

线性代数与几何

徐海翁

2024.1.6

0.1 教学团队介绍

教授联系方式：

陈璞教授邮箱：chenpu@pku.edu.cn

陈默涵邮箱：mohanchen@pku.edu.cn

主要内容：

1.代数部分（32次课左右）

2.几何内容（4-5次课）

期中：前三章+几何部分

期中考试（暂定周一）：11月6日上课时间（第九周）

期末考试时间：2024年1月7日下午

对应助教：王钰（055-087）三组

联系方式：2101111964@stu.pku.edu.cn

负责作业答案的助教：吴秉宪 2100011089@stu.pku.edu.cn

0.2 教学资料介绍

David C Lay 《线性代数及其应用》

建议：统一用第5th版

《向量代数、直线与平面》讲义会发

0.3 课程具体安排与学习方法介绍

侧重于数学概念

知识层面：探索新的知识层次：有限维空间

探索新的概念：

创造性

功夫 计算：手动求解简单问题

0.3.1 考试与成绩评定：

考试：80%为规定题难度，20%不高于思考题难度

出勤与课堂表现10%（含测验）

作业20%

期中考试20%

期终考试50%

附加分数：发现教材和讲义新的错误0.5分/错误

小论文5分/篇

课前预习非常重要!!! 提前预习，完成作业，作业共15次，取分数最好的12次计入作业分数

每周三布置作业（发布电子版作业），下周三课间上交（按照1-2组，3-4组，5-6组堆成三堆）

准备2个作业本

作业分为规定题（必做）与选做题（思考题）

规定题准时完成，主要用于复习课堂所讲的内容，加深理解

选做题可以不做或者延时完成，可能使用不同的方法

课上：按照自己的思路整理课堂笔记（思维导图）

课后：质疑，提出愚蠢的问题，比解决问题更为重要；

1 线性代数中的线性方程组

1.1 线性方程组

等价

两个线性方程组称为**等价的**，当第一个方程组的每一个解都是第二个方程组的解，第二个方程组的解都是第一个方程组的解

相容

我们称一个线性方程组是相容的，若其有一个解或无穷多个解；称它为不相容的，若其无解

行等价

我们称两个矩阵为行等价的，若其中一个矩阵可以经过一系列**初等行变换**成为另一个矩阵

关于行等价的事实

若两个线性方程组是行等价的，那他们具有相同的解集

1.2 行化简与阶梯形矩阵

1.2.1 定理1（简化阶梯形矩阵的唯一性）

每个矩阵行等价于唯一的简化阶梯型

主元位置

定义：矩阵中的**主元位置**是 A 中对应于它的简化阶梯形中先导元素1的位置。**主元列**是 A 的含有主元位置的列。

【因此我们可以看出主元位置和主元列是针对于系数矩阵 A 而言的】

主元

主元就是在主元位置上的非零元素

行化简的向前和向后步骤

向前步骤是向下获得阶梯形矩阵的过程

向后步骤是向上获得简化阶梯形矩阵的过程

基本变量和自由变量

举例：对于增广矩阵

$$\left[\begin{array}{cccc} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

对应于主元列的变量 x_1 和 x_2 称为**基本变量**，其他变量 x_3 称为**自由变量**

【由于变量是针对方程组而言的，所有我们一般说方程组 ($Ax = 0, x_1a_1 + x_2a_2 + \cdots + x_na_n = 0$) 有多少个自由变量】

1.2.2 存在与唯一性问题

线性方程组相容的**充要条件**是增广矩阵的最右列不是主元列，也就是说，增广矩阵的阶梯形中没有形如

$$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & b \end{array} \right], b \neq 0$$

的行.在已经知道线性方程组相容的情况下，其解集可能有两种情形：

- (i) 当没有自由变量时，有唯一解；
- (ii) 若至少有一个自由变量，则有无穷多解

1.3 向量方程

所有 n 个元素的向量的集记为 \mathbb{R}^n

【这里非常容易出错，比如之后列空间可能是3维的，但其中的向量可能仍

然有4个元素，那么这些向量仍然属于 \mathbb{R}^4 】

线性组合

给定 \mathbb{R}^n 中向量 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 和标量 c_1, c_2, \dots, c_n 向量

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$$

称为向量 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 以 c_1, c_2, \dots, c_n 为权的线性组合

生成的子集

定义：若 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 是 \mathbb{R}^n 中的向量，则 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 的所有线性组合所成的集合用记号 $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 表示，称为由 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 所生成（或张成）的 \mathbb{R}^n 的子集。也就是说 $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 是所有形如

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$$

的向量的集合，其中 c_1, c_2, \dots, c_n 为标量

1.4 矩阵方程 $A\mathbf{x} = \mathbf{b}$

1.4.1 定理3

若 A 是 $m \times n$ 矩阵，它的各列为 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ ，而 \mathbf{b} 属于 \mathbb{R}^m 则矩阵方程

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

与向量方程

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = \mathbf{0}$$

有相同的解集，它又与增广矩阵为

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$$

的线性方程组有相同的解集

解的存在性

方程 $A\mathbf{x} = \mathbf{b}$ 有解当且仅当 \mathbf{b} 是 A 的各列的线性组合

1.4.2 定理4

设 A 是 $m \times n$ 矩阵，则下列命题是逻辑上等价的。也就是说，对某个 A ，它们都成立或都不成立

- a. 对于 \mathbb{R}^m 中每个 \mathbf{b} ，方程 $A\mathbf{x} = \mathbf{b}$ 有解。
- b. \mathbb{R}^m 中的每个 \mathbf{b} 都是 A 的列的一个线性组合
- c. A 的各列生成 \mathbb{R}^m
- d. A 在每一行都有一个主元位置

【我们再次从书写规范性的层面了解到，主元位置、主元列是对于某个矩阵而言的，而基本变量、自由变量是针对某个方程组而言的】

1.4.3 定理5

若 A 是 $m \times n$ 矩阵， \mathbf{u} 和 \mathbf{v} 是 \mathbb{R}^n 中的向量， c 是标量，则

- a. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
- b. $A(c\mathbf{u}) = c(A\mathbf{u})$

1.5 线性方程组的解集

齐次线性方程组

依据定理2（存在与唯一性定理），可得出

齐次线性方程组 $A\mathbf{x} = \mathbf{0}$ 有非平凡解当且仅当方程至少有一个自由变量。

【注意！非齐次线性方程组 $A\mathbf{x} = \mathbf{b}$ 要先有解才能根据唯一性判据进行判断】

参数向量形式

方程 $10x_1 - 3x_2 - 2x_3 = 0$ 的解集是一个平面，对这个平面

方程 $10x_1 - 3x_2 - 2x_3 = 0$ 是这个平面的隐式描述

而方程 $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$ (s, t 为实数) 称为平面的参数向量方程，是这个平面

的显式描述

1.5.1 定理6

设方程 $Ax = b$ 对某个 b 是相容的, p 为一个特解, 则 $Ax = b$ 的解集是所有形如 $w = p + v_p$ 的向量的集, 其中 v_p 是齐次线性方程 $Ax = 0$ 的任意一个解

【注意: 仅有已经知道这个非齐次线性方程组有解的情况下, 才可以使用这样的方式!】

1.6 线性方程组的应用

经济学中的线性方程组
配平化学方程式

1.7 线性无关

定义

\mathbb{R}^n 中一组向量 $\{v_1, \dots, v_p\}$ 称为线性无关的, 若向量方程

$$x_1 v_1 + x_2 v_2 + \dots + x_p v_p = 0$$

仅有平凡解。向量组 $\{v_1, \dots, v_p\}$ 称为线性相关的, 若存在不全为零的权 c_1, \dots, c_p 使

$$c_1 v_1 + c_2 v_2 + \dots + c_p v_p = 0$$

矩阵 A 的各列线性无关, 当且仅当方程 $Ax = 0$ 仅有平凡解

1.7.1 定理7 (线性相关集的特征)

两个或更多个向量的集合 $S = \{v_1, \dots, v_p\}$ 线性相关, 当且仅当 S 中至少有一个向量使其他向量的线性组合。事实上, 若 S 线性相关, 且 $v_1 \neq 0$ 【否则后几个向量可以线性无关, 第一个向量为零向量, 那么后面任意一个向量都不能由前面的向量线性表出!!】 v_j 是它前面向量 v_1, \dots, v_{j-1} 的线性

组合

【注意：这只是说其中必定存在一个向量是前面向量的线性组合，而不是说任意一个向量都可以表示为前面向量的线性组合】

1.7.2 定理8

若一个向量组的向量个数超过了每个向量的元素数目，那么这个向量组线性相关。这是说 \mathbb{R}^n 中任意向量组 $\{v_1, \dots, v_p\}$ 当 $p > n$ 时线性相关。

1.7.3 定理8'

设 $\{v_1, \dots, v_p\}$ 为 \mathbb{R}^n 的一组线性无关向量，那么 $p \leq n$ 。

1.7.4 定理8''

如果一组向量线性无关，则向量数目 \leq 元素数目

【注意：定理8与定理8'互为逆否命题，而定理8的逆命题未必正确，即线性相关时未必向量数目要大于等于元素个数】

1.7.5 定理9

若 \mathbb{R}^n 中向量组 $S = \{v_1, \dots, v_p\}$ 包含零向量，则它线性相关

1.8 线性变换介绍

基本概念

由 \mathbb{R}^n 到 \mathbb{R}^m 的一个变换（或称函数、映射） T 是一个规则，它把 \mathbb{R}^n 中每个向量 x 对应以 \mathbb{R}^m 中的一个向量 $T(x)$ 。集 \mathbb{R}^n 称为 T 的定义域，而 \mathbb{R}^m 称为 T 的上域（或取值空间）。

符号 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 说明 T 的定义域是 \mathbb{R}^n 而上域是 \mathbb{R}^m 。对于 \mathbb{R}^n 中的向量 x ,

\mathbb{R}^m 中向量 $T(\boldsymbol{x})$ 称为 \boldsymbol{x} (在 T 作用下) 的像。所有像 $T((\boldsymbol{x}))$ 的集合称为 T 的值域。

矩阵变换

记为 $\boldsymbol{x} \rightarrow A\boldsymbol{x}$

当 A 为 $m \times n$ 矩阵时, 对应的线性变换 T 的上域是 \mathbb{R}^m , 值域是 A 的列的所有线性组合的集合。

1.8.1 线性变换

定义

变换 (或映射) T 被定义为线性的, 若 (i) 对 T 的定义域中一切 $\boldsymbol{u}, \boldsymbol{v}, T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$

(ii) 对 T 的定义域中一切 \boldsymbol{u} 和数 $c, T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ 【在很多情况下我们要证明某个变换是线性变化就可以从定义出发】

我们不难由上述两个性质推出:

若 T 是线性变换

$$T(\mathbf{0}) = \mathbf{0}$$

且对 T 的定义域中一切向量 $\boldsymbol{u}, \boldsymbol{v}$ 以及数 c, d 有:

$$T(c\boldsymbol{u} + d\boldsymbol{v}) = cT(\boldsymbol{u}) + dT(\boldsymbol{v})$$

1.9 线性变换中的矩阵

重要事实

从 \mathbb{R}^n 到 \mathbb{R}^m 的每一个线性变换实际上都是一个矩阵变换 $\boldsymbol{x} \rightarrow A\boldsymbol{x}$

【注意, 这里我们强调了从线性空间到线性空间的变换都是矩阵变换, 而有些线性变化比如导数、积分作用于函数, 不是矩阵变换】

【同样注意, 一切矩阵变换都是线性变化是正确的】

1.9.1 定理10

设 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 为线性变换，则存在唯一的矩阵 A 使得对 \mathbb{R}^n 中一切 x

$$T(x) = Ax$$

事实上 A 是 $m \times n$ 矩阵，它的第 j 列是向量 $T(e_j)$ ，其中 e_j 是 \mathbb{R}^n 中单位矩阵 I_n 的第 j 列

$$A = \begin{bmatrix} T(e_1) & \cdots & T(e_n) \end{bmatrix}$$

上面的矩阵 A 称为线性变换 T 的标准矩阵

1.9.2 常见的线性变换

旋转变换

$$\begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

对应了逆时针旋转一个角度 φ

对称

【我们这里只举出一个例子】

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

收缩与拉伸

【我们这里只举出一个例子】

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

其中当 $0 \leq k \leq 1$ 时，称为压缩变换 $k > 1$ 时称为拉伸变换

剪切变换

【我们这里只举出一个例子】

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

【剪切变换中其中一个基不发生变换】

投影

【我们这里只举出一个例子】

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

1.9.3 存在与唯一性问题

满射

定义：映射 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 称为到 \mathbb{R}^m 上的映射，若 \mathbb{R}^m 中每个 \mathbf{b} 是 \mathbb{R}^n 中至少一个 \mathbf{x} 的像。

【一个必要的条件是 $m \leq n$ 】

单射

定义：映射 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 称为一对一映射，若 \mathbb{R}^m 中每个 \mathbf{b} 是 \mathbb{R}^n 中至多一个 \mathbf{x} 的像。

【一个必要的条件是 $m \geq n$ 】

1.9.4 定理11

设 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 为线性变换，则 T 是一一对一的当且仅当方程 $A\mathbf{x} = \mathbf{0}$ 仅有平凡解。

1.9.5 定理12

设 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是线性变换， A 为线性变换 T 的标准矩阵。

a. T 把 \mathbb{R}^n 映上到 \mathbb{R}^m ，当且仅当 A 的列生成 \mathbb{R}^m 。方程 $A\mathbf{x} = \mathbf{b}$ 对 \mathbb{R}^m 中的任

意 **b** 有解。

b. T 是一对一的，当且仅当 A 的列线性无关。

【我们在下方从矩阵的视角进行直观的展示】

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

从左到右，分别是单射、双射（单射+满射）、满射

【以下是一些映射的**必要条件**】

单射：要求 $m \geq n$

满射：要求 $m \leq n$

1.10 商业、科学和工程中的线性模型

构建有营养的减肥食谱

线性方程与电路网络【基尔霍夫电压定律】

差分方程

对于一个向量序列 x_0, x_1, x_2, \dots 如果有矩阵 A 使 $x_1 = Ax_0, x_2 = Ax_1$ ，一般地

$$x_{k+1} = Ax_k, k = 0, 1, 2, \dots$$

则我们把上式称为**线性差分方程**（或**递归关系**）

2 矩阵代数

2.1 矩阵运算

请自己思考一下简单的概念，本笔记在此不做整理

对角矩阵

单位矩阵

初等矩阵

初等矩阵左乘某个矩阵是对于这个矩阵进行行变换，右乘某个矩阵是对这个矩阵进行列变换

具体变换的效果可以根据是左乘还是右乘进行判断

对于矩阵

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

如果将其用作左乘，那么我们思考这个初等矩阵是如何从单位矩阵通过行变换得到的，即第2行 k 倍加到第1行

如果将其用作右乘，那么我们思考这个初等矩阵是如何从单位矩阵通过列变换得到的，即第1列 k 倍加到第2列

零矩阵

矩阵的相等与和

2.1.1 定理1

设 A, B, C 是相同维数的矩阵, r, s 为数, 则有

a. $A + B = B + A$

b. $(A + B) + C = A + (B + C)$

c. $A + 0 = A$

d. $r(A + B) = rA + rB$

e. $(r + s)A = rA + sA$

f. $r(sA) = (rs)A$

2.1.2 矩阵乘法

定义

设 A 是 $m \times n$ 矩阵, B 是 $n \times p$ 矩阵, B 的列是 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$ 则乘积 AB 是 $m \times p$ 矩阵, 它的各列是 $A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_p$, 即

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

几种理解方式与视角

1. 对应元素相乘

设 A 是 $m \times n$ 矩阵, B 是 $n \times p$ 矩阵, 这里我们记 A 中第 i 行第 j 列元素为 a_{ij} , 记 B 中第 j 行第 k 列元素为 b_{jk} , 记 AB 中第 i 行第 k 列元素为 c_{ik} , 则

$$c_{ik} = \sum_{1 \leq j \leq n} a_{ij} \cdot b_{jk}$$

2. AB 中各列视作 A 中各列向量以 B 中对应列元素为权的线性组合

3. AB 中各行视作 B 中各行向量以 A 中对应行元素为权的线性组合

2.1.3 定理2

设 A 是 $m \times n$ 矩阵, B 和 C 的维数使下列各式的乘积有定义

a. $A(BC) = A(BC)$

b. $A(B + C) = AB + AC$

c. $(B + C)A = BA + CA$

$$\text{d. } r(AB) = (rA)B = A(rB)$$

$$\text{e. } I_m A = A = A I_n$$

矩阵的乘幂

对于方阵而言 A^k 表示 k 个 A 的乘积

若 A 不为零矩阵, A^0 被解释为单位矩阵

2.1.4 定理3 (矩阵的转置)

设 A 与 B 表示矩阵, 其维数使下列和与积有定义, 则

$$\text{a. } (A^T)^T = A$$

$$\text{b. } (A + B)^T = A^T + B^T$$

$$\text{c. 对任意实数 } r, (rA)^T = rA^T$$

$$\text{d. } (AB)^T = B^T A^T$$

对于d, 事实上有

若干个矩阵的乘积的转置等于它们转置的乘积, 但相乘的顺序相反

2.1.5 一些比较重要的习题

例1

若 B 的各列线性相关, 求证 AB 的各列线性相关

证:

由于 B 的各列线性相关, 故对于矩阵方程 $Bx = 0$ 有非平凡解

记一个这样的解为 m 由于我们有 $A(0) = 0$, 故 $AB(m) = A(Bm) = A(0) =$

0

故 $AB(x) = 0$ 有非平凡解, 故 AB 的各列线性相关

例2

设 $CA = I_n$, 证明 $Ax = 0$ 只有平凡解, 并证明 A 的列数不可以多于行数

证：

① 设 $A\mathbf{x} = \mathbf{0}$ 有非平凡解 \mathbf{m} , 则有

$$A(\mathbf{m}) = \mathbf{0}$$

$$I_n(\mathbf{m}) = CA(\mathbf{m}) = C(A\mathbf{m}) = C(\mathbf{0}) = \mathbf{0}$$

那么可证明 $\mathbf{m} = \mathbf{0}$, 这是矛盾的, 因而方程只有平凡解

② 由于 $A\mathbf{x} = \mathbf{0}$ 只有平凡解, 故 $A\mathbf{x} = \mathbf{0}$ 中无自由变量, A 中主元列有 n 列, 而 A 的每一行至多只能提供一个主元位置, 故 A 的列数不可以多于行数

例3

设 $AD = I_m$, 证明: 对于 \mathbb{R}^m 中的任意 \mathbf{b} , $A\mathbf{x} = \mathbf{b}$ 都有解, 并证明 A 的行数不能多于列数

证：

①

$$A(D\mathbf{x}) = AD(\mathbf{x}) = I_m(\mathbf{x}) = \mathbf{x}$$

故对于 \mathbb{R}^m 中任意一个 \mathbf{b} , 必然存在一个解 $D(\mathbf{b})$

② 由于对于 \mathbb{R}^m 中的任意 \mathbf{b} , $A\mathbf{x} = \mathbf{b}$ 都有解, 根据第一章定理4, 可知 A 的每一行都有主元位置, 由于 A 的每一列也至多可以提供一个主元位置, 故 A 的行数不可以多于列数

例4

设 A 是 $m \times n$ 矩阵, 若存在 $n \times m$ 矩阵 C 使得 $CA = I_n$, $n \times m$ 矩阵 D 使得 $AD = I_m$, 证明 $m = n$ 与 $C = D$

证：

这道题事实上需要结合前两题的结论

$$(1) \because m \leq n \text{ 且 } n \leq m$$

$$\therefore m = n$$

$$(2) \because (CA)D = CAD = C(AD)$$

$$\therefore I_n D = C I_m$$

$$\therefore C = D$$

事实上，我们学完了矩阵的秩之后，我们能够使用别的方法更快得出结果

首先，由于 $CA = I_m$ ，有

$$\text{rank} A \geq \text{rank}(CA) = m$$

然后，由于 $AD = I_n$ ，有

$$\text{rank} A \geq \text{rank}(AD) = n$$

然而对于 $m \times n$ 矩阵 A ，显然有 $\text{rank} A \leq n, \text{rank} A \leq m$

从而我们可以得出 $m = n$

2.2 矩阵的逆

定义

一个 $n \times n$ 矩阵 A 称为是**可逆**的，若存在一个 $n \times n$ 矩阵 C 使

$$CA = I \quad \text{与} \quad AC = I$$

这时称 C 是 A 的逆。若 A 可逆，它的逆是唯一的，我们将其记为 A^{-1} ，于是

$$AA^{-1} = I \quad \text{与} \quad A^{-1}A = I$$

不可逆矩阵有时称为**奇异矩阵**，可逆矩阵也称为**非奇异矩阵**

2.2.1 定理4

对于矩阵 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 若 $ad - bc \neq 0$, 则 A 可逆, 且

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

若 $ad - bc = 0$, 则 A 不可逆

2.2.2 定理5

若 A 是可逆 $n \times n$ 矩阵, 则对于 \mathbb{R}^m 每一 \mathbf{b} , 方程 $A\mathbf{x} = \mathbf{b}$ 有唯一解 $\mathbf{x} = A^{-1}\mathbf{b}$

【同类证明唯一解的核心】

1. 证明解是…… (只需要把解代入验证即可) (存在性)
2. 如果 \mathbf{a} 是解, 那么 \mathbf{a} 一定是 (把 \mathbf{a} 解出来) (唯一性)

例

设 A, B, C 是可逆 $n \times n$ 矩阵, 方程 $C^{-1}(A + X)B^{-1} = I_n$ 是否有解 X ? 若有, 请求解

首先先解出解是什么

在把解代回验证

这样就同时保证了唯一性与存在性 (尽管顺序有点颠倒了)

2.2.3 定理6

- a. 若 A 是可逆矩阵, 则 A^{-1} 也可逆而且 $(A^{-1})^{-1} = A$
- b. 若 A 和 B 都是 $n \times n$ 可逆矩阵, 则 AB 也可逆, 且其逆就是 A 和 B 的逆矩阵按相反顺序的乘积, 即

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

- c. 若 A 可逆, 则也 A^T 可逆, 且其逆是 A^{-1} 转置, 即 $(A^T)^{-1} = (A^{-1})^T$

【定理6 (b) 的推广: 若干个 $n \times n$ 可逆矩阵的积也是可逆的, 其逆等于这

些矩阵的逆按照相反顺序的乘积】

初等矩阵

把单位矩阵进行一次初等行变换就得到初等矩阵

例如：

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

若对 $m \times n$ 矩阵 A 进行某种初等行变换，所得的矩阵可写作 EA ，其中 E 是矩阵，是 I_m 经过同一行变换所得

【每个初等矩阵 E 是可逆的， E 的逆是同类型的初等矩阵，它把 E 变回 I 】

2.2.4 定理7

$n \times n$ 矩阵 A 是可逆的，当且仅当 A 行等价于 I_n ，这时，把 A 化简为 I_n 的一系列初等行变换同时把 I_n 变成 A^{-1}

