

线性代数 A 习题课讲义

原生生物

* 线性代数 A [丁一文老师班] 习题课讲义, 后续在下册习题课讲义中。

* 配套教材为丘维声《高等代数》上册, 个人还参考了王新茂《线性代数讲义》。

* 记号约定: 上标 T 表示转置, H 表示取转置并对每个元素取共轭。 O 表示零矩阵, I_a 表示 a 阶单位阵 (\mathcal{I} 为恒等映射 $\mathcal{I}x = x$), 下标 $m \times n$ 表示矩阵的行列数, \det 表行列式, diag 表将后方元素拼到主对角线上形成的对角阵。

目录

一 线性方程组与行列式	5
§1.1 习题解答	5
§1.2 线性方程组	5
1.2.1 算法视角	5
1.2.2 等价类与标准形	6
§1.3 行列式技巧 I	7
1.3.1 技巧概述	7
1.3.2 例题	7
§1.4 为什么是线性代数	10
1.4.1 从线性方程组到统计学习	10
1.4.2 线性代数学习指南	11
二 空间视角的线性方程组	12
§2.1 习题解答	12
§2.2 如何自学?	14
2.2.1 四个层面	14
2.2.2 关于证明书写	16
§2.3 行列式回顾与空间引入	16
2.3.1 Cramer 法则	16
2.3.2 线性表出	17
2.3.3 封闭性与空间	18
§2.4 秩	20
2.4.1 信息量的视角	20
2.4.2 从空间到向量组	20
2.4.3 视为函数的维数	22
§2.5 再看线性方程组	23
2.5.1 齐次线性方程组	23
2.5.2 非齐次线性方程组	24
2.5.3 秩的计算	25

三 向量空间与矩阵	26
§3.1 习题解答	26
§3.2 线性映射与矩阵	27
3.2.1 基的操作与坐标	27
3.2.2 线性映射	28
3.2.3 第一同构定理	29
§3.3 矩阵视角的行列变换	32
3.3.1 整体思想	32
3.3.2 秩的矩阵论定义	33
四 补充：通用技巧	36
§4.1 基本工具	36
4.1.1 限制映射	36
4.1.2 可逆与逆	37
4.1.3 相抵标准形	38
§4.2 习题讲解	39
4.2.1 秩相关的三种处理	39
4.2.2 方阵的特性	42
4.2.3 无穷维的线性算子	47
五 期中总结	48
§5.1 习题解答	48
§5.2 拾遗	51
5.2.1 分块矩阵	51
5.2.2 行列式技巧 II	51
§5.3 知识与技巧整理	52
5.3.1 线性方程组与行列式	52
5.3.2 向量空间	53
5.3.3 矩阵	55
六 补充：综合复习	56
§6.1 复习题	56
§6.2 解答	57
七 可逆矩阵	65
§7.1 期中考试	65
7.1.1 试题	65
7.1.2 解答	67
§7.2 习题解答	69
§7.3 可逆矩阵	70
7.3.1 基本性质	70
7.3.2 伴随方阵与摄动法	71
7.3.3 消去与拆分	73
§7.4 相抵与相似	74
7.4.1 相抵标准形	74
7.4.2 相似的动机	75

7.4.3 相似的基本结论	76
7.4.4 特征系统	78
八 补充：可逆的空间视角	78
§8.1 逆矩阵的计算	78
8.1.1 伴随方阵 vs 初等变换	79
8.1.2 整体思想	79
8.1.3 分块行列变换	81
§8.2 线性映射与相抵	84
8.2.1 线性映射的矩阵表示	84
8.2.2 作为基变换的可逆阵	86
8.2.3 不可区分的线性映射	87
§8.3 线性变换与相似	88
8.3.1 线性变换	88
8.3.2 复向量中的特征系统	89
8.3.3 可对角化 I	91
九 可逆阵的技巧	92
§9.1 习题解答	92
§9.2 矩阵论技巧	94
9.2.1 更复杂的分块	94
9.2.2 多项式相关	96
9.2.3 可对角化 II	97
§9.3 相似的特殊结论	99
9.3.1 相似三角化	99
9.3.2 置换相似	100
9.3.3 Caylay-Hamilton 定理	100
9.3.4 一般数域上的相似	101
§9.4 空间技巧	102
9.4.1 选取基的艺术	102
9.4.2 特征子空间与根子空间	104
9.4.3 可对角化 III	106
十 正交方阵	107
§10.1 习题解答	107
§10.2 内积	110
10.2.1 保内积的线性映射	110
10.2.2 正交化	111
10.2.3 合同	113
§10.3 正交相似	114
10.3.1 正交方阵	114
10.3.2 实对称阵	115
§10.4 正交相抵	116
10.4.1 反射变换	116
10.4.2 奇异值分解	117

10.4.3 矩阵等价关系再探	120
十一 补充：正定方阵	120
§11.1 合同下的标准化	120
11.1.1 惯性指数	120
11.1.2 成对初等变换	121
11.1.3 复对称方阵	122
§11.2 正定方阵	123
11.2.1 二次型	123
11.2.2 正定性	124
11.2.3 不等式问题	127
§11.3 最小二乘	129
11.3.1 二次型的最值	130
11.3.2 二次函数的极值与最值	130
11.3.3 最小二乘解	132
十二 期末总结	133
§12.1 习题解答	133
§12.2 拾遗	135
12.2.1 不变子空间	135
12.2.2 正交补空间	136
12.2.3 QR 分解	138
§12.3 知识与技巧整理	139
12.3.1 相抵	139
12.3.2 相似	140
12.3.3 正交阵与对称阵	141
十三 补充：期末复习	142
§13.1 复习题	142
§13.2 解答	144

— 线性方程组与行列式

§1.1 习题解答

1. 习题 1.1-1(5)

消元计算解得 $x_1 = -8, x_2 = 3, x_3 = 6, x_4 = 0$ 。

2. 习题 1.1-3(1)

消元计算解得

$$\begin{cases} x_1 = x_3 - x_4 - 3 \\ x_2 = x_3 + x_4 - 4 \end{cases}$$

3. 习题 1.2-4

直接消元可发现有解等价于 $a \neq -2$ ，此时解为

$$\begin{cases} x_1 = -x_3 - 2 \\ x_2 = 2x_3 + 5 \\ x_4 = -10 \end{cases}$$

4. 习题 2.1-4

由定义直接计算前/后的数量可知逆序数分别为 $k-1$ 与 $n-k$ 。

5. 习题 2.2-3

考虑完全展开式，无论每项如何选取，都要在后三行中选三个不同列的，一定会取到 0，由此行列式为 0。

6. 习题 2.3-3(1)

注意到三列和为 0，将二、三两列加到第一列即可得到一列 0，从而即得行列式为 0。

§1.2 线性方程组

1.2.1 算法视角

我们从构造算法的角度进行知识回顾。

目标：进行线性方程组的求解。

问题：何为求解？需要一个最简的能直接看出解的形式 [且解可能存在或无穷多]。

方法：之前学习过的高斯消元与代入，进一步规范为三种基本变换 [事实上，由于允许代入时某行已经变为了 $x_i = c$ ，代入也可看作某行减另一行倍数]。

* 将线性方程组描述为增广矩阵后，问题的求解就变为了矩阵上的三种基本行变换化简，但本质没有区别。

想法 I：只要某方程中有某个变量 x_i 有系数非零，就可以利用高斯消元使得只有该方程拥有此变量 [这里拥有代表系数非零，下同]。

操作 I：若方程组中拥有 x_1 ，可以将有 x_1 的交换到第一个方程，再利用高斯消元 [第一种初等变换] 消去其他，从而只剩第一个方程拥有 x_1 。

想法 II：上述过程不能完全重复 [消去 x_2 的过程可能会让 x_1 的系数重新出现]，但可以部分重复 [考虑第一个方程以外的方程，若拥有 x_2 ，则可以消去使得只有第二个方程有 x_2 ，否则考虑 x_3 ，以此类推]。

- 操作 II: 利用递归思想, 重复上述过程, 可以使得第 i 个方程拥有变量 $x_{j(i)}$, 且之后的方程均无 [可以将每个方程拥有的第一个变量称为主元]。根据我们的消元过程, 还能让 $j(i)$ 随 i 增加而严格单调增加。
 * 此时对应的矩阵称为阶梯形, 第 i 行的非零元素 $a_{i,j(i)}$ 称为第 i 个主元。
 * 上述过程里是用上面的方程消去下面的方程中的项, 称为前代 [向前代入]。

想法 III: 可以通过同除非零数使得主元前系数是 1, 从而写成更好看的形式。

操作 III: 利用第三种初等变换将主元前的系数都变为 1。

想法 IV: 接下来理应利用代入消去。

操作 IV: 若某个方程为 $x_i = c$, 则可以所有其他方程减去它的倍数以彻底消去 x_i 。

- * 但这样的想法和操作未必可实现, 阶梯形无法保证存在 $x_i = c$ 。

想法 V: 即使没有 $x_i = c$, 似乎也能消去其他方程中的 x_i , 但需要保证化为阶梯形中已经消去的部分不出现新的数。

操作 V: 其余方程减去最后一个方程的某倍数可以使得它们不拥有最后一个方程的主元, 且这个过程不会破坏阶梯形 [考虑阶梯形的形式]。

想法 VI: 仍然可以部分重复上述过程。

操作 VI: 自下而上, 先使其他方程不拥有最后一个方程的主元, 再使前面的方程不拥有倒数第二个方程的主元…… 最终得到的形式为阶梯形矩阵, 且主元前系数为 1、每个方程的主元其他方程均不拥有。

- * 此时对应的矩阵称为简化阶梯形。

- * 上述过程里是用下面的方程消去上面的方程中的项, 称为回代 [向回代入]。

回顾上述过程, 可以发现, 由于只使用了规定的基本变换, 简化阶梯形矩阵与原本的增广矩阵对应方程组的解相同。最后, 我们是否可以从中直接看出解? 书上的定理已经给出了答案, 从而我们给出了线性方程组求解的通用算法。

- * 上述的算法构造过程事实上既有自顶向下的分解问题分析思路, 也有自底向上的从可行操作出发化简思路。

1.2.2 等价类与标准形

* 本段内容为抽象代数中的部分内容, 相对偏抽象, 但解释这些是为了解释我们之后一直要使用的标准形思想。

* 省流版本: 我们按照可以某些要求 [解集相同、秩相等等] 将某些东西 [线性方程组、矩阵等] 分类, 每类里“最好看”的就是标准形。

关系: 集合 A 上的有序对到 $\{0, 1\}$ 的映射。

* 人话: 给两个 A 中元素 [有顺序], 可以确定它们是否有关系。对关系 \sim , 用 $a \sim b$ 代表有关系, $a \not\sim b$ 代表无关系。

等价关系: 满足反身性 [$a \sim a$]、对称性 [若 $a \sim b$ 则 $b \sim a$]、传递性 [若 $a \sim b, b \sim c$ 则 $a \sim c$] 的关系。

等价类: 若一个集合 A 上定义了等价关系 \sim , 则可划分 [即 A 为一些集合的并, 且它们两两交集为空] 为 $A_\alpha, \alpha \in I$, 使得同一个 A_α 中任取两元素等价, 不同的各任取一个不等价。

代表元: 从等价类中取出的某一个性质相对好的元素, 以代表整个等价类。

在上面这套定义下, 简化阶梯形矩阵事实上成为了线性方程组集合在可以做初等行变换得到的等价关系下选出的代表元, 用于直观看出解。

- * 由于初等行变换不改变解, 这样的等价关系构建确实可以用于寻找解。

问题: 同解的线性方程组 [假设增广矩阵行、列数相同] 一定能线性变换得到吗?

