. 令 $\mathbf{x} \in \mathbb{R}^n$，则我们有：

$$\mathbf{x}^{\mathsf{T}}S\mathbf{x} = \mathbf{x}^{\mathsf{T}}A^{\mathsf{T}}A\mathbf{x} = (A\mathbf{x})^{\mathsf{T}}(A\mathbf{x}) = A\mathbf{x} \cdot A\mathbf{x} = \|A\mathbf{x}\|^2$$

从而我们有：$\mathbf{x}^{\mathsf{T}}S\mathbf{x} \geqslant 0$ 并且：

$$\mathbf{x}^{\mathsf{T}}S\mathbf{x} = 0 \iff \|A\mathbf{x}\| = 0 \implies A\mathbf{x} = \mathbf{0}$$

由于 $\mathrm{rank}(A) = n$，从而：

$$\mathbf{x}^{\mathsf{T}}S\mathbf{x} = 0 \implies A\mathbf{x} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}$$

$\square$

我们先证 $(\Rightarrow)$ 方向。由于 S 是一个实对称矩阵，从而存在一个正交矩阵 Q 满足:

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = Q^{\mathsf{T}} S Q$$

这里 $\lambda_1, \cdots, \lambda_n$ 是 S 的特征值，并且由假设：$\lambda_i > 0$.

定义:

$$\mathbf{x} = Q^{\mathsf{T}}\mathbf{y}, \text{ 即: } \mathbf{y} = Q\mathbf{x}$$

从而:

$$\mathbf{x}^{\mathsf{T}} S \mathbf{x} = (Q\mathbf{y})^{\mathsf{T}} S (Q\mathbf{y}) = \mathbf{y}^{\mathsf{T}} Q^{\mathsf{T}} S Q \mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

注意到 Q 是可逆的，从而对于任意 $\mathbf{x} \neq \mathbf{0}$，我们有 $\mathbf{y} = Q\mathbf{x} \neq \mathbf{0}$，即:

$$\mathbf{x}^{\mathsf{T}} S \mathbf{x} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 > 0$$

定理209的证明 (II)

上海师范大学 Shanghai Normal University

另一方面，反设存在 $\lambda_i \leqslant 0$，注意到：

$$\mathbf{x}^{\mathsf{T}} S \mathbf{x} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = \lambda_i y_i^2 + \sum_{j \neq i} \lambda_j y_j^2$$

从而令 $y_i = 1$，$y_j = 0 (j \neq 0)$，由矩阵的可逆性我们有 $\mathbf{x} = Q\mathbf{y} \neq \mathbf{0}$，即：

$$\mathbf{x}^{\mathsf{T}} S \mathbf{x} = \lambda_i y_i^2 + \sum_{j \neq i} \lambda_j y_j^2 = \lambda_j \leqslant 0$$

与假设矛盾。 $\square$

通过上述的讨论，我们可以把正定的概念扩充到任何一个实矩阵上去。

定义 211.

给定一个 $n \times n$ 的实矩阵A，如果对任意向量 $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$:

1. $\mathbf{x}^\top A\mathbf{x} > 0$，则称 A 是正定的。
2. $\mathbf{x}^\top A\mathbf{x} \geqslant 0$，则称 A 是半正定的。
3. $\mathbf{x}^\top A\mathbf{x} < 0$，则称 A 是负定的。
4. $\mathbf{x}^\top A\mathbf{x} \leqslant 0$，则称 A 是半负定的。
5. 若不满足以上任何一种条件，则称矩阵 A 是不定的。

我们需要注意的是，对于非对称矩阵来说，即使其所有的特征值大于 0，它也不一定是正定的。考虑下列的例子：

$$\begin{bmatrix} 1 & 100 \\ 0 & 2 \end{bmatrix}$$

不难验证，其特征值为 $1, 2$。但是我们有：

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 100 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 1 + 2 - 100 = -97 < 0$$

二次型的转化

通过上述的证明，可以看到，对参数作一些变换可以使得对应的二次函数变得简单：

- $\mathbf{x}^\top S\mathbf{x} = s_{11}x_1^2 + \cdots + s_{nn}x_n^2 + 2\sum_{1\leqslant i\neq j\leqslant n} s_{ij}x_ix_j$
- 令 $\mathbf{x} = Q^\top\mathbf{y}$ 后我们有：

$$\mathbf{x}^\top S\mathbf{x} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

我们将后者这种只含平方项的二次型为标准二次型，特别的，如果其系数进一步化简为 $1, 1, \cdots, 1, -1, \cdots, -1$，我们称其为规范二次型。

定理209的证明已经给了我们一种转化二次型的方法，即找对应矩阵的谱分解。以下列二次函数为例：

$$f(x_1, x_2, x_3) = -2x_1x_2 + 2x_1x_3 + 2x_2x_3$$

其对应的对称矩阵为：

$$S = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

将其进行谱分解，则存在正交矩阵 P 和对角矩阵 Λ 使得

$$\begin{bmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \Lambda = P^{\mathsf{T}}SP = \begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}$$

从而令:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

我们有:

$$f = -2y_1^2 + y_2^2 + y_3^2$$

如果要进一步变成规范型，我们只需要在令:

$$\begin{cases} y_1 = \frac{1}{\sqrt{2}} z_1 \\ y_2 = z_2 \\ y_3 = z_3 \end{cases}$$

即有:

$$f = -z_1^2 + z_2^2 + z_3^2$$

二次型的转化-配方法 (I)

我们再介绍另一个方法，配方法。它不依赖于矩阵的运算。比如考虑:

$$f = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 + 2x_1x_3 + 6x_2x_3$$

我们考虑先将所有 x_1 的项集中起来:

$$x_1^2 + 2x_1x_2 + 2x_1x_3$$

对其配方可得:

$$x_1^2 + 2x_1x_2 + 2x_1x_3 = (x_1 + x_2 + x_3)^2 - x_2^2 - x_3^2 - 2x_2x_3$$

从而:

$$f = (x_1 + x_2 + x_3)^2 + x_2^2 + 4x_3^2 + 4x_2x_3 = (x_1 + x_2 + x_3)^2 + (x_2 + x_3)^2$$

$$f = (x_1 + x_2 + x_3)^2 + x_2^2 + 4x_3^2 + 4x_2x_3 = (x_1 + x_2 + x_3)^2 + (x_2 + 2x_3)^2$$

我们令

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \text{即:} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

则我们有:

$$f = \mathbf{x}^T \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 5 \end{bmatrix} \mathbf{x} = \mathbf{y}^T \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{y} = \mathbf{y}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}$$

需要注意的是，该方法并不一定满足系数矩阵是正交的。

如果只有类似 x_1x_2 的项，比如：

$$f = 2x_1x_2 + 2x_12x_3 - 6x_1x_3$$

我们也是可以用配方法的，只要注意到：

$$4x_1x_2 = (x_1 + x_2)^2 - (x_1 - x_2)^2$$

从而进行

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

代入再按之前配方的思路化简即可。

可以看到，通过一些可逆的系数变换，我们可以将一个二次型进行转换，即，令

$$f = \mathbf{x}^\top A\mathbf{x}$$

则可以通过一个可逆逆矩阵 C, 使得 $\mathbf{x} = C\mathbf{y}$，并且:

$$f = \mathbf{y}^\top C^\top AC\mathbf{y} = \mathbf{y}^\top B\mathbf{y}$$

我们称满足这样一个关系的矩阵 A, B 是合同的。

定义 212.

令 A, B 是 $n \times n$ 的矩阵，若存在可逆矩阵 C 使得:

$$B = C^\top AC$$

则称 A 和 B 是合同的。

正定性和行列式的关系

现在我们来看一下正定性和行列式的关系。

定理 213.

如果 S 是正定矩阵，则 $\det(S) > 0$.

证明. 由 S 是实对称的，从而存在正交矩阵 Q 使得:

$$\begin{bmatrix} \lambda_1 \cdots & 0 & \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} = Q^\top S Q$$

这里 $\lambda_1, \cdots, \lambda_n$ 是 S 的特征值。从而:

$$\begin{vmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{vmatrix} = \det(Q^\top)\det(S)\det(Q)$$

即:

$$\det(S) = \det(Q^\top)\det(S)\det(Q) = \lambda_1 \cdots \lambda_n > 0$$

行列式大于 0 的对称矩阵不一定正定

上述定理的逆命题并不一定正确。比如考察下列矩阵：

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

其行列式 $=4>0$，但是我们有：

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

定义 214 **[顺序主子式 (Leading Principal Minors)].**

给定一个 $n \times n$ 的矩阵:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

则对任意的 $k \in [n]$。定义下列行列式:

$$\Delta_k = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}$$

是 A 的一个 k 阶顺序主子式 (Leading Principal Minors).

定理 215.

给定一个 $n \times n$ 的对称矩阵 S，S 是正定的当且仅当所有的顺序主子式 $\Delta_k > 0$.

直观理解

直观上来讲，当每个顺序主子式都大于 0 时，我们有：

$$\lambda_k = \frac{\Delta_k}{\Delta_{k-1}}$$

这里可以不妨假设 $\Delta_0 = 1$.

定理215的证明 (I)

令

$$S_k = \begin{bmatrix} s_{11} & \cdots & s_{1k} \\ \vdots & \ddots & \vdots \\ s_{k1} & \cdots & s_{kk} \end{bmatrix}$$

则我们有: $\Delta_k = \det(S_k)$. 我们先证 $(\Leftarrow)$ 方向，即说明每个 S_k 都是正定的。令 $\mathbf{x} = (x_1, \cdots, x_k) \in \mathbb{R}^k \setminus \{\mathbf{0}\}$，注意到:

$$\begin{bmatrix} x_1 & \cdots & x_k \end{bmatrix} S_k \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_k & 0 & \cdots & 0 \end{bmatrix} S \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} > 0$$

从而 S_k 正定，因此 $\Delta_k = \det(S_k) > 0$.

定理215的证明 (II)

另一个方向我们使用归纳法。

BASE: $n = 1$ 时是显然的。

INDUCTION: 现在考虑 $n \geqslant 2$ 的时候，我们有对于任意的 $k \in [n]$:

$$\Delta_k = \begin{vmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ s_{21} & s_{22} & \cdots & s_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{k1} & s_{k2} & \cdots & s_{kk} \end{vmatrix} = \begin{vmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ 0 & s_{22} - \frac{s_{12}s_{21}}{s_{11}} & \cdots & s_{2k} - \frac{s_{1k}s_{21}}{s_{11}} \\ 0 & \vdots & \ddots & \vdots \\ 0 & s_{k2} - \frac{s_{1k}s_{12}}{s_{11}} & \cdots & s_{kk} - \frac{s_{1k}s_{k1}}{s_{11}} \end{vmatrix}$$

对任意的 $2 \leqslant i, j \leqslant k$ 定义:

$$t_{ij} = s_{ij} - \frac{s_{1i}s_{1j}}{s_{11}} (= s_{ij} - \frac{s_{1i}s_{j1}}{s_{11}})$$

从而我们有:

$$\Delta_k = \begin{vmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ 0 & t_{22} & \cdots & t_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & t_{k2} & \cdots & t_{kk} \end{vmatrix} = s_{11} \begin{vmatrix} t_{22} & \cdots & t_{2k} \\ \vdots & \ddots & \vdots \\ t_{k2} & \cdots & t_{kk} \end{vmatrix}$$

注意到 $s_{11} = \Delta_1 > 0$，$\Delta_k > 0$，从而我们有:

$$\begin{vmatrix} t_{22} & \cdots & t_{2k} \\ \vdots & \ddots & \vdots \\ t_{k2} & \cdots & t_{kk} \end{vmatrix} > 0$$

对任意的 $2 \leqslant k \leqslant n$ 是成立的。

定义 $(n-1)\times(n-1)$ 的矩阵 T:

$$T=\begin{bmatrix} t_{22} & \cdots & t_{2k} \\ \vdots & \ddots & \vdots \\ t_{k2} & \cdots & t_{kk} \end{bmatrix}$$

显然 T 是一个实对称矩阵，并且其所有的顺序主子式都是大于 0 的，从而有归纳假设 T 是正定的。从而:

$$\begin{aligned}
\mathbf{x}^T S\mathbf{x} &= \sum_{i\in[n]}\sum_{j\in[n]} s_{ij}x_i x_j \\
&= \frac{(s_{11}x_1+\cdots+s_{1n}x_n)^2}{s_{11}} + \sum_{i=2}^{n}\sum_{j=2}^{n} t_{ij}x_i x_j \\
&= \frac{(s_{11}x_1+\cdots+s_{1n}x_n)^2}{s_{11}} + \begin{bmatrix} x_2 & \cdots & x_n \end{bmatrix} T \begin{bmatrix} x_2 \\ \vdots \\ x_n \end{bmatrix}
\end{aligned}$$

最后我们分两种情况讨论:

1. $\begin{bmatrix} x_2 & \cdots & x_n \end{bmatrix}^\top \neq \mathbf{0}$，则由 T 是正定的，我们有:

$$\mathbf{x}^\top S \mathbf{x} \geqslant \begin{bmatrix} x_2 & \cdots & x_n \end{bmatrix} T \begin{bmatrix} x_2 \\ \vdots \\ x_n \end{bmatrix} > 0$$

2. $\begin{bmatrix} x_2 & \cdots & x_n \end{bmatrix}^\top = \mathbf{0}$，这意味着 $x_1 \neq 0$，注意到 $s_{11} > 0$，我们有:

$$\mathbf{x}^\top S \mathbf{x} = \frac{(s_{11}x_1)^2}{s_{11}} = s_{11}x_1^2 > 0$$

$\square$

我们讨论了矩阵一个性质，正定性。

正定矩阵的性质

令 S 是一个对称矩阵，

1. S 是正定的。
2. S 的特征值都大于 0.
3. S 的顺序主子式都大于 0。
4. 存在一个矩阵 A，使得 $S = A^{\top}A$.

我们还讨论了二次型的概念，以及变换成标准二次型的方法。

- 利用对称矩阵的谱分解。
- 配方法。

对称矩阵和正定矩阵 (Symmetric Matrices and Positive Definite Matrices)

特征空间、代数重数以及几何重数

给定一个 $n \times n$ 的矩阵 A 和其一个特征值 $\lambda \in \mathbb{C}$，若 $\mathbf{x}_1, \mathbf{x}_2$ 是其特征向量，则我们有：

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \lambda\mathbf{x}_1 + \lambda\mathbf{x}_2 = \lambda(\mathbf{x}_1 + \mathbf{x}_2)$$

$$A(c\mathbf{x}_1) = cA\mathbf{x}_1 = c\lambda\mathbf{x}_1 = \lambda(c\mathbf{x}_1)$$

即其特征向量组成的空间是一个向量空间，我们称为特征空间。

定义 216 **[特征空间 (Eigenspace)].**

给定一个 $n \times n$ 的矩阵 A 和其一个特征值 $\lambda \in \mathbb{C}$，定义：

$$V_\lambda = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = \lambda\mathbf{x}\}$$

为 A 的特征值 λ 对应的特征空间。

引理 217.

V_λ 是 $\mathbb{C}^n$ 的一个子空间，并且 $\dim(V_\lambda) \geqslant 1$。

考察旋转矩阵:

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

我们有:

$$\det(A - \lambda I) = \lambda^2 + 1 \implies \lambda = \pm i$$

从而我们有:

$$\lambda = i \implies V_i = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = i\mathbf{x}\} = \mathrm{span}\left\{\begin{bmatrix} 1 \\ i \end{bmatrix}\right\}$$

$$\lambda = -i \implies V_{-i} = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = -i\mathbf{x}\} = \mathrm{span}\left\{\begin{bmatrix} 1 \\ -i \end{bmatrix}\right\}$$

我们将 V_λ 的维数称为 λ 的几何重数 (Geometric Multiplicity)。

定义 218 [几何重数 (Geometric Multiplicity)].

给定一个 $n \times n$ 的矩阵 A 和其一个特征值 $\lambda \in \mathbb{C}$，λ 的几何重数是其特征空间 V_λ 的维数，即：

$$\begin{aligned}\dim(V_\lambda) &= \dim\left(\{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = \lambda\mathbf{x}\}\right)\\ &= \dim\left(\{\mathbf{x} \in \mathbb{C}^n \mid (A - \lambda I)\mathbf{x} = \mathbf{0}\}\right)\\ &= \dim\left(\mathrm{null}(A - \lambda I)\right) = n - \mathrm{rank}(A - \lambda I)\end{aligned}$$

例 219.

在上述的旋转矩阵 A 中，我们有 $\dim(V_i) = \dim(V_{-i}) = 1$，即 i 和 $-i$ 的几何重数为 1。

考虑 A 的特征多项式:

$$f_A(\lambda)=\det(A-\lambda I)=\prod_{i=1}^{n}(\lambda-\lambda_i)=\prod_{i=1}^{k}(\lambda-\lambda_i)^{m_i}(m_1+m_2+\cdots+m_k=n)$$

我们将 m_i 称为 λ_i 的代数重数 (Algebraic Multiplicity)。

定义 220 [代数重数 (Algebraic Multiplicity)].

给定一个 $n\times n$ 的矩阵 A 和其一个特征值 $\lambda\in\mathbb{C}$, λ 的代数重数是其特征多项式中 $(\lambda-\lambda_i)$ 的幂次 m_i，即:

$$m_i=\max\{m\mid(\lambda-\lambda_i)^m\text{是}f_A(\lambda)\text{的因子}\}$$

例 221.

在上述的旋转矩阵 A 中，我们有 $f_A(\lambda)=(\lambda+i)(\lambda-i)$，即 i 和 $-i$ 的代数重数为 1。

更多的例子

考察如下矩阵:

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

其特征多项式为:

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -1-\lambda & 1 & 0 \\ -4 & 3-\lambda & 0 \\ 1 & 0 & 2-\lambda \end{vmatrix} \\ &= (-1-\lambda)\begin{vmatrix} 3-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} - 1\begin{vmatrix} -4 & 0 \\ 1 & 2-\lambda \end{vmatrix} \\ &= (-1-\lambda)(3-\lambda)(2-\lambda) + 4(2-\lambda) \\ &= (1-\lambda)^2(2-\lambda) \end{aligned}$$

从而其特征值为 $\lambda = 1, 2$，代数重数分别为 $2, 1$。

$\lambda = 2$ 对应的特征空间

考察 $A - 2I$:

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

不难验证 $\operatorname{rank}(A - 2I) = 2$，从而 $\dim(V_2) = n - \operatorname{rank}(A - 2I) = 1$，即 2 的几何重数为 1，更精确的说:

$$V_2 = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = 2\mathbf{x}\} = \left\{ c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \middle| c \in \mathbb{C} \right\} = \operatorname{span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

考察 $A - I$:

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

不难验证 $\mathrm{rank}(A - I) = 2$，从而 $\dim(V_1) = n - \mathrm{rank}(A - I) = 1$，即 1 的几何重数为 1，更精确的说:

$$V_1 = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = \mathbf{x}\} = \left\{ c \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \middle| c \in \mathbb{C} \right\} = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \right\}$$

注意到1 的代数重数为 2，即代数重数和几何重数并不相等。

定理 222.