【求 A^{-1} 的算法】对于矩阵

$$\begin{bmatrix} A & I \end{bmatrix}$$

进行行化简直至将 A 化为 I ，此时 I 被行变换为 A^{-1}

逆矩阵的另一个观点

当我们只要求逆矩阵的其中某几列时，我们可以从一下视角理解上述求逆的算法

$$\begin{bmatrix} A & e_1 & e_2 & \cdots & e_n \end{bmatrix}$$

如果我们只需要逆的其中几列，那么我们只需要保留 I 中的对应几列即可

2.3 可逆矩阵的特征

2.3.1 定理8（可逆矩阵定理）

设 A 为 $n \times n$ 矩阵，则下列命题是等价的，即对于某一特定的 A ，它们同时为真或同时为假

- a. A 是可逆矩阵
- b. A 行等价于 $n \times n$ 单位矩阵
- c. A 有 n 个主元位置【也就是方程 $Ax = 0$ 有 n 个基本变量】
- d. 方程 $Ax = 0$ 仅有平凡解【也就是方程 $Ax = 0$ 自由变量数=0】
- e. A 的各列线性无关
- f. 线性变换 $x \rightarrow Ax$ 是一对一的
- g. 对 \mathbb{R}^n 中任意的 b ，方程 $Ax = b$ 至少有一个解【事实上，是唯一的解】
- h. A 的各列生成 \mathbb{R}^n
- i. 线性变换 $x \rightarrow Ax$ 把 \mathbb{R}^n 映上到 \mathbb{R}^n
- j. 存在 $n \times n$ 矩阵 C 使得 $CA = I$
- k. 存在 $n \times n$ 矩阵 D 使得 $AD = I$
- l. A^T 是可逆矩阵

【我们从几个层次分析这个定理】

- b和c侧重于讲 A 本身在形式上的特征（等价于单位矩阵、主元位置）
- d,e,f则侧重于单射或者线性无关的角度（本质上的要求有 $m \geq n$ ）
- g,h,i则侧重于满射的角度（在本质上的要求上也有 $m \leq n$ ）
- j和k则侧重于要能够找到一个矩阵能和原矩阵相乘得到单位矩阵
- l则侧重于矩阵的转置

【定理8的推论】

设 A 和 B 为方阵，若 $AB = I$ 则 A 和 B 都是可逆的，且 $B = A^{-1}, A = B^{-1}$

2.3.2 可逆线性变换

定义

线性变换 $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ 称为**可逆的**，若存在函数 $S: \mathbb{R}^n \rightarrow \mathbb{R}^n$ 使得
对所有 \mathbb{R}^n 中的 $x, S(T(x)) = x$ (1)

对所有 \mathbb{R}^n 中的 $x, T(S(x)) = x$ (2)

【这里我们其实可以联想到映射的左逆与右逆】

如果一个映射 f 存在左逆 g ，那么这个映射 f 一定是单射

如果一个映射 f 存在右逆 g ，那么这个映射 f 一定是满射

$$\text{左逆: } g(f(x)) = x$$

$$\text{右逆: } f(g(x)) = x$$

2.3.3 定理9

设 T 为线性变换, A 为 T 的标准矩阵, 则 T 可逆当且仅当 A 是可逆矩阵, 这时候由 $S(\mathbf{x}) = A^{-1}\mathbf{x}$ 定义的线性变换是满足上述 (1) (2) 两式的唯一函数

2.4 分块矩阵

加法与标量乘法

$A + B$ 的每一块恰好是 A 和 B 对应分块的矩阵和

分块矩阵的乘法

要求对 A 的列的分法与对 B 的行的分法一致

比如 A 是 3×5 矩阵, B 是 5×2 矩阵, 那么我们可以把 A 的列分为 $3 + 2$ 的两组, 把 B 的行对应地分为 $3 + 2$ 的两组【在组内如何划分是任意的】

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix}$$

2.4.1 定理10 (AB 的行列展开)

若 A 是 $m \times n$ 矩阵, B 是 $n \times p$ 矩阵, 则

$$\begin{aligned} AB &= \begin{bmatrix} col_1(A) & col_2(A) & \cdots & col_n(A) \end{bmatrix} \begin{bmatrix} row_1(B) \\ row_2(B) \\ \vdots \\ row_n(B) \end{bmatrix} \\ &= col_1(A)row_1(B) + col_2(A)row_2(B) + \cdots + col_n(A)row_n(B) \end{aligned}$$

2.4.2 分块矩阵的逆

例：形如 $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ 的矩阵称为分块上三角矩阵，设 A_{11} 是 $p \times p$ 矩阵， A_{22} 是 $q \times q$ 矩阵，且 A 为可逆矩阵，求 A^{-1} 的表达式

解：核心过程在于我们假设 A 的逆矩阵为 $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ ，之后利用分块乘法的方式进行运算，根据结果可表示为 $\begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}$ 的特性列出方程组

之后根据【可逆矩阵定理】和部分矩阵是方阵的事实，得出 $B_{11}, B_{12}, B_{21}, B_{22}$ 等矩阵

最终我们可以得到

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}$$

【重要结论：分块对角矩阵是一个分块矩阵，除了主对角线上各分块外，其余全是零分块。这样的矩阵是可逆的当且仅当主对角线上的各分块都是可逆的】

2.4.3 习题中的重要内容

舒尔补

这部分的交代我们后置到分块矩阵行列式的补充中

构造分块矩阵 M 使得 $M^2 = I$

例：构造一个 5×5 矩阵 $M = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix}$ 使得 $M^2 = I_5$ ，且 C 为 2×3 非零矩阵

构造

$$\begin{bmatrix} I_3 & 0 \\ C & I_2 \end{bmatrix}$$

即可

利用分块性质求逆

例：求矩阵 $A = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 5 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 7 & 8 \\ 0 & 0 & 0 & 5 & 6 \end{bmatrix}$ 我们可以考虑把矩阵进行连续两次分

块：

$$B = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}, C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 7 & 8 \\ 0 & 5 & 6 \end{bmatrix}$$

从而我们有

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

进而，我们在求 C 的逆的过程中，我们可以再次进行这样的分解，从而得到我们想要的分块矩阵

2.5 矩阵因式分解

2.6 列昂惕夫投入产出模型

2.7 计算机图形学中的应用

2.7.1 齐次坐标

引入原因：平移本身不对应线性变换，导致难以对图像进行平移处理

定义

对 \mathbb{R}^2 上每个点 (x, y) 可以对应于 \mathbb{R}^3 中的唯一点 $(x, y, 1)$ ，他们位于 xy 平面上方1单位的平面上，我们称 (x, y) 有齐次坐标 $(x, y, 1)$

注：点的齐次坐标不能相加，也不能乘以数，但可以通过 3×3 矩阵做变换

平移变换的实现

如果要实现 $(x, y) \rightarrow (x + h, y + k)$ ，可以通过左乘一个 3×3 矩阵来实现，即

$$\begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x + h \\ y + k \\ 1 \end{bmatrix}$$

其他 \mathbb{R}^2 变换的实现

可以通过左乘以分块矩阵

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

实现，其中 A 为 2×2 对应线性变换的矩阵
一个典型的例子是

$$\begin{bmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

2.7.2 齐次三维坐标

定义

类似于二维情形，我们称 $(x, y, z, 1)$ 是 \mathbb{R}^3 中点 (x, y, z) 的齐次坐标。一般地，若 $H \neq 0$ ，则 (X, Y, Z, H) 也是 (x, y, z) 的齐次坐标，且

$$x = \frac{X}{H}, y = \frac{Y}{H}, z = \frac{Z}{H}$$

$(x, y, z, 1)$ 的每一个非零的标量乘法得到一组 (x, y, z) 的齐次坐标，如 $(10, -6, 14, 2)$ 和 $(15, -9, 21, 3)$ 都是 $(-5, 3, -7)$ 的齐次坐标

2.7.3 透视投影

三维物体在二维计算机屏幕上的表示方式是将其投影在一个可视平面上，为简单起见，我们将 xy 平面表示计算机屏幕，假设一个观察者的眼

睛向着正 z 轴看去，眼睛的位置是 $(0, 0, d)$ ，投影透视将每个点 (x, y, z) 映射为 $(x^*, y^*, 0)$ ，这两点与透视中心在同一直线上

由相似三角形

$$x^* = \frac{x}{1 - z/d}, y^* = \frac{y}{1 - z/d}$$

使用齐次坐标，则可用矩阵表示透视投影，记此矩阵为 $P, (x, y, z, 1)$ 映射为

$$\left(\frac{x}{1 - z/d}, \frac{y}{1 - z/d}, 0, 1 \right)$$

把这个坐标乘以 $1 - z/d$ ，可用 $(x, y, 0, 1 - z/d)$ 表示齐次坐标的像

从而很容易求出 P ，

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1/d & 1 \end{bmatrix}$$

我们得到齐次坐标后，只需要将前三个元素除以第四个元素即可得到 \mathbb{R}^3 中的坐标了

2.7.4 一道有意义的练习题：图形绕 \mathbb{R}^2 中一点 p 旋转 φ 的实现方法

首先将图形平移 $-p$ ，然后绕原点旋转 φ ，再平移 p 回去

2.8 \mathbb{R}^n 的子空间

2.8.1 定义

\mathbb{R}^n 中的一个子空间是 \mathbb{R}^n 中的集合 H ，具有以下三个性质

- 零向量属于 H
 - 对于 H 中任意的向量 u 和 v ，向量 $u + v$ 属于 H
 - 对于 H 中任意的向量 u 和标量 c ，向量 cu 属于 H
- 【换句话说，子空间对加法和标量乘法封闭】

生成（或张成）的子空间

设 v_1, \dots, v_p 属于 \mathbb{R}^n , v_1, \dots, v_p 的所有线性组合是 \mathbb{R}^n 的子空间, 我们称 $\text{Span}\{v_1, \dots, v_p\}$ 为由 v_1, \dots, v_p 生成（或张成）的子空间

注意 \mathbb{R}^n 是它自身的子空间, 因为其满足三个性质。

另一个特殊的子空间是只包含零向量的集合, 它也满足子空间的条件, 称为零子空间（不等于零空间）

2.8.2 矩阵的列空间和零空间

定义

矩阵 A 的列空间是 A 各列的线性组合的集合, 记作 $\text{Col } A$

不难有 $\text{Col } A$ 与 $\text{Span}\{a_1, \dots, a_n\}$ 相同。且有 $m \times n$ 矩阵的列空间是 \mathbb{R}^n 的子空间

定义

矩阵 A 的零空间是齐次方程 $Ax = 0$ 的所有解的集合, 记为 $\text{Nul } A$

2.8.3 定理12

$m \times n$ 矩阵 A 的零空间是 \mathbb{R}^n 的子空间。等价地, n 个未知数的 m 个齐次线性方程组 $Ax = 0$ 的所有解的集合是 \mathbb{R}^n 的子空间

2.8.4 子空间的基

定义

\mathbb{R}^n 子空间 H 的一组基是 H 中一个线性无关集, 它生成 H

标准基

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

$\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ 称为 \mathbb{R}^n 的标准基

2.8.5 定理13

矩阵 A 的主元列构成 A 的列空间的基

【注意：要用 A 的主元列而不是阶梯形矩阵 B 的列。事实上行化简的过程会改变列空间!】

2.9 维数与秩

2.9.1 坐标系

定义

假设 $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ 是子空间 H 的一组基。对 H 中的每一个向量 \mathbf{x} ，相对于基 \mathcal{B} 的坐标是使 $\mathbf{x} = c_1\mathbf{b}_1 + \dots + c_p\mathbf{b}_p$ 成立的权 c_1, \dots, c_p ，且 \mathbb{R}^p 中的向量

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

称为 \mathbf{x} （相对于 \mathcal{B} ）的坐标向量，或 \mathbf{x} 的 \mathcal{B} -坐标向量

如果 H 的基 \mathcal{B} 包含 $\begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ 2个向量

虽然 H 中的点也在 \mathbb{R}^3 中，但它们完全由属于 \mathbb{R}^2 的坐标向量确定映射 $\mathbf{x} \rightarrow [\mathbf{x}]_{\mathcal{B}}$ 是 H 与 \mathbb{R}^2 之间保持线性组合关系的一一映射，我们称这种映射是同构

的, 且 H 与 \mathbb{R}^2 同构

2.9.2 子空间的维数

定义

非零子空间 H 的**维数** (用 $\dim H$ 表示) 是 H 的任意一个基的向量个数。零子空间 $\{\mathbf{0}\}$ 的维数定义为零。

定义

矩阵 A 的**秩** (记为 $\text{rank} A$) 是 A 的列空间的维数

2.9.3 定理14 (秩定理)

如果一个矩阵 A 有 n 列, 则 $\text{rank} A + \dim \text{Nul } A = n$

2.9.4 定理15 (基定理)

设 H 是 \mathbb{R}^n 的 p 维子空间, H 中的任何恰好由 p 个元素组成的线性无关集构成 H 的一个基。并且, H 中任何生成 H 的 p 个向量集也构成 H 的一个基

2.9.5 秩和可逆矩阵定理

设 A 是一 $n \times n$ 矩阵, 则下面的每个命题与 A 是可逆矩阵的命题等价:

m. A 的列向量构成 \mathbb{R}^n 的一个基

n. $\text{Col } A = \mathbb{R}^n$

o. $\dim \text{Col } A = n$

p. $\text{rank} A = n$

q. $\text{Nul } A = \{\mathbf{0}\}$

r. $\dim \text{Nul } A = 0$

3 行列式

3.1 行列式介绍

一些记号的说明

A_{ij} 表示通过删去 A 中第 i 行第 j 列得到的子矩阵
代数余子式 (余因子) $C_{ij} = (-1)^{i+j} \det A_{ij}$

3.1.1 行列式的递归定义

当 $n \geq 2$ 时, $n \times n$ 矩阵 $A = [a_{ij}]$ 的行列式是形如 $\pm a_{1j} \det A_{1j}$ 的 n 个项的和, 其中加号与减号交替出现, 元素 $a_{11}, a_{12}, \dots, a_{1n}$ 来自 A 的第一行, 用符号表示为

$$\begin{aligned}\det A &= a_{11} \cdot \det A_{11} - a_{12} \cdot \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \cdot \det A_{1n} \\ &= \sum_{j=1}^n (-1)^{1+j} a_{1j} \det A_{1j}\end{aligned}$$

3.1.2 定理1

行列式的余因子展开
按照第 i 行展开:

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

按照第 j 列展开

$$\det A = \sum_{i=1}^n a_{ij} C_{ij}$$

3.1.3 定理2

若 A 为三角阵 (上三角形矩阵、下三角形矩阵), 则 $\det A$ 等于 A 的主对角线上的元素的乘积

3.2 行列式的性质

3.2.1 定理3 (行(列)变换)

令 A 是一个方阵。

- a. 若 A 的某一行(列)的倍数加到另一行(列)得到矩阵 B , 则 $\det B = \det A$
- b. 若 A 的两行(列)互换得到矩阵 B , 则 $\det B = -\det A$
- c. 若 A 的某行(列)乘以 k 得到矩阵 B , 则 $\det B = k \cdot \det A$

3.2.2 定理4

方阵 A 是可逆的当且仅当 $\det A \neq 0$

3.2.3 定理5

若 A 为一个 $n \times n$ 矩阵, 则 $\det A = \det A^T$

3.2.4 定理6 (乘法的性质)

若 A 和 B 都为 $n \times n$ 矩阵, 则 $\det AB = (\det A)(\det B)$

3.2.5 行列式的一个线性性质

若 A 为 $n \times n$ 矩阵, 我们可以把 A 视作

$$\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{a}_j & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

假设 A 的第 j 列允许变化, 那么 A 可以写成

$$\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

定义由 \mathbb{R}^n 到 \mathbb{R} 的变换为 T

则有

$$T(c\mathbf{x}) = cT(\mathbf{x})$$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

【根据这个定理，我们可以在行列式运算中巧妙地将一个行列式“拆开”】

$$\begin{bmatrix} a_1 + c_1 & b_1 + d_1 \\ a_2 & b_2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} c_1 & d_1 \\ a_2 & b_2 \end{bmatrix}$$

对于一些（代数）余子式线性运算的处理方式

对于矩阵

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

，求代数余子式的和： $A_{11} + A_{12} + A_{13} + A_{14}$

首先固然是可以进行硬算的，然而，我们可以将矩阵的第一行替换为

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

【当然，如果线性组合是其他的也可以同理构造，如果要求的是余子式而非代数余子式也是同理，只需要将部分数变为相反数即可】

3.3 克拉姆法则、体积和线性变换

定义记号

定义 $A_i(\mathbf{b}) = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix}$ 表示 A 的第 i 列被向量 \mathbf{b} 替换得到的矩阵

3.3.1 定理7（克拉姆法则）

设 A 是一个 $n \times n$ 可逆矩阵，对 \mathbb{R}^n 任意向量 \mathbf{b} ，方程 $A\mathbf{x} = \mathbf{b}$ 的唯一解可由下式给出

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, i = 1, 2, \cdots, n$$

定义记号

$$\text{伴随矩阵} \operatorname{adj} A = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}, \text{ 其中 } C_{ij} \text{ 表示代数余子式 (余因子)}$$

3.3.2 定理8 (逆矩阵公式)

设 A 是一个 $n \times n$ 可逆矩阵, 则 $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$
事实上, 我们是先求得

$$\operatorname{adj}.A A = A \operatorname{adj}.A = (\det A) I$$

然后再利用 A 可逆, 在等式两边左乘一个 A 的逆得到定理8中的等式的

进而我们可以严格证明:

Cramer法则可解一个线性方程组, 也可解 n 个线性方程组, i 和 j 分别是行指标和列指标

我们考虑求逆的过程为

$$AX = I, \text{ i.e. } Ax^j = \varepsilon^j$$

对于上面每一个这样的线性方程组, 我们可以再使用Cramer法则进行求解

$$x_i^j = \frac{\det A(\mathbf{a}^i \leftarrow \varepsilon^j)}{\det A}$$

我们可以对第 i 列展开, 之后即可得到分子等于 C_{ij}

3.3.3 行列式求解的特殊技巧

利用矩阵的转置性质

例 (这个我们在向量代数这章中是重要的度量矩阵)

$$B = \begin{bmatrix} \mathbf{e}_1 \cdot \mathbf{e}_1 & \mathbf{e}_1 \cdot \mathbf{e}_2 & \mathbf{e}_1 \cdot \mathbf{e}_3 \\ \mathbf{e}_2 \cdot \mathbf{e}_1 & \mathbf{e}_2 \cdot \mathbf{e}_2 & \mathbf{e}_2 \cdot \mathbf{e}_3 \\ \mathbf{e}_3 \cdot \mathbf{e}_1 & \mathbf{e}_3 \cdot \mathbf{e}_2 & \mathbf{e}_3 \cdot \mathbf{e}_3 \end{bmatrix}$$

我们记 $A = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix}$ 则度量矩阵 $B = AA^T$

由于对于矩阵 $C = A^T A$ 有 $C = \mathbf{e}_1^2 + \mathbf{e}_2^2 + \mathbf{e}_3^2 > 0$

故 $\det C = \det A \cdot \det A^T > 0$

故而对于度量矩阵 B 我们有 $\det B = \det A \cdot \det A^T = \det A^T \cdot \det A > 0$

事实上，对于一些具有对称性的行列式我们采用这样的方式进行计算是非常常见的！

同理，我们可以使用二次方程判别式来进行证明对任意的实数 t ，向量 $t\mathbf{e}_1 + \mathbf{e}_2$ 与向量 \mathbf{e}_3 不共线，由 Cauchy-Schwarz 不等式

$$(t\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_3)(t\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_3) < (t\mathbf{e}_1 + \mathbf{e}_2, t\mathbf{e}_1 + \mathbf{e}_2)(\mathbf{e}_3, \mathbf{e}_3)$$

$$\therefore 0 < t^2(a_{11}a_{33} - a_{13}^2) + 2t(a_{12}a_{33} - a_{13}a_{23}) + (a_{22}a_{33} - a_{23}^2)$$

作为一个二次函数，其判别式 Δ 应当小于 0 而判别式 Δ 可以化简为 $-a_{33} \det A$ 又因为 $a_{33} > 0$ ，故可证明 $\det A > 0$

3.3.4 几个特殊的行列式

范德蒙行列式

$$D_2 = \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} = x_2 - x_1$$

$$D_3 = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ x_1^2 & x_2^2 & x_3^2 \end{bmatrix} = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$$

$$D_n = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & & \vdots \\ x_1^n & x_2^n & \cdots & x_n^n \end{bmatrix} = \prod_{1 \leq j < i \leq n} (x_i - x_j)$$

证明：数学归纳法

首先可以先证明 D_2 正确

进而假设 $(n-1)$ 阶正确，证明 n 阶正确

证明对于 n 阶的方式

用第 n 行减去 a_1 倍第 $n-1$ 行，用第 $n-1$ 行减去 a_1 倍 $n-2$ 行 \dots ，然后从后面 $(n-1)$ 列中提出部分式子，再利用归纳假设

3.3.5 分块行列式的求法点拨

事实上，在我们的教材里没有交代复杂的分块矩阵行列式的求法，但是，我们可以通过左乘与右乘初等矩阵的方式对行列式进行我们想要的变换，从而实现将分块矩阵对角化

命题

如果 $A = \begin{bmatrix} B & C \\ D & \end{bmatrix}$ ，则 $\det A = \det B \cdot \det D$

例题

求 $\det \begin{bmatrix} A & B \\ B & A \end{bmatrix}$

相对合适的方式是我们先通过左乘右乘初等矩阵于原矩阵，试图得到一个对角的分块矩阵

解：记原矩阵为 D

$$D = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix} = \begin{bmatrix} A & B \\ A+B & A+B \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} A & B \\ A+B & A+B \end{bmatrix}$$

然后就可以把目标矩阵用初等矩阵的逆和一个分块对角矩阵表示出来

基于舒尔补的强行解法

首先我们来看一下什么是舒尔补

对于矩阵 $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, 其中 A_{11}, A_{22} 都是方阵。若 A_{11} 可逆我们称 $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ 为 A_{11} 的舒尔补