答案: 错误, 考虑增广矩阵

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

两方程组均无解, 但无法通过我们定义的初等行变换得到。

* 由此同解事实上是比可行列变换得到更“大”的等价关系, 因为后者蕴含了前者。

标准形思想: 考虑某问题时, 若其有不变量, 可以将不变量相同作为等价关系, 并选取等价类中最易于研究的元素作为代表元, 称为标准形。

之后学习到矩阵时, Hermite 标准形/相抵标准形是同阶矩阵以秩相等作为等价关系下的代表元; Jordan 标准形是同阶方阵以相似作为等价关系下的代表元; 正/负惯性指数是同阶实对称阵以合同作为等价关系下的分类……

§1.3 行列式技巧 I

1.3.1 技巧概述

* 需要学完 Binet-Cauchy 公式才能获得完整的行列式技巧, 所以这里只有 I。

事实上只需要学习三种技巧:

1. 完全展开 [定义]

作用: 在行列式有很高对称性或大量位置为 0[稀疏] 的时候使用, 或某些情况下为了便于判断符号替代按行列展开。

举例: 第二章“应用小天地”例 2 [注意观察如何利用完全展开拆分], 或是对于行列变换性质的证明中应用的完全展开。

注意: 符号判断!

2. 按行列展开

作用: 在行列式稀疏或有一定自相似性时使用, 可用于化作低阶情况。

举例: 证明三角阵行列式为对角元乘积、习题 2.4-4 一类的三对角阵、2.4 节例 3 等。

* 本质上可以理解为某种动态规划的思想, 重复子问题不需要重复解决, 因此可简化。

注意: 还是符号判断, 尤其注意变换到代数余子式时的符号。

3. 行列变换

作用: 解决行列式的一般方法。

* 写成代码时复杂度为 $O(n^3)$, 比前两种理论复杂度 $O(n!)$ 的方法更普适

举例: 范德蒙德行列式的计算 (2.4 节)。

注意: 会有拆行、添行等较多变化, 见下一部分。

* 至于多行 Laplace 展开……个人觉得背都不用背, 总可以通过其他方法解决问题。

* 需要记住范德蒙德行列式的结果, 一些复杂问题可能会向这个方向化简。

1.3.2 例题

1. 习题 2.4-10

由于具有稀疏性, 考虑行列展开或完全展开。介绍两种思路:

- 利用按行列展开, 由于需要转化为低阶情况, 设其为 $f(a_1, \dots, a_{n-1})$, 按最后一行展开, 可发现 a_{n-1} 对应的代数余子式是 $f(a_1, \dots, a_{n-2})$, 而 1 对应的余子式最后一行只有一个 1, 按最后一行展开得到对角阵, 行列式 $a_1 \dots a_{n-2}$, 考虑 -1 的次数可知

$$f(a_1, \dots, a_{n-1}) = a_{n-1}f(a_1, \dots, a_{n-2}) + (-1)^{n+1}(-1)^{1+(n-1)}a_1a_2 \dots a_{n-2}$$

即

$$f(a_1, \dots, a_{n-1}) = a_{n-1}f(a_1, \dots, a_{n-2}) - a_1a_2 \dots a_{n-2}$$

不断展开最后一项 (可算几项找规律), 直到最后利用 $f(a_1) = a_1 - 1$, 可得到结果为

$$\prod_{i=1}^{n-1} a_i - \sum_{i=1}^{n-1} \prod_{j \neq i} a_j$$

- 利用完全展开, 第一行若取第一列, 则之后非零只能取 a_1, \dots, a_{n-1} , 得到 $a_1 \dots a_{n-1}$; 第一行若取第 i 列, $i > 1$, 考虑第 i 行可发现第 i 行只能取第 1 列, 其余每行 ($j \neq i$ 时) 只能取对角元, 由于实际上 i 和 1 进行了一次交换, 可发现应乘 -1 , 这样即为所有 $a_j, j \neq i$ 相乘后增添负号, 由此结果为

$$\prod_{i=1}^{n-1} a_i - \sum_{i=1}^{n-1} \prod_{j \neq i} a_j$$

2. 计算行列式 D , 其中第 i 行第 j 列的元素 $d_{ij} = \begin{cases} 1 + x_i y_j & i = j \\ x_i y_j & e_j \end{cases}$ 。

由于不稀疏, 一定需要行列变换解决, 然而, 若直接进行行列变换, 需要讨论 x_i, y_j 是否为 0 才能合理消去, 由此需要寻找更好的办法, 也介绍两种思路:

- 采用书上的拆行思路, 记 I 为 $a_{ii} = 1$, 其余为 0 的矩阵, 则其行列式为 1, 原行列式即为

$$I + P, \quad p_{ij} = x_i y_j$$

将第一行拆分为 I 的第一行与 P 的第一行得到两个行列式, 再将每个的第二行拆分为两个……以此拆分得到原行列式成为 2^n 个行列式的和, 其每行为 I 的行或 P 的行。

若选取了超过 1 行 P 的行, 则其中有两行为 $x_{i_1}(y_1, y_2, \dots, y_n)$ 与 $x_{i_2}(y_1, y_2, \dots, y_n)$, 若 $x_{i-1} = 0$ 则有一行为 0, 行列式为 0, 否则可以消去 $x_{i_2}(y_1, y_2, \dots, y_n)$, 行列式仍为 0, 由此只需考虑全为 I 的行的行列式 (即为 I , 行列式为 1) 与恰选取了一行 P 的行的行列式, 原行列式为这 $n+1$ 个行列式的和。

假设选取的是 P 的第 k 行, 不断按照第一行展开 (每次乘 1, 不改变行列式) 直到 P 的第 k 行成为第一行, 这时可发现行列式成为了下三角, 其对角元除了第一个对角元为 $x_k y_k$ 以外其他均为 1, 由此行列式为 $x_k y_k$ 。

综上可知结果为 $1 + \sum_k x_k y_k$ 。

- 注意到 $x_i y_j$ 相对容易消去, 利用添行的思路, 考虑右侧按第一列展开可知

$$\begin{vmatrix} 1 + x_1 y_1 & \cdots & x_1 y_n \\ \vdots & \ddots & \vdots \\ x_n y_1 & \cdots & 1 + x_n y_n \end{vmatrix} = \begin{vmatrix} 1 & y_1 & \cdots & y_n \\ 0 & 1 + x_1 y_1 & \cdots & x_1 y_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_n y_1 & \cdots & 1 + x_n y_n \end{vmatrix}$$

而每行减第一行 x_i 倍可知

$$\begin{vmatrix} 1 & y_1 & \cdots & y_n \\ 0 & 1 + x_1 y_1 & \cdots & x_1 y_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_n y_1 & \cdots & 1 + x_n y_n \end{vmatrix} = \begin{vmatrix} 1 & y_1 & \cdots & y_n \\ -x_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -x_n & 0 & \cdots & 1 \end{vmatrix}$$

类似习题 2.4-10 完全展开或行列展开即可得到行列式为 $1 + \sum_k x_k y_k$ 。

3. 习题 2.4-7

注意到每一行和前一行只相差两个元素, 第一行减第二行 1 倍、第二行减第三行 1 倍…… 可得到 (空白部分为 0)

$$D_n(x, y, z) = \begin{vmatrix} x - z & y - x & & & \\ & x - z & y - x & & \\ & & \ddots & \ddots & \\ & & & x - z & y - x \\ z & z & \cdots & z & x \end{vmatrix}$$

这回到了稀疏的情况, 考虑行列展开简化。对第一行展开, 则 $x - z$ 的代数余子式即为 $D_{n-1}(x, y, z)$, 而 $y - x$ 的代数余子式第一列只有 z , 按第一列展开后成为对角元均为 $y - x$ 的三角阵, 考虑 -1 的次方可知

$$D_n(x, y, z) = (x - z)D_{n-1}(x, y, z) + (-1)^{n+1}z(y - x)^{n-1} = (x - z)D_{n-1}(x, y, z) + z(x - y)^{n-1}$$

遇到这类递推时, 若无法直接看出求解思路, 可以算前几项找规律, 一般做法是两侧配成相似形式成为等比, 待定系数

$$D_n(x, y, z) + a(x - y)^n = b(D_{n-1}(x, y, z) + a(x - y)^{n-1})$$

即可得到

$$D_n(x, y, z) + \frac{z}{y - z}(x - y)^n = (x - z) \left(D_{n-1}(x, y, z) + \frac{z}{y - z}(x - y)^{n-1} \right)$$

由此可知

$$D_n(x, y, z) = \frac{y}{y - z}(x - z)^n - \frac{z}{y - z}(x - y)^n$$

4. [整体思想的应用] 设 $|A(t)|$ 为 A 的所有元素加 t 形成的行列式, 求证 $A(t)$ 的代数余子式之和与 A 相同。

由 2.4 节例 8

$$|A(t)| = |A| + t \sum_{i,j} A_{ij}$$

由此

$$|A(s + t)| = |A| + (s + t) \sum_{i,j} A_{ij}$$

另一方面将其看成 $A(t)$ 每个元素加 s 可知

$$|A(s + t)| = |A(t)| + s \sum_{i,j} A(t)_{ij} = |A| + t \sum_{i,j} A_{ij} + s \sum_{i,j} A(t)_{ij}$$

两式相减即得

$$s \sum_{i,j} A_{ij} = s \sum_{i,j} A(t)_{ij}$$

令 $s = 1$ 得证。

* 这里的做法即为矩阵看作整体的思想，不再纠结具体的细节元素行列变换，而是通过整体的性质来证明。这是一种代数的思想。

* 为什么能想到 $s + t$ ？因为 $A(t)$ 是 t 的线性函数，这本质上也是线性性的一种应用。

§1.4 为什么是线性代数

1.4.1 从线性方程组到统计学习

* 这段只是作为拓展说一说线性代数的应用。

线性代数对于现代计算机非常重要的核心原因在于，计算机本质上只能处理**有限精度的有限个数据**，而将它们排成向量或矩阵成为了一个非常自然的方式。

我们还是从最基本的线性方程组求解问题开始，虽然目前尚未学过矩阵乘法，不过我们姑且把增广矩阵为 (A, b) 的线性方程组写为

$$Ax = b$$

的形式。

注意到，在计算机中，无论是对 A 、 b 还是 x 的存储都只能达到有限精度，于是会存在一定误差，我们事实上想要的是误差允许范围内的解，也即

$$Ax = b + \varepsilon$$

且向量 ε 在某种意义下充分小，不妨考虑其**二范数**

$$\|\varepsilon\| = \sqrt{\varepsilon_1^2 + \cdots + \varepsilon_m^2}$$

利用此记号，我们可以将误差允许范围内的解看作，给定某 δ ，寻找 x 使得

$$\|Ax - b\| < \delta$$

当精度要求较低时，如何**更快**找到一个解就成为了关键的问题，这时，我们可以将原问题看作求 $\|Ax - b\|$ 的最小值问题，进行一定的迭代已快速减小其值（常用如梯度下降法、牛顿迭代法）。