特征值的代数重数大于等于其几何重数。

引理 223.

令 A 和 B 是相似矩阵，即存在可逆矩阵 P 使得 $B = P^{-1}AP$，则我们有：

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

即两个相似矩阵不仅有相同的特征值，而且其代数重数也相等。

引理223的证明

由于 $B = P^{-1}AP$，我们有：

$$\begin{aligned}
\det(B - \lambda I) &= \det(A - \lambda I) \\
&= \det(P^{-1}AP - P^{-1}(\lambda I)P) \\
&= \det(P^{-1}(A - \lambda I)P) \\
&= \det(P^{-1})\det(A - \lambda I)\det(P) \\
&= \det(A - \lambda I)
\end{aligned}$$

$\square$

$\det(P)\det(P^{-1}) = 1$

注意到，若 P 是可逆的，则我们有：

$$1 = \det(I) = \det(PP^{-1}) = \det(P)\det(P^{-1}) \implies \det(P^{-1}) = \frac{1}{\det(P)}$$

定理的直观思路

如果 $\dim(V_\lambda) = i$，意味着可以从中找出 i 个线性无关的特征向量 $\mathbf{x}_1, \cdots, \mathbf{x}_i$，使得：

$$A\begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_i\end{bmatrix} = \begin{bmatrix}\lambda\mathbf{x}_1 & \cdots & \lambda\mathbf{x}_i\end{bmatrix} = \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_i\end{bmatrix}\begin{bmatrix}\lambda & & \\ & \ddots & \\ & & \lambda\end{bmatrix}$$

从而我们可以将 A 转化成如下的形式：

$$A = P\begin{bmatrix}\lambda I_i & B \\ O & C\end{bmatrix}P^{-1}$$

由上述引理可知 λ 的代数重数一定大于等于 i。

定理222的证明 (II)

上海师范大学 Shanghai Normal University

现在令 A 是一个 $n \times n$ 的矩阵，λ_0 是其一个特征值，考虑其特征空间：

$$V_{\lambda_0} = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = \lambda_0 \mathbf{x}\}$$

令其的一组基为 $\mathbf{v}_1, \cdots, \mathbf{v}_m$，则 m 就是 λ_0 的几何重数。特别的，我们有：

$$\begin{aligned} A\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{bmatrix} &= \begin{bmatrix} \lambda_0 \mathbf{v}_1 & \cdots & \lambda_0 \mathbf{v}_m \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{bmatrix} \begin{bmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_0 \end{bmatrix} \end{aligned}$$

这里 $\begin{bmatrix} \lambda_0 & & \\ & \ddots & \\ & & \lambda_0 \end{bmatrix} = \lambda_0 I_m$ 是一个 $m \times m$ 的矩阵。

我们将 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 扩展成 $\mathbb{C}^n$ 的一组基，即:

$$\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{v}_{m+1}, \cdots, \mathbf{v}_n$$

从而我们有:

$$\begin{aligned} A\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_m & \mathbf{v}_{m+1} & \cdots & \mathbf{v}_n\end{bmatrix} &= \begin{bmatrix}\lambda_0\mathbf{v}_1 & \cdots & \lambda_0\mathbf{v}_m & A\mathbf{v}_{m+1} & \cdots & A\mathbf{v}_n\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_m & \mathbf{v}_{m+1} & \cdots & \mathbf{v}_n\end{bmatrix}\begin{bmatrix}\lambda_0 I_m & B \\ O & C\end{bmatrix}\end{aligned}$$

这里 B 是一个 $m \times (n-m)$ 的矩阵，C 是一个 $(n-m) \times (n-m)$ 的矩阵。

定理222的证明 (IV)

令 $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m & \mathbf{v}_{m+1} & \cdots & \mathbf{v}_n \end{bmatrix}$，则 P 是可逆矩阵，从而我们有:

$$A = P \begin{bmatrix} \lambda_0 I_m & B \\ O & C \end{bmatrix} P^{-1}$$

从而我们有

$$\det(A - \lambda I) = \det(\begin{bmatrix} \lambda_0 I_m & B \\ O & C \end{bmatrix} - \lambda I) = \begin{vmatrix} (\lambda_0 - \lambda) I_m & B \\ O & C - \lambda I_{n-m} \end{vmatrix}$$

不难验证:

$$\det(A - \lambda I) = (\lambda_0 - \lambda)^m g(\lambda)$$

这里 $g(\lambda)$ 是一个 $n-m$ 次的多项式，从而 λ_0 的代数重数至少为 m，即:

$$\lambda_0 \text{ 的代数重数} \geqslant \lambda_0 \text{ 的几何重数。}$$

$\square$

回顾矩阵可对角化的条件:

定理 224.

$n \times n$ 的矩阵可对角化当且仅当 A 有n 个线性无关的特征向量。

从而我们可以得到:

定理 225.

$n \times n$ 的矩阵可对角化当且仅当其每个特征值的代数重数等于几何重数。

因此对于实对称矩阵来说:

推论 226.

令 S 是一实对称矩阵，并且 λ 是其一个特征值，则 λ 的几何重数等于其代数重数。

阶段总结

Shanghai Normal University

- 特征空间。
- 代数重数与几何重数。
- 代数重数和几何重数之间的关系，及与对角化的联系。

线性变换 (Linear Transformation)

线性变换 (Linear Transformation)

线性变换的概念

”线性” 的回顾

Shanghai Normal University 上海师范大学

考虑函数:

$$y = 3x$$

这是一个线性函数 $f : \mathbb{R} \longrightarrow \mathbb{R}$.

- 几何上来讲，其定义了 $\mathbb{R}^2$ 上或者说是一个二维平面上的一条线。
- 代数上来讲，对任意的 $x, y, c \in \mathbb{R}$ 其满足:

$$f(x+y) = 3(x+y) = 3x + 3y = f(x) + f(y)$$

$$f(cx) = 3cx = cf(x)$$

定义 227 **[Linear Transformation].**

令 V 和 W 实两个向量空间。一个函数 $T : \mathbb{V} \longrightarrow \mathbb{W}$ 被称作是一个线性变换，如果其满足对任意的 $\mathbf{v}, \mathbf{w} \in \mathbb{V}$ 和实数 $c \in \mathbb{R}$:

$$T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}), \; T(c\mathbf{v}) \quad = cT(\mathbf{v})$$

术语说明!

在很多翻译当中，会将 $V \neq W$ 的情况称为线性映射，而只有 $V = W$ 的情况下称为线性变换。我们的课件则不作这类区分，但当 $V \neq W$ 时，我们会经常显示的表达出 V 和 W。

引理 228.

1. $T(\mathbf{0}) = \mathbf{0}$.
2. $T(c\mathbf{v} + d\mathbf{w}) = cT(\mathbf{v}) + dT(\mathbf{w})$.
3. $T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n)$

1. 内积 (Inner Product)： 固定一个向量 $\mathbf{a} = (1, 2, 3) \in \mathbb{R}^3$， 定义 $T_{\mathbf{a}}(\mathbf{v}) = \mathbf{a} \cdot \mathbf{v}$。则 $T_{\mathbf{a}}$ 是 $\mathbb{R}^3 \to \mathbb{R}$ 上的线性变换。
2. 长度 (Length): 定义 $T(\mathbf{v}) = \|\mathbf{v}\|$， 则 T不是$\mathbb{R}^n \to \mathbb{R}$ 上的线性变换。
3. 旋转 (Rotation)： 定义 $\mathbb{V}$ 上的映射:

$$T(\mathbf{v}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \mathbf{v}$$

则 T 是一个 $\mathbb{V}$ 到 $\mathbb{V}$ 上的线性变换， 我们也称其为 $\mathbb{V}$ 上的线性变换。

由上述旋转的例子可以看到，我们可以利用矩阵的乘法来定义一个线性变换。

引理 229.

令 A 是一个 $m \times n$ 的矩阵，定义 $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ 为：

$$T_A(\mathbf{x}) = A\mathbf{x}$$

则 T_A 是一个 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 上的线性变换。

一个线性变换的前瞻性视角

事实上我们将说明，对于任何一个如下的线性变换：

$$T : \mathbb{V} \longrightarrow \mathbb{W}$$

其中 $\dim(V) = n, \dim(W) = m$，我们都可以将其理解成如下的线性变换：

$$T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

我们还是先研究空间的基对线性变换的影响。

引理 230.

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是两个线性空间，并且 $\dim(\mathbb{V}) = \dim(\mathbb{W}) = n$。令 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是 $\mathbb{V}$ 的一组基，则：

1. 令 $T_1, T_2 : \mathbb{V} \longrightarrow \mathbb{W}$ 是两个其上的线性变换，满足对任意的 $i \in [n]$ 都有：

$$T_1(\mathbf{v}_i) = T_2(\mathbf{v}_i)$$

   则 $T_1 = T_2$。

2. 令 $\mathbf{w}_1, \cdots, \mathbf{w}_n$ 是 W 中的一组基，则存在唯一的线性变换 $T : \mathbb{V} \longrightarrow \mathbb{W}$ 使得对任意的 $i \in [n]$ 都有：

$$T(\mathbf{v}_i) = \mathbf{w}_i$$

这说明了，在给定的基下，线性变换是唯一确定的。

令 $\mathbf{v} \in \mathbb{V}$。由于 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是 $\mathbb{V}$ 的一组基，从而存在 $c_1, \cdots, c_n$ 使得:

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \sum_{i\in[n]} c_i\mathbf{v}_i$$

从而我们有:

$$\begin{aligned} T_1(\mathbf{v}) &= T_1\left(\sum_{i\in[n]} c_i\mathbf{v}_i\right) = \sum_{i\in[n]} c_i T_1(\mathbf{v}_i) \\ &= \sum_{i\in[n]} c_i T_2(\mathbf{v}_i) = T_2\left(\sum_{i\in[n]} c_i\mathbf{v}_i\right) = T_2(\mathbf{v}) \end{aligned}$$

另一方面，注意到 $c_1, \cdots, c_n$ 是唯一的，从而我们有:

$$T(\mathbf{v}) = \sum_{i\in[n]} c_i\mathbf{w}_i$$

即 T 满足对任意的 $i \in [n]$，有 $T(\mathbf{v}_i) = \mathbf{w}_i$。 $\square$

系数向量 (一组基下的坐标)

上海师范大学 Shanghai Normal University

令 $\mathbb{V}$ 是一个 n 维的向量空间，$\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是 $\mathbb{V}$ 的一组基，我们记作：

$$\bar{\mathbf{v}} = \mathbf{v}_1, \cdots, \mathbf{v}_n$$

注意到，对于任意的 $\mathbf{v} \in \mathbb{V}$，存在唯一的$c_1, \cdots, c_n \in \mathbb{R}$ 使得：

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

定义其系数构成的向量 (coordinate vector) 为：

$$(c_1, \cdots, c_n) \in \mathbb{R}^n$$

其被称作是在基 $\bar{\mathbf{v}}$ 下的坐标。

回顾我们之前提过的一个很奇怪的向量空间:

$$V = \mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$$

我们定义了其上的加法 $\oplus$ 和数乘 $\otimes$:

$$x \oplus x' = x \times x', \quad c \otimes x = x^c$$

由于 $\dim(V) = 1$,我们可以选取如下的一组基:

$$\mathbf{v}_1 = 2$$

则对于任意的 $x \in V$,我们有唯一的 $c \in \mathbb{R}$ 使得:

$$x = c \otimes \mathbf{v}_1 = 2^c$$

即: $c = \log_2 x$ 是 x 在基 $\mathbf{v}_1$ 下的坐标。

令 n 维空间 $\mathbb{V}$ 的一组基为 $\bar{\mathbf{v}}$，我们定义 $T_{\bar{\mathbf{v}}}: \mathbb{V} \longrightarrow \mathbb{R}^n$:

$$T_{\bar{\mathbf{v}}}(\mathbf{v}) = (c_1, \cdots, c_n)$$

这里 $(c_1, \cdots, c_n)$ 是 $\mathbf{v}$ 在基 $\bar{\mathbf{v}}$ 下的坐标。

定理 231.

$T_{\bar{\mathbf{v}}}$ 是一个 $\mathbb{V}$ 到 $\mathbb{R}^n$ 上的线性变换。并且 $T_{\bar{\mathbf{v}}}$ 是一一对应的 (bijiective)。

$$\mathsf{V} = \mathbb{R}^+ = \{x \in \mathbb{R} \mid x > 0\}$$

我们定义了其上的加法 $\oplus$ 和数乘 $\otimes$:

$$x \oplus x' = x \times x', \quad c \otimes x = x^c$$

由于 $\dim(\mathsf{V}) = 1$，我们可以选取如下的一组基:

$$\mathbf{v}_1 = 2$$

则对于任意的 $x \in \mathsf{V}$, $\log_2 x$ 是其在基 $\mathbf{v}_1$ 下的坐标。

下列映射:

$$x \mapsto \log_2 x$$

是 $\mathbb{R}^+$ 到 $\mathbb{R}$ 上的一一对应的线性变换。

我们只需证明:

1. 对任意的 $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ 有
$$T_{\bar{\mathbf{v}}}(\mathbf{u}+\mathbf{v}) = T_{\bar{\mathbf{v}}}(\mathbf{u}) + T_{\bar{\mathbf{v}}}(\mathbf{v})$$
2. 对任意的 $\mathbf{v} \in \mathbb{V}$ 和 $c \in \mathbb{R}$ 有
$$T_{\bar{\mathbf{v}}}(c\mathbf{v}) = cT_{\bar{\mathbf{v}}}(\mathbf{v})$$
3. $T_{\bar{\mathbf{v}}}$ 是一个双射:
    - 满射 (Surjection): 对任意的 $(c_1, \cdots, c_n) \in \mathbb{R}^n$, 存在 $\mathbf{v} \in \mathbb{V}$ 使得 $T_{\bar{\mathbf{v}}}(\mathbf{v}) = (c_1, \cdots, c_n)$。
    - 单射 (Injection): 对任意的 $\mathbf{v}, \mathbf{u} \in \mathbb{V}$, 如果 $T_{\bar{\mathbf{v}}}(\mathbf{v}) = T_{\bar{\mathbf{v}}}(\mathbf{u})$, 则 $\mathbf{v} = \mathbf{u}$。

定理231的证明 (II)

令 $\mathbf{v} = \sum_{i\in[n]} c_i\mathbf{v}_i$，$\mathbf{v}' = \sum_{i\in[n]} c_i'\mathbf{v}_i$，则我们有:

$$
\begin{aligned}
T_{\overline{\mathbf{v}}}(\mathbf{v}+\mathbf{v}') &= T_{\overline{\mathbf{v}}}\left(\sum_{i\in[n]} c_i\mathbf{v}_i + \sum_{i\in[n]} c_i'\mathbf{v}_i\right) \\
&= T_{\overline{\mathbf{v}}}\left(\sum_{i\in[n]} (c_i + c_i')\mathbf{v}_i\right) \\
&= (c_1 + c_1', \cdots, c_n + c_n') \\
&= (c_1, \cdots, c_n) + (c_1', \cdots, c_n') \\
&= T_{\overline{\mathbf{v}}}\left(\sum_{i\in[n]} c_i\mathbf{v}_i\right) + T_{\overline{\mathbf{v}}}\left(\sum_{i\in[n]} c_i'\mathbf{v}_i\right) = T_{\overline{\mathbf{v}}}(\mathbf{v}) + T_{\overline{\mathbf{v}}}(\mathbf{v}')
\end{aligned}
$$

定理231的证明 (III)

令 $\mathbf{v} = \sum_{i\in[n]} c_i \mathbf{v}_i$ 和 $c \in \mathbb{R}$，则我们有:

$$
\begin{aligned}
T_{\overline{\mathbf{v}}}(c\mathbf{v}) &= T_{\overline{\mathbf{v}}}(c \sum_{i\in[n]} c_i \mathbf{v}_i) \\
&= T_{\overline{\mathbf{v}}}\left(\sum_{i\in[n]} (cc_i)\mathbf{v}_i\right) \\
&= (cc_1, \cdots, cc_n) \\
&= c\,(c_1, \cdots, c_n) \\
&= cT_{\overline{\mathbf{v}}}(\sum_{i\in[n]} c_i \mathbf{v}_i) = cT_{\overline{\mathbf{v}}}(\mathbf{v})
\end{aligned}
$$

最后我们证明 $T_{\mathbb{V}}$ 是一个双射:

- 令 $\mathbf{v}, \mathbf{v}' \in \mathbb{V}$ 满足: $T_{\mathbb{V}}(\mathbf{v}) = T_{\mathbb{V}}(\mathbf{v}') = (c_1, \cdots, c_n)$, 则我们有:

$$\mathbf{v} = \sum_{i \in [n]} c_i \mathbf{v}_i = \mathbf{v}'$$

  即 $\mathbf{v} = \mathbf{v}'$。

- 令 $(c_1, \cdots, c_n) \in \mathbb{R}^n$, 则定义:

$$\mathbf{u} = c_1 \mathbf{v}_1 + \cdots + c_n \mathbf{v}_n = \sum_{i \in [n]} c_i \mathbf{v}_i \in \mathbb{V}$$

  从而我们有: $T_{\mathbb{V}}(\mathbf{u}) = (c_1, \cdots, c_n)$.

$\square$

我们可以看到, $\mathsf{T}_{\overline{\mathbf{v}}} : \mathbb{V} \longrightarrow \mathbb{R}^n$ 的定义取决于基 $\overline{\mathbf{v}}$ 的选择:

$$\mathbf{v}_1, \cdots, \mathbf{v}_n$$

如果我们选取不同的基 $\overline{\mathbf{v}}$ 和 $\overline{\mathbf{v}'}$, $\mathsf{T}_{\overline{\mathbf{v}}}$ 和 $\mathsf{T}_{\overline{\mathbf{v}'}}$ 有什么关系?

例 232.

考察 $\mathbb{R}^2$, 我们选取两组基:

$$\overline{\mathbf{v}_1} = \begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}, \quad \overline{\mathbf{v}_2} = \begin{bmatrix}1\\-1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix},$$

则对于 $\begin{bmatrix}2\\3\end{bmatrix}$ 来说，我们有:

$$\mathsf{T}_{\overline{\mathbf{v}_1}}(\begin{bmatrix}2\\3\end{bmatrix}) = \begin{bmatrix}2\\3\end{bmatrix}, \mathsf{T}_{\overline{\mathbf{v}_2}}(\begin{bmatrix}2\\3\end{bmatrix}) = \begin{bmatrix}2.5\\-0.5\end{bmatrix}$$

事实上，我们将证明其中存在一个基变换矩阵M 满足: $\mathsf{T}_{\overline{\mathbf{v}_1}}(\mathbf{v}) = M\mathsf{T}_{\overline{\mathbf{v}_2}}(\mathbf{v})$.