同理, 若 A_{22} 可逆我们称 $S' = A_{11} - A_{12}A_{22}^{-1}A_{21}$ 为 A_{22}

【注意只有对角线上的子矩阵才有舒尔补】

接下来我们给出 A 的行列式的通用表达式

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & Y \\ 0 & I \end{bmatrix}$$

其中

$$Y = A_{11}^{-1}A_{12}, X = A_{21}A_{11}^{-1}$$

并且 X, Y 都是可逆矩阵

命题

如果 $A = \begin{bmatrix} B & D \\ C & \end{bmatrix}$, $B \in P^{n \times n}, C \in P^{m \times m}, D \in P^{n \times m}$, 则 $\det A = \det B \cdot \det C$

这个命题的证明可以利用递归 (数学归纳法), 对 n 进行归纳, 归纳的过程中对始终对第1列展开即可

用行列式表示面积或体积

3.3.6 定理9

若 A 是一个 2×2 矩阵, 则由 A 的列确定的平行四边形的面积为 $|\det A|$ 。
若 A 为一个 3×3 矩阵, 则由 A 的列确定的平行六面体的体积为 $|\det A|$

线性变换

3.3.7 定理10

设 $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ 是由一个 2×2 矩阵 A 确定的线性变换, 若 S 为 \mathbb{R}^2 中的一个平行四边形, 则

$$\{T(S) \text{ 的面积} \} = |\det A| \cdot \{S \text{ 的面积} \}$$

设 $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ 是由一个 3×3 矩阵 A 确定的线性变换, 若 S 为 \mathbb{R}^3 中的一个平行六面体, 则

$$\{T(S) \text{ 的体积} \} = |\det A| \cdot \{S \text{ 的体积} \}$$

【定理10的推广: 定理10的结论对于 \mathbb{R}^2 中任何具有有限面积的区域或 \mathbb{R}^3 中任何具有有限体积的区域都成立】
(比如适用于椭圆、椭球等等)

我们在证明经过线性变换的元素组成的集合等于另一个集合时通常证明二者互为对方的子集

例

令 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是一个线性变换, \mathbf{p} 是 \mathbb{R}^n 中一个向量, S 是 \mathbb{R}^n 中一个集合, 证明: $\mathbf{p} + S$ 在 T 下的像等于 \mathbb{R}^m 中平移的集合 $T(\mathbf{p}) + T(S)$

证: 记 $\mathbf{p} + S = \{\mathbf{p} + \mathbf{v} | \mathbf{v} \in S\} \equiv M$
对于 $\forall \mathbf{u} \in T(M), \exists \mathbf{v} \in S \text{ s.t. } \mathbf{u} = T(\mathbf{p} + \mathbf{v}) = T(\mathbf{p}) + T(\mathbf{v})$

$$\therefore \mathbf{u} \in T(\mathbf{p}) + T(S), \text{ i.e. } T(M) \subset T(\mathbf{p}) + T(S)$$

同理对于 $\forall \mathbf{u} \in T(\mathbf{p}) + T(S), \mathbf{u} - T(\mathbf{p}) \in T(S), \exists \mathbf{v} \in S, \text{ s.t. } \mathbf{u} - T(\mathbf{p}) = T(\mathbf{v})$

$$\therefore \mathbf{u} = T(\mathbf{p}) + T(\mathbf{v}) = T(\mathbf{p} + \mathbf{v})$$

$$\therefore \mathbf{u} \in T(M), \text{ i.e. } T(\mathbf{p}) + T(S) \subset T(M)$$

补充章：向量代数、平面与直线

1. 向量及其线性运算

一系列定义

向量

向量的和

向量的差

向量的数乘

向量的共线与共面

定理1.1

向量 α, β 共线的充分必要条件是存在不全为零的实数 k, l ，使得

$$k\alpha + l\beta = \mathbf{0}$$

定理1.2

向量 α, β, γ 共面的充分必要条件是存在不全为零的实数 k, l, m ，使得

$$k\alpha + l\beta + m\gamma = \mathbf{0}$$

三角形特殊线相关运算问题（这个主要是为了应付考试，实际意义不大）

各种线的表示方式

中线

$$\overrightarrow{AD} = \frac{1}{2}\overrightarrow{AB} + \frac{1}{2}\overrightarrow{AC}$$

角平分线

$$\overrightarrow{AD} = \frac{AC}{AB + AC}\overrightarrow{AB} + \frac{AB}{AB + AC}\overrightarrow{AC}$$

另一种表述方式是利用

$$\overrightarrow{AD} = k\left(\frac{1}{AB}\overrightarrow{AB} + \frac{1}{AC}\overrightarrow{AC}\right)$$

高线

高线我们换一种角度思考：

$$\overrightarrow{BH} = \frac{(\overrightarrow{BC}, \overrightarrow{BA})}{(\overrightarrow{BC}, \overrightarrow{BC})} \overrightarrow{BC}$$

$$\overrightarrow{AH} = \overrightarrow{AB} + \overrightarrow{BH} = \overrightarrow{AB} - \frac{(\overrightarrow{AC} - \overrightarrow{AB}, \overrightarrow{AB})}{(\overrightarrow{AC} - \overrightarrow{AB}, \overrightarrow{AC} - \overrightarrow{AB})} (\overrightarrow{AC} - \overrightarrow{AB})$$

各种线交于一点的证明方式

首先，我们要明确证明的核心任务：

两条不平行的直线肯定交于一点，我们要证明的是第三条线过前两条线的交点

三条中线交于一点

我们假设前面两条中线 BE, CF 交于一点 G

不妨记第三条中线 AD

我们有

$$\overrightarrow{AD} = \frac{1}{2}\overrightarrow{AB} + \frac{1}{2}\overrightarrow{AC}$$

$$\overrightarrow{BE} = \frac{1}{2}\overrightarrow{BC} + \frac{1}{2}\overrightarrow{BA}$$

$$\overrightarrow{CF} = \frac{1}{2}\overrightarrow{CA} + \frac{1}{2}\overrightarrow{CB}$$

由于 G 在 BE 上，从而有

$$\overrightarrow{BG} = \frac{\lambda}{2}\overrightarrow{BC} + \frac{\lambda}{2}\overrightarrow{BA}$$

同理

$$\overrightarrow{CG} = \frac{\mu}{2}\overrightarrow{CA} + \frac{\mu}{2}\overrightarrow{CB}$$

$$\begin{aligned}
\overrightarrow{BC} &= \overrightarrow{BG} - \overrightarrow{CG} \\
&= \frac{\lambda}{2}(\overrightarrow{AC} - \overrightarrow{AB}) - \frac{\lambda}{2}\overrightarrow{AB} + \frac{\mu}{2}\overrightarrow{AC} - \frac{\mu}{2}(\overrightarrow{AB} - \overrightarrow{AC}) \\
&= (\frac{\lambda}{2} + \mu)\overrightarrow{AC} + (-\frac{\mu}{2} - \lambda)\overrightarrow{AB} \\
&= \overrightarrow{AC} - \overrightarrow{AB}
\end{aligned}$$

解得

$$\lambda = \mu = \frac{2}{3}$$

从而

$$\overrightarrow{AG} = \overrightarrow{AC} + \overrightarrow{CG} = \frac{1}{3}(\overrightarrow{AC} + \overrightarrow{AB})$$

证毕

三条角平分线交于一点

对于三角形ABC, 记 $\overrightarrow{CA} = \mathbf{a}$, $\overrightarrow{AB} = \mathbf{b}$, $\overrightarrow{BC} = \mathbf{c}$
有 $\mathbf{a} + \mathbf{b} + \mathbf{c} = \mathbf{0}$

我们可以构造单位向量

$$\mathbf{i} = \frac{\mathbf{a}}{|\mathbf{a}|}, \mathbf{j} = \frac{\mathbf{b}}{|\mathbf{b}|}, \mathbf{k} = \frac{\mathbf{c}}{|\mathbf{c}|}$$

记I为过C和B的两条角平分线的交点

设 $\overrightarrow{CI} = \lambda(\mathbf{i} - \mathbf{k})$, $\overrightarrow{BI} = \mu(\mathbf{k} - \mathbf{j})$

$$\begin{aligned}
\overrightarrow{BI} &= \overrightarrow{BC} + \overrightarrow{CI} \\
&= \mathbf{c} + \lambda(\mathbf{i} - \mathbf{k}) \\
&= c\mathbf{k} + \lambda\mathbf{i} - \lambda\mathbf{k}
\end{aligned}$$

将j用i和k表示出来, 然后反解 λ, μ , 然后表示出 \overrightarrow{AI} (用i和j)就结束了

三条高线交于一点

方法一: 利用高线的表达式, 相对麻烦一点

我们假设AD, BE, CF为三条高线, BE与CF的交点为H

这里我们不同的一点是只需要证明 $\overrightarrow{AH} \perp \overrightarrow{BC}$

记 $\overrightarrow{AB} = \alpha, \overrightarrow{AC} = \beta$

$$\overrightarrow{AE} = \frac{(\alpha, \beta)}{(\beta, \beta)} \beta$$

$$\overrightarrow{BE} = \frac{(\alpha, \beta)}{(\beta, \beta)} \beta - \alpha$$

同理

$$\overrightarrow{CF} = \frac{(\alpha, \beta)}{(\alpha, \alpha)} \alpha - \beta$$

$$\overrightarrow{BC} = \overrightarrow{BG} - \overrightarrow{CG}$$

$$\beta - \alpha = \lambda \frac{(\alpha, \beta)}{(\beta, \beta)} \beta - \lambda \alpha - \mu \frac{(\alpha, \beta)}{(\alpha, \alpha)} \alpha + \mu \beta$$

$$1 = \lambda + \mu \frac{(\alpha, \beta)}{(\alpha, \alpha)}, 1 = \mu + \lambda \frac{(\alpha, \beta)}{(\beta, \beta)}$$

进而我们表示 \overrightarrow{AH}

$$\overrightarrow{AH} = \alpha + \lambda \frac{(\alpha, \beta)}{(\beta, \beta)} \beta - \lambda \alpha$$

$$\overrightarrow{AH} = (1 - \lambda) \alpha + (1 - \mu) \beta$$

$$\overrightarrow{AH}(\beta - \alpha) = (1 - \lambda) \alpha \beta - (1 - \lambda) \alpha \alpha + (1 - \mu) \beta \beta - (1 - \mu) \beta \alpha$$

$$= (\mu - \lambda) \alpha \beta - (1 - \lambda) \alpha \alpha + (1 - \mu) \beta \beta$$

$$= (\mu - \lambda) \alpha \beta - \mu \alpha \beta + \lambda \alpha \beta$$

$$= 0$$

方法二：同三条角平分线的证法，构造一个和为零向量的环，并且假设其中一条高为另一个向量，之后同方法一，证明一个数量积为0即可

2.坐标系

仿射坐标系

定理2.1

设 e_1, e_2, e_3 为几何空间中三个不共面的向量，则对每个向量 a 都存在着唯一的三元有序实数组 x, y, z 使得

$$a = xe_1 + ye_2 + ze_3$$

定义2.1

空间中任意三个有序不共面的向量 e_1, e_2, e_3 称为空间的一组基，对向量 a ，若

$$a = xe_1 + ye_2 + ze_3$$

则称 (x, y, z) 为向量 a 在基 e_1, e_2, e_3 下的仿射坐标或简称坐标。可以记作

$$a = (e_1 \ e_2 \ e_3) \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

定义2.2

空间中任意一点 O 和一组基 e_1, e_2, e_3 合在一起称为空间的一个仿射坐标系，记为 $[O; e_1, e_2, e_3]$ 。点 O 称为坐标原点， (x, y, z) 称为坐标向量。 e_1, e_2, e_3 所在直线分别称为 x 轴， y 轴， z 轴，统称为坐标轴。三个坐标轴中的任意两个与原点 O 一起决定一个平面，称为坐标面，分别记为 Oxy, Oyz, Ozx 。

向量的坐标运算

定比分点公式

若 $\overrightarrow{AP} = \lambda \overrightarrow{PB}$ ，则

$$\overrightarrow{OP} = \frac{1}{1+\lambda} \overrightarrow{OA} + \frac{\lambda}{1+\lambda} \overrightarrow{OB} = \left(\frac{x_A + \lambda x_B}{1+\lambda}, \frac{y_A + \lambda y_B}{1+\lambda}, \frac{z_A + \lambda z_B}{1+\lambda} \right)$$

如何判定一个点在三角形内还是外

假设 A, B, C, P 四点共面，而 A, B, C 不共线那么 $\overrightarrow{OP} = s\overrightarrow{OA} + t\overrightarrow{OB} + (1-s-t)\overrightarrow{OC}$

要判断点P在三角形内还是外,只要判断以下三次:

$$t > 0, s > 0, 1 - t - s > 0 \quad (1)$$

在这里我们引入**计算机图形学**中判断的方式

我们定义一个三角形的有向面积矩阵(虽然实际面积是这个矩阵的行列式的值的一半)

对 $\triangle ABC, A(x_1, y_1), B(x_2, y_2), C(x_3, y_3)$

有向面积矩阵A可以表示为

$$A = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}$$

这个矩阵有一种理解方式是把A,B,C三点都放在 $z = 1$ 的平面上,那么这个矩阵对应的行列式就变成了3个向量的混合积,表征了平行六面体的体积,由于这个平行六面体的特殊性质,其体积等于三角形ABC的面积乘以高1乘上 $\frac{1}{3}$ (算出三棱锥的体积)再乘以3(乘上3倍才得到平行六面体的体积)事实上,这个矩阵的行列式的值等同于下面的矩阵的行列式,后者可能更为直观

$$A = \begin{bmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{bmatrix}$$

我们可以有性质:

$$\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} = \det \begin{bmatrix} 1 & 1 & 1 \\ x & x_2 & x_3 \\ y & y_2 & y_3 \end{bmatrix} + \det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x & x_3 \\ y_1 & y & y_3 \end{bmatrix} + \det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x \\ y_1 & y_2 & y \end{bmatrix}$$

我们成功地将一个三角形的面积表示为三个行列式的和

这时,如果我们可以根据这些行列式的符号来判断是否在三角形内

比如: $\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}$ 和 $\det \begin{bmatrix} 1 & 1 & 1 \\ x & x_2 & x_3 \\ y & y_2 & y_3 \end{bmatrix}$ 是同号的,那么可以说明,点P与点A在

直线BC的同侧

基于同样的方式,我们可以进行3次这样的判断,如果三次结果都为同号,则说

明点 P 在 $\triangle ABC$ 内

思考:为什么有这样的结果?

我们可以考虑在行列式等式的两侧同时除以 $\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}$,我们发现等式左边变成了1,而等式右边恰好是Cramer法则得出的三个根的形式

$$m_1 = \frac{\det \begin{bmatrix} 1 & 1 & 1 \\ x & x_2 & x_3 \\ y & y_2 & y_3 \end{bmatrix}}{\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}}, m_2 = \frac{\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x & x_3 \\ y_1 & y & y_3 \end{bmatrix}}{\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}}, m_3 = \frac{\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x \\ y_1 & y_2 & y \end{bmatrix}}{\det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}}$$

我们可以好好想想这样的三个比值式代表了什么:

这 $\mathbf{m} = (m_1, m_2, m_3)$ 可以视作线性方程组的解集

这三个分量满足的关系是和为1

对应的是增广矩阵如下的线性方程组

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x \\ y_1 & y_2 & y_3 & y \end{bmatrix}$$

你会发现这和我们最开始时候 $\overrightarrow{OP} = s\overrightarrow{OA} + t\overrightarrow{OB} + (1-s-t)\overrightarrow{OC}$ 的表达式实际上是一致的

这三个分量可以分别理解为 $s, t, 1-s-t$,因而分量都大于0或者说对应行列式同号就对应着 $s, t, 1-s-t$ 这三个系数大于0,也就是(1)式中的判定条件

如何判定一个点在四面体内还是外

这是原先的问题的进一步推广

有一个基础的引论:

设点A, B, C, D不共面, 那么空间中任何一点P, 可以表示为

$$\begin{aligned} \overrightarrow{DP} &= r\overrightarrow{DA} + s\overrightarrow{DB} + t\overrightarrow{DC} \\ \therefore \overrightarrow{OP} &= r\overrightarrow{OA} + s\overrightarrow{OB} + t\overrightarrow{OC} + (1-r-s-t)\overrightarrow{OD} \end{aligned}$$

在四面体内的条件是

$$\begin{cases} 0 < r < 1 \\ 0 < s < 1 \\ 0 < t < 1 \\ 0 < (1 - r - s - t) < 1 \end{cases}$$

这表明了 P 点在与 O 在面 ABC, ABD, ACD, BCD 的同侧

与前面判断是否在三角形内的情况类似，我们也可以构造行列式来判断，不妨假设 $A(x_1, y_1, z_1), B(x_2, y_2, z_2), C(x_3, y_3, z_3), D(x_4, y_4, z_4)$ ，那么可以构造行列式

$$\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix}$$

同理替换其中的的某一列为 $P(x, y, z)$ ，所得到的行列式应当与原先的行列式同号，那么则可以说明这个点在四面体 $ABCD$ 内

直角坐标系

特殊的仿射坐标系,三个坐标向量 $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 两两正交且为**单位向量**(我们还一般要求 $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 满足右手系)

定义2.3

在直角坐标系中,向量 $\overrightarrow{OA} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}$ 与三个坐标向量 $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 的夹角分别是 α, β, γ 称为向量 $\overrightarrow{OA} = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k}$ 的**方向角**, $\cos \alpha, \cos \beta, \cos \gamma$ 称为向量的**方向余弦**

不难看出

$$\cos \alpha = \frac{x_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} = \frac{x_1}{|\overrightarrow{OA}|}$$

$$\cos \beta = \frac{y_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} = \frac{y_1}{|\overrightarrow{OA}|}$$

$$\cos \gamma = \frac{z_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} = \frac{z_1}{|\vec{OA}|}$$

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1$$

八个卦限的记忆方式

前四个对应 z 为正, xy 分别对应平面直角坐标系中的四个象限
 后四个对应 z 为负, xy 分别对应平面直角坐标系中的四个象限

3.向量的数量积

定义3.1

两个向量 α, β 的数量积为一个实数

$$\alpha \cdot \beta = |\alpha| \cdot |\beta| \cos \langle \alpha, \beta \rangle$$

向量长度的定义

$$|\alpha| = \sqrt{\alpha^2} = \sqrt{(\alpha, \alpha)}$$

向量的正交分解

设 $\vec{OA} = \alpha, \vec{OB} = \beta$

$$\beta = \beta_{\alpha}^{\parallel} + \beta_{\alpha}^{\perp}$$

其中 $\beta_{\alpha}^{\parallel}$ 是 β 在 α 上的正交投影向量或关于向量 α 平行分量

β_{α}^{\perp} 称为向量 β 在 α 的垂直分量

$\beta_{\alpha}^{\parallel}$ 的代数模 $|\beta| \cos \theta$ 称为 β 的正交投影

数量积的性质

略

度量矩阵

在仿射坐标系 $[O; \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ 下向量 $\boldsymbol{\alpha} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3$ 和 $\boldsymbol{\beta} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$ 的数量积我们可以如下计算,

$$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = \left(\sum_{i=1}^3 u_i \mathbf{e}_i \right) \cdot \left(\sum_{j=1}^3 v_j \mathbf{e}_j \right) = \sum_{i=1}^3 \sum_{j=1}^3 u_i v_j (\mathbf{e}_i \cdot \mathbf{e}_j)$$

记基向量的数量积 $\mathbf{e}_i \cdot \mathbf{e}_j = a_{ij}$,写成矩阵的形式,就得到

$$A = \begin{bmatrix} \mathbf{e}_1 \cdot \mathbf{e}_1 & \mathbf{e}_1 \cdot \mathbf{e}_2 & \mathbf{e}_1 \cdot \mathbf{e}_3 \\ \mathbf{e}_2 \cdot \mathbf{e}_1 & \mathbf{e}_2 \cdot \mathbf{e}_2 & \mathbf{e}_2 \cdot \mathbf{e}_3 \\ \mathbf{e}_3 \cdot \mathbf{e}_1 & \mathbf{e}_3 \cdot \mathbf{e}_2 & \mathbf{e}_3 \cdot \mathbf{e}_3 \end{bmatrix}$$

这里的 A 为基 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 的度量矩阵,则

$$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 \cdot \mathbf{e}_1 & \mathbf{e}_1 \cdot \mathbf{e}_2 & \mathbf{e}_1 \cdot \mathbf{e}_3 \\ \mathbf{e}_2 \cdot \mathbf{e}_1 & \mathbf{e}_2 \cdot \mathbf{e}_2 & \mathbf{e}_2 \cdot \mathbf{e}_3 \\ \mathbf{e}_3 \cdot \mathbf{e}_1 & \mathbf{e}_3 \cdot \mathbf{e}_2 & \mathbf{e}_3 \cdot \mathbf{e}_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{u}^T A \mathbf{v}$$

可以证明,度量矩阵的顺序主子式全部大于零,即

$$a_{11} > 0, \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} > 0, \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} > 0$$

4.向量的向量积

定义4.1

两个向量 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 的向量积 $\boldsymbol{\alpha} \times \boldsymbol{\beta}$ 为一个向量,其方向与 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 都垂直,且使 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 与 $\boldsymbol{\alpha} \times \boldsymbol{\beta}$ 构成右手系

其模为以 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 为边的平行四边形的面积(这个面积是有向的)

$$\boldsymbol{\alpha} \times \boldsymbol{\beta} = |\boldsymbol{\alpha}| \cdot |\boldsymbol{\beta}| \sin \angle \boldsymbol{\alpha}, \boldsymbol{\beta}$$

命题4.1(向量积的性质)

设 $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ 为三个向量, k 为实数,则有

$$(1) \boldsymbol{\alpha} \times \boldsymbol{\beta} = -\boldsymbol{\beta} \times \boldsymbol{\alpha}$$

$$(2)(k\alpha) \times \beta = \alpha \times (k\beta) = k(\alpha \times \beta)$$

$$(3)\alpha \times (\beta + \gamma) = \alpha \times \beta + \alpha \times \gamma$$

在仿射坐标系 $[O; e_1, e_2, e_3]$ 下向量 $\alpha = u_1e_1 + u_2e_2 + u_3e_3$ 和 $\beta = v_1e_1 + v_2e_2 + v_3e_3$ 的数量积我们可以如下计算,

$$\alpha \times \beta = \left(\sum_{i=1}^3 u_i e_i \right) \times \left(\sum_{j=1}^3 v_j e_j \right) = \sum_{i=1}^3 \sum_{j=1}^3 u_i v_j (e_i \times e_j)$$

或者形式上写成行列式

$$\alpha \times \beta = \det \begin{bmatrix} e_2 \times e_3 & e_1 \times e_3 & e_1 \times e_2 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}$$

5.三向量积

混合积

给定向量 α, β, γ ,称数量 $(\alpha, \beta, \gamma) = (\alpha \times \beta) \cdot \gamma$ 为向量 α, β, γ 的**数量三向量积**或**混合积**

几何上,混合积 (α, β, γ) 以 α, β, γ 为棱的平行六面体的有向体积

混合积的运算性质

$$(1)\text{轮转对称性:}(\alpha, \beta, \gamma) = (\gamma, \alpha, \beta) = (\beta, \gamma, \alpha)$$

$$(2)\text{反对称性:}(\alpha, \beta, \gamma) = -(\beta, \alpha, \gamma)$$

$$(3)(k\alpha, \beta, \gamma) = (\alpha, k\beta, \gamma) = (\alpha, \beta, k\gamma) = k(\alpha, \beta, \gamma), \text{对任意的实数} k$$

$$(4)(\alpha_1 + \alpha_2, \beta, \gamma) = (\alpha_1, \beta, \gamma) + (\alpha_2, \beta, \gamma)$$

$$(5)(\alpha, \beta, \gamma) = 0 \text{ 当且仅当 } \alpha, \beta, \gamma \text{ 共面}$$

混合积是判断共面的工具!!!