至此，在计算机中，比起直接求解，事实上我们时常更倾向于解决

$$\min_x \|Ax - b\|$$

尤其是当 A 与 b 的规模 [行、列数] 极大时。

——那么，什么情况会面对规模极大的 A, b 呢？

考虑如下的情形：已知每个人的身高、体重、生理性别与年龄信息，希望找到体重与其他三者的关系，从而能对体重进行预测。

如果我们假设体重与其他要素成某个线性函数，并进行了大量统计 [N 个人]，我们最终得到的就是一个 N 行 4 列的矩阵 A ，与体重对应的列向量 b 。

* 4 列除了三个要素对应的 3 列外，还有一列全为 1，代表线性函数的截距。

很显然，上述方程一般是无解的，但利用 $\min_x \|Ax - b\|$ 的形式，我们事实上可以找到**相对最优的结果**！

这个做法称为**最小二乘法**。相对最优的结果有没有意义呢？答案是肯定的。若真实关系为 $Ax = b + e$ ，其中 e 为某个各个分量独立同正态分布的随机误差，则利用概率论知识可以计算得到，这样的最小二乘法找到的解的期望即为真实解。

由此，我们得到了一个非常基本的线性回归模型：假定结果 b 与 A 的每列呈线性关系，找到最优的系数。

如果并非线性关系呢？

若体重是其他三者的二次函数，事实上问题仍然可以化为某种线性回归，只是 A 从 4 列变为了 10 列： $1, x, y, z, xy, yz, xz, x^2, y^2, z^2$ 。根据分析中的魏尔斯特拉斯逼近定理，任何闭区间上的连续函数可以通过多项式进行逼近，而现实中要素的取值范围自然可以视为闭区间，现实中的关系函数也一般连续，由此我们得到：理论来说，通过取足够多的变量，即使是简单线性回归也能得到充分逼近的结果。

* 事实上，优化中最常用的方法之一，梯度下降法，就是将一般的方程局部近似为线性方程组以求解，因为线性方程组事实上几乎是最容易处理的结构。

这里的变量一般称为参数，而机器学习的一个基本思想即为，通过足够多参数的线性模型与一些非线性函数复合，可以拟合出任何连续函数。在这种思想下，通过不断扩充参数空间、放大模型，并加以一定的优化调整，方才得到了如今的 AI 技术。

1.4.2 线性代数学习指南

所以，如何学习线性代数？

- 从具体入手走向抽象

顾名思义，线性代数具有两个不同的视角：它是一门代数，因此势必有着抽象的结构与脱离直观的语言描述；但与此同时，它研究的是线性结构 [尤其是有限维] 这样最简单的结构，因此有着很多可以划归到向量、矩阵的计算的内容。

从代数的角度看，要学习线性代数必须力图掌握从抽象中描述的技巧。例如，如何通过向量空间的定义去推出向量空间的交与和还是向量空间？如何将矩阵看成整体 (而非逐个元素运算) 从 $A^2 = A$ 得到 A 可以写为 $P^{-1}\text{diag}(I, O)P$ 的形式？所谓的抽象，无非是，在最少的条件下去得到尽量多的结论。

* 当然，给出具体的数字在代数中是某种意义上最多的条件。

但是，在并未建立对代数的感觉时，直接去接触抽象概念是不可能有好的结果的，因此，必须回到具体的例子去计算、验证。

例如，当一个证明中一些以符号表示的过程并不清晰时，想要自己看懂，最基本的办法就是写一些具体的例子代入这个证明，去看看在具体的例子下是如何操作的。

- 不要排斥初等计算

正如上一部分所说，需要从具体的计算入手，但具体例子的计算恰恰是学习过程中容易被忽略的部分。

线性代数的很多证明都具有显著的操作性，例如学到用初等方阵描述行列变换后，证明的书写经常出现大量的矩阵描述。但是，由于其本质仍然是行列变换，必须回到如何操作行列变换的角度，而这又是在之前行列式的计算中训练的内容。

正是因为操作性的内容多，初等的计算反而变得非常重要，很多证明的思路也是从初等计算中得到的。虽然 Mathematica 等软件可以快速给出计算的结果，但发现计算过程中的不同特征，从而给出更一般的方法或操作，也是线性代数重要的一环。

- 不要妄想记住一切

最后，我们来聊聊怎样的模式能更容易想到证明的思路 [这里的证明是广义的，例如 n 阶行列式的计算也可视为某种证明题]。

由于线性代数的小结论非常之多，将其全部记忆是不可能的事。因此，虽然有的题目可以利用某个小结论很快就做出来，靠着翻书找到这个结论并顺理成章做出也不能代表学会了证明。与之相对，真正学会证明必然是，从自己能够熟练掌握的结论（不需要翻书也会用）出发，靠着分析和推导，得到最终结果。

在上述的过程里，很可能实际上证出来了那个并不熟悉的小结论，也可能绕了很大的弯子将看起来简单的题目做复杂，但只有这样，才算是自己能学会应用。

因此，学习过程中最关键的，不是记住一切小结论，而是有一些足够熟悉的小结论，并能用它们推出需要的体系。例如，拆行和添行技巧事实上很可能等价，只要能熟练某一种就足以做出行列变换计算题（虽然用另一种可能更简单）；又例如，秩的部分事实上将子矩阵定义、行列秩定义与相抵标准形记住，其他结论都可以较快推出。

当然，究竟哪些小结论是更本质的、自己更擅长用什么解决问题，都是不得不通过大量练习去完善的——对作业相对少的丁老师班来说，至少需要课后作业两倍量级的题目，才能练出属于自己的技巧。

从更之后的角度来说，学习任何数学课最终都是在培养直觉。即使后续已经忘记了具体的技巧，对于某个结构大概可以通过怎样的办法来处理的直觉也是一直保留的。

在课程学习中，直觉其实就对应着第一反应的做法可不可行。个人观点是，第一反应一定是重要的，如果它可行，就顺着它做完，如果做复杂了则应该思考为何能想到更简单的方案；如果它不可行，试着找到它不可行的根本原因，并且加以调整。

以上过程实际上是一种对直觉的训练，但有一个更本质的问题——自己没有做下去，且答案完全是另一种思路，并不意味着思路一定不可行。这时，需要通过讨论 [或问助教/老师] 来确定实际的可行性。在不断调整自己的直觉、以更快速度反应出正确思路后，方能将这种“印象”带到之后的类似情况中。

二 空间视角的线性方程组

§2.1 习题解答

1. 习题 2.4-4

对第一行展开，可发现 $2a$ 对应的代数余子式即为 D_{n-1} ，而 a^2 对应的代数余子式对第一列（此时这列只有一个非零元素）展开即成为 $-D_{n-2}$ ，由此可写出递推

$$D_n = 2aD_{n-1} - a^2D_{n-2}$$

此即为

$$D_n - aD_{n-1} = a(D_{n-1} - aD_{n-2})$$

于是利用 $D_1 = 2a, D_2 = 3a^2$ 可归纳证明 $D_n = (n+1)a^n$ 。

2. 习题 2.4-13

第二行减去第一行一倍、第三行减去第二行一倍……最后一行减去倒数第二行一倍，可将行列式化为 Vandermonde 行列式，从而结果即为

$$\prod_{i>j} (x_i - x_j)$$

3. 习题 2.5-4

齐次方程组一定有零解，由此有非零解等价于有无穷多解，根据 Cramer 法则，这当且仅当系数行列式为 0 时成立。

直接计算可知系数行列式为 $b - ab$ ，由此解为 $a = 1$ 或 $b = 0$ 。

4. 补充题二-3

其完全展开共有六项，且三项带正号的为 0 或 1、三项带负号的为 0 或 -1，因此理论最大值为 3。若三项带正号的全为 1，则所有元素只能均为 1，但此时行列式为 0，矛盾。

若两项带正号的为 1，可发现存在三项带负号的全为 0 的行列式，如

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

此时行列式为 2，即取到最大值。

5. 习题 3.1-2

直接计算知为 $(-21, 7, 15, 13)$ 。

6. 习题 3.1-4

将线性表出的系数写为线性方程组，利用 Cramer 法则即可直接计算出解为

$$\alpha = (a_1 - a_2)\alpha_1 + (a_2 - a_3)\alpha_2 + (a_3 - a_4)\alpha_3 + a_4\alpha_4$$

7. 习题 3.2-7

若 $\lambda_1(a_1\alpha_1 + b_2\alpha_2) + \lambda_2(a_2\alpha_2 + b_3\alpha_3) + \lambda_3(a_3\alpha_3 + b_1\alpha_1) = 0$ ，整理可得

$$(\lambda_1 a_1 + \lambda_3 b_1)\alpha_1 + (\lambda_2 a_2 + \lambda_1 b_2)\alpha_2 + (\lambda_3 a_3 + \lambda_2 b_3)\alpha_3 = 0$$

利用线性无关性，此方程为 0 当且仅当

$$\lambda_1 a_1 + \lambda_3 b_1 = \lambda_2 a_2 + \lambda_1 b_2 = \lambda_3 a_3 + \lambda_2 b_3 = 0$$

利用齐次方程组的 Cramer 法则（类似习题 2.5-4），此方程组只有零解当且仅当系数行列式非 0，而系数行列式为 $a_1 a_2 a_3 + b_1 b_2 b_3$ ，由此得证。

8. 习题 3.3-5

* 说明极大线性无关组相关问题时，由于其不唯一，一定要说清楚是取出某一个还是对任何一个。

考虑 $\alpha_1, \dots, \alpha_s$ 的一个极大线性无关组，不妨设为 $\{\alpha_1, \dots, \alpha_m\}$ ；考虑 β_1, \dots, β_r 的一个极大线性无关组，不妨设为 $\{\beta_1, \dots, \beta_n\}$ 。由此，等式右侧即为 $m+n$ ，将两极大线性无关组并集设为 A 。

由于 $\alpha_1, \dots, \alpha_m$ 可以表出 $\alpha_1, \dots, \alpha_s$ 、 β_1, \dots, β_n 可以表出 β_1, \dots, β_r ，根据线性表出定义即知 A 可以表出 $\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_r$ ，由此方程组左侧不超过 $\text{rank } A$ ，但 $\text{rank } A$ 不超过 A 中向量个数个数 $m+n$ ，即得证。

9. 习题 3.3-8

由条件已知 $\{\beta_i\}$ 可由 $\{\alpha_i\}$ 表出，另一方面有

$$\alpha_i = \frac{1}{m-1} \sum_{k=1}^m \beta_k - \beta_i$$

由此 $\{\alpha_i\}$ 也可被 $\{\beta_i\}$ 表出，从而秩相等。

10. 习题 3.4-4

直接求解线性方程组得到

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \left(\frac{2}{9}a_1 + \frac{1}{9}a_2 + \frac{2}{9}a_3 \right) \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} + \left(\frac{1}{9}a_1 + \frac{2}{9}a_2 - \frac{2}{9}a_3 \right) \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} + \left(-\frac{2}{9}a_1 + \frac{2}{9}a_2 + \frac{1}{9}a_3 \right) \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$$

11. 习题 3.5-4

通过列向量极大线性无关组计算秩可得秩为 3, 前三列构成列空间一组基。

利用行秩列秩相等可知行空间维数亦为 3。

12. 习题 3.5-11

利用秩的等价定义, 考虑 A 中的 $\text{rank } A$ 阶非零子式 A_1 与 B 中的 $\text{rank } B$ 阶非零子式 B_1 , 将 A_1 、 B_1 所在行、列构成的子式称为 X , 可发现其有

$$X = \begin{pmatrix} A_1 & C_1 \\ O & B_1 \end{pmatrix}$$

的形式, 再利用 2.6 节推论 1 可知 $\det X = \det(A_1) \det(B_1) \neq 0$, 这就找到了左侧矩阵的 $\text{rank } A + \text{rank } B$ 阶非零子式, 即得证。

* 也有行列变换与空间的思路, 将在之后介绍。

§2.2 如何自学?