令 $\mathsf{v}_1, \cdots, \mathsf{v}_n \in \mathbb{V}$，$A$ 是一个 $n \times m$ 的矩阵。则我们定义:

$$\begin{bmatrix} \mathsf{u}_1 & \cdots & \mathsf{u}_n \end{bmatrix} = \begin{bmatrix} \mathsf{v}_1 & \cdots & \mathsf{v}_n \end{bmatrix} A$$

为满足:

$$\mathsf{u}_i = \sum_{k \in n} A(k, i) \mathsf{v}_k$$

的 m 个 $\mathbb{V}$ 中的元素 $\mathsf{u}_1, \cdots, \mathsf{u}_m \in \mathbb{V}$。

说明

注意，这里的 $\mathsf{u}_i, \mathsf{v}_i$ 并不是列向量，而是 V 中的一个元素，从而 $\begin{bmatrix} \mathsf{u}_1 & \cdots & \mathsf{u}_m \end{bmatrix}$ 并不是矩阵，我们使用相同的写法是因为可以证明其有类似的含义。

引理 233.

令 $\mathbf{v}_1, \cdots, \mathbf{v}_n \in \mathbb{V}$，$A$ 是 $n \times m$ 的矩阵，B 是 $m \times l$ 的矩阵，则我们有：

$$\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} (AB) = \left(\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} A \right) B$$

证明. 定义：

$$\begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} A, \quad \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_l \end{bmatrix} = \left(\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} A \right) B$$

则对于任意的 $i \in [l]$，我们有：

$$\mathbf{w}_i = \sum_{k \in [m]} B(k, i) \mathbf{u}_k = \sum_{k \in [m]} B(k, i) \left(\sum_{j \in [n]} A(j, k) \mathbf{v}_j \right) = \sum_{j \in [n]} \left(\sum_{k \in [m]} A(j, k) B(k, i) \right) \mathbf{v}_j$$

即：

$$\left(\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} A \right) B = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_l \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} (AB)$$

$\square$

令 $\overline{\mathbf{v}}=\mathbf{v}_1,\cdots,\mathbf{v}_n$ 是 $\mathbb{V}$ 的一组基，则对于 $\mathbf{v}\in\mathbb{V}$ 我们有：

$$\mathbf{v}=\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} T_{\overline{\mathbf{v}}}(\mathbf{v})$$

引理 234.

令 $\overline{\mathbf{v}}=\mathbf{v}_1,\cdots,\mathbf{v}_n$ 和 $\overline{\mathbf{v}'}=\mathbf{v}'_1,\cdots,\mathbf{v}'_n$ 是 V 的两组基，则存在一个唯一的 $n\times n$ 矩阵 M满足：

$$\begin{bmatrix}\mathbf{v}'_1 & \cdots \mathbf{v}'_n\end{bmatrix}=\begin{bmatrix}\mathbf{v}_1 & \cdots \mathbf{v}_n\end{bmatrix}M$$

定义 235 [Change of Basis].

上述 M 被称作基 $\overline{\mathbf{v}}$ 到 $\overline{\mathbf{v}'}$ 的过渡矩阵。

由于 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是 $\mathbb{V}$ 的一组基，从而对于任意的 $\mathbf{v}_i'$，存在唯一的 $a_{i1}, \cdots, a_{in}$ 使得:

$$\mathbf{v}_i' = a_{i1}\mathbf{v}_1 + \cdots + a_{in}\mathbf{v}_n = \sum_{k\in[n]} a_{ik}\mathbf{v}_k$$

即 $(a_{i1}, \cdots, a_{in}) \in \mathbb{R}^n$ 是 $\mathbf{v}_i'$ 在基 $\bar{\mathbf{v}}$ 下的坐标。

定义下列矩阵:

$$M = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

则我们有:

$$\begin{bmatrix} \mathbf{v}_1' & \cdots & \mathbf{v}_n' \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} M$$

M 的唯一性由 $\mathbf{v}_i' = \sum_{k\in[n]} a_{ik}\mathbf{v}_k$ 保证。 $\square$

定理 236.

M 是可逆的。

证明. 由引理 234，存在唯一的 $n \times n$ 的矩阵 M' 满足:

$$\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}'_1 & \cdots & \mathbf{v}'_n \end{bmatrix} M'$$

从而:

$$\begin{bmatrix} \mathbf{v}'_1 & \cdots & \mathbf{v}'_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} M = (\begin{bmatrix} \mathbf{v}'_1 & \cdots & \mathbf{v}'_n \end{bmatrix} M')M = \begin{bmatrix} \mathbf{v}'_1 & \cdots & \mathbf{v}'_n \end{bmatrix} (M'M)$$

另一方面，我们有: $\begin{bmatrix} \mathbf{v}'_1 & \cdots & \mathbf{v}'_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}'_1 & \cdots & \mathbf{v}'_n \end{bmatrix} I$，即 I 也是 $\overline{\mathbf{v}'}$ 到 $\overline{\mathbf{v}'}$ 的过渡矩阵。根据过渡矩阵的唯一性，我们有:

$$M'M = I$$

$\square$

定理 237.

令 $\overline{\mathbf{v}}n$ 和 $\overline{\mathbf{v}'}$ 是 $\mathbb{V}$ 的两组基，则对于任意的 $\mathbf{u} \in \mathbb{V}$，我们有：

$$T_{\overline{\mathbf{v}}}(\mathbf{u}) = MT_{\overline{\mathbf{v}'}}(\mathbf{u})$$

证明. 注意到：

$$\begin{aligned}\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} T_{\overline{\mathbf{v}}}(\mathbf{u}) = \mathbf{u} = \begin{bmatrix}\mathbf{v}'_1 & \cdots & \mathbf{v}'_n\end{bmatrix} T_{\overline{\mathbf{v}'}}(\mathbf{u}) \\ = \left(\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} M\right) T_{\overline{\mathbf{v}'}}(\mathbf{u}) \\ = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} (MT_{\overline{\mathbf{v}'}}(\mathbf{u}))\end{aligned}$$

从而由 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是线性无关的，我们有：

$$T_{\overline{\mathbf{v}}}(\mathbf{u}) = MT_{\overline{\mathbf{v}'}}(\mathbf{u})$$

□

给定 $\mathbf{v} \in \mathbb{V}$:

- 其在基 $\overline{\mathbf{v}}$ 下的坐标为:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- 其在基 $\overline{\mathbf{v}'}$ 下的坐标为:

$$\begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$$

则 $T_{\overline{\mathbf{v}}}(\mathbf{v}) = MT_{\overline{\mathbf{v}'}}(\mathbf{v})$ 告诉我们:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = M \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$$

阶段总结

- 线性变换的基本概念。
- 利用坐标来表示向量空间的元素，线性变换 T_V.
- 利用矩阵来表示基变换下的不同坐标。

线性变换的矩阵形式

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是两个向量空间，并且 $\dim(\mathbb{V}) = n$，$\dim(\mathbb{W} = m)$ 特别的：

$$\overline{\mathsf{v}} = \mathsf{v}_1, \cdots, \mathsf{v}_n \text{ 和 } \overline{\mathsf{w}} = \mathsf{w}_1, \cdots, \mathsf{w}_m$$

分别是 $\mathbb{V}$ 和 $\mathbb{W}$ 的一组基。

定义 $T : \mathbb{V} \longrightarrow \mathbb{W}$ 是一个线性变换，这意味着对于 $i \in [n]$, $j \in [m]$，存在 $a_{ij} \in \mathbb{R}$ 使得：

$$\begin{aligned} T(\mathsf{v}_1) &= a_{11}\mathsf{w}_1 + \cdots + a_{1m}\mathsf{w}_m \\ T(\mathsf{v}_2) &= a_{21}\mathsf{w}_1 + \cdots + a_{2m}\mathsf{w}_m \\ &\vdots \\ T(\mathsf{v}_n) &= a_{m1}\mathsf{w}_1 + \cdots + a_{nm}\mathsf{w}_m \end{aligned}$$

给定 $\mathbf{v} \in \mathbb{V}$，则存在 $c_1, \cdots, c_n \in \mathbb{R}$:

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

从而:

$$\begin{aligned}
T(\mathbf{v}) &= T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\
&= c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) \\
&= c_1(a_{11}\mathbf{w}_1 + \cdots + a_{1m}\mathbf{w}_m) + \cdots + c_n(a_{m1}\mathbf{w}_1 + \cdots + a_{nm}\mathbf{w}_m) \\
&= (a_{11}c_1 + \cdots + a_{m1}c_n)\mathbf{w}_1 + \cdots + (a_{1m}c_1 + \cdots + a_{nm}c_n)\mathbf{w}_m \\
&= \sum_{j\in[m]} \left(\sum_{i\in[n]} c_i a_{ij} \right) \mathbf{w}_j
\end{aligned}$$

即 $T(\mathbf{v})$ 在基 $\mathbf{w}$ 下的坐标为:

$$\left(\sum_{i\in[n]} c_i a_{i1}, \cdots, \sum_{i\in[n]} c_i a_{im} \right)$$

定义下列矩阵:

$$A_T = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{nM} \end{bmatrix}$$

则 $T(\mathbf{v})$ 在基 $\overline{\mathbf{w}}$ 下的坐标为:

$$A_T \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

定义 238.

A_T 被称作线性变换 T 在基 $\overline{\mathbf{v}}$ 和 $\overline{\mathbf{w}}$ 下的矩阵。

引理 239.

1. $$\begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_m \end{bmatrix} A_T$$
2. A_T 是唯一的，即如果存在 $m \times n$ 的矩阵 B 使得:
$$\begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_m \end{bmatrix} B$$
则我们有 $B = A_T$。

证明.

1. 由 A_T 的定义可直接得到。
2. 由 $T(\mathbf{v}_j) = \sum_{i\in[n]} a_{ji}\mathbf{w}_i$ 的唯一性可以得到。

$\square$

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是两个向量空间，并且 $\dim(\mathbb{V}) = n$，$\dim(\mathbb{W} = m)$ 特别的:

$$\overline{\mathbf{v}} = \mathbf{v}_1, \cdots, \mathbf{v}_n \text{ 和 } \overline{\mathbf{w}} = \mathbf{w}_1, \cdots, \mathbf{w}_m$$

分别是 $\mathbb{V}$ 和 $\mathbb{W}$ 的一组基。我们有:

定理 240.

令 $\mathbf{v} \in \mathbb{V}$ 和 $\mathbf{w} \in \mathbb{W}$。若 $\mathbf{x} \in \mathbb{R}^n$ 是 $\mathbf{v}$ 在基 $\overline{\mathbf{v}}$ 下的坐标，$\mathbf{y} \in \mathbb{R}^m$ 是 $\mathsf{T}(\mathbf{v})$ 在基 $\overline{\mathbf{w}}$ 下的坐标，则我们有:

$$\mathsf{T}(\mathbf{v}) = \mathbf{w} \quad \Longleftrightarrow \quad \mathbf{y} = A_{\mathsf{T}}\mathbf{x}$$

这说明，在给定的基下，线性变换可以用一个大小确定的矩阵来表示。

定理240的证明

不妨令 $\mathbf{x}=(x_1,\cdots,x_n)$，$\mathbf{y}=(y_1,\cdots,y_n)$，从而：

$$\mathbf{v}=\sum_{i\in[n]} x_i\mathbf{v}_i,\quad \mathbf{w}=\sum_{j\in[m]} y_j\mathbf{w}_j$$

若 $\mathsf{T}(\mathbf{v})=\mathbf{w}$，则我们有：

$$\begin{aligned}\mathsf{T}(\mathbf{v})=\mathsf{T}\left(\sum_{i\in[n]} x_i\mathbf{v}_i\right)&=\sum_{i\in[n]} x_i\mathsf{T}(\mathbf{v}_i)\\ &=\begin{bmatrix}\mathsf{T}(\mathbf{v}_1) & \cdots & \mathsf{T}(\mathbf{v}_n)\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=\begin{bmatrix}\mathsf{T}(\mathbf{v}_1) & \cdots & \mathsf{T}(\mathbf{v}_n)\end{bmatrix}\mathbf{x}\\ &=\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_m\end{bmatrix}A_{\mathsf{T}}\mathbf{x}=\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_m\end{bmatrix}(A_{\mathsf{T}}\mathbf{x})\end{aligned}$$

另一方面，注意到 $\mathbf{w}_1,\cdots,\mathbf{w}_n$ 是线性无关的：

$$\mathbf{w}=\sum_{j\in[m]} y_j\mathbf{w}_j=\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_m\end{bmatrix}\mathbf{y}=(\begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_m\end{bmatrix}A_{\mathsf{T}})\mathbf{x}=\begin{bmatrix}\mathsf{T}(\mathbf{v}_1) & \cdots & \mathsf{T}(\mathbf{v}_n)\end{bmatrix}\mathbf{x}$$

令 $\mathbb{U},\mathbb{V},\mathbb{W}$ 是三个向量空间，其基分别为 $\bar{\mathbf{u}},\bar{\mathbf{v}},\bar{\mathbf{w}}$。

引理 241.

令 $S:\mathbb{U}\longrightarrow\mathbb{V}$ 和 $T:\mathbb{V}\longrightarrow\mathbb{W}$ 是两个线性变换。

- 定义 $TS:\mathbb{U}\longrightarrow\mathbb{W}$ 为:
$$TS(\mathbf{u})=T(S(\mathbf{u}))$$
则 TS 是一个线性变换。
- 若 A_S 和 A_T 分别是 S 和 T 在基 $\bar{\mathbf{u}},\bar{\mathbf{v}}$ 和 $\bar{\mathbf{v}},\bar{\mathbf{w}}$ 下的矩阵，则 A_{TS} 是 TS 在基 $\bar{\mathbf{u}},\bar{\mathbf{w}}$ 下的矩阵，则:
$$A_{TS}=A_T A_S$$
即 $A_T A_S$ 是 TS 在基 $\bar{\mathbf{u}},\bar{\mathbf{w}}$ 下的矩阵。

一个例子-旋转的复合 (I)

考察 $\mathbb{R}^2$ 上的旋转变换:

$$\mathsf{T} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

其将向量逆时针旋转 θ 角度，从而我们有:

$$\mathsf{T}(\mathbf{v}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \mathbf{v}$$

我们使用如下的标准基:

$$\bar{\mathbf{e}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

则:

$$A_{\mathsf{T}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

现在考察 $\mathbb{R}^2$ 上的两个旋转变换 (即令 $\mathbb{U}=\mathbb{V}=\mathbb{W}=\mathbb{R}^2$)，令 S 将向量逆时针旋转 θ 角度，T 将向量逆时针旋转 δ 角度，即:

$$S(\mathbf{v})=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\mathbf{v},\quad T(\mathbf{v})=\begin{bmatrix}\cos\delta & -\sin\delta\\ \sin\delta & \cos\delta\end{bmatrix}\mathbf{v}$$

从而:

$$\begin{aligned}TS(\mathbf{v})=T(S(\mathbf{v}))&=\begin{bmatrix}\cos\delta & -\sin\delta\\ \sin\delta & \cos\delta\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\mathbf{v}\\ &=\begin{bmatrix}\cos(\theta+\delta) & -\sin(\theta+\delta)\\ \sin(\theta+\delta) & \cos(\theta+\delta)\end{bmatrix}\mathbf{v}\end{aligned}$$

而在标准基 $\mathbf{e}$ 下我们有:

$$A_{TS}=\begin{bmatrix}\cos(\theta+\delta) & -\sin(\theta+\delta)\\ \sin(\theta+\delta) & \cos(\theta+\delta)\end{bmatrix}=\begin{bmatrix}\cos\delta & -\sin\delta\\ \sin\delta & \cos\delta\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}=A_TA_S$$

我们先证明 TS 是一个线性变换。给定 $\mathbf{u}, \mathbf{u}' \in \mathbb{U}$ 和 $c \in \mathbb{R}$:

$$\begin{aligned} TS(\mathbf{u}+\mathbf{u}') &= T(S(\mathbf{u}+\mathbf{u}')) = T(S(\mathbf{u})+S(\mathbf{u}')) = T(S(\mathbf{u}))+T(S(\mathbf{u}')) = TS(\mathbf{u})+TS(\mathbf{u}') \\ TS(c\mathbf{u}) &= T(S(c\mathbf{u})) = T(cS(\mathbf{u})) = cT(S(\mathbf{u})) = cTS(\mathbf{u}) \end{aligned}$$

从而 TS 是一个线性变换。

另一方面，令 $TS(\mathbf{u})$ 在 $\overline{\mathbf{w}}$ 下的坐标为 $\mathbf{z}$，$S(\mathbf{u})$ 在 $\overline{\mathbf{v}}$ 下的坐标为 $\mathbf{y}$，$\mathbf{u}$ 在 $\overline{\mathbf{u}}$ 下的坐标为 $\mathbf{x}$，则我们有:

$$\begin{aligned} TS(\mathbf{u}) = T(S(\mathbf{u})) &\implies \mathbf{z} = A_T\mathbf{y} \\ S(\mathbf{u}) = S(\mathbf{u}) &\implies \mathbf{y} = A_S\mathbf{x} \end{aligned}$$

从而我们有:

$$\mathbf{z} = A_TA_S\mathbf{x}$$

即:

$$A_{TS} = A_TA_S$$

□

目前为止对线性变换和矩阵的认知

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是两个向量空间，并且 $\dim(\mathbb{V}) = n$，$\dim(\mathbb{W} = m)$，特别的令其基分别为 $\overline{\mathbf{v}}$ 和 $\overline{\mathbf{w}}$。

- 我们定义了线性变换，特别的，$\mathbb{V}$ 到 $\mathbb{W}$ 上所有的线性变换为如下的集合：

$$\mathsf{T}(\mathbb{V}, \mathbb{W}) := \{\mathsf{T} \mid \mathsf{T} : \mathbb{V} \longrightarrow \mathbb{W} \text{ 是一个线性变换。}\}$$

  通过合适的加法运算和数乘运算，$\mathsf{T}(\mathbb{V}, \mathbb{W})$ 是一个向量空间。

- 考虑所有的 $m \times n$ 的实矩阵：

$$\mathbb{M}_{m\times n}(\mathbb{R}) = \{A \mid A\text{是一个 } m \times n \text{ 的实矩阵。}\}$$

- 令 $\mathsf{T} \in \mathsf{T}(\mathbb{V}, \mathbb{W})$，$A_{\mathsf{T}} \in \mathbb{M}_{m\times n}(\mathbb{R})$ 存在且唯一，满足：

$$\mathsf{T}(\mathbf{v}) = \mathbf{w} \quad \Longleftrightarrow \quad \mathbf{y} = A_{\mathsf{T}}\mathbf{x}$$

这其实说明了，在给定的基下，线性变换和矩阵是一一对应的。

定理 242.