设向量 $\alpha = u_1e_1 + u_2e_2 + u_3e_3$ 和 $\beta = v_1e_1 + v_2e_2 + v_3e_3, \gamma = w_1e_1 + w_2e_2 + w_3e_3$,再次运用矩阵的记号,有

$$\begin{aligned}
(\alpha, \beta, \gamma) &= (\alpha \times \beta) \cdot \gamma = \det \begin{bmatrix} \mathbf{e}_2 \times \mathbf{e}_3 & \mathbf{e}_1 \times \mathbf{e}_3 & \mathbf{e}_1 \times \mathbf{e}_2 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} (w_1 \mathbf{e}_1 + w_2 \mathbf{e}_2 + w_3 \mathbf{e}_3) \\
&= \det \begin{bmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \det \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)
\end{aligned}$$

我们得到的几何解释:

这里的混合积代表一个以 α, β, γ 为棱的平行六面体的体积与以 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 为棱的平行六面体体积的比值

复合积

给定三个向量 α, β, γ , 称向量 $(\alpha \times \beta) \times \gamma$ 为向量 α, β, γ 的**向量三向量积**或**复合积**

定理5.1

$$(\alpha \times \beta) \times \gamma = (\alpha \gamma) \beta - (\beta \gamma) \alpha$$

注:更有意思的是对偶关系,记 $\alpha' = \beta \times \gamma, \beta' = \gamma \times \alpha, \gamma' = \alpha \times \beta$, 则

$$\alpha' \times \beta' = (\alpha, \beta, \gamma) \gamma, \beta' \times \gamma' = (\alpha, \beta, \gamma) \alpha, \gamma' \times \alpha' = (\alpha, \beta, \gamma) \beta$$

在许多物理问题中,常常取

$$\alpha'' = \frac{\beta \times \gamma}{(\alpha, \beta, \gamma)}$$

$$\beta'' = \frac{\gamma \times \alpha}{(\alpha, \beta, \gamma)}$$

$$\gamma'' = \frac{\alpha \times \beta}{(\alpha, \beta, \gamma)}$$

为 α, β, γ 的对偶基.从矩阵的角度,如果 α, β, γ 是三阶矩阵 A 的列向量,那么 $\alpha'', \beta'', \gamma''$ 对应 A^{-1} 的行向量

(从Cramer法则角度解释)

一些更为复杂的恒等式及其代数与几何意义

Lagrange恒等式

$$(\alpha \times \beta) \cdot (\gamma \times \delta) = (\alpha\gamma)(\beta\delta) - (\alpha\delta)(\beta\gamma)$$

实际代数意义

我们首先可以假设

$$\alpha = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}, \beta = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$$

$$\gamma = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}, \delta = d_1\mathbf{i} + d_2\mathbf{j} + d_3\mathbf{k}$$

那么根据叉乘的定义

$$\alpha \times \beta = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \mathbf{k} + \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} \mathbf{i} - \begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix} \mathbf{j}$$

$$\gamma \times \delta = \begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \end{bmatrix} \mathbf{k} + \begin{bmatrix} c_2 & c_3 \\ d_2 & d_3 \end{bmatrix} \mathbf{i} - \begin{bmatrix} c_1 & c_3 \\ d_1 & d_3 \end{bmatrix} \mathbf{j}$$

当然事实上，我们在推导时为了完备性，我们一般可以先假设这四个向量共面，即没有 \mathbf{k} 的分量，然后再讨论有这个分量的情况，当没有这个分量的时候这个结论同样可以证明是成立的

$$(\alpha \times \beta) \cdot (\gamma \times \delta) = \begin{bmatrix} \alpha\gamma & \alpha\delta \\ \beta\gamma & \beta\delta \end{bmatrix}$$

接下来我们考虑这个式子与三个分量之间是否存在一定的关联

下面的左式可以用来表示形式上数量积的写法，右式则是通过Lagrange恒等式化简后的形式

$$\begin{aligned}
& \det \begin{bmatrix} a_1 & a_1 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} c_1 & c_1 \\ d_1 & d_2 \end{bmatrix} + \det \begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} \begin{bmatrix} c_2 & c_3 \\ d_2 & d_3 \end{bmatrix} + \det \begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix} \begin{bmatrix} c_1 & c_3 \\ d_1 & d_3 \end{bmatrix} \\
&= \det \begin{bmatrix} a_1c_1 + a_2c_2 + a_3c_3 & a_1d_1 + a_2d_2 + a_3d_3 \\ b_1c_1 + b_2c_2 + b_3c_3 & b_1d_1 + b_2d_2 + b_3d_3 \end{bmatrix} \\
&= \det \left(\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \\ c_3 & d_3 \end{bmatrix} \right)
\end{aligned}$$

这相当于我们把Lagrange恒等式转化为了代数上的两个矩阵相乘的形式，由此我们也可以进行推广，将这个推广到 \mathbb{R}^n 空间中，既然有 n 个分量，那么事实上这样的乘积可以表示为

$$\begin{aligned}
& \sum_{1 \leq i, j \leq n} \det \begin{bmatrix} a_1 & a_1 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} c_1 & c_1 \\ d_1 & d_2 \end{bmatrix} \\
&= \det \begin{bmatrix} \sum_{i=1}^n a_i c_i & \sum_{i=1}^n a_i d_i \\ \sum_{i=1}^n b_i c_i & \sum_{i=1}^n b_i d_i \end{bmatrix} \\
&= \det \left(\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix} \begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \\ \vdots & \vdots \\ c_n & d_n \end{bmatrix} \right)
\end{aligned}$$

这其中的第一个等号是Binet定理的特殊形式

Jacobi恒等式

$$(\alpha \times \beta) \times \gamma + (\gamma \times \alpha) \times \beta + (\beta \times \gamma) \times \alpha = \mathbf{0}$$

证明方式很简单，只需要分别点乘 α, β, γ 即可

几何意义

【这一点再ppt上表述的也不是很清楚，不过需要明确的一点是】

$\alpha \times \beta$ 与 α 和 β 正交

进而我们能确定这个 $(\alpha \times \beta) \times \gamma$ 与 α, β 共面的

6. 平面

平面的确定方法

- (1) 一个点 P_0 和在平面上的两个不共线向量 α, β 确定
- (2) 由一个点 P_0 和一个垂直于平面的(法)向量 n 确定
- (3) 由三个点 P_0, P_1, P_2 确定

平面的几种表示方式

参数方程

我们依照前面确定方式中的第一种,给出 $P_0 = (x_0, y_0, z_0)$ 和向量 $\alpha = (u_1, v_1, w_1)$ 和 $\beta = (u_2, v_2, w_2)$ 那么,我们可以得到点 $P(x, y, z)$ 的方程

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + s \begin{bmatrix} u_1 \\ v_1 \\ w_1 \end{bmatrix} + t \begin{bmatrix} u_2 \\ v_2 \\ w_2 \end{bmatrix}$$

一般方程

$$\det \begin{bmatrix} x - x_0 & u_1 & u_2 \\ y - y_0 & v_1 & v_2 \\ z - z_0 & w_1 & w_2 \end{bmatrix} = Ax + By + Cz + d = 0$$

法方程

在直角坐标系中 $[O; i, j, k]$,记 $n = (A, B, C)$.对于平面上任意点P,都有 $\overrightarrow{P_0P} \perp n$,由垂直条件给出点法式方程

$$\overrightarrow{P_0P} \perp \mathbf{n} \Leftrightarrow A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$$

我们继而把 \mathbf{n} 归一化,得到法方程

$$\frac{A(x - x_0) + B(y - y_0) + C(z - z_0)}{\sqrt{A^2 + B^2 + C^2}} = 0$$

三点方程

对于空间中三点 $P_0 = (x_0, y_0, z_0), P_1 = (x_1, y_1, z_1), P_2 = (x_2, y_2, z_2)$,我们可以得到三点式方程

$$\det \begin{bmatrix} x - x_0 & x_1 - x_0 & x_2 - x_0 \\ y - y_0 & y_1 - y_0 & y_2 - y_0 \\ z - z_0 & z_1 - z_0 & z_2 - z_0 \end{bmatrix} = 0$$

同理,我们可以改写为:

$$\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ x & x_0 & x_1 & x_2 \\ y & y_0 & y_1 & y_2 \\ z & z_0 & z_1 & z_2 \end{bmatrix} = 0$$

这里我们联想到4.2节对线段定比分点的拓展

定理6.1

平面方程是三元一次方程,而三元一次方程表示一个平面

截距式方程

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1$$

其中 a, b, c 称为平面在三个坐标轴上的截距

总结

平面的一般方程对应于线性方程组

参数方程与三点式方程对应线性方程组的解

其中参数方程是特解+齐次方程的通解

平面的位置关系

几何上,可能相交,重合,平行

几何	系数关系	解	$\text{rank}(A)$	$\text{rank}(A\mathbf{b})$	自由变量数
相交	$A_1 : B_1 : C_1 \neq A_2 : B_2 : C_2$	线	2	2	1
重合	$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} = \frac{D_1}{D_2}$	面	1	1	2
平行	$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} \neq \frac{D_1}{D_2}$	无	1	2	

没有解,谈论自由变量也没有意义

平面的位置关系——推广

之前我们仅仅讨论了两个平面的位置关系,事实上多个平面的位置关系我们也可仿照上面的方法进行讨论

可以利用矩阵

$$\begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_n & b_n & c_n & d_n \end{bmatrix}$$

只是说不定有些时候我们甚至要把矩阵拆成几块(选取一部分行)来考虑系数矩阵和增广矩阵的秩

7. 直线

直线的确定方法

- (1) 由两个点 A, B 确定
- (2) 由一个点 P_0 和一个方向向量 γ 确定
- (3) 由两个平面 π_1 与 π_2 确定

直线的集中表示方式

参数方程

过任意的两点 A, B 可以确定一条直线 l , l 上的点 P 与方向向量 $\gamma = \overrightarrow{AB}$ 共线, 因而有 $\overrightarrow{AP} = t\overrightarrow{AB}$

$$\overrightarrow{OP} = \overrightarrow{OA} + t\overrightarrow{AB}$$

这也称为直线 l 的方向向量

记 $A(x_0, y_0, z_0), \gamma(X, Y, Z)$ 也可以把方程重写为

$$\begin{cases} x = x_0 + tX \\ y = y_0 + tY \\ z = z_0 + tZ \end{cases}$$

两点式方程

对于 $A(x_0, y_0, z_0)B(x_1, y_1, z_1)$

$$\overrightarrow{OP} = (1-t)\overrightarrow{OA} + t\overrightarrow{OB}$$

进而可以改写为

$$\begin{cases} x = (1-t)x_0 + tx_1 \\ y = (1-t)y_0 + ty_1 \\ z = (1-t)z_0 + tz_1 \end{cases}$$

点向式方程(标准方程)(对称方程)

在上面的两点式方程中,我们可以对其中的两个方程两两消去参数 t ,于是就得到了点向式方程

$$\frac{x-x_0}{X} = \frac{y-y_0}{Y} = \frac{z-z_0}{Z}$$

其中 X, Y, Z 都不为0

假如 X, Y, Z 中含有0,那么方程为

$$x-x_0=0, \frac{y-y_0}{Y} = \frac{z-z_0}{Z}$$

事实上,这样的方程可以拆为两个平面方程的组合

$$\begin{cases} \frac{x-x_0}{X} = \frac{y-y_0}{Y} \\ \frac{x-x_0}{X} = \frac{z-z_0}{Z} \end{cases}$$

一般方程

直线可以看作两个平面的交线

$$\pi_1: A_1x + B_1y + C_1z + D_1 = 0, \pi_2: A_2x + B_2y + C_2z + D_2 = 0$$

直线 l 的方向向量垂直于两个平面的法向量(当然也平行于这两个法向量的公垂线方向),即

$$\gamma \parallel \mathbf{n}_1 \times \mathbf{n}_2$$

两点式方程的另一种形式

$$\frac{x-x_0}{x_1-x_0} = \frac{y-y_0}{y_1-y_0} = \frac{z-z_0}{z_1-z_0}$$

这个方程其实和标准方程的唯一区别就是没有使用标准的方向向量

平面束

$$L = \{\lambda(A_1x + B_1y + C_1z + D_1) + \mu(A_2x + B_2y + C_2z + D_2) = 0 \mid \lambda^2 + \mu^2 \neq 0\}$$

这个集合 L 由含轴 l 的所有平面组成

直线与平面的位置关系

直线与平面的位置关系可以转化为直线方向向量 $\gamma = (X, Y, Z)$ 与平面的法向量 $\mathbf{n} = (A, B, C)$ 之间的位置关系

对于给出的直线和平面:

$$l: \begin{cases} x = x_0 + tX \\ y = y_0 + tY \\ z = z_0 + tZ \end{cases} \quad \pi: Ax + By + Cz + D = 0$$

(1) 若 $AX + BY + CZ \neq 0$ 直线与平面交于一点

(2) 若 $AX + BY + CZ = 0$ 则需要区分点 (x_0, y_0, z_0) 是否在平面内

1' 若 $Ax_0 + By_0 + Cz_0 + D = 0$, 则对于任意的 t 都是解, 表明直线在平面内

2' 若 $Ax_0 + By_0 + Cz_0 + D \neq 0$, 则无解, 直线平行于平面, 而不在平面内

直线与直线的位置关系

对于直线 l_1 与 l_2

$$l_1: \begin{cases} x = x_1 + sX_1 \\ y = y_1 + sY_1 \\ z = z_1 + sZ_1 \end{cases} \quad l_2: \begin{cases} x = x_2 + tX_2 \\ y = y_2 + tY_2 \\ z = z_2 + tZ_2 \end{cases}$$

直线的异面与共面等价于向量 $\overrightarrow{P_1P_2}, \gamma_1, \gamma_2$ 的共面和异面
三个向量的坐标行列式

$$\Delta = \det \begin{bmatrix} x_2 - x_1 & X_1 & X_2 \\ y_2 - y_1 & Y_1 & Y_2 \\ z_2 - z_1 & Z_1 & Z_2 \end{bmatrix}$$

(1) l_1, l_2 异面当且仅当 $\overrightarrow{P_1P_2}, \gamma_1, \gamma_2$ 异面, 当且仅当 $\Delta \neq 0$

(2) l_1, l_2 共面当且仅当 $\overrightarrow{P_1P_2}, \gamma_1, \gamma_2$ 共面, 当且仅当 $\Delta = 0$
在这种情况下

1' l_1, l_2 相交当且仅当 $\overrightarrow{P_1P_2}, \gamma_1, \gamma_2$ 共面且 $\gamma_1 \nparallel \gamma_2$
2' l_1, l_2 平行当且仅当 $\overrightarrow{P_1P_2}, \gamma_1, \gamma_2$ 共面且 $\gamma_1 \parallel \gamma_2 \nparallel \overrightarrow{P_1P_2}$
3' l_1, l_2 重合当且仅当 $\overrightarrow{P_1P_2}, \gamma_1, \gamma_2$ 共面且 $\gamma_1 \parallel \gamma_2 \parallel \overrightarrow{P_1P_2}$

这里为什么不使用混合积而使用行列式? 本质上混合积的计算就基于行列式, 所以使用混合积只会增加复杂度

一道作业补充题中引发的思考

对于矩阵

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \\ z_1 & z_2 & \cdots & z_n \end{bmatrix}$$

当点共线时, 主元 = 2

当点共面时, 主元 ≤ 3

分析如下:

当点全部共线时, 任意两个点坐标相减可以得到这条直线的方向向量, 从而我们把后面3行减去第一列的 x_1, y_1, z_1 倍, 得到

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \\ 0 & y_2 - y_1 & \cdots & y_n - y_1 \\ 0 & z_2 - z_1 & \cdots & z_n - z_1 \end{bmatrix}$$

由于后面 $n - 1$ 列的后三个元素构成的向量都是直线的方向向量, 古其三个元素之间的倍数关系相同, 所以通过行倍加操作可以同时消去第3行,

第4行的所用元素，得到

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

从而我们成功证明了主元数为2

当点共面时，我们经过同样的操作可以得到

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \\ 0 & y_2 - y_1 & \cdots & y_n - y_1 \\ 0 & z_2 - z_1 & \cdots & z_n - z_1 \end{bmatrix}$$

此时后 $n-1$ 列的后三个元素构成的向量都是这个平面里的向量，于是它们共面。

显然共面的向量是可以共线的，如果共线，那么和前面一问一样，主元数=2

我们取右下角的子矩阵，由于这些列张成一个平面，因此这个子矩阵中有2个主元列，故也由2行有主元位置，我们可以通过行化简将这个矩阵化为简化阶梯型。

我们可以应用同样的行化简操作于原先的矩阵的后三行，从而在后面的3行也有2个主元位置，加上第一行一共有3个主元位置，从而就有3个主元

$$\begin{bmatrix} x_2 - x_1 & \cdots & x_n - x_1 \\ y_2 - y_1 & \cdots & y_n - y_1 \\ z_2 - z_1 & \cdots & z_n - z_1 \end{bmatrix}$$

8.距离与交角

点到直线的距离

记点为 $P(x, y, z)$ ，直线 l 方向向量为 $\gamma(X, Y, Z)$ 则可任取 l 上一点 $P_1(x_1, y_1, z_1)$

$$d = \sqrt{|\overrightarrow{PP_1}|^2 - \left(\frac{|\overrightarrow{PP_1} \cdot \gamma|}{|\gamma|}\right)^2}$$

点到平面的距离

记点为 $P(x, y, z)$, 平面 $\pi: Ax + By + Cz + D = 0$ 法向量为 $\mathbf{n}(A, B, C)$
则可任取 π 上一点 $P_1(x_1, y_1, z_1)$

$$d = \frac{|\overrightarrow{PP_1} \cdot \mathbf{n}|}{|\mathbf{n}|}$$

进一步可以优化为

$$d = \frac{|Ax + By + Cz + D|}{\sqrt{A^2 + B^2 + C^2}}$$

平面与平面间的距离

平面间的距离定义点对距离的极小值
对于平面

$$\pi_1: Ax + By + Cz + D_1 = 0, \pi_2: Ax + By + Cz + D_2 = 0$$

两平面之间的距离可以表示为

$$d = \frac{|D_1 - D_2|}{\sqrt{A^2 + B^2 + C^2}}$$

直线与平面间的距离

一个前提

直线的方向向量 γ 与平面的法向量 \mathbf{n} 垂直

计算方法

取直线上一点 $P_1(x_1, y_1, z_1)$ 和平面上一一点 $P_2(x_2, y_2, z_2)$

$$d = \frac{|\overrightarrow{P_1P_2} \cdot \mathbf{n}|}{|\mathbf{n}|}$$

直线与直线间的距离

共面直线之间的距离

转化为前面的点到直线的距离, 即可考虑 l_1 上一点 P_1 到另一条直线的距离

异面直线之间的距离

异面直线之间的距离定义为它们的点对之间距离的极小值

记这两条异面直线的方向向量分别为 γ_1, γ_2 ,

在这两条直线上分别取一点 $P_1(x_1, y_1, z_1), P_2(x_2, y_2, z_2)$

则异面直线之间的距离可以表示为

$$d = \frac{|(\gamma_1, \gamma_2, \overrightarrow{P_1P_2})|}{|\gamma_1 \times \gamma_2|}$$

平面与平面的夹角

对于平面

$$\pi_1: A_1x + B_1y + C_1z + D_1 = 0, \pi_2: A_2x + B_2y + C_2z + D_2 = 0$$

两平面的法向量分别为 $\mathbf{n}_1 = (A_1, B_1, C_1), \mathbf{n}_2 = (A_2, B_2, C_2)$, 则两平面的夹角 θ 满足

$$\cos \theta = \frac{|\mathbf{n}_1 \cdot \mathbf{n}_2|}{|\mathbf{n}_1| |\mathbf{n}_2|}$$

平面与直线的夹角

对于平面

$$\pi: Ax + By + Cz + D = 0$$

和直线

$$l: \frac{x - x_0}{X} = \frac{y - y_0}{Y} = \frac{z - z_0}{Z} = 0$$

平面的法向量与直线的方向向量分别为 $\mathbf{n} = (A, B, C), \gamma = (X, Y, Z)$, 则平面与直线的夹角 θ 满足

$$\sin \theta = \frac{|\mathbf{n} \cdot \gamma|}{|\mathbf{n}| |\gamma|}$$

直线与直线的夹角

两直线的方向向量分别为 $\gamma_1 = (X_1, Y_1, Z_1), \gamma_2 = (X_2, Y_2, Z_2)$, 则两直线的夹角 θ 满足

$$\cos \theta = \frac{|\gamma_1 \cdot \gamma_2|}{|\gamma_1| |\gamma_2|}$$

9.最小二乘法初步

向量代数知识补充

向量记号的简化

这部分事实上讲解了张量分析中的许多常用的记号，超出本书实际要求

在基矢量为 $[e_1, e_2, e_3]$ 的正交曲线坐标系中

标量积

我们知道以下标量积的结果

$$e_1 \cdot e_1 = 1, e_2 \cdot e_2 = 1, e_3 \cdot e_3 = 1$$

$$e_1 \cdot e_2 = 0, e_1 \cdot e_3 = 0, e_2 \cdot e_3 = 0$$

于是，我们可以引入记号 δ_{ij} ，这个记号称为Kronecker符号(delta)

$$e_i \cdot e_j = \delta_{ij}$$

显然有以下等式

$$\delta_{ij} = \begin{cases} 1, & \text{若 } i = j \\ 0, & \text{若 } i \neq j \end{cases}$$