2.2.1 四个层面

“自学”, 算是一个迟早会遇到的问题: 在不能保证所有课都能上课跟上、不能保证所有老师都讲得好的情况下, 总有一些课程是必须要通过自学解决的。

一般来说, 不同的课会有不同的思路和模式, 下述的方法主要针对数学课应该如何自学, 但对其他理科类的科目也有一定的参考价值。

对于一门数学课来说, 所谓的学会这件事事实上可以分为四个层面:

1. “知道”

所谓的知道有一个非常简单的标准: 能看懂题目。也即, 对所有的概念有基本的了解。

一般来说, 无论是上课还是自己看了一遍书, 只要没有太多的前置空缺——对前两年来说这基本是不可能的——都应该能对书上的概念与定理达到知道的层次。

不过, 这个层次距离自己实际操作还有很远。在了解书上大概讲了一些什么之后, 接下来就是细看书中的定理与例题了, 而这又分为两个部分。

2. “细节”

对证明的细节理解, 也即看懂每一步, 是建立更深理解的基础。这里所谓的看懂, 是每步之间的变化能看出是利用了什么结论, 或者有一些需要自己补全的简单步骤。

如果上课时能跟随老师的进度看懂每一步, 之后自己再看时完成接下来的层面是相对快的, 否则, 必须看书自己研究。这是一个相对繁琐但又必要的过程——如果没有搞清楚细节, 对全局的掌握只能停留在概念的层次, 无法做到数学的严谨。

3. “全局”

然而, 虽然细节理解是全局理解的基础, 全局理解事实上是比细节理解更为重要的。

全局理解, 指的是证明架构的拆分: 如何把一个问题分解为若干个小问题, 又如何把每个小问题分解成更小的问题, 最终选择合适的方法解决。如果说细节理解是以顺序结构读懂证明, 全局理解就是以树状结构明白一个证明的真正架构——对于困难的问题, 一般的思考方式都不可能以顺序写出一个证明, 而是先拆分解决了问题, 再整理成线性的形式。

举例: 拆分矩阵行秩等于列秩的证明 (基于我自己的感受)。

- 寻找核心问题：是什么让我自己写不出来这个证明？
- 发现：矩阵的行与列对应的向量组都不在同一个空间中，无法进行比较。
- 在证明中寻找：解决核心问题的方法是什么？
- 答案：在简化阶梯形矩阵中行秩、列秩都是易于确定的，先对这些矩阵确定相等，再证明任何矩阵可以在不改变行秩、列秩时化为它们即可。
- 对答案的抽象：仍然是利用标准形思想。
- 于是证明拆分为了三步：标准形中结论成立、任何矩阵可以通过一些方式化为标准形、化为标准形的过程不改变结论。
- 进一步划分为以下三块：
 - 简化阶梯形矩阵中行秩等于列秩；
 - 任何矩阵可以通过初等行变换化为简化阶梯形矩阵；
 - 初等行变换不改变行秩、列秩。
- 注意到，第二块直接由我们在第一章中证明的定理即可得到，由此只需解决剩下两块。
- 对第一块，新的核心问题是，如何计算行秩与列秩？
- 解答：只要说明简化阶梯形矩阵主元所在的行/列构成行/列向量组的极大线性无关组，即可通过定义得到秩，而前者也可以通过定义证明。
- 对第三块，它自然划分为不改变行秩与不改变列秩两部分，第一部分利用向量组等价可得到，第二部分利用线性相关/线性无关的定义可得到。[当然，若觉得这两部分的证明比较困难，可以进一步分解。]

通过上述的过程，我们完成了一个证明的拆分。拆分后，我们事实上可以明白这个证明是如何得到的。其中，越是在拆分后显得自然的做法，越代表它是直觉性的，而越是不自然则越偏向技巧性。对重要的技巧性内容，一方面需要在题目中不断训练以掌握，另一方面则是要尽量为它合理化，最终的目标仍然是纳入直觉。

在数学的学习中，“为何要引入这样的定义”与“为何能得到这样的证明”是非常重要的。因为所有概念都是在抽象中建构，引入新的定义、证明新的性质必然意味着它们具有更普遍的价值。找到能说服自己的理由后，这个定义才能进入自己构建的体系。

4. “应用”

最后，学习的一切知识都是为了解决问题，因此还是需要回到应用。

写出一道题的证明和之前看证明的过程本质上是相同的：先想办法通过寻找核心困难的方式将问题拆分出不同模块，再用自己学过的知识解决每个模块，从而合成整个问题的证明。

虽然对应用的训练必须靠做题，但只要看书时对细节和全局的理解充分，其实并不需要太大量练习就可以掌握对应的技巧。与之相对，若之前的看书理解不足，就必须靠更多题目进行理解，这样做即使总体上也能达到相同的效果，也会浪费更多的时间成本。

——事实上，“全局”和“细节”两部分是会互相影响的。若全局大致理解但处理细节的能力不足，很可能在已经拆分正确的情况下因为细节无法解决而怀疑自己拆分有误，绕更远的路；而若只有细节理解，没有全局的感受（也就是所谓的高观点），哪怕能解决很多细碎的问题，当题目的步骤变多、变复杂时，也容易摸不着头脑。

说回上课的话题，一般来说，靠课堂能建立“知道”，并理解大部分的“细节”，已经是学得不错的情况了。但是，即使在这种情况下，对全局的掌握仍然是缺乏的。知道在做什么而不知道为什么这样做，最终还是需要通过课下自己看书来补全——或者通过这份习题课讲义来了解。

因此，若是确实去听课，一定要注意，不要过度关注细节，而是应尽可能建立整体理解，因为后者才是靠自己比听课效率相对低的部分。反过来说，因为上课一定是按照老师的思路与节奏进行，若是无法跟着老师建立理解，也是非常正常的情况，不必过度担心。只要靠着自己看书按照之前所说的四个层面解决，终究还是能学会的。

——至少就这门课而言，自学不是太过困难的，按照我们班的讲课速度，也推荐大家多去自己研究、建立自己的理解。

2.2.2 关于证明书写

这个小块其实本没有准备在这次习题课讲，但因为看到了几个同学的小测证明书写，觉得还是有必要提一下，因为到现在终于出现非计算性的证明题了。

证明书写有两个完全不同的要求：一个是让会这道题的人知道你懂（如改作业的助教能看出你是对的），另一个则是让不会这道题的人能靠你的过程看懂。

前者是最基本的要求，也是所有的证明都至少应该达到的要求。其实，只要避免出现“伪证”和“如证”，前者还是容易达到的。

伪证：指证明过程中利用了错误的结论以构成错误的推导。想避免伪证，一定记得检查自己的前一步到后一步是否有依据。依据可以是某个定理、某个结论，也可以是计算得到的，但无论是哪种，一般都需要有出处（除非是过于简单的推论）。若只是直觉对但找不到出处，最好还是进行一些验证。

如证：跳过了关键步骤的证明。这个问题的出现事实上和之前所说的全局拆分非常有关系，通过拆分，可以发现要证的题目中的核心问题，而证明中最需要的就是说清楚核心问题如何解决。如果对核心问题没有概念，很可能会造成不必要的细节叙述太多、关键问题上含混过去的“如证”场面——含混也未必是真的不会，可能只是觉得这里用某个结论十分自然，但在助教的视角里，这就是不会。想避免这种情况，除了每步都写得尽量严谨、不要跳步之外，最重要的仍然是发现核心问题。

想完成后者，首先需要对问题的拆分有更精细的把握，将证明合理地分解为不同部分，其次需要预设看到这个证明时合理的知识基础，不使用过强的结论。由于这个讲义面向的对象不是数学系学生，一般并不会有这么高的要求，但有些思路仍然值得借鉴，例如，将每一步的依据清楚写出来，把“所以”改为“由……可知”，可大幅避免前面所说的问题。

§2.3 行列式回顾与空间引入

2.3.1 Cramer 法则

由于这次习题课主要内容为复习第三章知识，必须找到一个合理的组织结构。我们不妨认为，第三章所做的一切仍然是为了解决线性方程组问题。

但是，在第一章已经给出了一般算法时，究竟为何还要再次解决呢？

先回到第二章学过的 Cramer 法则，如果不关心解具体为何，Cramer 法则可以表述为：方程数与未知数个数相同的线性方程组有唯一解当且仅当系数行列式非零。

这事实上代表着，哪怕不看右边的 b ，左端的 A 可以某种意义上决定解的个数。

这种观察事实上可以推广：考虑任何线性方程组化为简化阶梯形的过程，可以发现，最后一列在这个过程中并不会影响其他列，因此，假设解存在时，自由未知量的个数只与 A 有关。

自由未知量的个数能否刻画解的多少？答案是肯定的。考虑如下三个通解的形式：

$$\begin{cases} x_1 = 1 \\ x_2 = 0 \\ x_3 = 1 \end{cases}, \quad \begin{cases} x_1 = 1 - x_3 \\ x_2 = 2 + x_3 \end{cases}, \quad x_1 = x_2 + x_3 + 5$$

第一个通解中并不存在自由未知量，也就是所谓唯一解。在三维空间中，解的点集构成了一个点；第二个通解有一个自由未知量，因此对 x_3 的每个取值，恰有一个 x_1, x_2 的取值与之对应，在三维空间中构成一条线；第三个通解有两个自由未知量，对 x_2, x_3 的每个取值，有唯一的 x_1 取值与之对应，在三维空间中构成一个面。

我们不妨按照以往的直观给出如下的“定义”：线性方程组的解的集合 [下称解集] 的维数定义为自由未知量的个数。

由此根据之前的观察可以马上得到，解存在时解集维数只与 A 有关，因此可以将解集维数称为系数矩阵的“维数”。

那么，这个定义究竟是否严谨呢？

容易想到，若需要这个定义是严谨的，必须自由未知量的个数唯一确定。在简化阶梯形中，它等价于列数减去主元个数，或是非零的行数。但是，化为简化阶梯形的过程不止一种，在简化阶梯形矩阵里定义的自由未知量个数未必对一个矩阵能化成的所有简化阶梯形矩阵相同。

但从直观上，解集是由线性方程组唯一确定的，它是点、线或面理应是确定的，因此，我们可以猜测上述定义实际上是良好的，但是，必须推广定义到任何矩阵中，才能有机会证明这样的唯一性。