下述映射：

$$T \mapsto A_T$$

是 $T(\mathbb{V},\mathbb{W})$ 到 $\mathbb{M}_{m\times n}(\mathbb{R})$ 的一一映射。

证明. 令 $T_1, T_2 \in T(\mathbb{V},\mathbb{W})$，并且假设：

$$A_{T_1} = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{nm} \end{bmatrix}, \quad A_{T_2} = \begin{bmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1m} & \cdots & b_{nm} \end{bmatrix}$$

即：

$$\begin{aligned} T_1(\mathbf{v}_1) &= a_{11}\mathbf{w}_1 + \cdots + a_{1m}\mathbf{w}_m \\ &\vdots \\ T_1(\mathbf{v}_n) &= a_{n1}\mathbf{w}_1 + \cdots + a_{nm}\mathbf{w}_m \end{aligned}, \quad \begin{aligned} T_2(\mathbf{v}_1) &= b_{11}\mathbf{w}_1 + \cdots + b_{1m}\mathbf{w}_m \\ &\vdots \\ T_2(\mathbf{v}_n) &= b_{n1}\mathbf{w}_1 + \cdots + b_{nm}\mathbf{w}_m \end{aligned}$$

定理242的证明续. 若 $T_1 \neq T_2$，则存在 $i \in [n]$ 使得

$$T_1(\mathbf{v}_i) \neq T_2(\mathbf{v}_i)$$

从而 $(a_{i1}, \cdots, a_{im}) \neq (b_{i1}, \cdots, b_{im})$，即 $A_{T_1} \neq A_{T_2}$。从而映射 $T \mapsto A_T$ 是单射。

另一方面，给定 $A \in \mathbb{M}_{m\times n}(\mathbb{R})$，定义：

$$\begin{aligned}
\mathbf{u}_i &= a_{11}\mathbf{w}_1 + \cdots + a_{1m}\mathbf{w}_m \\
&\vdots \\
\mathbf{u}_n &= a_{n1}\mathbf{w}_1 + \cdots + a_{nm}\mathbf{w}_m
\end{aligned}$$

定义 $T : \mathbb{V} \longrightarrow \mathbb{W}$ 满足对任意的 $i \in [n]$ 有 $T(\mathbf{v}_i) = \mathbf{u}_i$。则 $T \in \mathcal{T}(\mathbb{V}, \mathbb{W})$ 且 $A_T = A$。即映射 $T \mapsto A_T$ 是满射。 $\square$

可以看到 A_T 是依赖选择的基的。如果选择不同的基会发生什么?

定理 243.

令 $\mathbb{V}$ 是一个 n 维的向量空间，$\overline{\mathbf{v}}$ 和 $\overline{\mathbf{v}'}$ 是 $\mathbb{V}$ 的两组基。考虑 $\mathbb{V} \longrightarrow \mathbb{V}$ 的一个线性变换 T，令:

- M 是 $\overline{\mathbf{v}}$ 到 $\overline{\mathbf{v}'}$ 的过渡矩阵。
- A 是在基 $\overline{\mathbf{v}}$ 下 T 对应的矩阵。
- B 是在基 $\overline{\mathbf{v}'}$ 下 T 对应的矩阵。

则我们有:

$$B = M^{-1}AM$$

相似矩阵!

上述定理说明了，在同一个向量空间上不同基下的线性变换对应的矩阵是相似的。

定理243的证明 (I)

M 是 $\overline{\mathbf{v}}$ 到 $\overline{\mathbf{v}'}$ 的过渡矩阵，从而:

$$\begin{bmatrix}\mathbf{v}_1' & \cdots & \mathbf{v}_n'\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} M$$

M 是可逆的，因此:

$$\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1' & \cdots & \mathbf{v}_n'\end{bmatrix} M^{-1}$$

由 A, B 的定义我们有:

$$\begin{bmatrix}T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n)\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} A$$
$$\begin{bmatrix}T(\mathbf{v}_1') & \cdots & T(\mathbf{v}_n')\end{bmatrix} = \begin{bmatrix}\mathbf{v}_1' & \cdots & \mathbf{v}_n'\end{bmatrix} B$$

我们将先证明:

$$\begin{bmatrix}T(\mathbf{v}_1') & \cdots & T(\mathbf{v}_n')\end{bmatrix} = \begin{bmatrix}T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n)\end{bmatrix} M$$

令：

$$M = \begin{bmatrix} m_{11} & \cdots & m_{n1} \\ \vdots & \ddots & \vdots \\ m_{1n} & \cdots & m_{nn} \end{bmatrix}$$

从而对任意的 $i \in [n]$，我们有：

$$\mathbf{v}'_i = \sum_{j \in [n]} m_{ij} \mathbf{v}_j$$

从而：

$$T(\mathbf{v}'_i) = T\left(\sum_{j \in [n]} m_{ij} \mathbf{v}_j\right) = \sum_{j \in [n]} m_{ij} T(\mathbf{v}_j)$$

即：

$$\begin{bmatrix} T(\mathbf{v}'_1) & \cdots & T(\mathbf{v}'_n) \end{bmatrix} = \begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} \begin{bmatrix} m_{11} & \cdots & m_{n1} \\ \vdots & \ddots & \vdots \\ m_{1n} & \cdots & m_{nn} \end{bmatrix} = \begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} M$$

从而我们有:

$$\begin{aligned}\begin{bmatrix} T(\mathbf{v}_1') & \cdots & T(\mathbf{v}_n') \end{bmatrix} &= \begin{bmatrix} T(\mathbf{v}_1) & \cdots & T(\mathbf{v}_n) \end{bmatrix} M \\ &= \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} AM \\ &= \begin{bmatrix} \mathbf{v}_1' & \cdots & \mathbf{v}_n' \end{bmatrix} M^{-1}AM \\ &= \begin{bmatrix} \mathbf{v}_1' & \cdots & \mathbf{v}_n' \end{bmatrix} (M^{-1}AM)\end{aligned}$$

注意到:

$$\begin{bmatrix} T(\mathbf{v}_1') & \cdots & T(\mathbf{v}_n') \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1' & \cdots & \mathbf{v}_n' \end{bmatrix} B$$

由线性变换矩阵的唯一性可知:

$$B = M^{-1}AM$$

$\square$

我们知道基的选择会影响一个线性变换对应的矩阵。考察 n 维空间 $\mathbb{V}$ 上的一个线性变换，令其一组基为 $\bar{\mathbf{v}} = \mathbf{v}_1, \cdots, \mathbf{v}_n$. 若对于每一个 $i \in [n]$，都存在 $\lambda_i \in \mathbb{R}$ 使得:

$$T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$$

则我们有:

$$A_T = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

与矩阵相类似，我们也称满足 $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ 的 λ_i 和 $\mathbf{v}_i$ 为 T 上的特征值和特征向量。

若尔当标准型 (Jordan Form)

根据之前的学习，我们知道并不是所有矩阵都可以对角化。同样的，对于线性变换，我们也不是所有的线性变换都可以找到一组特征向量作为基。但我们可以寻找尽可能接近对角化的形式，这就是若尔当标准型 (Jordan Form)，但我们在这里不多叙述。。

阶段总结

线性变换的矩阵形式

1. 在给定的基下，线性变换可以用一个唯一的矩阵来表示。
2. 线性变换的复合就是矩阵的乘积。
3. 在给定的基下，线性变换和矩阵是一一对应的。
4. 不同基下的线性变换对应的矩阵是相似的。

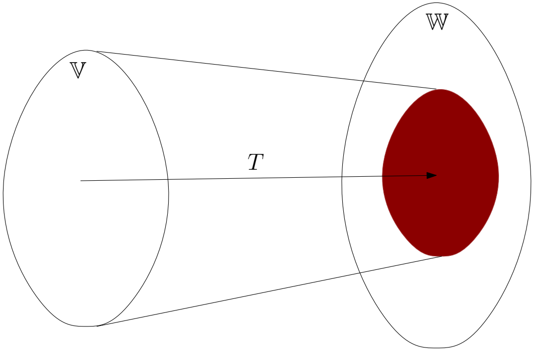
线性变换的像和核

本节我们固定两个向量空间 $\mathbb{V}$ 和 $\mathbb{W}$，并且考虑 $\mathbb{V}$ 到 $\mathbb{W}$ 上的一个线性变换 T。

定义 244 **[像 (Image)].**

T 的像 (Image) 是:

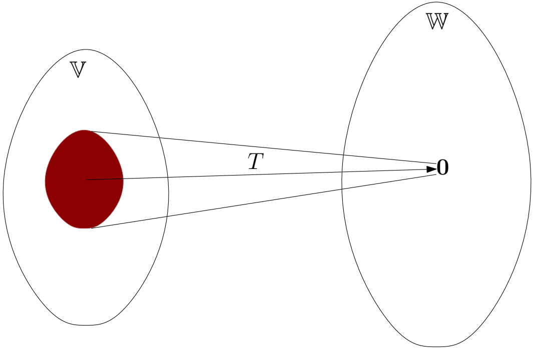
$$\mathrm{Im}(T) := \{T(\mathbf{v}) \mid \mathbf{v} \in \mathbb{V}\}$$



定义 245 **[核 (Kernel)].**

T 的核 (Kernel) 是:

$$\mathrm{Ker}(\mathrm{T}) := \{\mathbf{v} \in \mathbb{V} \mid \mathrm{T}(\mathbf{v}) = \mathbf{0}\}$$



引理 246.

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是两个向量空间，$\mathrm{T} : \mathbb{V} \longrightarrow \mathbb{W}$ 是一个线性变换。则：

1. $\mathrm{Im}(\mathrm{T})$ 是 $\mathbb{W}$ 的一个子空间。
2. $\mathrm{Ker}(\mathrm{T})$ 是 $\mathbb{V}$ 的一个子空间。

考虑一个 $m \times n$ 的矩阵 A:

$$T_A(\mathbf{x}) = A\mathbf{x}$$

是一个 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 上的线性变换。我们有:

$$\operatorname{Im}(T_A) = \{T_A(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\} = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} = \mathbf{C}(A)$$

$$\operatorname{Ker}(T_A) = \{\mathbf{x} \mid T_A(\mathbf{x}) = \mathbf{0}\} = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} = \mathbf{N}(A)$$

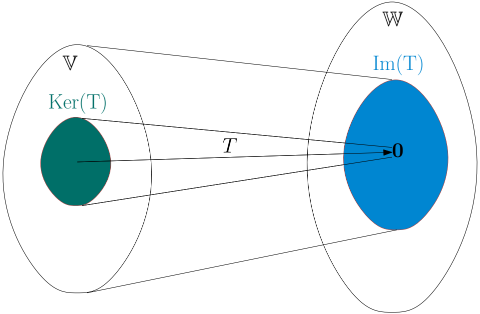
推论 247.

1. 线性变换 T 是单射当且仅当:
$$\operatorname{Ker}(T) = \{\mathbf{0}\}$$
2. 假设 $\mathbb{W}$ 是有限维的，线性变换 T 是满射当且仅当:
$$\dim(\operatorname{Im}(T)) = \dim(\mathbb{W})$$

定理 248 **[Rank-Nulity 定理].**

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是两个有限维的向量空间，$T : \mathbb{V} \longrightarrow \mathbb{W}$ 是一个线性变换。则：

$$\dim(\mathbb{V}) = \dim(\mathrm{Im}(T)) + \dim(\mathrm{Ker}(T))$$



推论 249.

令 A 是 $m \times n$ 的矩阵，则：

$$n = \dim(\mathbf{C}(A)) + \dim(\mathbf{N}(A)) = \mathrm{rank}(A) + \dim(\mathbf{N}(A))$$

Rank-Nullity 定理的证明 (I)

假设 $\mathbb{V}$ 存在一组基 $\overline{\mathbf{v}} = \mathbf{v}_1, \cdots, \mathbf{v}_n$，特别的 $\dim(\mathbb{V}) = n$。注意到:

$$\mathrm{Im}(T) = \mathrm{span}\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$$

从而 $\{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ 中的一组极大的线性无关的子集构成了 $\mathrm{Im}(T)$ 的一组基。令其为:

$$\overline{\mathbf{w}} = \mathbf{w}_1, \cdots, \mathbf{w}_m$$

从而 $\dim(\mathrm{Im}(T)) = m$，不妨令对于所有的 $i \in [m]$:

$$T(\mathbf{v}_i) = \mathbf{w}_i$$

否则我们重新排列 $\mathbf{v}_i$ 的顺序。

Ker(T) 是 $\mathbb{V}$ 的子空间，从而 Ker(T) 存在一组基：

$$\mathbf{u}_1, \cdots, \mathbf{u}_p$$

其中 $p \leqslant n$。

我们下面证明：

$$\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{u}_1, \cdots, \mathbf{u}_p \text{是 } \mathbb{V} \text{ 中的一组基。}$$

若上述命题成立，则立马可以得到：

$$\dim(\mathbb{V}) = \dim(\mathrm{Im}(T)) + \dim(\mathrm{Ker}(T))$$

引理 250.

$\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{u}_1, \cdots, \mathbf{u}_p$ 是 $\mathbb{V}$ 中的一组基。

证明. 令 $\mathbf{v} \in \mathbb{V}$，则 $\mathrm{T}(\mathbf{v}) \in \mathrm{Im}(\mathrm{T})$，从而存在 $x_1, \cdots, x_n \in \mathbb{R}$ 使得:

$$\mathrm{T}(\mathbf{v}) = \sum_{i=1}^{m} x_i \mathbf{w}_i = \sum_{i=1}^{m} x_i \mathrm{T}(\mathbf{v}_i) = \mathrm{T}\left(\sum_{i=1}^{m} x_i \mathbf{v}_i\right)$$

从而:

$$\mathrm{T}\left(\mathbf{v} - \sum_{i=1}^{m} x_i \mathbf{v}_i\right) = \mathbf{0} \implies \mathbf{v} - \sum_{i=1}^{m} x_i \mathbf{v}_i \in \mathrm{Ker}(\mathrm{T})$$

即:

$$\mathbf{v} - \sum_{i=1}^{m} x_i \mathbf{v}_i \in \mathrm{span}\{\mathbf{u}_1, \cdots, \mathbf{u}_p\} \implies \mathbf{v} = \sum_{i=1}^{m} x_i \mathbf{v}_i + \sum_{i=1}^{p} y_i \mathbf{u}_i$$

从而我们有:

$$\mathbf{v} \in \mathrm{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{u}_1, \cdots, \mathbf{u}_p\} \implies \mathbb{V} = \mathrm{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{u}_1, \cdots, \mathbf{u}_p\}$$

上海师范大学 Shanghai Normal University

引理的证明续. 我们还需要证明$\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{u}_1, \cdots, \mathbf{u}_p$ 是线性无关的。假设存在 $x_1, \cdots, x_m, y_1, \cdots, y_p \in \mathbb{R}$ 使得:

$$\sum_{i=1}^{m} x_i \mathbf{v}_i + \sum_{i=1}^{p} y_i \mathbf{u}_i = \mathbf{0}$$

则:

$$\begin{aligned}\mathbf{0} &= T\left(\sum_{i=1}^{m} x_i \mathbf{v}_i + \sum_{i=1}^{p} y_i \mathbf{u}_i\right)\\ &= \sum_{i=1}^{m} x_i T(\mathbf{v}_i) + \sum_{i=1}^{p} y_i T(\mathbf{u}_i) = \sum_{i=1}^{m} x_i T(\mathbf{v}_i) = \sum_{i=1}^{m} x_i \mathbf{w}_i\end{aligned}$$

由 $\mathbf{w}_1, \cdots, \mathbf{w}_m$ 是线性无关的，从而对任意的 $i \in [m]$ 都有 $x_i = 0$，因此我们有:

$$\sum_{i=1}^{p} y_i \mathbf{u}_i = \mathbf{0}$$

由 $\mathbf{u}_1, \cdots, \mathbf{u}_j$ 是线性无关的，我们也可以得到 $y_1 = \cdots = y_p = 0$。从而 $\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{u}_1, \cdots, \mathbf{u}_p$ 是线性无关的。 □

我们来展现一个 Rank-Nullity 定理的应用。

令 A 是一个 $m \times n$ 的矩阵，B 是一个 $n \times l$ 的矩阵。我们有：

$$\operatorname{rank}(AB) \leqslant \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}$$

我们进一步将证明：

$$\operatorname{rank}(A) + \operatorname{rank}(B) \leqslant \operatorname{rank}(AB) + n$$

一个简短但神秘的证明

$$\begin{aligned}
\operatorname{rank}(AB) + n &= \operatorname{rank}\left(\begin{bmatrix} AB & O \\ O & I \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} AB & O \\ B & I \end{bmatrix}\right) \\
&= \operatorname{rank}\left(\begin{bmatrix} O & -A \\ B & I \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} B & I \\ O & -A \end{bmatrix}\right) \\
&= \operatorname{rank}\left(\begin{bmatrix} B & -I \\ O & A \end{bmatrix}\right) \geqslant \operatorname{rank}(A) + \operatorname{rank}(B)
\end{aligned}$$

从 Rank-Nullity 定理角度的证明 (I)

由 Rank-Nullity 定理我们有:

$$\mathrm{rank}(A) + \mathrm{rank}(B) \leqslant \mathrm{rank}(AB) + n$$

等价于:

$$(n - \dim(\mathbf{N}(A))) + (l - \dim(\mathbf{N}(B))) \leqslant (l - \dim(\mathbf{N}(AB))) + n$$

即要证:

$$\dim(\mathbf{N}(A)) + \dim(\mathbf{N}(B)) \geqslant \dim(\mathbf{N}(AB))$$

更加清楚的直观

其实从上述式子可以看到，这个不等式是比较显然的，因为 $\mathbf{N}(AB)$ 的维数显然不可能超过 $\mathbf{N}(A)$ 与 $\mathbf{N}(B)$ 的维数之和。

令 $\mathbf{x} \in \mathbb{R}^l$，则我们有：

$$\mathbf{x} \in \mathbf{N}(AB) \iff AB\mathbf{x} = \mathbf{0} \iff B\mathbf{x} \in \mathbf{N}(A)$$

定义：

$$W = \{B\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^l,\ B\mathbf{x} \in \mathbf{N}(A)\} \implies W = \mathbf{C}(B) \cap \mathbf{N}(A) \subseteq \mathbb{R}^n$$

从而 W 是 $\mathbf{N}(A)$ 的一个子空间。定义线性变换 $\mathrm{T} : \mathbf{N}(AB) \implies \mathbb{R}^n$：

$$\mathrm{T}(\mathbf{x}) = B\mathbf{x}$$

从而我们有：

$$\mathrm{Im}(\mathrm{T}) = W$$

$$\mathrm{Ker}(\mathrm{T}) = \{\mathbf{x} \in \mathbf{N}(AB) \mid B\mathbf{x} = 0\} = \mathbf{N}(B)$$

从而由 Rank-Nullity 定理我们有：

$$\dim(\mathbf{N}(AB)) = \dim(\mathrm{Im}(\mathrm{T})) + \dim(\mathrm{Ker}(\mathrm{T})) = \dim(W) + \dim(\mathbf{N}(B)) \leqslant \dim(\mathbf{N}(A)) + \dim(\mathbf{N}(B))$$

$\square$

阶段总结

- 线性变换的核与像，矩阵的零空间与列空间。
- Rank-Nullity 定理。

线性变换 (Linear Transformation)

对偶性 (Duality)

我们固定一个向量空间 $\mathbb{V}$。我们考察其到 $\mathbb{R}$ 上的线性映射

定义 251 **[线性泛函].**

一个 $\mathbb{V}$ 上的泛函 L 是一个线性映射 $f : \mathbb{V} \longrightarrow \mathbb{R}$，即：$L \in T(\mathbb{V}, \mathbb{R})$。

例 252.