向量积

同样，我们也知道下列向量积的结果

$$e_1 \times e_1 = 0, \quad e_1 \times e_2 = e_3, \quad e_1 \times e_3 = -e_2$$

$$e_2 \times e_1 = -e_3, \quad e_2 \times e_2 = 0, \quad e_2 \times e_3 = e_1$$

$$e_3 \times e_1 = e_2, \quad e_3 \times e_2 = -e_1, \quad e_3 \times e_3 = 0$$

我们同样可以简洁地写为

$$e_i \times e_j = \sum_{k=1}^3 \epsilon_{ijk} e_k$$

其中 ϵ_{ijk} 称为Levi-Civita三秩全反对称张量

Levi-Civita三秩全反对称张量的定义可以等价地写作

$$\epsilon_{ijk} = \begin{cases} 1, & \text{若}(ijk)\text{是}(123)\text{及其偶排列} \\ -1, & \text{若}(ijk)\text{是}(123)\text{的奇排列} \\ 0, & \text{其他情形(包括但不限于}i=j) \end{cases}$$

ϵ_{ijk} 可以表示为

$$\epsilon_{ijk} = \delta_{i1}\delta_{j2}\delta_{k3} + \delta_{i2}\delta_{j3}\delta_{k1} + \delta_{i3}\delta_{j1}\delta_{k2} - \delta_{i3}\delta_{j2}\delta_{k1} - \delta_{i2}\delta_{j1}\delta_{k3} - \delta_{i1}\delta_{j3}\delta_{k2}$$

【怎么记？每个 δ 第一个角标都是以 i, j, k 按照顺序，第二个角标为123的所有排列，其中逆序数为偶数的整项为正，反之为负】

ϵ_{ijk} 也可以表示为如下的行列式

$$\det \begin{bmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{bmatrix}$$

代数展开

假设两个矢量 \vec{A}, \vec{B} 在本坐标系中的分量表达式分别为

$$\vec{A} = A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3, \vec{B} = B_1\mathbf{e}_1 + B_2\mathbf{e}_2 + B_3\mathbf{e}_3$$

那么，这两个矢量的运算可以如下表示：

$$\vec{A} + \vec{B} = \sum_{i=1}^3 (A_i + B_i)\mathbf{e}_i$$

$$\vec{A} \cdot \vec{B} = \sum_{i=1}^3 A_i B_i$$

$$\vec{A} \times \vec{B} = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{bmatrix}$$

$$\begin{aligned}
&= \mathbf{e}_1(A_2B_3 - B_2A_3) + \mathbf{e}_2(A_3B_1 - B_3A_1) + \mathbf{e}_3(A_1B_2 - B_1A_2) \\
&= \mathbf{e}_1(\epsilon_{123}A_2B_3 + \epsilon_{132}A_3B_2) + \mathbf{e}_2(\epsilon_{231}A_3B_1 + \epsilon_{213}A_1B_3) + \mathbf{e}_3(\epsilon_{312}A_1B_2 + \epsilon_{321}A_2B_1) \\
&= \mathbf{e}_1 \sum_{j=1, k=1}^3 \epsilon_{1jk} A_j B_k + \mathbf{e}_2 \sum_{j=1, k=1}^3 \epsilon_{2jk} A_j B_k + \mathbf{e}_3 \sum_{j=1, k=1}^3 \epsilon_{3jk} A_j B_k \\
&= \sum_{i=1, j=1, k=1}^3 \mathbf{e}_i \epsilon_{ijk} A_j B_k
\end{aligned}$$

尽管我们费了九牛二虎之力才化简出这样的等式，显然这样的表达式还是有点太复杂了，我们应当思考如何再次简化我们的表达式

矢量 \vec{A} 在对应直角坐标系中可以表示为 $\vec{A} = A_1\mathbf{e}_1 + A_2\mathbf{e}_2 + A_3\mathbf{e}_3$ ，这实在有点不方便，我们可以简化为

$$\vec{A} = \sum_{i=1}^3 A_i \mathbf{e}_i$$

这样还是不方便，我们考虑继续简化，略去繁琐的求和号，采用哑指标的方式

$$\vec{A} = A_i \mathbf{e}_i$$

【这是我们在电动力学中常用的一种表达方式，不过要注意以下几点：】

- (1)重复的指标代表求和
- (2)在方程的任意一项中，表示求和的重复指标只可以出现2次

为什么这样处理之后就可以极大提高效率呢？原因在于以下几个重要性质：

由于

$$\delta_{ij}\delta_{mj} = \delta_{i1}\delta_{m1} + \delta_{i2}\delta_{m2} + \delta_{i3}\delta_{m3} = \delta_{im}$$

即有

$$(1)\delta_{ij}\delta_{mj} = \delta_{im}$$

思考为什么会等于 δ_{im} ，因为如果 $i=m$ 那么三个乘积项中必定有一个为1，总和为1，恰好与 δ_{im} 一致；如果 $i \neq m$ ，那么三个乘积项肯定都为零，【(1,2,3)中的任何一个数都不可能同时等于两个不相等的数】，总和也为0这与此时

的 δ_{im} 是一致的!

同时有

$$\delta_{ii} = \delta_{11} + \delta_{22} + \delta_{33} = 3$$

即有

$$(2)\delta_{ii} = 3$$

更为重要的是下面的等式:

$$(3)\epsilon_{ijk}\epsilon_{mnk} = \delta_{im}\delta_{jn} - \delta_{in}\delta_{jm}$$

我们可以记作【两交叉减两内两外 (与高中的“两外两内减交叉”恰好相反)】

下面我们给出这个定理的证明:

首先根据之前对Levi-Civita三秩全反对称张量的定义

$$\det \begin{bmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{bmatrix}$$

由于转置不改变行列式的值

$$= \det \begin{bmatrix} \delta_{i1} & \delta_{j1} & \delta_{k1} \\ \delta_{i2} & \delta_{j2} & \delta_{k2} \\ \delta_{i3} & \delta_{j3} & \delta_{k3} \end{bmatrix}$$

因此我们有 (可以注意一下这并不等于(3)式的左式, 在这里我们把 ϵ_{mnk} 换成了 ϵ_{mnl})

$$\epsilon_{ijk}\epsilon_{mnl} = \det \begin{bmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{bmatrix} \det \begin{bmatrix} \delta_{m1} & \delta_{n1} & \delta_{k1} \\ \delta_{m2} & \delta_{n2} & \delta_{k2} \\ \delta_{m3} & \delta_{n3} & \delta_{k3} \end{bmatrix}$$

根据行列式的运算法则以及我们前面的(1)式

$$\det \begin{bmatrix} \delta_{im} & \delta_{in} & \delta_{il} \\ \delta_{jm} & \delta_{jn} & \delta_{jl} \\ \delta_{km} & \delta_{kn} & \delta_{kl} \end{bmatrix}$$

在这里，我们令 $k = l$ 于是

$$\epsilon_{ijk}\epsilon_{mnk} = \det \begin{bmatrix} \delta_{im} & \delta_{in} & \delta_{ik} \\ \delta_{jm} & \delta_{jn} & \delta_{jk} \\ \delta_{km} & \delta_{kn} & \delta_{kk} \end{bmatrix} = \det \begin{bmatrix} \delta_{im} & \delta_{in} & \delta_{ik} \\ \delta_{jm} & \delta_{jn} & \delta_{jk} \\ \delta_{km} & \delta_{kn} & 3 \end{bmatrix}$$

$$= 3\delta_{im}\delta_{jn} + \delta_{in}\delta_{jk}\delta_{km} + \delta_{ik}\delta_{jm}\delta_{kn} - \delta_{ik}\delta_{jn}\delta_{km} - 3\delta_{in}\delta_{jm} - \delta_{im}\delta_{jk}\delta_{kn}$$

我们在化简过程中谨记重复指标代表求和的规则！

$$\begin{aligned} &= 3\delta_{im}\delta_{jn} + \delta_{in}\delta_{jm} + \delta_{in}\delta_{jm} - \delta_{im}\delta_{jn} - 3\delta_{in}\delta_{jm} - \delta_{im}\delta_{jn} \\ &= \delta_{im}\delta_{jn} - \delta_{in}\delta_{jm} \end{aligned}$$

得证！真神奇！（）

这个等式有以下两个推论：

$$\textcircled{1} \epsilon_{ijk}\epsilon_{mjk} = \delta_{im}\delta_{jj} - \delta_{ij}\delta_{jm} = 3\delta_{im} - \delta_{im} = 2\delta_{im}$$

$$\textcircled{2} \epsilon_{ijk}\epsilon_{ijk} = \delta_{ii}\delta_{jj} - \delta_{ij}\delta_{ji} = 9 - 3 = 6$$

事实上，这两个推论不用专门去记，记住等式(3)运用记忆规则应该就好了

实际应用

运用新引入的简化表达式，矢量运算的基本计算公式可以归纳为

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$$

$$\mathbf{e}_i \times \mathbf{e}_j = \epsilon_{ijk}\mathbf{e}_k$$

在前面简单的推导后，我们成功地简化了以下向量运算的表达形式：

$$\begin{aligned}
\vec{A} + \vec{B} &= (A_i + B_i)\mathbf{e}_i \\
\vec{A} \cdot \vec{B} &= A_i B_i \\
\vec{A} \times \vec{B} &= A_i \mathbf{e}_i \times B_j \mathbf{e}_j \\
&= A_i B_j (\mathbf{e}_i \times \mathbf{e}_j) \\
&= \epsilon_{ijk} A_i B_j \mathbf{e}_k
\end{aligned}$$

是不是非常简洁呢

下面我们用这些符号来推导向量运算中的常见等式

$$(1) \vec{A} \cdot (\vec{B} \times \vec{C}) = \epsilon_{ijk} A_i B_j C_k = \vec{B} \cdot (\vec{C} \times \vec{A}) = \vec{C} \cdot (\vec{A} \times \vec{B})$$

$$\begin{aligned}
(2) \vec{A} \times (\vec{B} \times \vec{C}) &= \mathbf{e}_i \epsilon_{ijk} A_j (\vec{B} \times \vec{C})_k = \mathbf{e}_i \epsilon_{ijk} A_j (\epsilon_{mnk} B_m C_n) \\
&= \mathbf{e}_i \epsilon_{ijk} \epsilon_{mnk} A_j B_m C_n = \mathbf{e}_i (\delta_{im} \delta_{jn} - \delta_{in} \delta_{jm}) A_j B_m C_n \\
&= \mathbf{e}_i (\delta_{im} \delta_{jn} A_j B_m C_n - \delta_{in} \delta_{jm} A_j B_m C_n) \\
&= \mathbf{e}_i (\delta_{im} B_m (\vec{A} \cdot \vec{C}) - \delta_{in} C_n (\vec{A} \cdot \vec{B})) \\
&= (\vec{A} \cdot \vec{C}) \cdot \vec{B} - (\vec{A} \cdot \vec{B}) \cdot \vec{C}
\end{aligned}$$

终于证完了(事实上，中间有很多详细步骤是可以不写的)

在最后的最后，我们最后证明一个相对复杂一点的式子

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{d}) - (\mathbf{b} \cdot \mathbf{c})$$

证：

$$\begin{aligned}
(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) &= (\mathbf{e}_k \epsilon_{ijk} a_i b_j) \cdot (\mathbf{e}_l \epsilon_{mnl} c_m d_n) \\
&= (\mathbf{e}_k \epsilon_{ijk} a_i b_j) \cdot (\mathbf{e}_k \epsilon_{mnk} c_m d_n) \\
&= (\delta_{im} \delta_{jn} - \delta_{in} \delta_{jm}) a_i b_j c_m d_n \\
&= a_i c_i b_j d_j - a_i d_i b_j c_j = (\mathbf{a} \cdot \mathbf{d}) - (\mathbf{b} \cdot \mathbf{c})
\end{aligned}$$

4 向量空间

4.1 向量空间与子空间

向量空间的定义略，主要包含
加法封闭性

加法交换律

加法结合律

存在零向量

存在负元

标量乘法封闭性

标量乘法分配律（对标量、向量）

标量乘法结合律

存在单位元

4.1.1 子空间的定义

向量空间 V 的一个子空间 H 是 V 的满足以下三个性质的子集

- a. V 中的零向量在 H 中
- b. H 对标量加法封闭，即对 H 中任意向量 u, v ，向量 $u + v$ 仍在 H 中
- c. H 对标量乘法封闭，即对 H 中任意向量 u 和任意标量 c ，向量 cu 仍在 H 中

注意:一个平面未必是一个二维子空间!因为这个平面必须过平面!

4.1.2 定理1

若 v_1, \dots, v_p 在向量空间 V 中，则 $\text{Span}\{v_1, \dots, v_p\}$ 是 V 的一个子空间

我们称 $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ 是由 $\mathbf{v}_1, \dots, \mathbf{v}_p$ 张成的子空间，给定 V 的任意一个子空间 H ， H 的生成集是集合 $\{\mathbf{v}_1, \dots, \mathbf{v}_p\} \in H$ ，满足 $H = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$

从中我们可以知道的每个子空间都是一个向量空间

4.2 零空间、列空间和线性变换

4.2.1 矩阵的列空间和零空间

定义

矩阵 A 的零空间是齐次方程 $A\mathbf{x} = \mathbf{0}$ 的所有解的集合，记为 $\text{Nul } A$ ，用集合表示为

$$\text{Nul } A = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{0}\}$$

4.2.2 定理2

$m \times n$ 矩阵 A 的零空间是 \mathbb{R}^n 的子空间。等价地， n 个未知数的 m 个齐次线性方程组 $A\mathbf{x} = \mathbf{0}$ 的所有解的集合是 \mathbb{R}^n 的子空间

定义

矩阵 A 的列空间是 A 各列的线性组合的集合，记作 $\text{Col } A$ ，若 $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ ，则 $\text{Col } A = \text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$

4.2.3 定理3

$m \times n$ 的矩阵 A 的列向量是 \mathbb{R}^m 的一个子空间

4.2.4 线性变换的定义

由向量空间 V 映射到向量空间 W 内的线性变换 T 是一个规则，它将 V 中的每个向量 \mathbf{v} 映射为 W 中的唯一向量 $T(\mathbf{x})$ ，且满足

(i) 对 V 的定义域中一切 \mathbf{u}, \mathbf{v} , $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

(ii) 对 V 的定义域中一切 \mathbf{u} 和数 c , $T(c\mathbf{u}) = cT(\mathbf{u})$

如果线性映射是由某个矩阵得到的, 那么线性映射 T 的核和值域分别为对应矩阵 A 的零空间和列空间

例 4.2.1 设 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是一一对一的映射, 求证 T 的值域的维数为 n

证明 由于 T 是一一对一映射, 故 A 的列线性无关, 从而 $\text{Nul}A = \{0\}$, 因而也有 $\text{Col}A = n - 0 = n$, 而上面我们提到了线性映射 T 的值域对应 A 的列空间, 故值域为 n 维的

4.3 线性无关集与基

4.3.1 定理4

两个或更多个向量的集合 $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ 线性相关, 当且仅当 S 中至少有一个向量使其他向量的线性组合。事实上, 若 S 线性相关, 且 $\mathbf{v}_1 \neq \mathbf{0}$ 【否则后几个向量可以线性无关, 第一个向量为零向量, 那么后面任意一个向量都不能由前面的向量线性表出!!】 \mathbf{v}_j 是它前面向量 $\mathbf{v}_1, \dots, \mathbf{v}_{j-1}$ 的线性组合

4.3.2 基的定义

令 H 是向量空间 V 的一个子空间, V 中的一组向量 $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ 称为 H 的一个基, 如果

(i) B 是一个线性无关集

(ii) 由 B 生成的子空间与 H 相同, 即 $H = \text{Span}\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$

4.3.3 定理5 (生成集定理)

若 $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ 是 V 中的向量集, $H = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$

a. 若 S 中某一个向量 \mathbf{v}_k 是其余向量的线性组合, 则 S 中去掉 \mathbf{v}_k 后形成的集合仍然可以生成 H

b.若 $H \neq \{\mathbf{0}\}$ ，则 S 的某一子集是 H 的一个基

4.3.4 定理6

矩阵 A 的主元列构成 $\text{Col}A$ 的一个基

4.4 坐标系

4.4.1 定理7（唯一表示定理）

令 $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ 为向量空间 V 的一个基，则对于 V 中每一个向量 \mathbf{x} ，均存在唯一的实数对 c_1, \dots, c_n 使得对

$$\mathbf{x} = c_1 \mathbf{b}_1 + \dots + c_n \mathbf{b}_n$$

4.4.2 定义

设 $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ 为向量空间 V 的一个基， \mathbf{x} 对于 V 中， \mathbf{x} 相对于基 \mathcal{B} 的坐标是使得 $\mathbf{x} = c_1 \mathbf{b}_1 + \dots + c_n \mathbf{b}_n$ 的权 c_1, \dots, c_n 。若 c_1, \dots, c_n 是 \mathbf{x} 的 \mathcal{B} -坐标，则 \mathbb{R}^n 中的向量

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

是 \mathbf{x} 相对于基 \mathcal{B} 的坐标向量或者 \mathbf{x} 的 \mathcal{B} -坐标向量，映射 $\mathbf{x} \rightarrow [\mathbf{x}]_{\mathcal{B}}$ 称为由 \mathcal{B} 确定的坐标映射

4.4.3 坐标变换矩阵[顾名思义这个矩阵作用于坐标向量，得到一个新的坐标向量]

$$\mathbf{x} = P_{\mathcal{B}} [\mathbf{x}]_{\mathcal{B}}$$

中的 $P_{\mathcal{B}}$ 称为从 \mathcal{B} 到 \mathbb{R}^n 中的标准基 \mathcal{E} 中的坐标变换矩阵

4.4.4 定理8

令 $\mathcal{B} = \{b_1, \dots, b_n\}$ 为向量空间 V 的一个基，则坐标映射映射 $x \rightarrow [x]_{\mathcal{B}}$ 是一个由 \mathcal{B} 映上到 \mathbb{R}^n 的一对一的线性变换

一般来说这样一个由向量空间 V 映上到 W 的一对一的线性变换称为从 V 到 W 的一个同构比如 \mathbb{R}^3 空间是 \mathbb{P}_2 空间的同构

4.5 向量空间的维数

4.5.1 定理9

若向量空间 V 具有一组基 $\mathcal{B} = \{b_1, \dots, b_n\}$ ，则 V 中任意包含多余 n 个向量的集合一定线性相关

4.5.2 定理10

若向量空间 V 的一组基含有 n 个向量，则其每一组基都含有 n 个向量

4.5.3 有限维与无限维的定义

若 V 由一个有限集生成，则我们称其为**有限维的**， V 的维数写作 $\dim V$ ，是 V 的基中向量的数目。

零向量空间 $\{0\}$ 的维数定义为零。

若 V 不能由有限集生成，则定义其维数为无限维

4.5.4 定理11

令 H 为有限维向量空间 V 的子空间，若有必要的话， H 中任何一个线性无关集都可以扩充成为 V 的一个基， H 也是有限维的并且

$$\dim H \leq \dim V$$

注意这里的子空间可以是其本身，从而我们知道一个子空间中任何一个线性无关集都可扩充称为这个子空间的一个基，这恰好是生成集定理的反面

4.5.5 定理12（基定理）

令 V 是一个 p 维向量空间，则 V 中任何一个 p 个向量生成 V 的集合为 V 的一组基，任何一个 p 个向量的线性无关集也是 V 的一组基

4.5.6 Nul A 和 Col A 的维数

Nul A 的维数是方程 $A\mathbf{x} = \mathbf{0}$ 的自由变量的数目，Col A 的维数为矩阵 A 中主元列的个数

4.6 秩

4.6.1 定理13

若两个矩阵 A, B 行等价，则它们的行空间相同。若 B 是简化阶梯形矩阵，则 B 的非零行构成 A 的行空间的一组基的同时也是 B 的行空间的一组基

这里我们要注意以下变换的特点

行变换	列变换
不改变行空间	改变行空间
改变列空间	不改变列空间
改变行相关关系	不改变行相关关系
不改变列相关关系	改变列相关关系

4.6.2 定义

矩阵 A 的秩定义为其列空间的维数

4.6.3 定理14 (秩定理)

如果一个 $m \times n$ 矩阵 A 的列空间和行空间的维数相同, 这个公共的维数(A 的秩) 还等于 A 的主元位置数, 且满足方程

$$\text{rank} A + \dim \text{Nul} A = n$$

4.6.4 秩和可逆矩阵定理

设 A 是一 $n \times n$ 矩阵, 则下面的每个命题与 A 是可逆矩阵的命题等价:

m. A 的列向量构成 \mathbb{R}^n 的一个基

n. $\text{Col } A = \mathbb{R}^n$

o. $\dim \text{Col } A = n$

p. $\text{rank} A = n$

q. $\text{Nul } A = \{\mathbf{0}\}$

r. $\dim \text{Nul } A = 0$

4.6.5 矩阵的满秩分解

设 $m \times n$ 的矩阵 A 的主元数为 $r > 0$, 证明存在 $m \times r$ 矩阵 B 和 $r \times n$ 矩阵 C 使得 $A = BC$, 其 $\text{rank} B = \text{rank} C = r$

证明 首先假设 A 的前 r 列是线性无关的. 因此对矩阵 A 做初等行变换即可得到 $\begin{bmatrix} I_r & D \\ 0 & 0 \end{bmatrix}$, 这说明存在可逆矩阵 $P_{m \times m}$ 使得

$$PA = \begin{bmatrix} I_r & D \\ 0 & 0 \end{bmatrix}$$

进而

$$A = P^{-1} \begin{bmatrix} I_r & D \\ 0 & 0 \end{bmatrix} = P^{-1} \begin{bmatrix} I_r \\ 0 \end{bmatrix} \begin{bmatrix} I_r & D \end{bmatrix} = BC$$

其中 $B = P^{-1} \begin{bmatrix} I_r \\ 0 \end{bmatrix}$, $C = \begin{bmatrix} I_r & D \end{bmatrix}$

现在假设 A 的前 r 列是线性相关的, 只需要将 A 做一个列变换即可使得前 r 列线性无关, 用刚才的方法证明

$$PAQ = \begin{bmatrix} I_r & D \\ 0 & 0 \end{bmatrix}$$

说明存在可逆矩阵 $P_{m \times m}$ 和 $Q_{n \times n}$ 使得

$$\begin{aligned} A &= P^{-1} \begin{bmatrix} I_r & D \\ 0 & 0 \end{bmatrix} Q^{-1} \\ &= P^{-1} \begin{bmatrix} I_r \\ 0 \end{bmatrix} \begin{bmatrix} I_r & D \end{bmatrix} Q^{-1} \\ &= BC \end{aligned}$$