2.3.2 线性表出

注意到，在讨论解集时，由于不关心 b 的矩阵内容，我们可以将它作为一个整体的向量来看待。与之对应，我们也希望将左端拆分为向量，也即把线性方程组写成

$$b = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$$

这里 α_i 是 A 的第 i 列构成的向量，也即列向量。我们假设 α_i 与 b 都是 \mathbb{R}^m 中的向量

对于一个线性方程组，我们会关心解的存在性、唯一性，或是唯一性进行推广得到的解集结构问题——目前，我们可以把结构简单理解为自由未知量个数。

在之前的讨论里， b 无法确定时，我们并没有很好的刻画解的存在性的方法。由于给每个向量 α_i 乘系数 x_i 后再相加是线性的组合，我们可以直接将“存在性”定义为新的概念：若上述方程组的解存在，则称 b 能够被 $\alpha_1, \dots, \alpha_n$ 线性表出，将 $\alpha_1, \dots, \alpha_n$ 能线性表出的某个向量称为它们的一个线性组合。

接下来讨论唯一性，如果解不唯一，即代表存在不全相等的 x_i 与 y_i 使得

$$b = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$$

$$b = y_1\alpha_1 + y_2\alpha_2 + \cdots + y_n\alpha_n$$

作差得到

$$0 = (x_1 - y_1)\alpha_1 + (x_2 - y_2)\alpha_2 + \cdots + (x_n - y_n)\alpha_n$$

可以发现，记 $z_i = x_i - y_i$ ，只要关于 z_i 的方程组有不全为 0 的解，原方程组的解就不可能唯一：若其无解，自然不是唯一解，而只要有解，将其解加上一组不全为 0 的 z_i 能成为新的解。

因此，我们将唯一性概念定义为：若关于 z_i 的方程组

$$0 = z_1\alpha_1 + z_2\alpha_2 + \cdots + z_n\alpha_n$$

没有不全为 0 的解，则称 $\alpha_1, \dots, \alpha_n$ 线性无关，否则称为线性相关。

利用这几个新定义的概念，我们可以给上方问题一个简洁的描述：

- 若 b 不能被 $\alpha_1, \dots, \alpha_n$ 线性表出，则原方程组无解；
- 若 b 能被 $\alpha_1, \dots, \alpha_n$ 线性表出，且 $\alpha_1, \dots, \alpha_n$ 线性无关，则原方程组有唯一解；

- 若 b 能被 $\alpha_1, \dots, \alpha_n$ 线性表出, 且 $\alpha_1, \dots, \alpha_n$ 线性相关, 则原方程组有不止一个解。

* 第三种情况的不止一个解实际上是无穷多解: 假设原问题有解 x , $0 = z_1\alpha_1 + z_2\alpha_2 + \dots + z_n\alpha_n$ 有不全为零的解 z , 则对任何 $\lambda \in \mathbb{R}$, 可发现 $x + \lambda z$ 都是不同的解。

* 到现在, 或许大家会有一个疑问: 我们看起来什么也没干, 只是新定义了一些概念重新表述出问题。但是, 在新定义的概念下, 我们可以更好推进整个体系, 从而解决问题, 而接下来进行的推进就是不进行这些定义时很难说清楚的部分了。

注意到, 我们事实上已经将问题拆分成了两个部分: 和 b 有关系的 **存在性问题**、与和 b 没关系的**唯一性问题**。这样的拆分是符合我们之前的直觉性分析的: 只要知道解存在, 解的多少就和 b 没关系了。

不过, 由于这里的线性相关、线性无关定义只涉及了是否存唯一解, 我们仍然无法从定义感受到解的多少从何而来, 可以试着先从存在性问题开始寻找思路。

2.3.3 封闭性与空间

* 本章中所有的讨论都对 \mathbb{R} 上的线性方程组与矩阵进行, 其他数域上完全类似。这里所说的向量空间即为书上的 \mathbb{R}^n 或其线性子空间。

在数学中, 一个相对自然的想法是, 若我们要讨论一个问题解的存在性, 可以定义出所有使解存在的集合, 观察集合的性质, 再对特定的 b 研究是否存在。因此, 我们定义所有能被 $\alpha_1, \dots, \alpha_n$ 线性表出的 b 构成集合

$$V = \left\{ b \in \mathbb{R}^m \mid \exists x_1, \dots, x_n \in \mathbb{R}, \quad b = \sum_{i=1}^n x_i \alpha_i \right\}$$

我们希望能对这个集合的结构进行一定的刻画, 并给出一个可以确定某个 b 是否落在其中的更简单方式。

一个很重要的观察是, 虽然其中一般有无穷多个向量, 但其实只需要有限个向量就可以 $\alpha_1, \dots, \alpha_n$ 就可以完全刻画出它。这是因为空间中恰收录了所有 $\alpha_1, \dots, \alpha_n$ 可以表出的向量。我们试着直接将此作为 V 的特性定义:

若对 $V \subset \mathbb{R}^m$, 存在向量 $\alpha_1, \dots, \alpha_n \in \mathbb{R}^m$ 使得 (n 可以取 0, 此时其中只有零向量)

$$V = \left\{ b \in \mathbb{R}^m \mid \exists x_1, \dots, x_n \in \mathbb{R}, \quad b = \sum_{i=1}^n x_i \alpha_i \right\}$$

则称其为一个**向量空间**。

* 可喜可贺, 我们又说了一堆废话, 然后把无法研究的东西作为新的定义。不过, 化归到这一步, 我们必须研究向量空间的定义能不能表述为别的方式, 否则研究将无法进行。接下来, 我们一步步将向量空间的定义变为大家熟悉的样子。

由于向量空间的定义是以**线性表出** [或简称“表出”] 作为基础的, 我们希望知道能不能在其中谈论**线性表出问题**。所谓的“谈论”某个运算 [线性表出可视为一堆向量和数到另一个向量的运算], 指的是, 如果参与运算的元素都在其中, 运算结果也在其中, 也就是**封闭性**。对这个问题, 其实我们希望知道的即:

若 $\beta_1, \dots, \beta_k \in V$, $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, 是否有 $\sum_{i=1}^k \lambda_i \beta_i \in V$?

答案是肯定的, 证明过程是, 设

$$\beta_i = \sum_{j=1}^n x_{ij} \alpha_j$$

则通过乘法分配律可以得到 (若觉得这里求和符号的交换并不直观, 可以取定具体的 k, n 尝试操作)

$$\sum_{i=1}^k \lambda_i \beta_i = \sum_{i=1}^k \lambda_i \sum_{j=1}^n x_{ij} \alpha_j = \sum_{j=1}^n \left(\sum_{i=1}^k \lambda_i x_{ij} \right) \alpha_j \in V$$

反过来, 我们也希望知道, 可以谈论线性表出问题的结构是否一定是**向量空间**?

答案仍然是肯定的。若 V 满足，只要 $\beta_1, \dots, \beta_k \in V, \lambda_1, \dots, \lambda_k \in \mathbb{R}$ ，则 $\sum_{i=1}^k \lambda_i \beta_i \in V$ 。考虑如下过程：

假设 V 中只有零向量，则结论成立；否则，取出某非零向量 α_1 ，若其可以表出 V 中全部向量，则结论亦已经成立 [根据定义， α_1 能表出的向量一定在 V 中，而又假设了 V 中的向量都可以由 α_1 表出，于是 V 恰为所有 α_1 能表出的向量集合]；若不能表出全部向量，取出其无法表出的向量 α_2 ，若 α_1, α_2 可以表出 V 中全部向量则结论已经成立；若还是不能表出全部向量，再取它们不能表出的 α_3 ……

我们只需要证明，这个过程至多进行有限次，就一定在某次后可以找到 $\alpha_1, \dots, \alpha_n$ 表出全空间，而它们所有能表出的向量又在空间中，因此 V 即为 $\alpha_1, \dots, \alpha_n$ 能表出的所有向量。

下面证明，在 \mathbb{R}^m 中，最多能找到 m 个向量，使得每一个不能被前面的向量表出。

先证明引理：未知数个数多于方程个数的齐次线性方程组必有非零解。考虑其化为简化阶梯形后，主元至多等于方程个数，于是一定存在自由变量。而又由于齐次性，全为 0 必然为解，因此其解存在，从而有非零解。

于是，若能找到 $m+1$ 个向量 $\alpha_1, \dots, \alpha_{m+1}$ ，根据刚才讨论考虑齐次线性方程组

$$x_1 \alpha_1 + x_2 \alpha_2 + \dots + x_{m+1} \alpha_{m+1} = 0$$

其有 $m+1$ 个未知数、 m 个方程，必然存在 x_i 不全为 0 的解。考虑非零的最大下标 N ，即 $x_N \neq 0$ ，而 $x_{N+1}, x_{N+2}, \dots, x_{m+1} = 0$ ，将 $x_N \alpha_N$ 移项到右侧并同除以 $-x_N$ ，可发现 α_N 能被 $\alpha_1, \dots, \alpha_{N-1}$ 表出 ($\alpha_{N+1}, \dots, \alpha_{m+1}$ 前的系数均为 0)，与取法中每一个向量无法被前面的向量表出矛盾。

到这里，我们通过一些复杂的操作证明了 \mathbb{R}^m 子集 V 是向量空间当且仅当其中的任何向量线性组合后仍在其中，这个定义已经与教材上的定义非常接近了。

* 在习题课中事实上默认了两者的等价性，这段证明算是作为一个严谨性的补充。

最后，若任何向量线性组合仍在其中，两个向量的加法与一个向量的数乘是线性组合的特例，因此必然在其中。而只要保证了两个其中向量的加法与一个向量的数乘在其中，对一些 $\alpha_1, \dots, \alpha_n \in V$ ，利用数乘封闭性可知

$$x_1 \alpha_1, \dots, x_n \alpha_n \in V$$

再利用加法封闭性可以归纳得到

$$x_1 \alpha_1 + \dots + x_n \alpha_n \in V$$

于是只通过加法和数乘的封闭性，我们就得到了线性组合的封闭性。

最终的结论即 \mathbb{R}^m 子集 V 满足其中的任何向量线性组合后仍在其中等价于其中任何两向量相加、任何向量数乘仍在其中。这就是书上对于向量空间的定义。

现在，是时候回顾本节所做的事了。我们从研究线性方程组解的存在性出发，定义 $V \subset \mathbb{R}^m$ 为向量空间，当且仅当

$$\exists \alpha_1, \dots, \alpha_n \in \mathbb{R}^m, \quad V = \left\{ b \in \mathbb{R}^m \mid \exists x_1, \dots, x_n \in \mathbb{R}, \quad b = \sum_{i=1}^n x_i \alpha_i \right\}$$

由于这个定义过于复杂，我们试着找到更简洁的表述，最终得到的表述为

$$\forall \alpha, \beta \in V, \quad \alpha + \beta \in V$$

$$\forall \alpha \in V, \lambda \in \mathbb{R}, \quad \lambda \alpha \in V$$

后者的确比前者简单很多，且易于验证，但就研究线性方程组而言，直接定义后者并没有清晰的意义，而前者的含义是明确的，即给定系数矩阵 A ，所有使得线性方程组有解的 b 构成的集合。

若将后者作为向量空间的定义，我们事实上得到了：任何向量空间可以被有限个向量线性表出，更进一步地， \mathbb{R}^m 中的向量空间可以被至多 m 个向量表出；任何有限个向量线性组合而成的所有向量构成向量空间。