1. $F : \mathbb{R}^3 \Longrightarrow \mathbb{R}$:
$$F(x, y, z) = 4x - 5y + 2z$$
是一个线性泛函。
2. 固定 $(c_1, \cdots, c_n) \in \mathbb{R}^n$，定义 $F : \mathbb{R}^n \Longrightarrow \mathbb{R}$:
$$F(\mathbf{x}) = \sum_{i=1}^{n} c_i x_i$$
是一个线性泛函。

定义 253 **[对偶空间 (Dual Space)].**

$\mathbb{V}$ 的对偶空间$\mathbb{V}'$是 $\mathbb{V}$ 上所有线性泛函的集合。即：

$$\mathbb{V}' = \mathsf{T}(\mathbb{V}, \mathbb{R})$$

引理 254.

$\mathbb{V}'$ 是一个向量空间。

我们先给出对偶基 (Dual Basis) 的概念。

定义 255 **[对偶基 (Dual Basis)].**

令

$$\mathbf{v}_1, \cdots, \mathbf{v}_n$$

是 V 中的一组基。则 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 的对偶基是如下的系列:

$$L_1, \cdots, L_n$$

其中，L_i 是 V 上的线性泛函，满足:

$$L_i(\mathbf{v}_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

我们下面证明，我们给出的对偶基的概念，恰恰是对偶空间的一组基。

定理 256.

假设 $\mathbb{V}$ 是有限维的，则 $\mathbb{V}'$ 中一组基的对偶基是 $\mathbb{V}'$ 的一组基。

推论 257.

$\dim(\mathbb{V}) = \dim(\mathbb{V}')$。

证明. 令 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是 $\mathbb{V}$ 的一组基，$L_1, \cdots, L_n$ 是其对偶基。假设存在 $c_1, \cdots, c_n \in \mathbb{R}$ 使得：

$$\mathsf{T} = c_1 L_1 + \cdots + c_n L_n = 0$$

则对任意的 $i \in [n]$：

$$0 = \mathsf{T}(\mathbf{v}_i) = (c_1 L_1 + \cdots + c_n L_n)(\mathbf{v}_i) = c_i L_i(\mathbf{v}_i) = c_i$$

从而 $c_1 = \cdots = c_n = 0$，即 $L_1, \cdots, L_n$ 是线性无关的。

证明续. 另一方面，考虑 $T \in \mathbb{V}' = T(\mathbb{V}, \mathbb{R})$，即 T 是一个 $\mathbb{V}$ 到 $\mathbb{R}$ 的线性映射。

对任意的 $i \in [n]$，令:

$$c_i = T(\mathbf{v}_i)$$

则定义:

$$T' = c_1 L_1 + \cdots + c_n L_n$$

显然 $T' \in T(\mathbb{V}, \mathbb{R}) = \mathbb{V}'$，并且对任意的 $i \in [n]$:

$$T'(\mathbf{v}_i) = c_i L_i(\mathbf{v}_i) = c_i = T(\mathbf{v}_i)$$

即 $T' = T$，从而:

$$\mathbb{V}' = \text{span}\{L_1, \cdots, L_n\}$$

$\square$

令 $\mathbb{V}$ 和 $\mathbb{W}$ 是两个向量空间。

定义 258 **[对偶变换 (Dual Transformation)].**

对于一个线性变换 $\mathsf{T} \in \mathcal{T}(\mathbb{V}, \mathbb{W})$，定义其对偶变换 $\mathsf{T}' \in (\mathbb{W}', \mathbb{V}')$:

$$\mathsf{T}'(\mathrm{L}) = \mathrm{L}\mathsf{T}, \text{ 即对任意的 } \mathbf{v} \in \mathbb{V} \text{ 有 } \mathsf{T}'(\mathrm{L})(\mathbf{v}) = \mathrm{L}(\mathsf{T}(\mathbf{v}))$$

我们要说明两件事:

1. 对任意的 $\mathrm{L} \in \mathbb{W}'$，$\mathsf{T}'(\mathrm{L})$ 是 $\mathbb{V}'$ 上的线性泛函，即是 $\mathbb{V}$ 到 $\mathbb{R}$ 上的线性变换。
2. T' 是一个 $\mathbb{W}'$ 到 $\mathbb{V}'$ 上的线性变换。

注意到:

$$T'(L) = LT, \text{ 即对任意的 } \mathbf{v} \in \mathbb{V} \text{ 有 } T'(L)(\mathbf{v}) = L(T(\mathbf{v}))$$

我们只需验证 $T'(L)$ 是 $\mathbb{V}'$ 到 $\mathbb{R}$ 上的线性变换即可。

- 向量加法: 令 $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{V}$, 则:

$$\begin{aligned} T'(L)(\mathbf{v}_1 + \mathbf{v}_2) &= L(T(\mathbf{v}_1 + \mathbf{v}_2)) = L(T(\mathbf{v}_1) + T(\mathbf{v}_2)) \\ &= L(T(\mathbf{v}_1)) + L(T(\mathbf{v}_2)) = T'(L)(\mathbf{v}_1) + T'(L)(\mathbf{v}_2) \end{aligned}$$

- 数乘: 令 $\mathbf{v} \in \mathbb{V}$, $c \in \mathbb{R}$, 则:

$$T'(L)(c\mathbf{v}) = L(T(c\mathbf{v})) = L(cT(\mathbf{v})) = cL(T(\mathbf{v})) = cT'(L)(\mathbf{v})$$

注意到：

$$T'(L) = LT\text{，即对任意的 } \mathbf{v} \in \mathbb{V} \text{ 有 } T'(L)(\mathbf{v}) = L(T(\mathbf{v}))$$

- 向量加法：令 $L_1, L_2 \in \mathbb{W}' = T(\mathbb{W}, \mathbb{R})$，则对于任意的 $\mathbf{v} \in \mathbb{V}$:

$$\begin{aligned} T'(L_1 + L_2)(\mathbf{v}) &= (L_1 + L_2)(T(\mathbf{v})) = L_1(T(\mathbf{v})) + L_2(T(\mathbf{v})) \\ &= T'(L_1)(\mathbf{v}) + T'(L_2)(\mathbf{v}) \end{aligned}$$

即 $T'(L_1 + L_2) = T'(L_1) + T'(L_2)$.

- 数乘：令 $L \in \mathbb{W}' = T(\mathbb{W}, \mathbb{R})$，$c \in \mathbb{R}$，则对于任意的 $\mathbf{v} \in \mathbb{V}$:

$$T'(cL)(\mathbf{v}) = (cL)(T(\mathbf{v})) = cL(T(\mathbf{v})) = cT'(L)(\mathbf{v})$$

即 $T'(cL) = cT'(L)$.

上海师范大学 Shanghai Normal University

给定两个向量空间 $\mathbb{V}$ 和 $\mathbb{W}$:

- 令 $\bar{\mathbf{v}} = \mathbf{v}_1 \cdots \mathbf{v}_n$ 是 $\mathbb{V}$ 的一组基，$L_1^{\bar{\mathbf{v}}}, \cdots, L_n^{\bar{\mathbf{v}}}$ 是其对偶基，也是 $\mathbb{V}'$ 的一组基。
- 令 $\bar{\mathbf{w}} = \mathbf{w}_1 \cdots \mathbf{w}_m$ 是 $\mathbb{W}$ 的一组基，$L_1^{\bar{\mathbf{w}}}, \cdots, L_m^{\bar{\mathbf{w}}}$ 是其对偶基，也是 $\mathbb{W}'$ 的一组基。

定理 259.

令 $\mathsf{T} \in \mathsf{T}(\mathbb{V}, \mathbb{W})$ 并且 A_{T} 是 T 在基 $\bar{\mathbf{v}}$ 和 $\bar{\mathbf{w}}$ 下的矩阵。令 $A_{\mathsf{T}'}$ 是其对偶变换 $\mathsf{T}' \in \mathsf{T}(\mathbb{W}', \mathbb{V}')$ 在基 $L_1^{\bar{\mathbf{w}}}, \cdots, L_m^{\bar{\mathbf{w}}}$ 和 $L_1^{\bar{\mathbf{v}}}, \cdots, L_n^{\bar{\mathbf{v}}}$ 下的矩阵。则:

$$A_{\mathsf{T}'} = A_{\mathsf{T}}^{\mathsf{T}}$$

记

$$A_T = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{nm} \end{bmatrix}, \qquad A_{T'} = \begin{bmatrix} b_{11} & \cdots & b_{m1} \\ \vdots & \ddots & \vdots \\ b_{1n} & \cdots & b_{mn} \end{bmatrix}$$

即:

$$\begin{aligned} T(\mathbf{v}_1) &= a_{11}\mathbf{w}_1 + \cdots + a_{1m}\mathbf{w}_m \\ T(\mathbf{v}_2) &= a_{21}\mathbf{w}_1 + \cdots + a_{2m}\mathbf{w}_m \\ &\vdots \\ T(\mathbf{v}_n) &= a_{n1}\mathbf{w}_1 + \cdots + a_{nm}\mathbf{w}_m \end{aligned}, \qquad \begin{aligned} T'(L_1^{\mathbf{w}}) &= b_{11}L_1^{\mathbf{v}} + \cdots + b_{1n}L_n^{\mathbf{v}} \\ T'(L_2^{\mathbf{w}}) &= b_{21}L_1^{\mathbf{v}} + \cdots + b_{2n}L_n^{\mathbf{v}} \\ &\vdots \\ T'(L_m^{\mathbf{w}}) &= b_{m1}L_1^{\mathbf{v}} + \cdots + b_{mn}L_n^{\mathbf{v}} \end{aligned}$$

注意到对偶基的定义:

$$L_i^{\mathbf{V}}(\mathbf{v}_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

从而对于任意的 $i \in [m]$, $j \in [n]$ 有:

$$T'(L_i^{\mathbf{W}})(\mathbf{v}_j) = (b_{i1}L_1^{\mathbf{V}} + \cdots + b_{in}L_n^{\mathbf{V}})(\mathbf{v}_j) = b_{ij}L_j^{\mathbf{V}}(\mathbf{v}_j) = b_{ij}$$

另一方面, 由定义可得:

$$\begin{aligned} T'(L_i^{\mathbf{W}})(\mathbf{v}_j) = L_i^{\mathbf{W}}(T(\mathbf{v}_j)) &= L_i^{\mathbf{W}}(a_{j1}\mathbf{w}_1 + \cdots + a_{jm}\mathbf{w}_m) \\ &= a_{j1}L_i^{\mathbf{W}}(\mathbf{w}_1) + \cdots + a_{jm}L_i^{\mathbf{W}}(\mathbf{w}_m) \\ &= a_{ji}L_i^{\mathbf{W}}(\mathbf{w}_i) = a_{ji} \end{aligned}$$

从而:

$$b_{ij} = T'(L_i^{\mathbf{W}})(\mathbf{v}_j) = a_{ji}$$

$\square$

现在我们来考虑一下对偶变换中的像与核。

定义 260 **[零化子 (Annihilator)].**

令 $\mathbb{V}$ 是一个向量空间，$\mathbb{U} \subseteq \mathbb{V}$ 是其子空间。$\mathbb{U}$ 的零化子，记作 $\mathbb{U}^0$，定义如下：

$$\mathbb{U}^0 = \{L \in \mathbb{V}' \mid L(\mathbf{u}) = 0, \ \forall \mathbf{u} \in \mathbb{U}\}$$

引理 261.

$\mathbb{U}^0$ 是 $\mathbb{V}'$ 的一个子空间。

$\mathbb{U}^0$ 是子空间的证明

- 显然零函数 $\mathbf{0} \in \mathbb{U}^0$。
- 零 L_1，$L_2 \in \mathbb{U}^0$, 则对任意的 $\mathbf{u} \in \mathbb{U}$，有:

$$(L_1 + L_2)(\mathbf{u}) = L_1(\mathbf{u}) + L_2(\mathbf{u}) = 0 + 0 = 0$$

  即 $L_1 + L_2 \in \mathbb{U}^0$。
- $L \in \mathbb{U}^0$，$c \in \mathbb{R}$，则对任意的 $\mathbf{u} \in \mathbb{U}$，有:

$$(cL)(\mathbf{u}) = cL(\mathbf{u}) = c \cdot 0 = 0$$

  即 $cL \in \mathbb{U}^0$。

$\square$

定理 262.

令 $\mathbb{V}$ 是一个有限维的向量空间，$\mathbb{U} \subseteq \mathbb{V}$ 是其子空间。则：

$$\dim(\mathbb{U}) + \dim(\mathbb{U}^0) = \dim(\mathbb{V})$$

证明. 我们使用如下的包含映射 (Inclusion Linear Transformation)：$\mathsf{T}_{\text{incl}} : \mathbb{U} \longrightarrow \mathbb{V}$：

$$\text{对任意的 } \mathbf{u} \in \mathbb{U} \text{ 有} \mathsf{T}_{\text{incl}}(\mathbf{u}) = \mathbf{u}$$

则其对偶映射为 $\mathsf{T}'_{\text{incl}} \in \mathsf{T}(\mathbb{V}', \mathbb{U}')$，则由 Rank-Nullity 定理有：

$$\dim(\mathbb{V}') = \dim(\mathsf{Im}(\mathsf{T}'_{\text{incl}})) + \dim(\mathsf{Ker}(\mathsf{T}'_{\text{incl}}))$$

我们将证明：

- $\mathsf{Im}(\mathsf{T}'_{\text{incl}}) = \mathbb{U}'$。
- $\mathsf{Ker}(\mathsf{T}'_{\text{incl}}) = \mathbb{U}^0$。

从而：

$$\dim(\mathbb{V}) = \dim(\mathbb{V}') = \dim(\mathbb{U}') + \dim(\mathbb{U}^0) = \dim(\mathbb{U}) + \dim(\mathbb{U}^0)$$

$\mathrm{Im}(T'_{\mathrm{incl}}) = \mathbb{U}'$

上海师范大学 Shanghai Normal University

令 $L \in \mathbb{U}' = T(\mathbb{U}, \mathbb{R})$。注意到 $\mathbb{V}$ 是有限维的，不妨令其一组基为：

$$\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{v}_{m+1}, \cdots, \mathbf{v}_n$$

注意到 $\mathbb{U}$ 是 $\mathbb{V}$ 的子空间，不妨进一步假设 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 是 $\mathbb{U}$ 的一组基。我们定义 $L_V \in T(\mathbb{V}, \mathbb{R})$：

$$L_V(\mathbf{v}_i) = \begin{cases} L(\mathbf{v}_i) & i \in [m] \\ 0 & i \in [m+1, n] \end{cases}$$

则对任意的 $\mathbf{u} \in \mathbb{U}$ 有：

$$T'_{\mathrm{incl}}(L_V)(\mathbf{u}) = L_V(T_{\mathrm{incl}}(\mathbf{u})) = L_V(\mathbf{u}) = L(\mathbf{u})$$

这里最后一个等号是因为 $\mathbf{u} \in \mathbb{U}$ 可以写成：

$$\mathbf{u} = u_1\mathbf{v}_1 + \cdots + u_m\mathbf{v}_m + 0\mathbf{v}_{m+1} + \cdots + 0\mathbf{v}_n$$

的形式，从而：

$$\mathrm{Im}(T'_{\mathrm{incl}}) = \mathbb{U}'$$

$\mathrm{Ker}(T'_{\mathrm{incl}}) = \mathbb{U}^0$

上海师范大学 Shanghai Normal University

对于 $\mathrm{Ker}(T'_{\mathrm{incl}})$ 有:

$$\begin{aligned}
\mathrm{Ker}(T'_{\mathrm{incl}}) &= \{L \in \mathbb{V}' \mid \mathrm{Ker}(T'_{\mathrm{incl}})(L) = \mathbf{0}\} \\
&= \{L \in \mathbb{V}' \mid \text{对任意的 } \mathbf{u} \in \mathbb{U} \text{ 有 } T'_{\mathrm{incl}}(L)(\mathbf{u}) = 0\} \\
&= \{L \in \mathbb{V}' \mid \text{对任意的 } \mathbf{u} \in \mathbb{U} \text{ 有 } L(T_{\mathrm{incl}}(\mathbf{u})) = 0\} \\
&= \{L \in \mathbb{V}' \mid \text{对任意的 } \mathbf{u} \in \mathbb{U} \text{ 有 } L(\mathbf{u}) = 0\} \\
&= \mathbb{U}^0
\end{aligned}$$

$\square$

定理 263.

假设 $\mathbb{V}$ 和 $\mathbb{W}$ 是有限维的向量空间，令 $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$，则：

1. $\operatorname{Ker}(T') = (\operatorname{Im}(T))^0$。
2. $\dim(\operatorname{Ker}(T')) = \dim(\operatorname{Ker}(T)) + \dim(\mathbb{W}) - \dim(\mathbb{V})$。

证明. 我们先利用第一个结论来证明第二个，注意到：

$$
\begin{aligned}
\dim(\operatorname{Ker}(T')) &= \dim((\operatorname{Im}(T))^0) \\
&= \dim(\mathbb{W}) - \dim(\operatorname{Im}(T)) \\
&= \dim(\mathbb{W}) - (\dim(\mathbb{V}) - \dim(\operatorname{Ker}(T))) \\
&= \dim(\operatorname{Ker}(T)) + \dim(\mathbb{W}) - \dim(\mathbb{V})
\end{aligned}
$$

$\mathrm{Ker}(T') = (\mathrm{Im}(T))^0$

- 令 $L \in \mathrm{Ker}(T')$，即 $T'(L) = \mathbf{0} \in \mathbb{V}' = T(\mathbb{V}, \mathbb{R})$，也就是对任意的 $\mathbf{v} \in \mathbb{V}$ 有

$$T'(L)(\mathbf{v}) = L(T(\mathbf{v})) = 0$$

从而对于任意的 $\mathbf{w} \in \mathrm{Im}(T)$，我们都有 $L(\mathbf{w}) = 0$，即 $L \in (\mathrm{Im}(T))^0$，即：

$$\mathrm{Ker}(T') \subseteq (\mathrm{Im}(T))^0$$

- 另一方面，考察 $L \in (\mathrm{Im}(T))^0$，这意味着对任意的 $\mathbf{w} \in \mathrm{Im}(T)$ 都有

$$L(\mathbf{w}) = 0$$

从而对任意的 $\mathbf{u} \in \mathbb{V}$:

$$T'(L)(\mathbf{u}) = L(T(\mathbf{u})) = 0$$

即：$L \in \mathrm{Ker}(T)'$，也就是$(\mathrm{Im}(T))^0 \subseteq \mathrm{Ker}(T')$。

$\square$

定理 264.