其中 $B = P^{-1} \begin{bmatrix} I_r \\ 0 \end{bmatrix}$, $C = \begin{bmatrix} I_r & D \end{bmatrix} Q^{-1}$ 不难得知

$$\begin{aligned} \text{rank} B &= \text{rank} \left(P^{-1} \begin{bmatrix} I_r \\ 0 \end{bmatrix} \right) = \text{rank} \left(\begin{bmatrix} I_r \\ 0 \end{bmatrix} \right) = r \\ \text{rank} C &= \text{rank} \left(\begin{bmatrix} I_r & D \end{bmatrix} Q^{-1} \right) = \text{rank} \left(\begin{bmatrix} I_r & D \end{bmatrix} \right) = r \end{aligned}$$

例 4.6.1 利用矩阵的满秩分解 $A = CR$ 证明

$$\text{rank}(A + B) \leq \text{rank} A + \text{rank} B$$

证明 不妨假设 $\text{rank} A = r_1$, $\text{rank} B = r_2$, 那么分别存在 $m \times r_1$, $r_1 \times n$, $m \times r_2$, $r_2 \times n$ 矩阵 C_1, R_1, C_2, R_2 使得 $A = C_1 R_1$, $B = C_2 R_2$, 利用分块矩阵的分解,

$$A + B = C_1 R_1 + C_2 R_2 = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = CR$$

其中 C 为 $m \times (r_1 + r_2)$ 矩阵, R 为 $(r_1 + r_2) \times n$ 矩阵, 它们的秩都小于等于 $r_1 + r_2$, 故

$$\text{rank}(A + B) \leq \text{rank} C \leq \text{rank} A + \text{rank} B$$

$$\text{rank}(A + B) \leq \text{rank} R \leq \text{rank} A + \text{rank} B$$

4.7 基的变换

设 $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ 为向量空间 V 的基, 则存在一个 $n \times n$ 矩阵 $P_{\mathcal{C} \leftarrow \mathcal{B}}$ 使得

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C} \leftarrow \mathcal{B}} [\mathbf{x}]_{\mathcal{B}}$$

其中 $P_{C \leftarrow B}$ 的列为基 B 中向量的 C -坐标, 即

$$P_{C \leftarrow B} = \begin{bmatrix} [b_1]_C & [b_2]_C & \cdots & [b_n]_C \end{bmatrix}$$

这个定理中间的部分可以类比我们之前从标准基到基 B 的过程, 只是起点和终点的坐标表示发生了变化, 起点不再是标准基而是 B , 终点不再是 B 而是 C

4.7.1 \mathbb{R}^n 中基的变换

事实上 $P_{C \leftarrow B}$ 可以用 $P_C^{-1}P_B$ 来计算, 我们有一种理解方式, 即我们可以将基 B 中的坐标先转化为 \mathcal{E} 中的坐标, 再转化为 C 中的坐标。

效果上, 有

$$[x]_C = P_C^{-1}x = P_{C \leftarrow B}[x]_B$$

另一种更快的计算方法是把 B, C 中的基向量都放在一个矩阵中, 然后将基 C 对应的部分化为单位矩阵, 那么基 B 行变换后的部分就变为了坐标变换矩阵的列

这一点的理解方式与上面类似, 行化简的过程本质上是实现了左乘一个将 \mathcal{E} 变为 C 的矩阵 P_C^{-1} , 从而我们有效地得到了乘积矩阵

然而, 如果我们并不像变换坐标, 而是想要得知两组基之间如何相互转换, 那么我们乘法的位置是不一样的

$$CP_{C \leftarrow B} = B$$

4.8 差分方程中的应用

4.9 马尔可夫链中的应用

一些习题

引理 4.9.1 假设 A 是一个 $m \times n$ P 是一个 $n \times n$ 可逆矩阵, Q 是一个 $n \times n$ 可逆矩阵, 则有

$$\text{rank} A = \text{rank} PA = \text{rank} AQ$$

证明 这个证明非常显然,由于矩阵 P, Q 的可逆性,可以把他们想象成为初等行变换和初等列变换,它们都不改变矩阵的秩

已知 A 是 $m \times n$ 矩阵, B 为 $n \times p$ 矩阵,求证

$$\min(\text{rank}A, \text{rank}B) \geq \text{rank}AB \geq \text{rank}A + \text{rank}B - n$$

等式的左侧是显然的,我们考虑证明右侧的不等式,主要有两种思路

证明 思路(1)转化为零空间的维数关系

$$\text{rank}AB = p - \dim \text{Nul}AB, \text{rank}A = n - \dim \text{Nul}A, \text{rank}B = p - \dim \text{Nul}B$$

转化为证明

$$\dim \text{Nul}A + \dim \text{Nul}B \geq \dim \text{Nul}AB$$

然而这样的结论直接去想象还是非常困难,我们考虑利用基的性质方便证明
我们设 B 零空间的一组基为 $\{\alpha_1, \dots, \alpha_k\}$,显然并非 A 零空间中的所有向量都可以由这个基中的向量表示出来.

由生成集定理,我们可以不断添加向量直到这个集合可以生成 AB 的零空间,我们假设添加若干向量后得到的基为 $\{\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_t\}$,那么很显然 $B\beta_i$ 属于 A 的零空间,从而 $t \leq \dim \text{Nul}A$,那么事实上我们就不那么严谨的证明了不等式

思路(2)利用可逆矩阵乘积的性质 我们假设 A 的秩为 r_1, B 的秩为 r_2, AB 的秩为 r

首先,通过初等行变换和列变换,我们可以将 A 矩阵化为如下形式

$$PAQ = \begin{bmatrix} I_{r_1} & 0 \\ 0 & 0 \end{bmatrix}$$

然后,我可以将 Q^{-1} 这个变换作用于 B 得到 $Q^{-1}B = \begin{bmatrix} M_{r_1 \times p} \\ M_{(n-r_1) \times p} \end{bmatrix}$

$$(PAQ)Q^{-1}B = \begin{bmatrix} I_{r_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} M_{r_1 \times p} \\ M_{(n-r_1) \times p} \end{bmatrix} = \begin{bmatrix} M_{r_1 \times p} & 0 \\ 0 & 0 \end{bmatrix}$$

从而有 $r = \text{rank}(M_{r_1 \times p})$, 另一方面, 由于我们 $r_2 = \text{rank}\left(\begin{bmatrix} M_{r_1 \times p} \\ M_{(n-r_1) \times p} \end{bmatrix}\right)$ 而后面的 $M_{(n-r_1) \times p}$ 至多有 $n - r_1$ 个向量线性无关, 因此

$$r_2 \leq \text{rank}(M_{r_1 \times p}) + n - r_1$$

综合上面的两个等式, 我们可以得到要证明的式子

5 特征值与特征向量

5.1 特征向量与特征值

5.1.1 定义

A 为 $n \times n$ 矩阵， x 为**非零向量**，若存在数 λ 使得 $Ax = \lambda x$ 有**非平凡解** x ，则称 λ 为 A 的特征值， x 为 A 的特征向量

特征向量是非零的，而特征值是可以为0的

警告：对矩阵 A 进行行化简一般会改变矩阵的特征值

5.1.2 特征空间的定义

由前面的定义我们知道，对应特征值 λ 特征向量所在的集合即为

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

的解的集合，即为矩阵 $A - \lambda I$ 的零空间，因而其是 \mathbb{R}^n 的子空间，称为 A 的对应于 λ 的**特征空间**

5.1.3 定理1

三角矩阵主对角线上的元素是其特征值

2×2 矩阵的特征值计算公式

对于矩阵 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ，我们记 $m = \frac{a+d}{2}$, $p = ad - bc$ ，则矩阵的特征值 $\lambda = m \pm \sqrt{m^2 - p}$

其中的 $a + d$ 称为**矩阵的迹**，对一般矩阵，对角线元素的和称为矩阵的迹

5.1.4 定理2

若 $\lambda_1, \dots, \lambda_p$ 为矩阵 $n \times n$ 矩阵 A 相异的特征值, v_1, \dots, v_p 是与 $\lambda_1, \dots, \lambda_p$ 对应的特征向量, 则向量集合 v_1, \dots, v_p 线性无关

5.2 特征方程

5.2.1 可逆矩阵定理的补充

矩阵 A 是可逆矩阵, 当且仅当

s.0不是 A 的特征值

t. A 的行列式不为0

5.2.2 定理3 (行列式的性质)

这些定理内容全部来自第三章, 在此不再赘述

数值方程 $\det(A - \lambda I) = 0$ 称为 A 的特征方程

数 λ 是 $n \times n$ 矩阵 A 的充分必要条件是 λ 是特征方程 $\det(A - \lambda I) = 0$ 的根

5.2.3 特征多项式的定义

对于 $n \times n$ 矩阵 A , $\det(A - \lambda I)$ 是 n 次多项式, 称为矩阵 A 的特征多项式

5.2.4 (代数)重数与(几何)重数的定义

把特征值 λ 作特征方程的根的重数称为特征值 λ 的代数重数

特征值 λ 对应的特征子空间的维数称为 λ 的几何重数, 我们有以下关系

$$1 \leq \text{几何重数} \leq \text{代数重数}$$

5.2.5 相似性的定义

对于 $n \times n$ 矩阵 A , 若存在可逆矩阵 P 与矩阵 B 使得

$$A = PBP^{-1}$$

则我们称 A 相似于 B , 因为我们很容易依据 P 的可逆性得到 $B = P^{-1}AP$, 而我们知道可逆矩阵 P 的逆 P^{-1} 也是一个可逆矩阵, 从而 B 也相似于 A , 所以我们说 A 与 B 是相似的, 将 A 变成 PBP^{-1} 的变换为相似变化

警告: 许多矩阵有相同的特征值也不相似, 比如矩阵 $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$, $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

5.2.6 定理4

若 $n \times n$ 矩阵 A 和 B 是相似的, 那么他们有相同的特征多项式, 也因此有相同的特征值 (和相同的重数)

注意: 尽管两个相似矩阵有相同的特征值, 但它们未必有相同的特征向量, 比如对于 $A = PDP^{-1}$, 其中 D 的特征向量可以为标准基, 但 A 的特征向量未必是标准基

例 5.2.1 如果 A 和 B 均为 $n \times n$ 可逆矩阵, 则 AB 详细与 BA

证明 这是正确的, 因为 $BA = B(AB)B^{-1}$, 且其中 B 是可逆矩阵

5.3 对角化

5.3.1 定理5

$n \times n$ 矩阵 A 可对角化的充分必要条件是 A 有 n 个线性无关的特征向量
事实上, $A = PDP^{-1}$, D 为对角矩阵的充分必要条件是 P 的列向量为 A 的 n 个线性无关的特征向量, 此时 D 的主对角线上的元素分别是 A 的对应于 P 中特征向量的特征值 (即要有一一对应的关系)

换句话说 A 可对角化的充分必要条件是足够的特征向量形成 \mathbb{R}^n 的基, 我们称这样的基为**特征向量基**

5.3.2 定理6

矩阵可对角化的充分条件是其有 n 个相异的特征值

注意两点：

- ①这是充分条件，不是必要的
- ②特征值必须相异才能是充分的，不能考虑重数

[矩阵 D 和矩阵 A 未必是可逆的,因为如果 A 的特征值包含0的话,那么 D 对角线上含有0元, D 不可逆,同理 A 的行列式为0,也是不可逆的]

注意:矩阵的可逆和可对角化既不充分也不必要!

5.3.3 定理7

设 A 是 $n \times n$ 矩阵，其相异的特征值为 $\lambda_1, \dots, \lambda_p$

a.对于 $1 \leq k \leq p$ ， λ_k 的特征空间的维数小于或等于特征根 λ_k 的代数重数（特征根的重数）

b.矩阵 A 可对角化的充分必要条件是矩阵所有不同特征空间的维数之和为 n ，即

- (i)特征多项式可以完全分解为线性因子（换言之，不能分解不出实根而是得到 $\lambda^2 + 1 = 0$ 这种式子）
- (ii)每个 λ_k 的特征空间维数等于其代数重数

c.若 A 可对角化， \mathcal{B}_k 是对应于 λ_k 的特征空间的基，则集合 $\mathcal{B}_1, \dots, \mathcal{B}_p$ 中所有向量的集合为 \mathbb{R}^n 的特征向量集

5.4 特征向量与线性变换

首先理解线性变换本身是唯一的,只不过我们可以选取不同的基来描述这个线性变换,因此可以得到不同的线性变换的矩阵,但他们事实上表示的线性变换是相同的

比如,我们假定 D_B, D_C 分别是线性变换 T 在基 B, C 下的矩阵,则

$$(P_{C \leftarrow B})^{-1} D_B P_{C \leftarrow B} = D_C$$

这事实上反映了线性变换实质的一致性

5.4.1 线性变换的矩阵

设 V 是 n 维向量空间, W 是 m 维向量空间, T 是 V 到 W 的线性变换, 为了把 T 与矩阵关联起来, 我们指定 B 和 C 分别为 V 和 W 的基

如果 \mathbf{x} 是 V 中的向量, 那么坐标向量 $[\mathbf{x}]_B \in \mathbb{R}^n$, 线性映射的结果的坐标向量 $[T(\mathbf{x})]_C \in \mathbb{R}^m$

假设 V 的基为 $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, 若 $\mathbf{x} = r_1 \mathbf{b}_1 + \dots + r_n \mathbf{b}_n$, 则有

$$[\mathbf{x}]_B = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

由于 T 是一个线性的变换, 从而有 $T(\mathbf{x}) = T(r_1 \mathbf{b}_1 + \dots + r_n \mathbf{b}_n) = r_1 T(\mathbf{b}_1) + \dots + r_n T(\mathbf{b}_n)$

又由于坐标映射也是线性的, 所以

$$[T(\mathbf{x})]_C = r_1 [T(\mathbf{b}_1)]_C + \dots + r_n [T(\mathbf{b}_n)]_C$$

还记得我们之前 \mathbf{x} 对应的 \mathbb{R}^n 中的坐标向量吗? 不难发现上式可以写成矩阵的形式

$$[T(\mathbf{x})]_C = M [\mathbf{x}]_B$$

其中的

$$M = \begin{bmatrix} [T(\mathbf{b}_1)]_C & \dots & [T(\mathbf{b}_n)]_C \end{bmatrix}$$

矩阵 M 是线性映射 T 的矩阵表示, 我们称 M 为 T 相对于基 B 和 C 的矩阵

特殊的情况: 如果基 B 和 C 是统一空间 V 的基, T 是恒等变换, 那么矩阵 M 恰好是坐标变化矩阵 (将 B 下的坐标转化为 C 下的坐标)

5.4.2 从 V 到 V 的线性变换

当 $W = V, C = \mathcal{B}$ 时，前面提到的 M 称为 T 相对于 \mathcal{B} 的矩阵，或简称为 T 的 \mathcal{B} -矩阵，记为 $[T]_{\mathcal{B}}$

$V \rightarrow V$ 的线性变换对于 V 中所有的 \mathbf{x} 都有

$$[T(\mathbf{x})]_{\mathcal{B}} = [T]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

我们怎样理解这样的变换矩阵？这个变换矩阵的每一列的向量相当于，我们把 V 中的基映射到 W 中的像【这里不说是基是因为这个线性变换未必可逆，像未必线性无关】，这些像分别转化以 C 为基的 \mathbb{R}^m 中的向量， \mathbf{x} 的 \mathcal{B} -坐标相当于这些向量的系数。

而这一小节中如果有向量空间的等同，那么我们同样可以这么理解，这个变换矩阵的每一列相当于在 V 中将 \mathcal{B} 的基映射为像，再将这些像重新对应到这组基下坐标向量， \mathbf{x} 的 \mathcal{B} -坐标相当于这些向量的系数

【一个例子： \mathbb{P}_2 空间中的求导变换】

5.4.3 定理8（对角矩阵表示）

设 $A = PDP^{-1}$ ，其中 D 为 $n \times n$ 对角矩阵，若 \mathbb{R}^n 的基 \mathcal{B} 由 P 的列向量组成，那么 D 是变换 $\mathbf{x} \rightarrow A\mathbf{x}$ 的 \mathcal{B} -矩阵

这个定理又怎么理解呢？我们要实现一个线性变换（如果这个线性变换的矩阵 A 可以对角化），我们可以找到一个合适的基 \mathcal{B} ，我们先将原先的向量通过左乘一个逆矩阵 $P_{\mathcal{B}}^{-1}$ 转化为这个基下的坐标向量，然后我们在这个基下进行相对于这个基的线性变换 $[T]_{\mathcal{B}}$ ，最后我们再通过左乘矩阵 $P_{\mathcal{B}}$ 将这个矩阵恢复回我们原来的基。这里的相对于这个基的线性变换，或者说 T 相对于 \mathcal{B} 的矩阵（ T 的 \mathcal{B} -矩阵），也就是我们对角化出的矩阵 D

这个定理其实让我们能够从线性变换的角度理解我们对角化出的矩阵的实际意义

5.4.4 矩阵表示的相似性

事实上在定理8的证明中并没有使用 D 是对角矩阵这一点，事实上如果 A 相似于 C ，即有 $A = PCP^{-1}$ ，且有 \mathcal{B} 由 P 的列向量组成，则 C 是变换 $\mathbf{x} \rightarrow A\mathbf{x}$ 的 \mathcal{B} -矩阵。

相反的，若 $\mathbb{R}^n \rightarrow \mathbb{R}^n$ 的变换 $T: T(\mathbf{x}) = A\mathbf{x}$ ，而 \mathcal{B} 是 \mathbb{R}^n 的任意一个基，那么 T 的 \mathcal{B} -矩阵相似于 A

进而有，相似于 A 的所有矩阵的集合与线性变换 $\mathbf{x} \rightarrow A\mathbf{x}$ 相对于所有基的矩阵的集合是同一集合

5.4.5 进一步的理解

我们设 $f(x)$ 为向量空间的向量:函数,多项式,显然有

$$f(x) = B[f]_{\mathcal{B}} = C[f]_{\mathcal{C}} \quad (1)$$

从这个式子中,我们可以得到前面4.7节中的坐标变换矩阵

$$[f]_{\mathcal{C}} = P_{\mathcal{C} \leftarrow \mathcal{B}} [f]_{\mathcal{B}} \quad (2)$$

也可得到两组基之间的关系

$$CP_{\mathcal{C} \leftarrow \mathcal{B}} = B \quad (3)$$

如果我们再加入这一节中引入的线性变换的概念,我们发现

$$Df(x) = BD_{\mathcal{B}}[f]_{\mathcal{B}} = CD_{\mathcal{C}}[f]_{\mathcal{C}} \quad (4)$$

5.5 复特征值

5.5.1 定理9

假设 A 为 2×2 矩阵，且其特征值为 $a-bi$ ，特征向量为 \mathbf{v} 则令 $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, $P = \begin{bmatrix} \text{Re} \mathbf{v} & \text{Im} \mathbf{v} \end{bmatrix}$ （为了我们后面讲解的方便，我们把这两个向量组成的基记为 \mathcal{B} ），则

$$A = PCP^{-1}$$

这个怎么理解？相当于一个具有复特征值的矩阵的作用效果，相当于在以特征向量的实部与虚部为基向量的基的一个旋转变换

假设我们有 $\mathbf{v} = \text{Re}\mathbf{v} + i\text{Im}\mathbf{v}$ ，由于特征向量的性质 $A\mathbf{v} = (a - bi)\mathbf{v}$ ，即 $A(\text{Re}\mathbf{v} + i\text{Im}\mathbf{v}) = (a - bi)(\text{Re}\mathbf{v} + i\text{Im}\mathbf{v}) = a\text{Re}\mathbf{v} + a\text{Im}\mathbf{v} + (-b\text{Re}\mathbf{v} + a\text{Im}\mathbf{v})i$ 即

$$A(\text{Re}\mathbf{v}) + A(\text{Im}\mathbf{v})i = (a\text{Re}\mathbf{v} + b\text{Im}\mathbf{v}) + (-b\text{Re}\mathbf{v} + a\text{Im}\mathbf{v})i$$

相当于 A 的 \mathcal{B} -矩阵就是 $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$

5.6 离散动力系统

在这一讲中，我们假定矩阵 A 可对角化，有 n 个线性无关的特征向量 $\mathbf{v}_1, \dots, \mathbf{v}_n$ ，分别对应于特征值 $\lambda_1, \dots, \lambda_n$ ，我们将特征向量按照 $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$ 的顺序排列好，由于 $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ 是 \mathbb{R}^n 的一组基，故 \mathbb{R}^n 中任何一个初始向量 \mathbf{x}_0 均可以表示成为

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$$

我们有

$$\mathbf{x}_k = c_1(\lambda_1)^k\mathbf{v}_1 + \dots + c_n(\lambda_n)^k\mathbf{v}_n$$

5.6.1 解的几何意义

吸引点

对于任意一个初始向量，所有的轨迹都趋于原点，一个合适的矩阵是

$$\begin{bmatrix} 0.6 & 0 \\ 0 & 0.8 \end{bmatrix}$$

排斥点

对于任意一个初始向量，所有的轨迹都远离原点，一个合适的矩阵是

$$\begin{bmatrix} 1.6 & 0 \\ 0 & 2 \end{bmatrix}$$

鞍点

对于任意一个初始向量,原点在一部分方向吸引解,在一部分方向排斥解,一个合适的矩阵是

$$\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$$

5.6.2 变量代换

当矩阵不再是特殊的对角矩阵而是一般的矩阵时,我们可以采用对角化的方式进行变量代换.令 $P = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, 令 $y = P^{-1}x$ 或 $x = Py$, 从而有

$$y_{k+1} = Dy_k$$

从这个矩阵方程的意义来看, y_k 相当于 x_k 在基 $\{v_1, \dots, v_n\}$ 下的坐标向量, 我们有效地解耦了这个动力系统

5.6.3 复特征值

这个让我们联想到前一节中对于复特征值的矩阵的意义的讨论, 我们知道这样的 2×2 矩阵的效果是将一个向量乘以一个值(复特征值的模长)并且旋转一个角度

5.7 微分方程中的应用

5.8 特征值的迭代估计

5.8.1 幂法

估算严格占优的特征值, 即我们假设矩阵 A 有特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$, 且满足 $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$. (事实上, 当特征值差异越大的时候, 所需要的迭代次数越少).