此外，我们证明过程里其实得到了若干附属结论，这里仅作部分列举，大家可以看看如何从证明过程里直接或间接得到，接下来我们将直接引用这些结论：

1. \mathbb{R}^m 中无法取出 $m+1$ 个线性无关的向量；
2. 若向量 $\alpha_1, \dots, \alpha_k$ 线性相关，则一定存在其中一个能被它之前的向量线性表出，反之，若存在其中一个能被它之前的向量表出，移项可知线性相关；
3. 向量组 $\{\gamma_i\}$ 能被向量组 $\{\beta_i\}$ 线性表出、向量组 $\{\beta_i\}$ 能被向量组 $\{\alpha_i\}$ 线性表出，则向量组 $\{\gamma_i\}$ 能被向量组 $\{\alpha_i\}$ 线性表出；
4. 任何向量空间中线性无关的向量组可以扩充为向量空间的一组基。（若在看讲义时并未学完书，可以暂时忽略这个结论，我们之后会再次提及。）

§2.4 秩

2.4.1 信息量的视角

下面称

$$V = \left\{ b \in \mathbb{R}^m \mid \exists x_1, \dots, x_n \in \mathbb{R}, \quad b = \sum_{i=1}^n x_i \alpha_i \right\}$$

为 $\alpha_1, \dots, \alpha_n$ 生成的向量空间。

根据刚才的证明过程，我们可以发现一个有趣的事：即使 $n > m$ ，我们仍然可以在 V 中找到 β_1, \dots, β_k ，使得 $k \leq m$ ，且 V 是 β_1, \dots, β_k 生成的向量空间。

这也就意味着，虽然在原始定义里 x_1, \dots, x_n 看似都是自由变量，实际却可以取出更少的自由变量对应同一个空间，也即自由变量未必是真实的。

因此，为了如开始一样研究空间的“维数”作为结构表征，我们无法直接把 n 个向量生成的空间维数定义为 n ，而是应该如此定义：

对向量空间 V ，若存在 β_1, \dots, β_k 使得 V 由 β_1, \dots, β_k 生成，且不存在 $\gamma_1, \dots, \gamma_{k-1}$ 使得 V 由 $\gamma_1, \dots, \gamma_{k-1}$ 生成，则称其维数为 k ，记作 $\dim V = k$ 。

此定义事实上即是寻找最少的自由变量个数作为维数。因为一定存在有限个向量能表出 V 中的全部向量，这个最小值当然存在，且我们知道对 \mathbb{R}^m 的子集 V ，其维数不会超过 m 。

* 某种意义上，维数其实代表着信息量，因为其至少能被 k 个向量刻画，则代表着存储它至少需要存储 k 个向量。

* 非常自然地，若 U 是向量空间 V 的子集，且 U 是向量空间，可称 U 是 V 的子空间。可验证 \mathbb{R}^m 是向量空间，以后即可称 V 是 \mathbb{R}^m 的子空间。

可是，问题并没有在这里结束，我们虽然知道这个最小值一定存在，但如何找到它呢？例如，用之前的算法可以找到一组向量 $\{\beta_i\}$ 生成 V ，但 $\{\beta_i\}$ 的个数一定是 V 的维数吗？是否能够更小？

另一方面，如果给出 $\alpha_1, \dots, \alpha_n$ ，能不能从这些向量自身中找到能够刻画向量空间维数的方式，而不是从头去寻找 $\{\beta_i\}$ ？

2.4.2 从空间到向量组

为了找到合适的刻画方式，我们可以先做一个简单的观察：只要 β_1, \dots, β_r 能够线性表出 $\alpha_1, \dots, \alpha_n$ ，它们就能线性表出整个 V 。

* 利用 2.3.3 最后的结论 3 即可证明。

于是, 我们其实将表出向量空间所有向量的问题变为了表出有限个向量的问题, 显然后者是更加简单的。

更进一步地, 我们进行一个大胆的推测: 我们希望能从 $\alpha_1, \dots, \alpha_n$ 中选出一些, 使得它们能表出 $\alpha_1, \dots, \alpha_n$, 并且, 若从中最少需要选出 r 个, 则任何能表出 $\alpha_1, \dots, \alpha_n$ 的向量组都至少有 r 个向量。

* 这个推测不管从什么角度来说都是非常大胆的, 因为我们将 V 中任取局限到了这 n 个向量中选择。不过, 如果进行一些具体例子的实验, 会发现它存在一定的道理。当然, 之后的部分我们会证明这个推测。

由此, 我们试图先解决如下的问题: 给定向量组 $\alpha_1, \dots, \alpha_n$, 希望从中选出 r 个, 使得它们能表出 $\alpha_1, \dots, \alpha_n$, 且 r 最小。我们不妨设 $\alpha_1, \dots, \alpha_r$ 满足上述的要求, 来观察一下它们的性质:

- 由于 r 最小, 无法从其中去掉一个使得仍然能表出 $\alpha_1, \dots, \alpha_n$ 。仍利用 2.3.3 最后的结论 3, 若去掉一个能表出 $\alpha_1, \dots, \alpha_r$, 则必然能表出 $\alpha_1, \dots, \alpha_n$ 。

因此, r 的最小性可以推出, α_1 到 α_r 中不存在 $r-1$ 个向量能表出剩下那个, 而利用 2.3.3 最后的结论 2, 这即能推出线性无关。

于是, $\alpha_1, \dots, \alpha_r$ 必须是线性无关的。

- 由于它们能表出 $\alpha_1, \dots, \alpha_n$, 任意在其中增添一个向量, 都能被 $\alpha_1, \dots, \alpha_r$ 所表出, 由此利用 2.3.3 最后的结论 2, $\alpha_1, \dots, \alpha_r$ 增添任何向量都将线性相关, 于是线性无关性是极大的。

我们将 $\alpha_1, \dots, \alpha_n$ 中线性无关、且任意增添一个就线性相关的向量组称为极大线性无关组。刚才说明了, 满足上述性质的向量组一定是极大线性无关组, 而接下来说明, 极大线性无关组一定满足上述性质。仍然设极大线性无关组为 $\alpha_1, \dots, \alpha_r$ 。

- 由于线性无关性是极大的, 任意添加一个向量 $\alpha_j, j > r$, 都存在不全为 0 的 λ_i 使得

$$\lambda_1\alpha_1 + \lambda_2\alpha_2 + \dots + \lambda_r\alpha_r + \lambda_j\alpha_j = 0$$

若 $\lambda_j = 0$, 则意味着 $\alpha_1, \dots, \alpha_r$ 线性相关, 矛盾, 于是可以移项使得 α_j 能被 $\alpha_1, \dots, \alpha_r$ 表出。另一方面, $\alpha_i, i \leq r$ 当然可以由 $\alpha_1, \dots, \alpha_r$ 表出 ($i = k$ 时系数为 1, 否则为 0), 因此得证 $\alpha_1, \dots, \alpha_r$ 能表出 $\alpha_1, \dots, \alpha_n$ 。

- 为了说明 r 最小, 我们只需要证明所有极大线性无关组的元素个数相同即可。这样, 由于取到最小 r 的一定是某个极大线性无关组, 所有的便都能取到最小。

先证明引理: 若向量 $\alpha_1, \dots, \alpha_r$ 能表出线性无关的向量 β_1, \dots, β_s , 则 $s \leq r$ 。

为证引理, 我们只需要说明 $s > r$ 时线性相关即可。仿照之前设

$$\beta_i = \sum_{j=1}^r c_{ij}\alpha_j$$

则

$$\lambda_1\beta_1 + \dots + \lambda_s\beta_s = 0$$

可以化为

$$\sum_{i=1}^s \lambda_i \sum_{j=1}^r c_{ij}\alpha_j = 0$$

即

$$\sum_{j=1}^r \left(\sum_{i=1}^s \lambda_i c_{ij} \right) \alpha_j = 0$$

由于未知数个数多于方程个数, 与 2.3.3 相同知

$$\forall j = 1, \dots, r, \quad \sum_{i=1}^s \lambda_i c_{ij} = 0$$

有非零解。而这就导致了 $\lambda_1, \dots, \lambda_s$ 有非零解, 从而 β_1, \dots, β_s 线性相关, 矛盾。

根据证明的第一部分, 若有两个个数为 r, s 极大线性无关组, 它们一定能互相表出, 再由刚才的引理即可得到 $r \geq s, s \geq r$, 于是 $r = s$, 得证。

上述的证明过程给出了研究极大线性无关组的动机: 极大线性无关组就是“最小线性表出组”, 而后者对我们刻画结构是重要的。

将向量组的极大线性无关组元素个数称为其秩, 设 $\alpha_1, \dots, \alpha_n$ 的秩为 r , 且不妨设 $\alpha_1, \dots, \alpha_r$ 是一个极大线性无关组, 我们下面说明, 即使取 $\alpha_1, \dots, \alpha_n$ 外的向量, 也至少需要 r 个才能表出。

设这些向量为 β_1, \dots, β_s , 由于它们能表出 $\alpha_1, \dots, \alpha_n$, 则必然能表出 $\alpha_1, \dots, \alpha_r$, 而利用引理即得 $r \leq s$, 从而得证。

因此, 我们立刻得到, $\alpha_1, \dots, \alpha_n$ 至少需要 r 个向量才能表出, 从而 V 至少需要 r 个向量才能表出。另一方面, $\alpha_1, \dots, \alpha_n$ 可以由 $\alpha_1, \dots, \alpha_r$ 表出, 从而 V 可以由它们表出, 由此, 我们证明了向量组的秩等于其生成向量空间的维数。

另一方面, 我们将个数最少的 (根据本节中的证明, 最小性即等价于线性无关性)、能表出向量空间中全部向量的向量组称为它的一组基, 在刚才的假设下, $\alpha_1, \dots, \alpha_r$ 即成为 V 的一组基, 由此, 我们证明了向量组的极大线性无关组构成了其生成向量空间的一组基。

本节最后, 我们来说明 2.3.3 最后的结论 4 为何成立。若在向量空间 V 中找到了一些线性无关的向量 β_1, \dots, β_t , 利用不断找无法表出的加入可以最终得到 β_1, \dots, β_r , 且 $r \geq t$ 。与 2.3.3 证明过程相同, β_1, \dots, β_r 能表出 V 的全部向量, 且利用 2.3.3 最后的结论 2 可知它们线性无关, 从而得证其为基。

此外, 利用完全相同的思路, 某向量组的一个线性无关的子集可以扩充为极大线性无关组。

2.4.3 视为函数的维数

至此, 我们事实上已经可以利用之前的工具完全解决线性方程组的解问题了。不过, 在这之前, 为了之后讨论向量空间的方便, 我们需要证明一些基本的性质。由于这部分的证明基本可以参考书上, 我们只介绍相对重要的性质, 并给出证明思路。

1. 若 U, V 为向量空间 W 的子空间, 则 $U \cap V$ 为 W 的子空间。

思路: 只需说明 $U \cap V$ 中两向量的和/某向量的数乘在 V 与 U 中, 利用定义验证。

2. 若 U, V 为向量空间 W 的子空间, 则

$$U + V = \{w \in W \mid \exists u \in U, v \in V, w = u + v\}$$

为 W 的子空间。

思路: 利用交换律、结合律、分配律等基本性质与 U, V 的封闭性。

3. 任何包含 U, V 的向量空间必然包含 $U + V$ 。

思路: 利用线性组合的封闭性即可知。另一方面, 可以发现, $U + V$ 包含 $U \cup V$ 中元素任意线性组合的结果, 因此称其为 $U \cup V$ 生成的向量空间。