假设 $\mathbb{V}$ 和 $\mathbb{W}$ 是有限维的向量空间，令 $T \in T(\mathbb{V}, \mathbb{W})$，则：

1. $\dim(\mathrm{Im}(T')) = \dim(\mathrm{Im}(T))$.
2. $\mathrm{Im}(T') = (\mathrm{Ker}(T))^0$。

证明.

1. $\dim(\mathrm{Im}(T')) = \dim(\mathbb{W}') - \dim(\mathrm{Ker}(T')) = \dim(\mathbb{W}) - \dim((\mathrm{Im}(T))^0) = \dim(\mathrm{Im}(T))$.
2. 令 $L \in \mathrm{Im}(T') \subseteq \mathbb{V}'$，则存在 $L_w \in \mathbb{W}'$ 使得：$T'(L_w) = L$

   注意到对任意的 $\mathbf{u} \in \mathrm{Ker}(T)$ 我们有：

   $$L(\mathbf{u}) = T'(L_w)(\mathbf{u}) = L_w(T(\mathbf{u})) = L_w(\mathbf{0}) = 0$$

   从而$L \in (\mathrm{Ker}(T))^0$，即 $\mathrm{Im}(T)' \subseteq (\mathrm{Ker}(T))^0$，另一方面，我们有：

   $$\dim(\mathrm{Im}(T)') = \dim(\mathrm{Im}(T)) = \dim(\mathbb{V}) - \dim(\mathrm{Ker}(T)) = \dim((\mathrm{Ker}(T))^0)$$

   从而$\mathrm{Im}(T)' = (\mathrm{Ker}(T))^0$。

□

定理 265.

令 A 是一个 $m \times n$ 的矩阵，则：$\text{row-rank}(A) = \text{column-rank}(A)$.

证明. 定义 $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$:

$$T(\mathbf{v}) = A\mathbf{v}$$

特别的，在标准基 $\mathbf{e}_1, \cdots, \mathbf{e}_n$ 下的 T 的矩阵为 A。从而其对偶变换 T' 在对应的对偶基下的矩阵为

$$A^{\top}$$

因此：

$$\begin{aligned}\text{column-rank}(A) &= \dim(\mathbf{C}(A)) = \dim(\text{Im}(T)) \\ &= \dim(\text{Im}(T')) = \dim(\mathbf{C}(A^{\top})) = \text{column-rank}(A^{\top}) = \text{row-rank}(A)\end{aligned}$$

$\square$

阶段总结

- 线性泛函的概念。
- 对偶空间、对偶基。
- 对偶变换，和转置矩阵的关系。
- 对偶变换的核与像的性质。

奇异值分解 (Singular Value Decomposition)

奇异值分解 (Singular Value Decomposition)

SVD 基础

对图片的存储

假设一张图片被如下的矩阵表示:

$$A = \begin{bmatrix} a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \\ a & a & b & b & c & c \end{bmatrix}$$

直接存储需要 $6 \times 6 = 36$ 个数。注意到该矩阵的秩 $\mathrm{rank}(A) = 1$，从而:

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

从而我们只需要存储 12 个数。

注意到实对称矩阵可以对角化，即：

定理 266.

令 S 是一个实对称矩阵，$\lambda_1, \cdots, \lambda_n \in \mathbb{R}$ 是其特征值 (可能重复)。从而存在 n 个正交的向量 $\mathbf{v}_1, \cdots, \mathbf{v}_n \in \mathbb{R}^n$，使得：

$$S = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix} = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^\top + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^\top$$

从而：

引理 267.

令 S 是一个 $n \times n$ 的实对称矩阵，S 恰好有 $\mathrm{rank}(S)$ 个非零的特征值 (计算代数重数)。从而 S 可以被 $(n+1)\mathrm{rank}(S)$ 个数字表示。

引理的证明. 注意到 S 是实对称的，从而存在正交矩阵 Q 使得:

$$Q^\top SQ = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

因此有:

$$\operatorname{rank}(S) = \operatorname{rank}(\Lambda) = |\{i \in [n] \mid \lambda_i \neq 0\}|$$

令 $s = \operatorname{rank}(S)$，即存在 s 个正交的向量 $\mathbf{v}_1, \cdots, \mathbf{v}_s \in \mathbb{R}^n$ 使得:

$$S = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^\top + \cdots + \lambda_s \mathbf{v}_s \mathbf{v}_s^\top$$

即存储 S 需要 s 个特征值和 s 个特征向量，共 $(n+1)\operatorname{rank}(S)$ 个数字。 $\square$

实对称矩阵可以完成谱分解，那么对于任意的矩阵 $A \in \mathbb{R}^{m\times n}$，是否存在类似的分解呢？

答案就是奇异值分解 (Singular Value Decomposition, SVD).

定理 268.

令 A 是一个 $m \times n$ 的矩阵，$\mathrm{rank}(A) = r$，定义下列 $m \times n$ 的矩阵 Σ:

$$\Sigma = \begin{bmatrix} D & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix}, \text{其中} D = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

这里, $\sigma_1 \geqslant \cdots \geqslant \sigma_r$ 是 A 的奇异值。则存在 $m \times m$ 的正交矩阵 U 和 $n \times n$ 的正交矩阵 $V =$，使得:

$$A = U\Sigma V^T$$

特别的，存在 $\mathbf{u}_1, \cdots, \mathbf{u}_r \in \mathbb{R}^m$ 和 $\mathbf{v}_1, \cdots, \mathbf{v}_r \in \mathbb{R}^n$ 使得:

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

给定一个任意的矩阵 A，我们有：

$$(AA^\top)^\top = AA^\top,\ (A^\top A)^\top = A^\top A$$

即其都是对称矩阵，同时我们有：

引理 269.

$$\mathrm{rank}(AA^\top) = \mathrm{rank}(A^\top A) = \mathrm{rank}(A)$$

证明. 我们只需证明：

$$\mathrm{rank}(A^\top A) = \mathrm{rank}(A)$$

因为：

$$\mathrm{rank}(AA^\top) = \mathrm{rank}((A^\top)^\top A^\top) = \mathrm{rank}(A^\top) = \mathrm{rank}(A) = \mathrm{rank}(A^\top A)$$

def

$\text{rank}(A^\top A) = \text{rank}(A)$ 的证明

由 Rank-Nullity 定理，我们有：

$$\text{rank}(A^\top A) + \dim(\mathbf{N}(A^\top A)) = n = \text{rank}(A) + \dim(\mathbf{N}(A))$$

从而我们只需证明：

$$\dim(\mathbf{N}(A^\top A)) = \dim(\mathbf{N}(A))$$

- 任取 $\mathbf{x} \in \mathbf{N}(A)$，即 $A\mathbf{x} = \mathbf{0}$，从而有：

  $$A^\top A\mathbf{x} = A^\top \mathbf{0} = \mathbf{0}$$

  即$\mathbf{N}(A) \subseteq \mathbf{N}(A^\top A)$。
- 任取 $\mathbf{x} \in \mathbf{N}(A^\top A)$，即 $A^\top A\mathbf{x} = \mathbf{0}$，从而有：

  $$\mathbf{x}^\top A^\top A\mathbf{x} = 0 \implies \|A\mathbf{x}\| = 0 \implies A\mathbf{x} = \mathbf{0}$$

  即$\mathbf{N}(A^\top A) \subseteq \mathbf{N}(A)$。 $\square$

$A^{\top}A$ 和 $AA^{\top}$ 的特征值 (I)

现在我们来关注一下 $A^{\top}A$ 和 $AA^{\top}$ 的特征值的性质。

引理 270.

1. $A^{\top}A$ 和 $AA^{\top}$ 的特征值都是非负的。
2. $A^{\top}A$ 和 $AA^{\top}$ 是半正定矩阵。

证明. 这里只证明 $A^{\top}A$ 的特征值是非负的。令 λ 是 $A^{\top}A$ 的一个特征值，$\mathbf{v}$ 是对应的特征向量，即：

$$A^{\top}A\mathbf{v} = \lambda\mathbf{v}$$

从而有：

$$\lambda\|\mathbf{v}\|^2 = \lambda\mathbf{v}^{\top}\mathbf{v} = \mathbf{v}^{\top}(\lambda\mathbf{v}) = \mathbf{v}^{\top}A^{\top}A\mathbf{v} = (A\mathbf{v})^{\top}(A\mathbf{v}) = \|A\mathbf{v}\|^2 \geqslant 0$$

即：

$$\lambda \geqslant 0$$

□

引理 271.

令 $\operatorname{rank}(A) = r$，则 $A^\top A$ 和 $AA^\top$ 有相同的 r 个非零特征值 (计算重数)。

证明. 由前面的引理可知:

$$\operatorname{rank}(A^\top A) = \operatorname{rank}(A) = \operatorname{rank}(AA^\top) = r$$

令 $A^\top A$ 有如下 n 个特征值:

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$$

其对应的一组标准正交的特征向量为 $\mathbf{v}_1, \cdots, \mathbf{v}_n$。从而:

1. $\lambda_1, \cdots, \lambda_r > 0$。
2. $\lambda_{r+1} = \cdots = \lambda_n = 0$。
3. 对任意的 $i \in [n]$ 有 $A^\top A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ 且 $\|\mathbf{v}_i\| = 1$.
4. 对任意的 $1 \leqslant i < j \leqslant n$ 有 $\mathbf{v}_i \cdot \mathbf{v}_j = \mathbf{v}_i^\top \mathbf{v}_j = 0$。

考察如下的一组向量 $\bar{\mathbf{v}}$:

$$\bar{\mathbf{v}} := \frac{1}{\sqrt{\lambda_1}}A\mathbf{v}_1, \cdots, \frac{1}{\sqrt{\lambda_r}}\mathbf{v}_r$$

我们接下来证明:

1. $\bar{\mathbf{v}}$ 是标准正交的。
2. $\bar{\mathbf{v}}$ 是 $AA^{\top}$ 的特征值 $\lambda_1, \cdots, \lambda_r$ 的特征向量。
3. $AA^{\top}$ 和 $A^{\top}A$ 的特征值的代数重数相同。

从而最终得到:

$A^{\top}A$ 和 $AA^{\top}$ 由相同的 r 个非零特征值 (计算重数)。

$\bar{\mathbf{v}}$ 是标准正交的

上海师范大学 Shanghai Normal University

我们先证明:

$$\frac{1}{\sqrt{\lambda_1}}A\mathbf{v}_1, \cdots, \frac{1}{\sqrt{\lambda_r}}\mathbf{v}_r$$

是标准正交的。

- 对于 $i \in [r]$. 有:

$$\|\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\|^2 = \left(\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\right)^\top \frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i = \frac{1}{\lambda_i}\mathbf{v}_i^\top A^\top A\mathbf{v}_i = \frac{1}{\lambda_i}\mathbf{v}_i^\top \lambda_i \mathbf{v}_i = \|\mathbf{v}_i\|^2 = 1$$

- 对于 $1 \leqslant i < j \leqslant r$. 有:

$$\begin{aligned}\left(\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\right) \cdot \left(\frac{1}{\sqrt{\lambda_j}}A\mathbf{v}_j\right) &= \left(\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\right)^\top \frac{1}{\sqrt{\lambda_j}}A\mathbf{v}_j \\ &= \frac{1}{\sqrt{\lambda_i\lambda_j}}\mathbf{v}_i^\top A^\top A\mathbf{v}_j = \frac{1}{\sqrt{\lambda_i\lambda_j}}\mathbf{v}_i^\top \lambda_j \mathbf{v}_j = \sqrt{\frac{\lambda_j}{\lambda_i}}\mathbf{v}_i \cdot \mathbf{v}_j = 0\end{aligned}$$

我们再证明，对任意的 $i \in [r]$ 有：

$$AA^\top\left(\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\right) = \lambda_i\left(\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\right)$$

事实上，

$$\begin{aligned} AA^\top\left(\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\right) &= \frac{1}{\sqrt{\lambda_i}}A(A^\top A\mathbf{v}_i) \\ &= \frac{1}{\sqrt{\lambda_i}}A(\lambda_i\mathbf{v}_i) \\ &= \lambda_i\left(\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i\right) \end{aligned}$$

从而对 $A^\top A$ 任意的非零特征值 λ_i，λ_i 也是 $AA^\top$ 的特征值，$\frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i$ 是对应的特征向量。

现在令 $\lambda_{i_1},\cdots,\lambda_{i_t}$ 是 $A^\top A$ 的两两不同的非零特征值，其中:

$$1 = i_1 < i_2 < \cdots < i_t \leqslant r$$

注意到其也是 $AA^\top$ 的特征值，定义:

1. a_i 是 $A^\top A$ 的特征值 λ_i 的代数重数, a_i' 是 $AA^\top$ 的特征值 λ_i 的代数重数。
2. g_i 是 $A^\top A$ 的特征值 λ_i 的几何重数, g_i' 是 $AA^\top$ 的特征值 λ_i 的几何重数。

并且我们有:

$$\sum_{i\in[t]} a_i = r,\ \sum_{i\in[t]} a_i' \leqslant r$$

注意到 $\mathbf{v}_1, \cdots, \mathbf{v}_r$ 是一组标准正交的向量，也是 $A^\top A$ 的特征值 $\lambda_1, \cdots, \lambda_r$ 的特征向量。从而我们对任意的 $i \in [t]$ 有：

$$g_i \leqslant g_i'$$

从而由代数重数和几何重数的关系我们有：

$$a_i = g_i \leqslant g_i' \leqslant a_i' \Longrightarrow \sum_{i\in[t]} a_i' \geqslant \sum_{i\in[t]} a_i = r$$

从而 $\sum_{i\in[t]} a_i' = r$，即对所有的 $i \in [t]$:

$$a_i = a_i'$$

从而 $A^\top A$ 和 $AA^\top$ 有相同的 r 个非零特征值 (计算重数)。 $\square$

现在我们来直观的理解一下一般矩阵的分解。令 A 是一个 $m \times n$ 的矩阵，则 $A^{\top}A$ 是一个对称阵，从而存在一个 $n \times n$ 的正交矩阵 Q 使得:

$$A^{\top}A = Q\Lambda Q^{\top}$$

这里 $\Lambda = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$，其中 $D = \mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}$ 是一个对角阵，$r = \mathrm{rank}(A)$。另一方面，在定理268中我们有:

$$A = U\Sigma V^{\top}$$

从而:

$$A^{\top}A = V\Sigma^{\top}U^{\top}U\Sigma V^{\top} = V\Sigma^{\top}\Sigma V^{\top}$$

我们可以看到:

$$\Lambda = \Sigma^{\top}\Sigma = \begin{bmatrix} D & O_{(m-r)\times r} \\ O_{r\times(n-r)} & O_{(n-r)\times(m-r)} \end{bmatrix}\begin{bmatrix} D & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_r^2 \end{bmatrix}$$

从而我们有，对任意的 $i \in [r]$ 有：

$$\lambda_i = \sigma_i^2$$

我们称其为奇异值 (Singular Value)。

定义 272.

令 A 是一个 $m \times n$ 的矩阵，$\lambda_1, \cdots, \lambda_r$ 是对称矩阵 $A^\top A$ 的非零特征值，从而其都是非负的。对任意的 $i \in [r]$，定义：

$$\sigma_i = \sqrt{\lambda_i}$$

我们将 $\sigma_1, \cdots, \sigma_r$ 称为矩阵 A 的奇异值 (Singular Value)。

我们现在开始证明定理268。再次回顾一下定理的内容:

定理 268.
令 A 是一个 $m \times n$ 的矩阵，rank$(A) = r$，定义下列 $m \times n$ 的矩阵 Σ:

$$\Sigma = \begin{bmatrix} D & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix}, \text{其中} D = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

这里, $\sigma_1, \cdots, \sigma_r$ 是 A 的奇异值。则存在 $m \times m$ 的正交矩阵 U 和 $n \times n$ 的正交矩阵 $V =$, 使得:

$$A = U\Sigma V^\top$$

特别的，存在 $\mathbf{u}_1, \cdots, \mathbf{u}_r \in \mathbb{R}^m$ 和 $\mathbf{v}_1, \cdots, \mathbf{v}_r \in \mathbb{R}^n$ 使得:

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top$$

定理268的证明 (I)

令

$$\lambda_1, \cdots, \lambda_r, \lambda_{r+1}, \cdots, \lambda_n$$

是 $A^\top A$ 的特征值，并且我们有：

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$$

并且令 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是其相应的标准正交的特征向量，从而矩阵：

$$V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$$

是一个 $n \times n$ 的正交矩阵。

现在对于任意的 $i \in [r]$，定义:

$$\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i \in \mathbb{R}^m$$

则由之前的引理:

$$\mathbf{u}_1, \cdots, \mathbf{u}_r$$

是 $AA^\top$ 的标准正交的特征向量。利用 Gram–Schmidts 正交化将其扩展成一组标准正交基:

$$\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$$

从而矩阵 $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}$ 是 $m \times m$ 的正交矩阵。

定理268的证明 (III)

现在对于任意的 $i \in [n]$，考虑 $A\mathbf{v}_i$：

- 若 $i \leqslant r$。则由定义：

$$A\mathbf{v}_i = \sqrt{\lambda_i}\mathbf{u}_i$$

- 若 $i \geqslant r$，注意到 $(A^\top A)\mathbf{v}_i = \lambda_i \mathbf{v}_i = \mathbf{0}$，从而：

$$0 = \mathbf{v}_i^\top (A^\top A)\mathbf{v}_i = (A\mathbf{v}_i)^\top (A\mathbf{v}_i) = \|A\mathbf{v}_i\|^2 \implies A\mathbf{v}_i = \mathbf{0}$$

从而：

$$AV = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1}\mathbf{u}_1 & \cdots & \sqrt{\lambda_r}\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & & & \\ & \ddots & & & \\ & & \sqrt{\lambda_r} & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} = U\Sigma$$

注意到 V 是正交矩阵，从而 $A = U\Sigma V^\top$。

上海师范大学 Shanghai Normal University

让我们再来关注一下奇异值究竟想表示的内容，注意到：

$$A = U\Sigma V^T，即：AV = U\Sigma$$

则对于 $i \in [n]$ 有：

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i = \sqrt{\lambda_i}\mathbf{u}_i$$

- $\mathbf{u}_i$ 是 $\mathbb{R}^m$ 的一组标准正交基。
- $\mathbf{v}_i$ 是 $\mathbb{R}^n$ 的一组标准正交基。
- $AV = U\Sigma$ 与再方阵中 $AX = X\Lambda$ 类似，表达的是不改变方向的变换，只是对一般矩阵而言，其对应的是 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换，而不是 $\mathbb{R}^n$ 到 $\mathbb{R}^n$。

由定理268我们知道，对于任意的矩阵 A，我们有：

$$A = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_r^\top \\ \mathbf{v}_{r+1}^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix}$$

$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top$$

推论 274.