原理

首先这个矩阵可以对角化,即有足够多的特征向量可以张成 \mathbb{R}^n ,那么任意一个初始向量可以表示为

$$\mathbf{x}_0 = c_1 \mathbf{v}_1 + \cdots + c_n \mathbf{v}_n$$

经过若干次迭代后,我们可以得到

$$\mathbf{x}_k = c_1 (\lambda_1)^k \mathbf{v}_1 + \cdots + c_n (\lambda_n)^k \mathbf{v}_n$$

等式两边可以同时除去 λ_1 这个最大的特征值,即可得到

$$\mathbf{x}_k = c_1 \mathbf{v}_1 + \cdots + c_n \left(\frac{\lambda_n}{\lambda_1}\right)^k \mathbf{v}_n$$

很显然,当 $k \rightarrow \infty$ 时,我们有 $\mathbf{x}_k \rightarrow c_1 \mathbf{v}_1$,即可以无限趋近于特征向量(的倍数)

然而事实上,我们不知道这样的特征值究竟是多少,即除去 λ_1 是不可能的,但为了让迭代后的向量不过大,我们可以采用“归一化”的方法,即将其最大分量化为1,同时当我们的 \mathbf{x}_k 接近于 λ_1 时,每一次迭代向量的最大分量变化的倍数接近于特征值 λ_1 ,我们可以把这个倍数 μ_k 作为我们特征值的一个估计值

使用方法

主要思路,任意选取起始向量 \mathbf{x}_0 ,经过迭代的运算利用 \mathbf{x}_k 得到 $A\mathbf{v}_x$,然后将得到的这个向量进行归一化,如此得到 \mathbf{x}_{k+1} ,当 $k \rightarrow \infty$ 时,这个向量的方向逐渐接近于这个最大特征值对应的特征向量,比值 μ_k 逐渐接近于特征值 λ_1 .

\mathbf{x}_k									
$A\mathbf{x}_k$									
μ_k									

5.8.2 逆幂法

这个算法可以用来求解任意特征值的估计值,不仅限于最大的特征值,前提是我们知道这样的一个矩阵的特征值的估计值,比如我们已知某个估计值 x 要求 λ_i 的某个估计值,我们考虑这一系列的值

$$\frac{1}{\lambda_1 - x}, \frac{1}{\lambda_n - x}, \cdots, \frac{1}{\lambda_n - x}$$

其中由于 x 理应最接近于要求的 λ_i ,因此我们对应的式子应当是最大的,我们可以证明矩阵 $B = (A - xI)^{-1}$ 的特征值为上面这些式子的形式

(如果 B 不可逆呢?那么 x 就可以认为是一个合适的特征值的估计了)

6 正交性与最小二乘法

6.1 内积、长度和正交性

6.1.1 定理1

设 $\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}$ 为 \mathbb{R}^n 中的向量, c 是一个数,则

a. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$

b. $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$

c. $(c\boldsymbol{u}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot \boldsymbol{v}) = \boldsymbol{u} \cdot (c\boldsymbol{v})$

d. $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$, 并且 $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ 当且仅当 $\boldsymbol{u} = \mathbf{0}$

向量的长度

向量 \boldsymbol{v} 的长度(或范数)定义为

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + \cdots + v_n^2} \text{ 且 } \|\boldsymbol{v}\|^2 = \boldsymbol{v} \cdot \boldsymbol{v}$$

\mathbb{R}^n 中的距离 $dis(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|$

正交的定义 $\boldsymbol{u} \cdot \boldsymbol{v} = 0$

6.1.2 定理2(毕达哥拉斯(勾股)定理)

两个向量 $\boldsymbol{u}, \boldsymbol{v}$ 正交当且仅当

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$$

正交补

如果 \boldsymbol{z} 与 \mathbb{R}^n 的子空间 W 中任意向量都正交,则称 \boldsymbol{z} 正交于 W ,与子空间 W 正交的全体向量组成的集合称为 W 的正交补,记为 W^\perp

以下是关于正交补的重要性质

1. 向量 \boldsymbol{x} 属于 W 的充分必要条件是其与 W^\perp 中所有向量都正交
2. W^\perp 是 \mathbb{R}^n 的一个子空间

6.1.3 定理3

假设 A 是 $m \times n$ 矩阵,那么 A 的行空间的正交补是 A 的零空间, A 的列空间的正交补是 A^T 的零空间(即 A 的左零空间):

$$(\text{Row}A)^\perp = \text{Nul } A, (\text{Col}A)^\perp = \text{Nul } A^T$$

正交补是相互的证明 即需要证明:对于 \mathbb{R}^n 的子空间任意 W

$$(W^\perp)^\perp = W$$

引理:

(1) W^\perp 是 \mathbb{R}^n 的子空间

(2) $\dim W + \dim W^\perp = n$

我们假设 W 不为 $\{\mathbf{0}\}$ (如果是很容易得证),那么 W 的一组基为 $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$,同样 W^\perp 的一组基为 $\{\mathbf{u}_{p+1}, \dots, \mathbf{u}_n\}$,

W^\perp 中的向量可以表示为

$$\mathbf{y} = c_{p+1}\mathbf{u}_{p+1} + \dots + c_n\mathbf{u}_n$$

根据上面的引理, $(W^\perp)^\perp$ 的维数和 W 相同,显然 $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ 是 $(W^\perp)^\perp$ 的一组基(这 p 个向量均在 $(W^\perp)^\perp$ 中,而且它们线性无关, p 维空间中 p 个向量的线性无关集构成这个空间的一组基),而这组基同样是 W 的基,故

$$(W^\perp)^\perp = W$$

6.2 正交集

正交集 当一个集合中的向量两两正交时,称这个集合为正交集

6.2.1 定理4

如果向量集合 $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ 是 \mathbb{R}^n 中**非零向量**构成的一组正交集,那么 S 是线性无关集,且构成 S 生成的子空间的的一组基

注意:

1. 并非 \mathbb{R}^n 中的所有正交集都是线性无关的,必须要强调非零,因为 $\{0\}$ 是线性相关的,但也是正交集

2. 一个集合是正交集,也不代表这个集合线性无关,因为里面可能有零向量

正交基 一个生成 \mathbb{R}^n 的子空间 W 正交向量集构成一组基,这组基是 W 的一个基,也是正交集

6.2.2 定理5

在 W 的正交基 $\{v_1, \dots, v_p\}$ 下, \mathbb{R}^n 子空间 W 中向量 y 在各个基向量 v_i 上的分量可以通过 $c_i = \frac{y \cdot v_i}{v_i \cdot v_i}$ 求得

正交投影 设 L 为 \mathbb{R}^n 的一个子空间,由 $\{u\}$ 生成, \mathbb{R}^n 中的向量 y 在子空间 L 上的正交投影可以表示为

$$\text{proj}_L y = \hat{y} = \frac{y \cdot u}{u \cdot u} u$$

单位正交集(单位正交基) 一个正交集集中的向量全部化为单位向量,则这个集合称为单位正交基,同时也是这个空间的单位正交基

6.2.3 定理6

一个 $m \times n$ 矩阵 U 具有单位正交列向量当且仅当 $U^T U = I_n$

6.2.4 定理7

假设 U 是一个具有单位正交列的 $m \times n$ 矩阵,且 x, y 是 \mathbb{R}^n 中的向量,那么

a. $\|Ux\| = \|x\|$

b. $(Ux) \cdot (Uy) = x \cdot y$

c. $(Ux) \cdot (Uy) = 0$ 的充分必要条件是 $x \cdot y = 0$

正交矩阵 正交矩阵是一个可逆的方阵 U ,且满足 $U^{-1} = U^T$,这样的矩阵具有单位正交列

[这里易错的一点是认为正交矩阵的列并不是单位化的,而认为存在”单位正交矩阵”这一概念]

容易验证:有单位正交列的方阵是正交矩阵,这样的矩阵同样具有正交行

正交矩阵与线性映射

例 6.2.1 设 $\mathbb{R}^n \rightarrow \mathbb{R}^n$ 为一个保长的线性变换,即对 \mathbb{R}^n 上所有向量 \mathbf{x} 有 $\|\mathbf{x}\| = \|T(\mathbf{x})\|$, 则

- (1) T 同时保持正交性
- (2) T 的标准矩阵是一个正交矩阵

证明 (1) 对于 \mathbf{x}, \mathbf{y} 满足 $\mathbf{x} \cdot \mathbf{y} = 0$ 由勾股定理

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

由于保长性

$$\|T(\mathbf{x}) + T(\mathbf{y})\|^2 = \|T(\mathbf{x})\|^2 + \|T(\mathbf{y})\|^2$$

因而也说明了 $T(\mathbf{x}) \cdot T(\mathbf{y}) = 0$

(2) T 的标准矩阵为 $\begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$, 由于其每一列既是正交的,也是单位化的,故其标准矩阵是一个正交矩阵

一个结论

正交矩阵 A 的 (i, j) 元的代数余子式是 $\pm a_{ji}$

证明

由于 A 是正交矩阵, $A^T A = I$, 从而 $\det A \det A^T = 1, i.e. \det A = \pm 1$

$$\text{adj.} A = A^{-1} \det A = \pm A^{-1} = \pm A^T$$

当 $\det A = 1$ 时, 我们利用伴随矩阵的定义

$$\text{adj.} A_{ij} = a_{ji}$$

当 $\det A = -1$ 时, 同理可以得到

$$\text{adj.} A_{ij} = -a_{ji}$$

6.3 正交投影

6.3.1 定理8(正交投影定理)

如果 W 是 \mathbb{R}^n 的一个子空间,那么 \mathbb{R}^n 中的向量 \mathbf{y} 可以唯一表示为

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

其中 $\hat{\mathbf{y}}$ 属于 W 而 \mathbf{z} 属于 W^\perp 如果 $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ 是 W 的任意正交基,那么

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p$$

其中 $\hat{\mathbf{y}}$ 称为 \mathbf{y} 在 W 上的正交投影,记作 $\text{proj}_W \mathbf{y}$

6.3.2 定理9(最佳逼近定理)

如果 W 是 \mathbb{R}^n 的一个子空间,那么 \mathbb{R}^n 中的向量 \mathbf{y} , $\hat{\mathbf{y}}$ 称为 \mathbf{y} 在 W 上的正交投影,那么 $\hat{\mathbf{y}}$ 是 W 中最接近 \mathbf{y} 的点,也就是

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$$

对所有属于 W 但又异于 $\hat{\mathbf{y}}$ 的向量 \mathbf{v} 成立

其中 $\hat{\mathbf{y}}$ 称为 W 中元素对 \mathbf{v} 的最佳逼近

6.3.3 定理10

如果 $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ 是 \mathbb{R}^n 中子空间 W 的单位正交基,则 \mathbb{R}^n 中向量 \mathbf{y} 在 W 的正交投影可以表示为

$$\text{proj}_W \mathbf{y} = (\mathbf{u}_1 \cdot \mathbf{y}) \mathbf{u}_1 + (\mathbf{u}_2 \cdot \mathbf{y}) \mathbf{u}_2 + \dots + (\mathbf{u}_p \cdot \mathbf{y}) \mathbf{u}_p$$

如果 $U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_p \end{bmatrix}$,那么我们有

$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y}, \forall \mathbf{y} \in \mathbb{R}^n$$

6.4 格拉姆-施密特方法

6.4.1 定理11(格拉姆-施密特方法)

对 \mathbb{R}^n 子空间 W 的一个基 $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$,定义

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1 \\ \mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 \\ &\vdots \\ \mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 - \dots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}} \mathbf{v}_{p-1} \end{aligned}$$

那么 $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ 为 W 的一组正交基,且满足

$$\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\}, \forall 1 \leq k \leq p$$

这说明了 \mathbb{R}^n 的非零子空间 W 一定存在一个正交基,因为 W 的一个基始终是存在的.

6.4.2 定理12(QR分解)

如果 $m \times n$ 矩阵 A 的列线性无关,那么 A 可以分解为 $A = QR$,其中 Q 为一个 $m \times n$ 矩阵,其列为 $\text{Col}A$ 的一个**标准**正交基, R 是一个 $n \times n$ 上三角可逆矩阵且在对角线上的元素为正数

注意:这里强调了 A 的列是线性无关的!

由于 Q 是一个正交矩阵,因而有 $Q^T Q = I$,我们计算 R 的方法是

$$\begin{aligned} A &= QR \\ Q^T A &= Q^T QR \\ R &= Q^T A \end{aligned}$$

6.5 最小二乘问题

最小二乘解是使得 $\|\mathbf{b} - A\mathbf{x}\|$ 最小的向量 $\hat{\mathbf{x}}$

6.5.1 定义

若 A 是 $m \times n$ 矩阵, \mathbf{b} 是 \mathbb{R}^m 中向量, 则 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解是 \mathbb{R}^n 中的 $\hat{\mathbf{x}}$, 使得

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

[回顾前面的最小逼近定理, 那里的结果中不等式是不能取等的, 而这个定理中是可以取等的, 这是为什么? 因为对于最佳逼近的 $\hat{\mathbf{b}}$, 可能存在多个 $\hat{\mathbf{x}}$ 使得 $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$]

6.5.2 定理13

$A\mathbf{x} = \mathbf{b}$ 的最小二乘解集和法方程 $A^T A\mathbf{x} = A^T \mathbf{b}$ 的非空解集一致

6.5.3 定理14

设 A 为 $m \times n$ 矩阵, 则下面的条件是等价的

- (a) 对于 \mathbb{R}^m 空间中的每个向量 \mathbf{b} , 方程 $A\mathbf{x} = \mathbf{b}$ 有唯一最小二乘解
- (b) 矩阵 A 的列是线性无关的
- (c) 矩阵 $A^T A$ 是可逆的

在上述条件成立的情况下, $A\mathbf{x} = \mathbf{b}$ 的唯一最小二乘解可以表示为

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

$\|\mathbf{b} - A\hat{\mathbf{x}}\|$ 的值称为近似的最小二乘误差

6.5.4 定理15

给定一个 $m \times n$ 矩阵 A , 其具有线性无关的列, 取 $A = QR$, 那么对于每一个属于 \mathbb{R}^m 的 \mathbf{b} , 方程 $A\mathbf{x} = \mathbf{b}$ 有唯一的最小二乘解, 且解为

$$\hat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}$$

注意:在数值计算中,我们要知道求一个矩阵的逆比解一个线性方程组更花时间,因此我们采用行化简解方程

$$R\hat{x} = Q^T b$$

的方法

6.5.5 Gram 矩阵

由一组向量的内积构成的对称矩阵,例如对于 \mathbb{R}^3 中的向量 u_1, u_2, u_3 ,其对应的Gram矩阵为

$$G = U^T U = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{bmatrix}$$

如果 u_1, u_2, u_3 是线性无关的,那么

$$U = QR, G = U^T U = (QR)^T (QR) = R^T (Q^T Q) R = R^T R$$

6.6 线性模型中的应用

我们把 $Ax = b$ 写作 $X\beta = y$,其中 X 称为设计矩阵, β 称为参数向量, y 为观测向量

对于最简单的二乘直线,我们的矩阵方程可以写作

$$X\beta = y, \text{ 其中 } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

对于一般的线性模型,我们引入残差向量 ϵ ,定义为 $\epsilon = y - X\beta$

6.7 内积空间

6.7.1 定义

向量空间 V 上的内积是一个映射 $V \times V \rightarrow \mathbb{R}$,对于每一对属于 V 的向量 u, v ,存在一个实数 $\langle u, v \rangle$ 满足下面公理

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
 2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
 3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$
 4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, 且 $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ 的充分必要条件是 $\mathbf{u} = \mathbf{0}$
- 一个赋予上面内积的向量空间称为内积空间

有了内积的定义后,我们很容易就可以得到范数的定义

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

一个**单位向量**是长度(范数)为1的向量.

同样我们可以得到距离的定义

$$dis(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

两个向量正交当且仅当 $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ 成立

6.7.2 定理16(柯西不等式)

对于 V 中任意向量 \mathbf{u}, \mathbf{v} , 有

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

6.7.3 定理17(三角不等式)

对于 V 中任意向量 \mathbf{u}, \mathbf{v} , 有

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

6.7.4 $C[a, b]$ 上的一个内积

定义内积为一个黎曼和

$$\langle p, q \rangle = \frac{1}{n+1} \sum_{i=0}^n p(t_i) q(t_i) \Delta t_i$$

$$\langle p, q \rangle = \frac{1}{b-a} \int_a^b p(t) q(t) dt$$

6.8 内积空间的应用

7 对称矩阵与二次型

7.1 对称矩阵的对角化

7.1.1 定理1

如果 A 是对称矩阵,那么两个不同特征空间中的特征向量相互正交
[事实上,不仅这样的特征向量相互正交,而且对称矩阵的特征值一定是实数]

7.1.2 定理2

一个 $n \times n$ 矩阵 A 可正交对角化当且仅当 A 是对称矩阵

但在我们构建一组正交基的时候,在一个特征子空间可能是2维或者更高维度的,于是我们要构建一组正交的向量

7.1.3 定理3(谱定理)

一个对称的矩阵 A 具有以下性质:

- a. A 有 n 个实特征值(包含重复的特征值)
- b. 对于每一个特征值,其特征根的几何重数等于其代数重数
- c. A 的特征空间相互正交,这仅在它们对应于不同的特征值时成立
- d. A 可正交对角化

这里有一点容易弄错的是,一个对称矩阵虽然可以正交对角化,但其不一定是满秩的,因为其零特征值对应的子空间可能是一(多)维的(即 $\text{Nul}A$ 是一(多)维的)

7.1.4 谱分解

$$\begin{aligned} A = PDP^T &= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T \end{aligned}$$

我们称 $\mathbf{u}_j \mathbf{u}_j^T$ 为投影矩阵,注意!构成投影矩阵形成的每个向量 \mathbf{u}_j 必须是单位向量!,否则就不是投影向量了

7.2 二次型

7.2.1 定义

\mathbb{R}^n 中的二次型是一个定义在 \mathbb{R}^n 上的函数, 其在 \mathbf{x} 处的值可以由表达式 $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ 确定, 其中 A 是一个 $n \times n$ 的对称矩阵

对称矩阵和二次型形成的是一一对应的双射

对于二次型, 可以进行如下形式的变量代换

$$\mathbf{x} = P\mathbf{y} \text{ 或 } \mathbf{y} = P^{-1}\mathbf{x}$$

由于对称矩阵可以正交对角化为 $A = PDP^T$ 的形式, 故二次型 $\mathbf{x}^T A \mathbf{x}$ 可以化为 $\mathbf{x}^T PDP^T \mathbf{x} = \mathbf{y}^T D \mathbf{y}$ 的形式

7.2.2 定理4(主轴定理)

设 A 是一个 $n \times n$ 对称矩阵, 那么存在一个变量代换 $\mathbf{x} = P\mathbf{y}$, 它将二次型 $\mathbf{x}^T A \mathbf{x}$ 变换为不含交叉乘积项的二次型 $\mathbf{y}^T D \mathbf{y}$

定理中矩阵 P 的列称为二次型 $\mathbf{x}^T A \mathbf{x}$ 的主轴, 向量 \mathbf{y} 为向量 \mathbf{x} 在这些主轴构造的 \mathbb{R}^n 空间的单位正交基下的坐标向量

注意: 这里的 P 必须是正交矩阵, 而不是任意可以将 A 对角化的矩阵, 为什么? 把 P 任意乘上一个常数即可

同样注意: 如果 A 存在一个高于1维的特征子空间, 那么主轴不是唯一确定的

7.2.3 定义

如果对于任意 $\mathbf{x} \neq \mathbf{0}$, 有二次型 $Q(\mathbf{x}) > 0$ 则称二次型是正定的

如果对于任意 $\mathbf{x} \neq \mathbf{0}$, 有二次型 $Q(\mathbf{x}) < 0$ 则称二次型是负定的

如果 $Q(\mathbf{x})$ 既有正值又有负值, 则称二次型是不定的

如果对于任意 $\mathbf{x} \neq \mathbf{0}$, 有二次型 $Q(\mathbf{x}) \geq 0$, 则称二次型是半正定的

如果对于任意 $\mathbf{x} \neq \mathbf{0}$, 有二次型 $Q(\mathbf{x}) \leq 0$, 则称二次型是半负定的

7.2.4 定理5(二次型与特征值)

设 A 是 $n \times n$ 矩阵,那么一个二次型是:

- a. 正定的,当且仅当 A 的特征值所有特征值是正数
- b. 负定的,当且仅当 A 的特征值所有特征值是负数
- a. 不定的,当且仅当 A 既有正特征值,又有负特征值

同理,我们可以对**对称矩阵**进行定义

一个正定矩阵是相应的二次型 $\mathbf{x}^T A \mathbf{x}$ 正定的对称矩阵

半正定等概念可以同理定义

从正定等信息中,我们可以得知什么?

如果一个矩阵 A 是正定的,那么其所有特征值大于0,因而其不存在零特征值,因而矩阵 A 是可逆的!

7.3 条件优化

7.3.1 多种角度理解约束

Rayleigh商函数 对于 \mathbb{R}^n 上的单位球面

$$S_n := \{\mathbf{x} | \mathbf{x}^T \mathbf{x} = 1, \mathbf{x} \in \mathbb{R}^n\}$$

问题: 在 S_n 上找出二次函数 Q 达到最大最小的点.

使用Rayleigh商函数,对于齐二次函数,将二次约束藏匿

多元微积分带约束的极值, 我们还没学, 但不妨碍用一下

考虑将 \mathbb{R}^n 中的球面约束转化为无约束

$$\begin{cases} \min / \max Q(\mathbf{x}) \\ s.t. \mathbf{x}^T \mathbf{x} = 1 \end{cases} \quad \text{我们构造} \quad \begin{cases} \min / \max_{\mathbf{x} \neq \mathbf{0}} R(\mathbf{x}) = \frac{Q(\mathbf{x})}{\mathbf{x}^T \mathbf{x}} \end{cases}$$

我们就得到了这个无约束函数 $R(\mathbf{x})$

$$R(\mathbf{x}) = \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

不难发现,我们此时再对 \mathbf{x} 进行等比放缩是不会改变这个函数的值的,因而我们的定义域范围就变成了 $\mathbb{R}^n (\mathbf{x} \neq \mathbf{0})$

我们利用微积分的性质,极值点: 变量的微小改变, 函数不变.

$$\delta R = \frac{(\mathbf{x} + \delta \mathbf{x})^T A (\mathbf{x} + \delta \mathbf{x})}{(\mathbf{x} + \delta \mathbf{x})^T (\mathbf{x} + \delta \mathbf{x})} - \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

最后我们通过舍弃一些二阶小量可以得到(这里的过程目前太复杂了)

$$\delta R = 0 \Leftrightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$$

于是我们成功地把有约束的问题转化成为了矩阵 A 的特征值问题
同理,如果我们进行变量代换 $\mathbf{y} = P^{-1}\mathbf{x}$,带入上面的式子得到

$$R(\mathbf{y}) = \frac{\mathbf{y}^T D \mathbf{y}}{\mathbf{y}^T \mathbf{y}}$$

我们不妨假设

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad (0 < \lambda_1 \leq \dots \leq \lambda_n)$$

从而我们得到

$$R(\mathbf{y}) = \frac{\mathbf{y}^T D \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2}$$

$$\frac{\sum_{i=1}^n \lambda_1 y_i^2}{\sum_{i=1}^n y_i^2} \leq \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \leq \frac{\sum_{i=1}^n \lambda_n y_i^2}{\sum_{i=1}^n y_i^2}$$

$$\lambda_1 \leq \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \leq \lambda_n$$

因此我们得知Rayleigh商函数的取值范围

题外话: Rayleigh商算法对于厄密矩阵 $A = A^H$ 在复数域内也成立.为什么?