4. 若 U 为向量空间 V 的子空间, 则 $\dim U \leq \dim V$ 。

思路: 考虑 U 的基扩充为 V 的基即可。

5. $\dim V + \dim U = \dim(U + V) + \dim U \cap V$ 。

思路: 考虑 $U \cap V$ 的一组基 $\{\alpha_i\}$, 扩充为 U 的一组基 $\{\alpha_i\}, \{\beta_j\}$, 再扩充为 V 的一组基 $\{\alpha_i\}, \{\gamma_k\}$, 说明这些基的并集是 $U + V$ 的一组基即可。

* 这里证明过程有一个十分重要的思想, 也即线性相关和线性表出等效。具体来说, 若这些基的并集线性相关, 可以将 α_i, β_j 与对应系数得到等式一边、 γ_k 与对应系数移到等式另一边, 从而得到 V 中某元素与 U 中某元素相等, 这个元素必然属于 $U \cap V$ 。当存在线性相关时, 可以如此移项使得式子成为某一部分表出另一部分的形式, 从而左右必然在两边可表出的交集中, 再进一步进行分析。

* 一些秩 (记为 rank) 相关的不等式可以从上述空间相关的维数等式中得到, 例如, 要证

$$\text{rank}(\alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_t) \leq \text{rank}(\alpha_1, \dots, \alpha_s) + \text{rank}(\beta_1, \dots, \beta_t)$$

只需设 V 是 $\alpha_1, \dots, \alpha_s$ 生成的空间, U 是 β_1, \dots, β_t 生成的空间, 则问题即转化为 $\dim(U + V) \leq \dim U + \dim V$, 通过最后一个等式知成立。

§2.5 再看线性方程组

2.5.1 齐次线性方程组

本节延续 2.4.2 的讨论。

我们已经知道, 向量组 $\alpha_1, \dots, \alpha_n$ 的“结构”某种意义上就是其秩与一个极大线性无关组。至此, 我们先讨论与 b 无关的部分, 也就是所谓的唯一性, 或者更进一步的, 解集的结构。

既然与 b 无关, 我们也希望有解, 不妨取 $b = 0$, 考虑方程组

$$x_1\alpha_1 + \dots + x_n\alpha_n = 0$$

利用定义可发现, 其解集对加法与数乘封闭, 因此构成向量空间, 对其结构的讨论归于寻找基与维数。

为了引入更多的信息, 我们假设 $\alpha_1, \dots, \alpha_n$ 的极大线性无关组是 $\alpha_1, \dots, \alpha_r$, 从而其秩为 r 。我们希望消去尽量多未知的 α_i , 因此选择用 $\alpha_1, \dots, \alpha_r$ 表出剩下的, 也即设

$$\forall i > r, \quad \alpha_i = \sum_{j=1}^r c_{ij}\alpha_j$$

由此原方程组可以化为

$$\sum_{j=1}^r x_j\alpha_j + \sum_{i=r+1}^n x_i \sum_{j=1}^r c_{ij}\alpha_j = 0$$

整理可得

$$\sum_{j=1}^r \left(x_j + \sum_{i=r+1}^n x_i c_{ij} \right) \alpha_j = 0$$

由于 α_j 线性无关, 我们发现, 这个方程组为 0 事实上可以归结为每个系数为 0, 也即

$$\forall j = 1, \dots, r, \quad x_j + \sum_{i=r+1}^n x_i c_{ij} = 0$$

* 上面这步是一个非常重要的技巧, 即利用线性无关性将针对向量组的方程转化为系数的方程, 从而可以消去 α_j 。某种意义上, 对线性无关的定义就是为了这样的步骤得以执行。

举例观察可发现此方程组对应的矩阵已经是简化阶梯形, 由此其通解即为

$$x_j = - \sum_{i=r+1}^n x_i c_{ij}, \quad j = 1, \dots, r$$

不过, 我们并不满足于得到通解, 因为 c_{ij} 实际上是未知的, 但我们希望在向量空间的视角刻画结构。

我们将通解重新写为下方的形式:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_r \\ x_{r+1} \\ x_{r+2} \\ \vdots \\ x_n \end{pmatrix} = x_{r+1} \begin{pmatrix} -c_{r+1,1} \\ \vdots \\ -c_{r+1,r} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_{r+2} \begin{pmatrix} -c_{r+2,1} \\ \vdots \\ -c_{r+2,r} \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} -c_{n,1} \\ \vdots \\ -c_{n,r} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

对比每个分量的系数可以发现, 这确实蕴含了通解, 而由于左侧是解向量 x , 记右侧为 $x_{r+1}\gamma_{r+1} + \cdots + x_n\gamma_n$, 解集构成的向量空间 W 即为

$$W = \left\{ x \in \mathbb{R}^n \mid \exists x_{r+1}, \dots, x_n \in \mathbb{R}, x = \sum_{i=r+1}^n x_i \gamma_i \right\}$$

注意到, $\gamma_{r+1}, \dots, \gamma_n$ 可以表出 W , 只要再证明它们的线性无关性, 即可以说明它们是一组基, 而这是利用定义容易得到的: 若 $\sum_{i=r+1}^n \lambda_i \gamma_i = 0$, 考虑第 $r+1$ 个分量可发现对应方程即 $\lambda_{r+1} = 0$, 考虑第 $r+2$ 个分量可发现对应方程即 $\lambda_{r+2} = 0$, 由此可推得所有 λ_i 为 0。

综合以上, 我们最终得到的结论是: 齐次线性方程组的解空间维数等于变量个数减去系数矩阵列向量组的秩 (不妨记为 $\text{rank } A$), 也即

$$\dim W = n - \text{rank } A$$

其一组基可以通过 $\alpha_1, \dots, \alpha_n$ 的极大线性无关组表出 $\alpha_1, \dots, \alpha_n$ 的系数计算得到。

* 若 $\alpha_1, \dots, \alpha_n$ 的极大线性无关组不是前 r 个向量呢? 注意到, 线性方程组中的下标其实是没有实际意义的虚指标, $x_1\alpha_1 + x_2\alpha_2 = 0$ 与 $x_2\alpha_2 + x_1\alpha_1 = 0$ 表示的是同一个线性方程组。因此, 我们可以将 x 与 α 的下标同时做一些交换, 使得极大线性无关组变为 $\alpha_1, \dots, \alpha_r$, 最后再交换回来即可。具体情况的通解可以写为

$$\begin{pmatrix} x_{p_1} \\ \vdots \\ x_{p_r} \\ x_{p_{r+1}} \\ x_{p_{r+2}} \\ \vdots \\ x_{p_n} \end{pmatrix} = x_{p_{r+1}} \begin{pmatrix} -c_{p_{r+1},p_1} \\ \vdots \\ -c_{p_{r+1},p_r} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_{p_{r+2}} \begin{pmatrix} -c_{p_{r+2},p_1} \\ \vdots \\ -c_{p_{r+2},p_r} \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_{p_n} \begin{pmatrix} -c_{p_n,p_1} \\ \vdots \\ -c_{p_n,p_r} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

这里 p_1 到 p_n 为 1 到 n 的一个排列, $\alpha_{p_1}, \dots, \alpha_{p_r}$ 构成列向量组的极大线性无关组, 系数 c 满足

$$\forall i > r, \quad \alpha_{p_i} = \sum_{j=1}^r c_{p_i p_j} \alpha_{p_j}$$

2.5.2 非齐次线性方程组

解决了齐次线性方程组后, 非齐次线性方程组要解决的问题分为了存在性问题和存在时解集的结构。

由于我们已经知道, 非齐次线性方程组的解存在当且仅当 b 在 $\alpha_1, \dots, \alpha_n$ 生成的子空间中, 也即 β 可以被 $\alpha_1, \dots, \alpha_n$ 表出。这事实上等价于 $\text{rank } A = \text{rank}(A, b)$: 若 β 可以被 $\alpha_1, \dots, \alpha_n$ 表出, 则其可以被它

们的极大线性无关组表出，于是极大线性无关组无法再增添 b ，秩不变；反之，若秩不变，极大线性无关组增添 b 后线性相关，与 2.4.2 中的证明相同可知 b 可被极大线性无关组表出。

于是，我们得到了线性方程组有解的充要条件， $\text{rank } A = \text{rank}(A, b)$ ，即系数矩阵的秩等于增广矩阵的秩 [这里秩指列向量组的秩]。

而只要知道存在某个解 x ，设对应齐次方程组的解空间 W ，可发现非齐次方程组的全部解 X 可写为

$$\{z \in \mathbb{R}^n \mid \exists w \in W, z = x + w\}$$

* 一方面，另一个解与 x 的差是齐次线性方程组的解；另一方面， x 加某个齐次线性方程组的解代入可知仍然为非齐次线性方程组的解。

* 可以将这样在向量空间 W 的基础上平移某个 x 的结构称为一个线性流形，记为 $x + W$ ，其维数定义为向量空间的维数，这就符合了我们开始的讨论。

于是，在解决了齐次线性方程组的问题以后，我们很快就能得到非齐次线性方程组问题的解答。

2.5.3 秩的计算

最后，我们来谈谈算法。上述的过程说明，如果要确定一个线性方程组的解，找到其列向量组的线性无关组是本质的做法。但是，如何寻找呢？我们下面称矩阵的列向量组的极大线性无关组个数为列秩，行向量组为行秩。

我们现在完成 2.2.1 节中行秩等于列秩的证明，算法就蕴含在其中：

1. 先说明一个基本的结论：初等行变换不改变列向量组的极大线性无关组下标。例如，若 $\alpha_1, \dots, \alpha_r$ 是 $\alpha_1, \dots, \alpha_n$ 的极大线性无关组，则将矩阵 A 进行某些初等行变换后成为矩阵 B ，列向量组为 β_1, \dots, β_n ，则 β_1, \dots, β_r 为其一个极大线性无关组。

证明：注意到初等行变换不会改变对应线性方程组的解，对某些下标 j_1, \dots, j_k ，方程组（相当于对应 A 中选出某些列构成的矩阵）

$$x_1 \alpha_{j_1} + \dots + x_k \alpha_{j_k} = 0$$

可以与 A 到 B 利用相同的初等行变换变换至

$$x_1 \beta_{j_1} + \dots + x_k \beta_{j_k} = 0$$

由此，若变换前有非零解则变换后有非零解，反之亦然。既然线性相关、线性无关的关系不会改变，极大线性无关组的下标亦不会改变。

2. 另一个基本的结论是：初等列变换不改变列向量组生成的向量空间。这是由于初等列变换是可逆的，变换前后的列向量组必然可以互相表出 [可分三种变换进行讨论]，由此生成的向量空间相同。
3. 无论是不改变下标还是不改变生成的向量空间，根据定义都有初等行/列变换不改变列秩，而利用矩阵行列的对偶性可以完全相似证明初等列变换不改变行向量组的极大线性无关组下标与初等行变换不改变行向量组生成的向量空间，由此最终得到初等行/列变换不改变行/列秩。
4. 为了完成行秩等于列秩的证明，我们只需要对任何矩阵可以初等行变换到的简化阶梯形矩阵说明即可。