正交向量组	作为正交基的对应的向量空间
$\mathbf{u}_1, \cdots, \mathbf{u}_r$	$C(A)$
$\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$	$N(A^\top)$
$\mathbf{v}_1, \cdots, \mathbf{v}_r$	$C(A^\top)$
$\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n$	$N(A)$

推论690的证明

由对称性我们只需要证明 $C(A)$ 和 $\mathbf{N}(A)$ 的性质。

- 注意到 $\dim(C(A)) = \operatorname{rank}(A)$，我们只要证明：

$$C(A) \subseteq \operatorname{span}\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$$

  而由 $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$ 可知，A 的每一列都是 $\mathbf{u}_1, \cdots, \mathbf{u}_r$ 的线性组合。

- 注意到 $\dim(\mathbf{N}(A)) = n - \operatorname{rank}(A) = n - r$，我们只要证明对任意的 $r+1 \leqslant i \leqslant n$ 有：

$$A\mathbf{v}_i = \mathbf{0}$$

  而此时有：

$$\begin{aligned} A\mathbf{v}_i &= (\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T)\mathbf{v}_i \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T \mathbf{v}_i + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T \mathbf{v}_i + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T \mathbf{v}_i \\ &= \sigma_1 \mathbf{u}_1 (\mathbf{v}_1 \cdot \mathbf{v}_i) + \sigma_2 \mathbf{u}_2 (\mathbf{v}_2 \cdot \mathbf{v}_i) + \cdots + \sigma_r \mathbf{u}_r (\mathbf{v}_r \cdot \mathbf{v}_i) \\ &= \mathbf{0} \end{aligned}$$

$\square$

阶段总结

- 矩阵的低秩分解-更少的存储方式。
- 矩阵的奇异值分解 (SVD)、奇异值的概念。
- 奇异值分解的证明。

奇异值分解 (Singular Value Decomposition)

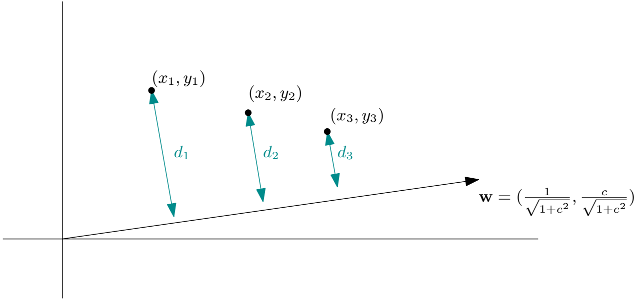
用线来拟合数据

用线来拟合数据 (I)

假设我们现在在 $\mathbb{R}^2$ 中有 m 个数据：

$$(x_1, y_1), \cdots, (x_m, y_m)$$

我们希望找到一条直线：$y = cx (c \in \mathbb{R})$ 使得每个数据点到直线的距褙尽可能的小。

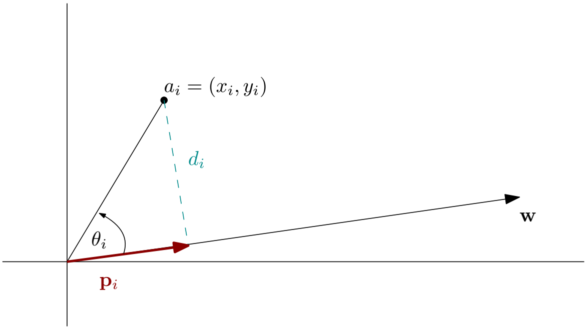


这等价于找到一个单位向量 $\mathbf{w} = \begin{bmatrix} \frac{1}{\sqrt{1+c^2}} \\ \frac{c}{\sqrt{1+c^2}} \end{bmatrix}$ 使得：

$$d_1^2 + \cdots d_m^2 \text{ 最小}$$

用线来拟合数据 (II)

对于每个数据点 $\mathbf{a}_i(x_i, y_i) \in \mathbb{R}^2$，我们希望寻找到一个单位向量 $\mathbf{w} \in \mathbb{R}^2$：



最小化:

$$\sum_{i=1}^{m} d_i^2 = \sum_{i=1}^{m} (\|\mathbf{a}_i\|^2 - \mathbf{p}_i^2) = \sum_{i=1}^{m} \|\mathbf{a}_i\|^2 - \sum_{i=1}^{m} \mathbf{p}_i^2$$

即最大化:

$$\sum_{i=1}^{m} \mathbf{p}_i^2 = \sum_{i=1}^{m} (\|\mathbf{a}_i\| \cos\theta_i)^2 = \sum_{i=1}^{m} \left(\|\mathbf{a}_i\| \frac{\mathbf{w} \cdot \mathbf{a}_i}{\|\mathbf{w}\|\|\mathbf{a}_i\|}\right)^2 = \sum_{i=1}^{m} (\mathbf{w} \cdot \mathbf{a}_i)^2$$

假设现在数据有 n 个特征，即：

$$\mathbf{a}_1, \cdots, \mathbf{a}_m \in \mathbb{R}^n$$

我们依旧是希望找到一个单位向量 $\mathbf{w} \in \mathbb{R}^n$ 最小化：

$$\sum_{i=1}^{m} \mathbf{w} \text{ 和 } \mathbf{a}_i \text{ 的距离}$$

即最大化：

$$\sum_{i=1}^{m} (\mathbf{w} \cdot \mathbf{a}_i)^2$$

矩阵的表示

定义 $m \times n$ 的矩阵：

$$A = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}$$

则我们目标是找到一个单位向量 $\mathbf{w} \in \mathbb{R}^n$ 使得：

$$\sum_{i=1}^{m} (\mathbf{w} \cdot \mathbf{a}_i)^2 = \sum_{i=1}^{m} \left(\mathbf{a}_i^\top \mathbf{w} \right)^2 = \|A\mathbf{w}\|^2 \text{ 最大}$$

由矩阵的奇异值分解可知，事实上这就是矩阵 A 最大的奇异值的平方。

定理 275.

令 A 是 $m \times n$ 的矩阵，σ_1 是其最大的奇异值，则：

$$\max_{\mathbf{w} \in \mathbb{R}^n,\ \|\mathbf{w}\|=1} \|A\mathbf{w}\|^2 = \sigma_1^2$$

特比的，该值取到最大的时候恰好是 A 的 SVD 分解中 V 对应 σ_1 的向量，即：

$$A = U\Sigma V^\top = U\Sigma \begin{bmatrix} \mathbf{v}_1^\top \\ \mathbf{v}_2^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix}$$

在证明这个定理之前，我们先讨论回顾一下基于标准正交基表示的向量的长度。

考虑 $\mathbb{R}^n$ 上的一组标准正交基 $\mathbf{v}_1, \cdots, \mathbf{v}_n$。则任何一个 $\mathbf{w} \in \mathbb{R}^n$ 都可以被其线性表示：

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

即 $\mathbf{w}$ 关于基 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 的坐标为 $(c_1, \cdots, c_n) \in \mathbb{R}^n$。注意到对所有的 $i \in [n]$：

$$\mathbf{v}_i \cdot \mathbf{w} = \mathbf{v}_i \cdot (c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = c_i(\mathbf{v}_i \cdot \mathbf{v}_i) = c_i$$

从而我们有：

$$\begin{aligned}
\mathbf{w} &= (\mathbf{v}_1 \cdot \mathbf{w})\mathbf{v}_1 + \cdots + (\mathbf{v}_n \cdot \mathbf{w})\mathbf{v}_n = \mathbf{v}_1\mathbf{v}_1^\top\mathbf{w} + \cdots + \mathbf{v}_n\mathbf{v}_n^\top\mathbf{w} \\
&= \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top\mathbf{w} \\ \vdots \\ \mathbf{v}_n^\top\mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix} \mathbf{w} = QQ^\top\mathbf{w}
\end{aligned}$$

这其实就是：

$$Q\mathbf{x} = \mathbf{b} \text{ 的最小二乘解就是 } \hat{\mathbf{x}} = Q^\top\mathbf{b}$$

回顾：使用正交矩阵投影

令 $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^n$ 是标准正交的，并且：

$$Q = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix}$$

显然 Q 是一个正交矩阵。从而 $Q\mathbf{x} = \mathbf{b}$ 的最小二乘解为：$\hat{\mathbf{x}} = Q^\top \mathbf{b}$，对应的投影矩阵为 $QQ^\top = I$，从而：

$$\mathbf{b} = QQ^\top \mathbf{b} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^\top \mathbf{b} \\ \vdots \\ \mathbf{q}_n^\top \mathbf{b} \end{bmatrix} = \mathbf{q}_1 \mathbf{q}_1^\top \mathbf{b} + \cdots + \mathbf{q}_n \mathbf{q}_n^\top \mathbf{b}$$

也就是：

$\mathbf{p}$ 是 $\mathbf{b}$ 分别到每条线 $span(\{\mathbf{q}_i\})$ 上的投影的和。

考虑向量空间 $\mathbb{V}=\mathbb{R}^n$ 和其两组基:

$$\bar{\mathbf{e}}=\mathbf{e}_1,\cdots,\mathbf{e}_n \text{ 和 } \bar{\mathbf{v}}=\mathbf{v}_1,\cdots,\mathbf{v}_n$$

从而 $\bar{\mathbf{e}}$ 到 $\bar{\mathbf{v}}$ 的基变换矩阵 M 满足:

$$\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}=\begin{bmatrix}\mathbf{e}_1 & \cdots & \mathbf{e}_n\end{bmatrix}M$$

并且可以得出:

$$M=\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$$

即: 对于任意的 $\mathbf{w}=(w_1,\cdots,w_n)\in\mathbb{V}$, 其在基 $\bar{\mathbf{e}}$ 下的坐标就是其自己 $\mathbf{w}$。

回顾坐标变换公式：

定理 276.

令 $\overline{\mathbf{v}}n$ 和 $\overline{\mathbf{v}'}$ 是 $\mathbb{V}$ 的两组基，则对于任意的 $\mathbf{u} \in \mathbb{V}$，我们有：

$$T_{\overline{\mathbf{v}}}(\mathbf{u}) = MT_{\overline{\mathbf{v}'}}(\mathbf{u})$$

从而 $\mathbf{w}$ 在基 $\overline{\mathbf{v}}$ 下的坐标为：

$$M^{-1}\mathbf{w}$$

如果基 $\overline{\mathbf{v}}$ 是标准正交的，即：

$$M = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \text{是正交矩阵}$$

则我们有：

$$M^{-1}\mathbf{w} = M^{\mathsf{T}}\mathbf{w}, \quad \text{即：} \; \mathbf{w} = MM^{\mathsf{T}}\mathbf{w} = \mathbf{v}_1\mathbf{v}_1^{\mathsf{T}}\mathbf{w} + \cdots + \mathbf{v}_n\mathbf{v}_n^{\mathsf{T}}\mathbf{w}$$

关于 w 的长度

注意到:

定理 277.

令 Q 是 $n \times n$ 的正交矩阵，并且 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，则:

1. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$，特别的 $\|Q^T\mathbf{x}\| = \|\mathbf{x}\|$。
2. $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$，从而 $Q^T\mathbf{x} \cdot Q^T\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$。

从而考虑任意的 $\mathbf{w} \in \mathbb{R}^n$，我们有:

1. 其在标准正交基 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 下的坐标为 $Q^T\mathbf{w}$，这里 $Q = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$，即:
$$(\mathbf{v}_1^T\mathbf{w}, \cdots, \mathbf{v}_n^T\mathbf{w})$$
2. 对应坐标的长度和 $\mathbf{w}$ 的长度是相同的，即:
$$\|\mathbf{w}\| = \|Q^T\mathbf{w}\| = \sqrt{(\mathbf{v}_1^T\mathbf{w})^2 + \cdots + (\mathbf{v}_n^T\mathbf{w})^2}$$

现在我们来证明定理275。回顾 A 的 SVD 分解，令 $A^\top A$ 的正特征值为 $\lambda_1 > \cdots > \lambda_r$。对任意的 $i \in [r]$，我们定义 $\sigma_i = \sqrt{\lambda_i}$。则 $\sigma_1, \cdots, \sigma_r$ 是 A 的奇异值。并且存在标准正交基 $\mathbf{u}_1, \cdots, \mathbf{u}_m \in \mathbb{R}^m$ 和 $\mathbf{v}_1, \cdots, \mathbf{v}_n \in \mathbb{R}^n$:

$$A = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_r^\top \\ \mathbf{v}_{r+1}^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix}$$

$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top$$

考虑 $\mathbb{R}^n$ 的一个单位向量 $\mathbf{w}$，存在 $c_1, \cdots, c_n$ 使得 $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$，并且由前面的讨论:

$$c_1^2 + \cdots + c_n^2 = \|\mathbf{w}\|^2 = 1$$

从而:

$$\begin{aligned} A\mathbf{w} &= (\sigma_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T)(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\ &= \sum_{i\in[r],j\in[n]} \sigma_i c_j \mathbf{u}_i\mathbf{v}_j^T\mathbf{v}_j = \sum_{i\in[r]} \sigma_i c_i \mathbf{u}_i \end{aligned}$$

从而 $A\mathbf{w}$ 在标准正交基 $\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$ 下的坐标为 $(\sigma_1c_1, \cdots, \sigma_rc_r, 0, \cdots, 0)$，因此我们有:

$$\|A\mathbf{w}\| = \sigma_1^2c_1^2 + \cdots + \sigma_r^2c_r^2 \leqslant \sigma_1^2\sum_{i\in[r]} c_i^2 \leqslant \sigma_1^2\sum_{i\in[n]} c_i^2 = \sigma_1^2$$

当 $\mathbf{w} = \mathbf{v}_1$ 时取到上述不等式等号。 $\square$

奇异值分解 (Singular Value Decomposition)

用 k 维子空间拟合数据

用 k 维子空间去拟合数据

固定一个 $k \geqslant 1$。

$$\mathbf{a}_1, \cdots, \mathbf{a}_m \in \mathbb{R}^n$$

我们希望找到一个 k 维的子空间 $\mathbb{W} \subseteq \mathbb{R}^n$ 最小化:

$$\sum_{i \in [m]} (\mathbf{a}_i \text{ 到 } \mathbb{W} \text{ 的距离})^2$$

我们可以选取 $\mathbb{W}$ 中的一组标准正交基 $\mathbf{w}_1, \cdots, \mathbf{w}_k \in \mathbb{W}$，从而 $\mathbf{a}_i$ 到 $\mathbb{W}$ 的距离是:

$\mathbf{a}_i$ 和 $\mathbf{a}_i$ 到 $\mathbb{W}$ 的投影 $\mathbf{p}_i$ 的距离。

回顾：投影的做法

上海师范大学 Shanghai Normal University

1. 我们的目标是计算 $\mathbf{b}$ 到下列空间：

$$V = span(\{\mathbf{a}_1, \ldots, \mathbf{a}_n\})$$

   的投影 $\mathbf{p}$，其中 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 是线性无关的，$\mathbf{p} \in V$.

2. 我们令 $\mathbf{p} \in V$ 是满足其误差 $\mathbf{e} = \mathbf{b} - \mathbf{p}$ 与 V 垂直的向量。我们证明了，对于任意的 $\mathbf{v} \in V$:

$$\|\mathbf{b} - \mathbf{v}\| = \min_{\mathbf{u} \in V} \|\mathbf{b} - \mathbf{u}\| \quad \Longleftrightarrow \quad \mathbf{v} = \mathbf{p}$$

3. 我们得到了相应的投影矩阵 $P = A(A^T A)^{-1} A^T$，即:

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

   并且我们证明了当 $rank(A) = n$ 时 $(A^T A)^{-1}$ 是存在的，这也说明了 $\mathbf{p}$ 的唯一性。

我们利用如下的标准正交的向量来表示 $\mathbb{W}$:

$$\mathbf{w}_1, \cdots, \mathbf{w}_k \in \mathbb{R}^n$$

特别的，令

$$Q = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_k \end{bmatrix}$$

则:

$$Q^\top Q = \begin{bmatrix} \mathbf{w}_1^\top \\ \vdots \\ \mathbf{w}_k^\top \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_k \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1^\top \mathbf{w}_1 & \cdots & \mathbf{w}_1^\top \mathbf{w}_k \\ \vdots & \ddots & \vdots \\ \mathbf{w}_k^\top \mathbf{w}_1 & \cdots & \mathbf{w}_k^\top \mathbf{w}_k \end{bmatrix} = I_{k \times k}$$

$$QQ^\top = \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_k \end{bmatrix} \begin{bmatrix} \mathbf{w}_1^\top \\ \vdots \\ \mathbf{w}_k^\top \end{bmatrix} = \mathbf{w}_1 \mathbf{w}_1^\top + \cdots + \mathbf{w}_k \mathbf{w}_k^\top$$

则 $\mathbf{a}_i$ 到 $\mathbb{W}(=\text{span}(\{\mathbf{w}_1,\cdots,\mathbf{w}_k\}))=C(Q)$ 的投影为:

$$\begin{aligned}
\mathbf{p}_i &= Q(Q^\top Q)Q^\top \mathbf{a}_i \\
&= QQ^\top \mathbf{a}_i \\
&= \mathbf{w}_1\mathbf{w}_1^\top \mathbf{a}_i + \cdots + \mathbf{w}_k\mathbf{w}_k^\top \mathbf{a}_i \\
&= (\mathbf{w}_1\cdot\mathbf{a}_i)\mathbf{w}_1 + \cdots + (\mathbf{w}_k\cdot\mathbf{a}_i)\mathbf{w}_k \\
&= \sum_{j\in[k]} (\mathbf{w}_j\cdot\mathbf{a}_i)\mathbf{w}_j
\end{aligned}$$

即:

$\mathbf{a}_i$ 到 $\mathbb{W}$ 的投影 $=$ $\mathbf{a}_i$ 到直线 $\text{span}(\{\mathbf{w}_1\}),\cdots,\text{span}(\{\mathbf{w}_k\})$ 的投影之和

另一方面，注意到 $\mathbf{a}_i$ 到 $\mathbb{W}$ 的距离可以表示为：

$$d_i^2 = \|\mathbf{a}_i - \mathbf{p}_i\|^2 = \|\mathbf{a}_i\|^2 - \|\mathbf{p}_i\|^2 = \|\mathbf{a}_i\|^2 - \sum_{j\in[k]} (\mathbf{w}_j \cdot \mathbf{a}_i)\mathbf{w}_j$$

从而最小化 $d_1^2 + \cdots + d_m^2$ 等价于最大化：

$$\sum_{i\in[m]} \|\mathbf{p}_i\|^2 = \sum_{i\in[m]} \left\| \sum_{j\in[k]} (\mathbf{w}_j \cdot \mathbf{a}_i)\mathbf{w}_j \right\|^2 = \sum_{i\in[m]} \sum_{i\in[k]} (\mathbf{w}_j \cdot \mathbf{a}_i)^2$$

注意到之前用线去拟合的情况就是该例中 $k = 1,\ \mathbf{w}_1 = \mathbf{w}$ 的特殊情况。

我们依旧用一个矩阵来表示这 m 个数据:

$$A = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix} \in \mathbb{R}^{m \times n}$$

从而我们的目标是寻找到一组标准正交的向量 $\mathbf{w}_1, \cdots, \mathbf{w}_k \in \mathbb{R}^n$ 最大化:

$$\sum_{i \in [m]} \sum_{j \in [k]} (\mathbf{w}_j \cdot \mathbf{a}_i)^2 = \sum_{j \in [k]} \sum_{i \in [k]} (\mathbf{w}_j \cdot \mathbf{a}_i)^2 = \sum_{j \in [k]} \|A\mathbf{w}_j\|^2$$

利用 SVD 最大化

注意到：

$$A = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_r^\top \\ \mathbf{v}_{r+1}^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix}$$

$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top$$

定理 278.