Lagrange函数 构造Lagrange乘子函数

$$f(\mathbf{x}, \mu) = \mathbf{x}^T A \mathbf{x} - \mu(\mathbf{x}^T \mathbf{x} - 1)$$

求极值要求 $f(\mathbf{x}, \mu)$ 对 \mathbf{x} 和 μ 的微商(梯度)为0,即

$$\frac{\partial f(\mathbf{x}, \mu)}{\partial \mathbf{x}} = 0, \frac{\partial f(\mathbf{x}, \mu)}{\partial \mu} = 0$$

同理可以得到两个约束条件的式子

$$A\mathbf{x} - \mu\mathbf{x} = 0, \mathbf{x}^T\mathbf{x} - 1 = 0$$

这里的一种特殊情况是如果二次型对应的矩阵是一个 2×2 的矩阵,那么可以建立参数方程(关于极角 θ),求导求极值即可

谱分解 我们对 A 进行正交对角化得到正交矩阵 U

$$A = UDU^T = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T$$

从而使用正交矩阵得到 $\mathbf{y} = U^T\mathbf{x}$ 约束条件 $\mathbf{x}^T\mathbf{x} = 1$ 转化成为 $\mathbf{y}^T\mathbf{y} = 1$

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

从这个角度来看,最大值和最小值就对应着矩阵 A 特征值的最大值和最小值,对应的 \mathbf{x} 就是对应特征值的特征向量

7.4 奇异值分解

7.4.1 定义

A 的特征值是矩阵 $A^T A$ 的特征值的平方根,记为 $\sigma_1, \cdots, \sigma_n$, A 的奇异值是向量 $A\mathbf{v}_1, \cdots, A\mathbf{v}_r$ 的长度

7.4.2 定理9

假若 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 是包含 $A^T A$ 的特征向量的 \mathbb{R}^n 上的单位正交基,重新整理 $A^T A$ 的特征值使得 $\lambda_1 \geq \cdots \geq \lambda_n$,假若 A 有 r 个奇异值,那么 $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_r\}$ 是 $\text{Col}A$ 的一个正交基,且 $\text{rank}A = r$

7.4.3 定理10

设 A 是秩为 r 的 $m \times n$ 矩阵,那么存在一个 $m \times n$ 的矩阵 Σ ,其中 D 的对角线元素为 A 的前 r 个奇异值, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$,并且存在一个 $m \times m$ 正交矩阵 U 和 $n \times n$ 正交矩阵 V 使得 $A = U\Sigma V^T$

上述的分解称为矩阵 A 的奇异值分解, U 中的列称为 A 的左奇异向量, V 中的列称为 A 的右奇异向量

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & O_{(m-r) \times (n-r)} \end{bmatrix}, V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$$

$A^T A$ 与 AA^T 的特征值的关联

$$\begin{aligned} \det \begin{bmatrix} I_m & A \\ A^T & \lambda I_n \end{bmatrix} &= \det \left(\begin{bmatrix} I_m & 0 \\ -A^T & I_n \end{bmatrix} \begin{bmatrix} I_m & A \\ A^T & \lambda I_n \end{bmatrix} \right) = \det(\lambda I_n - A^T A) \\ &= \det \left(\begin{bmatrix} I_m & -A/\lambda \\ 0 & I_n \end{bmatrix} \begin{bmatrix} I_m & A \\ A^T & \lambda I_n \end{bmatrix} \right) = \lambda^{n-m} \det(\lambda I_m - AA^T) \end{aligned}$$

因此,对于 $A^T A$ 和 AA^T 的非零特征值是相同的,它们的特征值只相差若干个零特征值

那么我们既然知道了它们的非零特征值是一一对应的,那么我们会想:它们的特征向量有什么关系呢?

7.4.4 奇异值分解相关公式

由于 $A^T A$ 是一个对称矩阵,因此其可以正交对角化
故存在 $n \times n$ 矩阵 V, Ξ ,使得

$$VV^T = V^T V = I_n \quad (5)$$

$$\Xi = \begin{bmatrix} \mu_1 & & & \\ & \ddots & & \\ & & \mu_r & \\ & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \quad (6)$$

$$A^T A = V \Xi V^T \quad (7)$$

$$VA^TAV^T = \Xi \quad (8)$$

$$A^TAV = V^T\Xi \quad (9)$$

其中我们发现 V 的前 r 列张成了 $\text{Row}A = \text{Row}A^TA = \text{Col}A^TA = \text{Col}A^T$
我们观察到方程8可以得到

$$(Av_i) \cdot (Av_j) = \mu_i \delta_{ij} \quad (10)$$

由10中可以知道 Av_i 正交但不归一($i = 1, \dots, r$),它们构成了 $\text{Col}A$ 的一组正交但不归一的基

同时我们也可以得到关于其长度的信息

$$\|Av_i\| = \begin{cases} \sqrt{\mu_i} & , 1 \leq i \leq r \\ 0 & , r+1 \leq i \leq n \end{cases} \quad (11)$$

根据我们前面对于奇异值的定义,我们可以得到

$$A^TAv_i = \mu_i v_i, \quad 1 \leq i \leq r \quad (12)$$

因此 v_i 是 A^TA 的特征向量

在12中我们在等式两边同乘以 A ,我们得到

$$AA^TAv_i = \mu_i Av_i, \quad 1 \leq i \leq r \quad (13)$$

从这个式子中我们可以看出 Av_i 是 AA^T 的特征向量

同样我们也可以从12得到有关长度的信息

$$\|A^TAv_i\| = \mu_i \quad (14)$$

我们也可以进行归一化的操作

$$u_i = \frac{Av_i}{\sqrt{\mu_i}}, \quad 1 \leq i \leq r \quad (15)$$

因此我们得到的 u_i 可以形成 $\text{Col}A$ 正交归一的基

我们可以把归一化的结果带入12,从而得到

$$A^Tu_i = \sqrt{\mu_i} v_i, \quad 1 \leq i \leq r \quad (16)$$

在16中,我们在等式两边同时左乘一个 A 矩阵即可得到

$$AA^Tu_i = \mu_i u_i, \quad 1 \leq i \leq r \quad (17)$$

注意,上述很多方程中虽然我们限制了 $1 \leq i \leq r$,但事实上对于 $i > r$ 也成立,我们发现一个重要的关于 $A^T A$ 和 AA^T 矩阵的性质:

V 的列为 $A^T A$ 的特征向量(不过前 r 列对应了非零的特征值,后面 $n - r$ 列对应了零特征值)

U 的列为 AA^T 的特征向量(不过前 r 列对应了非零的特征值,后面 $m - r$ 列对应了零特征值)

7.4.5 $A^T A$ 和 AA^T 矩阵的再讨论

映射特性

首先我们先明确,在SVD过程中,我们事实上构造了 v_1, \dots, v_n 作为 \mathbb{R}^n 的基,其中 v_1, \dots, v_r 为Row A 的基,而构造了 u_1, \dots, u_m 作为 \mathbb{R}^m 的基,其中 u_1, \dots, u_r 为Col A 的基

行空间的像

$$\text{Col}A^T A = \text{Col}A \Rightarrow A : \text{Row}A \rightarrow \text{Col}A \text{ 双射}$$

列空间的像

$$\text{Col}AA^T = \text{Row}A \Rightarrow A^T : \text{Col}A \rightarrow \text{Row}A \text{ 双射}$$

双射是因为两个子空间维数相等,而且基存在一个一一对应的关系

秩

$$\text{rank}AA^T = \text{rank}A^T A = \text{rank}A = r$$

在列出了上面的方程之后,我们可以重新思考 $A, A^T, A^T A, AA^T$ 这四个矩阵具有的意义

$$A : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$A^T : \mathbb{R}^m \rightarrow \mathbb{R}^n$$

$$A^T A : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$AA^T : \mathbb{R}^m \rightarrow \mathbb{R}^m$$

上面的映射关系中,我们也可以将 \mathbb{R}^n 理解为Row A, \mathbb{R}^m 理解为Col A

从向量模长的角度,前两个将向量模长变为 $\sqrt{\mu_i}$ 倍,后两个将向量模长变为 μ_i 倍

几个空间的关系 如下:

$\mathbf{v}_1, \dots, \mathbf{v}_r$ 属于 $\text{Row}A$

$\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ 属于 $\text{Nul}A$

$\mathbf{u}_1, \dots, \mathbf{u}_r$ 属于 $\text{Col}A$

$\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$ 属于 $\text{Nul}A^T$

7.4.6 SVD奇异值和特征值的讨论

奇异值在数值上相对特征值更加稳定,不容易受到矩阵小变化的影响.

例 7.4.1 对于矩阵

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \frac{1}{6000} & 0 & 0 & 0 \end{bmatrix}$$

方阵的特征值从 $0, 0, 0, 0$ 变为 $\frac{1}{10}, \frac{i}{10}, -\frac{i}{10}, -\frac{1}{10}$, 变化幅度相当大

然而,奇异值则从 $1, 2, 3$ 变化为 $1, 2, 3, \frac{1}{6000}$, 变化十分有限

实际应用中,我们应避免 $A^T A$ 的计算,原因是任何 A 中元素的误差在 $A^T A$ 中被平方.存在快速迭代的方法,可计算精确到很多位数的矩阵 A 的奇异值和奇异向量.

7.4.7 可逆矩阵定理的最后补充

设 A 为 $n \times n$ 矩阵,则下列命题与 A 是可逆矩阵等价

u. $(\text{Col}A)^\perp = \{\mathbf{0}\}$

v. $(\text{Nul}A)^\perp = \mathbb{R}^n$

w. $\text{Row}A = \mathbb{R}^n$

x. A 有 n 个非零奇异值

7.4.8 简化奇异值分解

首先 A 可以写成谱分解的形式

$$A = U \Sigma V^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

$$U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}, V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

从而有

$$A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r & V_{n-r} \end{bmatrix} = U_r D V_r^T$$

7.4.9 Moore-Penrose逆(伪逆)与最小二乘解

Moore-Penrose逆具有的性质包含以下几点

(1) $AXA = A$

(2) $XAX = X$

(3) $(AX)^H = AX$

(4) $(XA)^H = XA$

则称 X 为 A 的M-P逆

可以证明 A^\dagger 是唯一的解(这实质上依赖于(3)(4)两个条件的成立)

$$A = U\Sigma V^T$$

$$A^\dagger = V\Sigma^\dagger U^T$$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & O_{(m-r) \times (n-r)} \end{bmatrix}, \Sigma^\dagger = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \ddots & & \\ & & \frac{1}{\sigma_r} & \\ & & & O_{(n-r) \times (m-r)} \end{bmatrix},$$

按照前面的奇异值分解的简化,我们同样可以把伪逆写成对应形式

$$A^\dagger = V_r D^{-1} U_r^T$$

接下来我们来考虑这个M-P逆和最小二乘解的关系,我们有 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解为

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

证明 利用最小二乘解是法方程的解,我们可以带入这个解 $\hat{\mathbf{x}} = A^\dagger \mathbf{b}$ 到法方程左侧

$$A^T A \hat{\mathbf{x}} = A^T A A^\dagger \mathbf{b} = A^T (U \Sigma V^T) (V \Sigma^\dagger U^T) \mathbf{b} = A^T \mathbf{b}$$

同理,我们可以利用简化的P-M逆,验证这个最小二乘解就是 \mathbf{b} 在 $\text{Col}A$ 上的投影,

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

我们知道 $\text{Col}U_r = \text{Col}A$,因而等式右边就表示了投影

7.4.10 对Moore-Penrose逆几何意义的理解

引理 7.4.1 首先,若方程 $A\mathbf{x} = \mathbf{b}$ 是相容的,令 $\mathbf{x}^\dagger = A^\dagger \mathbf{b}$,则可以证明 \mathbf{x}^\dagger 是方程 $A\mathbf{x} = \mathbf{b}$ 的最小长度解

证明 首先,

$$A^\dagger A \mathbf{x} = V \Sigma^\dagger U^T U \Sigma V^T \mathbf{x} = V V^T \mathbf{x}$$

也就是 \mathbf{x} 在 $\text{Row}A$ 上的正交投影

我们在等式 $\mathbf{x}^\dagger = A^\dagger \mathbf{b}$ 两侧同乘 $A^\dagger A$

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

从而有

$$\text{proj}_{\text{Row}A} \mathbf{x}^\dagger = A^\dagger \mathbf{b} = \mathbf{x}^\dagger$$

因此 $\mathbf{x}^\dagger \in \text{Row}A$

之后,对于任意的解 \mathbf{u} 满足 $A\mathbf{x} = \mathbf{b}$,

可知

$$\mathbf{u} = \mathbf{x}^\dagger + (\mathbf{u} - \mathbf{x}^\dagger)$$

其中 $\mathbf{u} - \mathbf{x}^\dagger \in \text{Nul}A$,从而利用这两个子空间的正交性和勾股定理可知

$$\|\mathbf{u}\|^2 = \|\mathbf{x}^\dagger\|^2 + \|\mathbf{u} - \mathbf{x}^\dagger\|^2 \geq \|\mathbf{x}^\dagger\|^2$$

从而我们可知

$$\|\mathbf{u}\| \geq \|\mathbf{x}^\dagger\|$$

证明了 \mathbf{x}^\dagger 是方程 $A\mathbf{x} = \mathbf{b}$ 的最小长度解

引理 7.4.2 我们把 \mathbf{b} 推广到 \mathbb{R}^m 中的任意向量,此时方程 $A\mathbf{x} = \mathbf{b}$ 未必相容,我们同样可以证明 $A^\dagger \mathbf{b}$ 是最小长度的最小二乘解

证明 记 $\hat{\mathbf{b}}$ 为向量 \mathbf{b} 在 $\text{Col}A$ 上的正交投影,不妨假设

$$\mathbf{b} = \hat{\mathbf{b}} + \mathbf{z}, \mathbf{z} \in \text{Nul}A^T$$

不难知道 $A\mathbf{x} = \hat{\mathbf{b}}$ 的解即为原方程的最小二乘解

根据上面的结论我们知道 $\mathbf{p} = A^\dagger \hat{\mathbf{b}}$ 是最小长度的最小二乘解

然而

$$A^\dagger \mathbf{z} = V\Sigma^\dagger U^T \mathbf{z} = V\Sigma^\dagger (U^T \mathbf{z}) = V\Sigma^\dagger \mathbf{0} = \mathbf{0}$$

故

$$A^\dagger \mathbf{b} = A^\dagger \hat{\mathbf{b}} + A^\dagger \mathbf{z} = A^\dagger \hat{\mathbf{b}}$$

因而我们证明了 $A^\dagger \mathbf{b}$ 就是最小长度的最小二乘解

$A^\dagger \mathbf{b}$ 其实也是任意最小二乘解在 $\text{Row}A$ 上唯一的正交投影(这点不难理解,因为最小二乘解在 $\text{Nul}A$ 上的分量是任意的,而由于 $\text{Nul}A = (\text{Row}A)^\perp$,其分量在 $\text{Row}A$ 上投影一定)

对于 $\forall \mathbf{b} \in \mathbb{R}^m$,都有唯一的 \mathbf{x}^\dagger 与之对应,这定义了一个从 \mathbb{R}^m 到 \mathbb{R}^n 的线性映射 A^\dagger ,这个线性映射对应的矩阵其实就是矩阵 A 的M-P逆,根据M-P的逆的定义可知

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

7.4.11 奇异值分解与线性变换

例 7.4.2 令 $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是线性变换,描述如何求 \mathbb{R}^n 和 \mathbb{R}^m 中的基 \mathcal{B}, \mathcal{C} ,使得 T 相对于 \mathcal{B} 和 \mathcal{C} 的矩阵是一个 $m \times n$ “对角”矩阵

证明 设 A 为线性变换 T 的标准矩阵,对 A 进行奇异值分解得到 $A = U\Sigma V^T$,我们取 \mathcal{B} 为 V 的各列, \mathcal{C} 为 U 的各列

同时,我们记 \mathbf{e}_j 为单位矩阵 I_n 的第 j 列, \mathbf{e}'_j 为单位矩阵 I_m 的第 j 列

$$T(\mathbf{v}_j) = \begin{cases} A\mathbf{v}_j = U\Sigma V^T \mathbf{v}_j = U\Sigma \mathbf{e}_j = U\sigma_j \mathbf{e}'_j = \sigma_j U\mathbf{e}'_j = \sigma_j \mathbf{u}_j & , 1 \leq j \leq m \\ \mathbf{0} & , m < j \leq n \end{cases}$$

从而

$$[T(\boldsymbol{v}_j)]_{\mathcal{C}} = \begin{cases} \sigma_j \boldsymbol{e}'_j & , 1 \leq j \leq m \\ 0 & , m < j \leq n \end{cases}$$

故结论得证

7.5 图像处理和统计学中的应用

8 向量空间的几何学

8.1 仿射组合

8.2 仿射无关性

8.3 凸组合

8.4 超平面

8.5 多面体

8.6 曲线和曲面

一个有意思的内容拓展

群、环、域

集合(Set)

无序性

互异性

确定性

原群(Magma)

集合与在该集合上的一个二元封闭性运算，构成了原群(Magma)

例如：自然数与加法构成原群

半群(Semigroup)

结合律(Associative)

$$x \circ (y \circ z) = (x \circ y) \circ z, \forall x, y, z \in S$$

在原群的基础上添加约束条件结合律即可得到半群

么半群(Monoid)

单位元(Identity)

对于集合 S ， e 为其中的元素，如果对于 S 中任何一个元素 x 有

$$e \circ x = x \text{ 则称 } e \text{ 为左单位元}$$

$$x \circ e = x \text{ 则称 } e \text{ 为右单位元}$$

$$e \circ x = x \circ e = x \text{ 则称 } e \text{ 为单位元}$$

么半群是包含单位元的半群

群(Group)

逆元(Inverse)

对于集合 S , e 为其中的单位元, 对于 S 中元素 x ,

如果集合中存在元素 y 和运算 \circ 使得 $y \circ x = e$ 则称 y 为 x 的左逆元

如果集合中存在元素 y 和运算 \circ 使得 $x \circ y = e$ 则称 y 为 x 的右逆元

如果集合中存在元素 y 和运算 \circ 使得 $y \circ x = x \circ y = e$ 则称 y 为 x 的逆元

如果在么半群的基础上加上其中的任何元素都存在逆元, 那么则得到群

阿贝尔群(Abelian Group/commutative group)

交换律(commutative)

对于集合 S 中的元素 x, y 有一运算 \circ , 如果满足 $x \circ y = y \circ x$, 则称 x 和 y 在运算 \circ 下可交换

如果在群的基础上加上可交换性就得到了阿贝尔群

半环(Semi Ring)

分配律(Distributive)

对于集合 S 中的元素 x, y , 有运算 $\circ, *$, 这两个运算有如下关系如果 $x * (y \circ z) = (x * y) \circ (x * z), \forall x, y, z \in S$ 则称运算 $*$ 在 \circ 上满足左分配律

如果 $x * (y \circ z) = (x * y) \circ (x * z), \forall x, y, z \in S$ 则称运算 $*$ 在 \circ 上满足右分配律

如果运算 $*$ 在 \circ 上同时满足左分配律和右分配律, 则称 $*$ 在 \circ 上满足分配律

半环 $(S, \circ, *)$ 符合如下条件:

1. 由一个阿贝尔么半群 (S, \circ) 和么半群 $(S, *)$ 组成
2. 运算 $*$ 在 \circ 上满足分配律

3. 运算 $*$ 在 S 上有零元

$$x * e' = e' * x = e', e' \in S, \forall x \in S$$

注意这里要求的是 $*$ 有零元

理解时可以考虑把 \circ 想象为加法， $*$ 想象为乘法，可以想象乘法在加法上满足分配律

环(Ring)

环是在半环的基础上，将作为阿贝尔么半群（通常代表加法）的 (S, \circ) 变为阿贝尔群 (S, \circ)

注意这里变为阿贝尔群的是那个通常代表加法的 (S, \circ)
在 \circ 运算中引入逆元不会造成问题（如果对于乘法引入逆元，而半环的3.中又要求有零元就会有问题）

域(Field)

域在环的基础上引入了以下约束：

1. 乘法么半群 $(S, *)$ 满足交换律，满足这一条件，这时环变为阿贝尔环或交换环

2. 乘法除零元之外都有逆元，也就是除了零元外的集合与乘法都可以构成阿贝尔群，满足这两个条件，环就升级成域

Module

假设 $(R, +, *)$ 是一个阿贝尔环， (M, \oplus) 是一个阿贝尔群，如果存在一个运算 $\cdot : R \times M \rightarrow M$ 满足如下条件，我们称 (M, \oplus) 是 $(R, +, *)$ 的Module，

需要满足的条件如下：

$$1. (r_1 + r_2) \cdot m = r_1 \cdot m + r_2 \cdot m$$

$$2. r \cdot (m_1 \oplus m_2) = r \cdot m_1 \oplus r \cdot m_2$$

$$3. r_1 \cdot (r_2 \cdot m) = (r_1 * r_2)m$$

$$4. 1_R \cdot m = m$$

以下几点需要注意：

R 和 M 上存在各自的加法运算，分别为 $+: R \times R \rightarrow R$ 和 $M \times M \rightarrow M$

R 上存在自身的乘法运算，为 $*: R \times R \rightarrow R$

M 不要求存在自身的乘法运算，这里定义的乘法运算要求第一个元素是 R 的，第二个元素是 M 的。即 $\cdot: R \times M \rightarrow M$

例子：线性空间是域的Module。这时交换环 R 就是域，交换环 M 则是向量的集合