假设 $k \leqslant r$，则对任意的标准正交的 $\mathbf{w}_1, \cdots, \mathbf{w}_k \in \mathbb{R}^n$ 有：

$$\sum_{j\in[k]} \|A\mathbf{w}_j\|^2 \leqslant \sum_{j\in[k]} \sigma_j^2$$

当 $\mathbf{w}_1 = \mathbf{v}_1, \cdots, \mathbf{w}_k = \mathbf{v}_k$ 时等号成立。

定理278的证明思路

由定理275，我们知道存在 $\mathbf{w}_1 \in \mathbb{R}^n$，使得对于任意的 $\mathbf{w} \in \mathbb{R}^n$：

$$\|A\mathbf{w}\|^2 \leqslant \|A\mathbf{w}_1\|^2 = \sigma_1^2$$

从而 $\mathbf{w}_2$ 的选取应当满足 $\mathbf{w}_2 \perp \mathbf{w}_1$，之后的情况可以类似归纳。为此我们先证明两个引理。

引理 279.

令 $k \in [r]$，并且 $\mathbf{w} \in \mathbb{R}^n$ 是一个单位向量，并且满足对任意的 $i \in [k-1]$ 有 $\mathbf{w} \perp \mathbf{v}_i$，则我们有：

$$\|A\mathbf{w}\|^2 \leqslant \|A\mathbf{v}_k\|^2 = \sigma_k^2$$

引理 280.

令 $2 \leqslant k \leqslant n$，$\mathbb{W} \subseteq \mathbb{R}^n$ 是一个 k 维的子空间，则存在一个单位向量 $\mathbf{w} \in \mathbb{R}^n$，使得：对任意的 $i \in [k-1]$ 有 $\mathbf{w} \perp \mathbf{v}_i$

引理 279.

令 $k \in [r]$，并且 $\mathbf{w} \in \mathbb{R}^n$ 是一个单位向量，并且满足对任意的 $i \in [k-1]$ 有 $\mathbf{w} \perp \mathbf{v}_i$，则我们有：

$$\|A\mathbf{w}\|^2 \leqslant \|A\mathbf{v}_k\|^2 = \sigma_k^2$$

证明. 令 $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$，则由题目假设我们有：

1. $c_1 = c_2 = \cdots = c_{k-1} = 0$
2. $c_k^2 + \cdots + c_n^2 = 1$

从而：

$$A\mathbf{w} = (\sigma_1\mathbf{u}_1\mathbf{v}_1^\top + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^\top)(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = \sigma_k c_k\mathbf{u}_k + \cdots + \sigma_r c_r\mathbf{u}_r$$

因此：

$$\|A\mathbf{w}\|^2 = \|(0, \cdots, \sigma_k c_k, \cdots, \sigma_r c_r, 0, \cdots, 0)\| = \sum_{k\leqslant j\leqslant r} \sigma_j^2 c_j^2 \leqslant \sigma_k^2 \sum_{k\leqslant j\leqslant r} c_j^2 \leqslant \sigma_k^2 \sum_{j\in[n]} c_j^2 = \sigma_k^2$$

$\square$

引理 281.

令 $2 \leqslant k \leqslant n$，$\mathbb{W} \subseteq \mathbb{R}^n$ 是一个 k 维的子空间，则存在一个单位向量 $\mathbf{w} \in \mathbb{R}^n$，使得：对任意的 $i \in [k-1]$ 有 $\mathbf{w} \perp \mathbf{v}_i$

证明. 对任意的 $i \in [k-1]$，定义 $\mathbf{v}_i'$ 是 $\mathbf{v}_i$ 到 $\mathbb{W}$ 的投影，显然我们有 $(\mathbf{v}_i - \mathbf{v}_i') \perp \mathbb{W}$。注意到：

$$\mathbb{W}_0 = \mathrm{span}(\{\mathbf{v}_1', \cdots, \mathbf{v}_{k-1}'\}) \subseteq \mathbb{W}(\text{由于 } \dim(\mathbb{W}_0) \leqslant k-1)$$

令 $\dim(\mathbb{W}_0) = l$，则我们可以构造出一组 $\mathbb{W}$ 的标准正交基：

$$\mathbf{w}_1, \cdots, \mathbf{w}_l, \cdots, \mathbf{w}_{k-1}, \mathbf{w}_k$$

使得 $\mathbf{w}_1, \cdots, \mathbf{w}_l$ 是 $\mathbb{W}$ 中的一组标准正交基，定义 $\mathbf{w} = \mathbf{w}_k$，则：

1. $\|\mathbf{w}\| = \|\mathbf{w}_k\| = 1$。
2. $\mathbf{w} \perp \mathbb{W}_0$，从而对任意的 $i \in [k-1]$ 有：

$$\mathbf{w} \cdot \mathbf{v} = \mathbf{w}^\mathsf{T}\mathbf{v}_i = \mathbf{w}^\mathsf{T}\mathbf{v}_i' + \mathbf{w}^\mathsf{T}(\mathbf{v}_i - \mathbf{v}_i') = 0 + 0 = 0$$

□

定理 278.

假设 $k \leqslant r$，则对任意的标准正交的 $\mathbf{w}_1, \cdots, \mathbf{w}_k \in \mathbb{R}^n$ 有：

$$\sum_{j\in[k]} \|A\mathbf{w}_j\|^2 \leqslant \sum_{j\in[k]} \sigma_j^2$$

当 $\mathbf{w}_1 = \mathbf{v}_1, \cdots, \mathbf{w}_k = \mathbf{v}_k$ 时等号成立。

证明. 定义：

$$\mathbb{V}_k = \mathrm{span}(\{\mathbf{v}_1, \cdots, \mathbf{v}_k\})$$

从而：

$$\begin{aligned}\sum_{i\in[m]} (\mathbf{a}_i \text{ 到 } \mathbb{V}_k \text{ 的投影})^2 &= \sum_{j\in[k]} \|A\mathbf{v}_j\|^2 \\ &= \sum_{j\in[k]} \left\|(\sigma_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T)\mathbf{v}_j\right\|^2 \\ &= \sum_{j\in[k]} \|\sigma_j\mathbf{u}_j\|^2 = \sum_{j\in[k]} \sigma_j^2\end{aligned}$$

下面我们对 k 作归纳，证明对任意的 k 维子空间 $\mathbb{W}$, 我们有:

$$\sum_{i\in[m]}(\mathbf{a}_i \text{ 到 } \mathbb{W} \text{ 的投影})^2 \leqslant \sum_{i\in[m]}(\mathbf{a}_i \text{ 到 } \mathbb{V}_k \text{ 的投影})^2 = \sum_{j\in[k]}\sigma_j^2$$

即 $\mathbb{V}_k$ 是最佳的 k 维子空间。

$k=1$ 的时候就是定理 275，即对任意的 $\mathbf{w}\in\mathbb{R}^n$ 且 $\|\mathbf{w}\|=1$，我们有:

$$\|A\mathbf{w}^2\|^2 \leqslant \|A\mathbf{v}_1\|^2 = \sigma_1^2$$

当 $k \geqslant 2$ 的时候，令 $\mathbb{W} \subseteq \mathbb{R}^n$，且 $\dim(\mathbb{W}) = k$。由引理280存在一个单位向量 $\mathbf{w} \in \mathbb{W}$，使得对任意的 $i \in [k-1]$ 有 $\mathbf{w} \perp \mathbf{v}_i$。利用 Gram–Schmidts 正交化，我们可以获得 $\mathbb{W}$ 的一组标准正交基：

$$\mathbf{w}_1, \cdots, \mathbf{w}_{k-1}, \mathbf{w}_k = \mathbf{w}$$

从而由归纳假设我们有：

$$\sum_{j \in [k-1]} \|A\mathbf{w}_j\|^2 \leqslant \sum_{j \in [k-1]} \sigma_j^2$$

利用引理279我们有：$\|A\mathbf{w}\|^2 \leqslant \sigma_k^2$，从而：

$$\sum_{j \in [k]} \|A\mathbf{w}_j\|^2 = \|A\mathbf{w}\|^2 + \sum_{j \in [k-1]} \|A\mathbf{w}_j\|^2 \leqslant \sigma_k^2 + \sum_{j \in [k-1]} \sigma_j^2 = \sum_{j \in [k]} \sigma_j^2$$

$\square$

奇异值分解 (Singular Value Decomposition)

再看 $A\mathbf{x} = \mathbf{b}$ 的近似解

令 A 是 $m \times n$ 的矩阵，$\mathbf{b} \in \mathbb{R}^n$，考虑方程组：

$$A\mathbf{x} = \mathbf{b}$$

可能有无数解，也可能没有任何解。

- $A\mathbf{x} = \mathbf{b}$ 当且仅当：

$$\text{rank}(A) = \text{rank}(\begin{bmatrix} A & b \end{bmatrix})$$

- 其最小二乘解 $\hat{\mathbf{x}}$ 满足：

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|$$

上海师范大学 Shanghai Normal University

1. 我们知道如果 $\hat{\mathbf{x}} \in \mathbb{R}$ 满足:

$$A^{\mathsf{T}}(A\hat{x} - \mathbf{b}) = \mathbf{0} \quad \text{即} \quad A^{\mathsf{T}}A\hat{\mathbf{x}} = A^{\mathsf{T}}\mathbf{b}$$

   则 $A\hat{\mathbf{x}} - \mathbf{b}$ 与 $C(A)$ 正交。

2. 我们可以得到 $\hat{\mathbf{x}}$ 的表达式:

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

   并且我们知道 $\hat{x}$ 是唯一的。

3. 我们称 $\hat{x}$ 就是最小二乘解（least square solution），因为其误差的长度 $\|\mathbf{e}\|$

$$\mathbf{e} = \mathbf{b} - A\hat{x}$$

   是所有 $\mathbf{b} - A\mathbf{x}$ 中最小的。

最小二乘解-$\text{rank}(A) < n$

Shanghai Normal University

1. 选择 $C(A)$ 的一组基 $\{\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_r}\}$，其中 $r = \text{rank}(A)$。
2. 定义 $m \times r$ 的矩阵：

$$A' = \begin{bmatrix} \mathbf{a}_{i_1} & \cdots & \mathbf{a}_{i_r} \end{bmatrix}$$

显然我们有：$C(A) = C(A')$

3. $\text{rank}(A')$ 是列满秩的，所以我们可以利用前面的方法来找到 $A'\mathbf{x}' = \mathbf{b}$ 的最优近似解，即：

$$\hat{\mathbf{x}}' = (A'^{T}A')^{-1}A'^{T}\mathbf{b} \in \mathbb{R}^n$$

显然其误差 $\mathbf{e} = \mathbf{b} - A'\hat{\mathbf{x}}'$ 是所有 $\mathbf{b} - A\mathbf{x}$ 中长度最小的，即：

$$\|\mathbf{e}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\| = \min\{\|\mathbf{b} - \mathbf{v}\| \mid \mathbf{v} \in C(A')\} == \min\{\|\mathbf{b} - \mathbf{v}\| \mid \mathbf{v} \in C(A)\}$$

4. 我们需要的 $\hat{\mathbf{x}} \in \mathbb{R}^n$ 只要满足：

$$A\hat{\mathbf{x}} = A'\hat{\mathbf{x}}'$$

注意此时 $\hat{\mathbf{x}}$ 并不是唯一的。

现在我们利用 SVD 来获取最小二乘解。令 $m \times n$ 的矩阵 A 的 SVD 分解为:

$$A = U\Sigma V^\top = U \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} V^\top$$

定义 284 **[A 的广义逆 (Pseudoinverse)].**

定义 A 的广义逆 A^+ 为:

$$A^+ = V\Sigma^+U^\top = V \begin{bmatrix} \sigma_1^{-1} & & & & & \\ & \ddots & & & & \\ & & \sigma_r^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} U^\top$$

这里 Σ^+ 是一个 $n \times m$ 的矩阵。

定义:

$$\mathbf{x}^{+} = A^{+}\mathbf{b}$$

我们说明 $\mathbf{x}^{+}$ 就是 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解。

定理 285.

$\mathbf{x}^{+}$ 是 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解，即:

$$\|\mathbf{b} - A\mathbf{x}^{+}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|$$

特别的，对于该方程的每个最小二乘解 $\mathbf{x} \neq \mathbf{x}^{+}$，即:

$$\|\mathbf{b} - A\mathbf{x}^{+}\| = \|\mathbf{b} - A\mathbf{x}\|$$

我们有:

$$\|\mathbf{x}^{+}\| < \|\mathbf{x}\|$$

注意到:

$$A^+ = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \sigma_1^{-1} & & & & & \\ & \ddots & & & & \\ & & \sigma_r^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^\top \\ \vdots \\ \mathbf{u}_m^\top \end{bmatrix} = \sigma_1^{-1}\mathbf{v}_1\mathbf{u}_1^\top + \cdots + \sigma_r^{-1}\mathbf{v}_r\mathbf{u}_r^\top$$

从而:

$$\mathbf{x}^+ = A^+\mathbf{b} = \sigma_1^{-1}\mathbf{v}_1\mathbf{u}_1^\top\mathbf{b} + \cdots + \sigma_r^{-1}\mathbf{v}_r\mathbf{u}_r^\top\mathbf{b}$$

同时注意到:

$$A = \sigma\mathbf{u}_1\mathbf{v}_1^\top + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^\top$$

因此:

$$\begin{aligned} A\mathbf{x}^+ &= (\sigma\mathbf{u}_1\mathbf{v}_1^\top + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^\top)(\sigma_1^{-1}\mathbf{v}_1\mathbf{u}_1^\top\mathbf{b} + \cdots + \sigma_r^{-1}\mathbf{v}_r\mathbf{u}_r^\top\mathbf{b}) \\ &= \mathbf{u}_1\mathbf{u}_1^\top\mathbf{b} + \cdots + \mathbf{u}_r\mathbf{u}_r^\top\mathbf{b} \\ &= (\mathbf{u}_1 \cdot \mathbf{b})\mathbf{u}_1 + \cdots + (\mathbf{u}_r \cdot \mathbf{b})\mathbf{u}_r \end{aligned}$$

注意到 $\mathbf{u}_1, \cdots, \mathbf{u}_r$ 是标准正交的，从而：

$$A\mathbf{x}^+ = \mathbf{b} \text{ 到 } \mathrm{span}(\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}) \text{ 的投影}$$

由推论690:

正交向量组	作为正交基的对应的向量空间
$\mathbf{u}_1, \cdots, \mathbf{u}_r$	$\mathrm{C}(A)$
$\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$	$\mathrm{N}(A^\top)$
$\mathbf{v}_1, \cdots, \mathbf{v}_r$	$\mathrm{C}(A^\top)$
$\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n$	$\mathrm{N}(A)$

$A\mathbf{x}^+$ 是 $\mathbf{b}$ 到 $\mathrm{C}(A)$ 上的投影，即 $\mathbf{x}^+$ 是 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解。

现在考虑 $A\mathbf{x}=\mathbf{b}$ 的另一个最小二乘解 $\mathbf{x}(\neq \mathbf{x}^+)$，即：

$$\|\mathbf{b}-A\mathbf{x}\| = \|\mathbf{b}-A\mathbf{x}^+\|$$

我们有：

$$A\mathbf{x}=A\mathbf{x}^+ \Longrightarrow A(\mathbf{x}-\mathbf{x}^+)=\mathbf{0},\ \text{即}\ \mathbf{x}-\mathbf{x}^+ \in \mathbf{N}(A)$$

从而

$$\mathbf{x}^+ = \sigma_1^{-1}\mathbf{v}_1\mathbf{u}_1^\top\mathbf{b}+\cdots+\sigma_r^{-1}\mathbf{v}_r\mathbf{u}_r^\top\mathbf{b} \in \mathrm{span}(\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}) = \mathbf{C}(A^\top)$$

$$\mathbf{x}-\mathbf{x}^+ \in \mathbf{N}(A) = \mathrm{span}(\{\mathbf{v}_{r+1},\cdots,\mathbf{v}_n\})$$

从而 $\mathbf{x}^+ \perp (\mathbf{x}-\mathbf{x}^+)$，并且：

$$\|\mathbf{x}\|^2 = \|\mathbf{x}^+\|^2 + \|\mathbf{x}-\mathbf{x}^+\|^2 > \|\mathbf{x}^+\|^2$$

最后一个严格大于号是由于 $\mathbf{x}-\mathbf{x}^+ \neq \mathbf{0}$. $\square$

矩阵 A 的奇异值分解并不是唯一的，即：

$$A = U_1 \Sigma_1 V_1^T, \qquad A = U_2 \Sigma_2 V_2^T$$

则从定义上来讲，其可能存在两个广义逆：

$$A_1^+ = V_1 \Sigma_1^+ U_1^+, \qquad A_2^+ = V_2 \Sigma_2^+ U_2^+$$

但我们可以证明，这两个是相等的，即 A 的广义逆是唯一的。

定理 286.

$A_1^+ = A_2^+$

由定理285，对于该方程的每个最小二乘解 $\mathbf{x} \neq \mathbf{x}^+$，即：

$$\|\mathbf{b} - A\mathbf{x}^+\| = \|\mathbf{b} - A\mathbf{x}\|$$

我们有：

$$\|\mathbf{x}^+\| < \|\mathbf{x}\|$$

从而对任意的 $\mathbf{b} \in \mathbb{R}^m$，我们有：

$$A_1^+\mathbf{b} = A_2^+\mathbf{b}$$

考察所有的 $\mathbf{b} = \mathbf{e}_i$，则有：

$$A_1^+ = A_2^+$$

$\square$

阶段总结

我们介绍了 SVD 的一些应用，以及其几何意义。

- 利用 k 维子空间拟合数据-SVD 的最大 k 个奇异值是最好的结果。
- 矩阵的广义逆，最小的二乘解。

谢谢聆听。