考虑其主元所在的行/列，与 2.5.1 寻找 W 一组基相同可知主元所在行线性无关，而其他行全为 0，因此其构成行向量组的极大线性无关组。另一方面，主元所在列根据相同理由可知线性无关，而直接求解可知它们可以线性表出其他列，从而其构成列向量组的极大线性无关组。因此，矩阵的行秩/列秩都等于主元个数，得证。

* 这同样代表了开始时我们对“解空间维数”的直观定义是合理的，自由变量数与主元个数和为未知数个数，恰好对应 $\dim W + \text{rank } A = n$ 。

由此, 若我们想知道列向量组的极大线性无关组, 应利用初等行变换不改变列向量组的极大线性无关组下标, 变换为简化阶梯形后, 主元所在的列的下标即为原本的极大线性无关组下标。这是求一个极大线性无关组的通用算法。

* 将矩阵的行秩/列秩定义为矩阵的秩, 它事实上可以脱离线性相关、线性无关的概念而定义, 这就是下一次习题课要介绍的内容了。

三 向量空间与矩阵

§3.1 习题解答

1. 习题 3.6-5

设其增广矩阵为 (A, b) , 由于增广矩阵 (A, b) 为 A 增添最后一列 b , 考虑列秩可知 $\text{rank}(A, b) \geq \text{rank } A$; 又由于 B 为 (A, b) 增添最后一行, 考虑行秩可知 $\text{rank } B \geq \text{rank}(A, b)$, 从而由 $\text{rank } B = \text{rank } A$ 可知 $\text{rank}(A, b) = \text{rank } A$, 即有解。

2. 习题 3.7-2

利用本讲义第二章或教材中已经证明的, 齐次线性方程组的解构成向量空间, 而基础解系即为一组基。

设与其等价的线性无关向量组为 S , 则类似极大线性无关组个数相等的证明可知 S 的个数必然也为 t , 且由等价性它们可以被基础解系生成, 从而在解空间中, 因此构成一组基。

3. 习题 3.7-5

* 看到前一题作为提示后, 这题的想法就比较自然了。

分类讨论。由条件可知 $\text{rank } A < n$, 若 $\text{rank } A = n - 2$, 根据最大非零子式阶数定义可知一切 A_{ij} 为 0, 从而成比例。

否则, $\text{rank } A = n - 1$, 考虑线性方程组

$$\forall i = 1, \dots, n, \quad \sum_{j=1}^n a_{ij}x_j = 0$$

根据 3.7 节定理 1 可知, 此方程组的解空间为 1 维, 由此必然能写成 $\alpha x, \alpha \in \mathbb{K}$ 的形式, 从而任何两个解成比例。

另外一方面, 根据 2.4 节定理 1 与定理 3, 结合 $\det A = 0$ 可知对任何 k , $x_j = A_{kj}$ 构成此方程的一组解, 而这即为第 k 列, 从而任意两列成比例。对两行考虑转置完全同理。

4. 习题 4.7-5

直接根据定义计算, $\text{Im } A$ 为其列向量生成的向量空间, 前两列构成一组基; $\text{Ker } A$ 为其解空间, 计算可知维数为 2, 一个基础解系为 $(-2, -\frac{3}{2}, 1, 0)', (-1, -2, 0, 1)'$ 。

5. 习题 4.1-5

由定义计算得结果为 $a_{11}x^2 + a_{22}y^2 + 2a_{12}xy + 2a_1x + 2a_2y + 1$ 。

6. 习题 4.1-11

利用乘法对加法的分配律与加法交换、结合律可知

$$(I - B)(I + B + B^2) = I + B + B^2 - B - B^2 - B^3 = I - B^3 = I$$

§3.2 线性映射与矩阵

3.2.1 基的操作与坐标

目前为止的内容里，最重要的技巧事实上已经在本讲义第二章中书写了，也即扩充。

由于任何线性无关向量组可以扩充出极大线性无关组、任何子空间基可以扩充为原空间一组基，很多时候我们可以将向量空间问题的讨论化归到对基的讨论以大大简化。

对于更多的技巧，我们以补充题三-2 为例进行介绍。

问题：任给某线性方程组，其系数矩阵为 A 、常数项为 b ，并假设解集 X 非空。若一些线性无关的向量 $\gamma_1, \dots, \gamma_m$ 能将 W 中所有向量表出，讨论 m 的最小值与 A, b 的关系。

1. 先用空间语言描述线性方程组问题。

设其任何一个解为 x_0 ，对应齐次方程组的解为 W ，则有 $X = W + x_0$ ；找能将其中向量都表出的向量个数，即为找其极大线性无关组个数。

由此，问题暂时变为讨论 $W + x_0$ 的极大线性无关组个数（之后还要还原为 A, b 的性质）。

2. 由于齐次线性方程组为特殊情况，分类讨论。

若其为齐次线性方程组，则 $b = 0$ ，解集即为 W ，从而最小的 m 即为其维数 $n - \text{rank } A$ 。

* 熟悉核心定理，如解空间维数。

3. 非齐次情况，仍然需要分其是否为向量空间讨论，否则难以进行观察。（下面假设出现的 w_i 均为 W 中元素）

若 $W + x_0$ 为向量空间，写出定义，有若 $a = w_1 + x_0, b = w_2 + x_0$ ，则 $a + b = w_3 + x_0$ ，从而 $w_1 + w_2 + 2x_0 = w_3 + x_0$ ，即可发现 $x_0 \in W$ 。然而， $x_0 \in W$ ，且 x_0 为原方程解意味着原方程必为齐次线性方程组（ W 的定义），矛盾。

4. 已知其不为向量空间，分析极大线性无关组。

设 W 一组基 w_1, \dots, w_k ，则由于 $x_0 \notin W$ ，可知 x_0, w_1, \dots, w_k 线性无关。

* 熟知一些常用小结论：若线性无关的向量组加入一个向量后线性相关，则新向量可被原有向量表出，证明见本讲义第二章。

此外，由于 W 中向量可被 w_1, \dots, w_k 表出，即可知 $W + x_0$ 中向量可被 w_1, \dots, w_k, x_0 表出，从而它们构成极大线性无关组（熟悉用表出叙述线性无关相关内容）， $m = k + 1 = n - \text{rank } A + 1$ 。

事实上，由于线性映射与矩阵相关的内容都并未开始，向量空间暂时还并无太多的技巧。一些关键内容，如限制映射相关的操作，也无法在此处介绍。

不过，在进入下一部分前，希望大家能关注到一件事：相同维数的向量空间是无法区分的（这里指同一个数域上的，我们还是讨论 \mathbb{R} 上）。

——什么？ $(a, 2a, 50a), a \in \mathbb{R}$ 和 \mathbb{R} 为何会无法区分？

务必注意，这里所谓的“区分”是一种内蕴的区分，也即把元素当成黑箱后无法通过允许的操作（加法、数乘）区分。我们不妨假设维数同为 k 。

例如，两个空间中都可以找到 k 个线性无关的向量（不妨称为 $\alpha_1, \dots, \alpha_k$ 与 β_1, \dots, β_k ），且找不到 $k + 1$ 个线性无关的向量。不止如此，第一个空间中的任何向量都可以用其中的 k 个线性无关的向量唯一表示成

$$\lambda_1 \alpha_1 + \dots + \lambda_k \alpha_k$$

第二个空间中的任何向量都可以用其中的 k 个线性无关的向量唯一表示成

$$\lambda_1 \beta_1 + \dots + \lambda_k \beta_k$$

如果第一个空间中两个向量分别表示成 (注意 $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, 不会因为元素黑箱而无法取)

$$\lambda_1\alpha_1 + \dots + \lambda_k\alpha_k, \mu_1\alpha_1 + \dots + \mu_k\alpha_k$$

第二个空间中两个向量分别表示成

$$\lambda_1\beta_1 + \dots + \lambda_k\beta_k, \mu_1\beta_1 + \dots + \mu_k\beta_k$$

则两个向量的和与第一个向量数乘 λ 在第一个空间分别为

$$(\lambda_1 + \mu_1)\alpha_1 + \dots + (\lambda_k + \mu_k)\alpha_k, \lambda\lambda_1\alpha_1 + \dots + \lambda\lambda_k\alpha_k$$

两个向量的和与第一个向量数乘 λ 在第二个空间分别为

$$(\lambda_1 + \mu_1)\beta_1 + \dots + (\lambda_k + \mu_k)\beta_k, \lambda\lambda_1\beta_1 + \dots + \lambda\lambda_k\beta_k$$

由此可发现, 一个 k 元组 $(\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ 唯一确定了维数为 k 的空间中的元素和运算方式。

对任何一组基, 由于表示唯一性, 任何一个元素都可以在这组基下唯一看成 $(\lambda_1, \dots, \lambda_k)$, 称为其在这组基下的坐标。

我们把两个通过内蕴性质无法区分的结构称为同构, 由此可知向量空间维数相同时同构。

反之, 若向量空间同构, 它们必然能取出相同数量的极大线性无关组 (否则将可以区分), 于是向量空间维数同构当且仅当维数相同。

完全类似地, 由于维数更低的向量空间会与维数更高的向量空间的一个子空间同构, 我们可以认为前者的结构被后者包含, 而后者的结构可以某种意义上“覆盖”前者。

* 但是, 什么是通过内蕴性质无法区分? 什么又是结构被包含? 除了上述直观方法, 有没有更严谨的定义?

3.2.2 线性映射

我们回到大家唯一学过, 也是最简单的代数结构: 集合。

集合的内蕴性质可以简单理解为交、并、补运算与取子集操作等。那么, 什么样的两个集合之间无法区分呢?

对于有限集合, 这个问题的答案是元素个数相同。只要元素个数相同, 就能将它们编号记为成 a_1 到 a_k 、 b_1 到 b_k , 具有完全相同的操作结果。

但对于无限集合, 上述的简单思路无法适用, 我们需要更进一步对两个集合进行关联来定义无法区分。

关联集合的方式是映射。利用映射, 只要两个集合之间存在一个一一对应的映射, 它们就无法区分了——无论如何考虑子集、交、并、补, 只要针对的元素被一一对应了, 运算的结果也是一一对应的。

类似“结构被包含”, 若一个集合 A 与另一个集合 B 的子集能建立一一对应, 就可以看作其结构被另一个结构包含。这样的一一对应称为单射, 更具体来说, 其定义为 $\forall x \neq y \in A, f(x) \neq f(y)$ 的映射 $f: A \rightarrow B$ 。

* 这里 B 为陪域, 也即讨论结构的集合, 而 $f(A)$ 为值域, 也即“实际映射到”的部分。

反之, 若一个集合可以覆盖另一个集合, 我们可以定义为, 存在一个映射 f 使得 $\forall \beta \in B, \exists x \in A, f(x) = \beta$ 。这样的映射称为满射。

* 因此, 单射和满射本质是在刻画结构的包含与覆盖。

由此, 我们可以将上述的一一对应称为双射, 定义为既是单射又是满射的映射。其存在唯一的逆映射满足 $f^{-1}(f(x)) = f(f^{-1}(x)) = x$ 。此外, 可以证明若 f 并非双射, 则不存在这样的逆映射。

那么, 自然的疑问是, 如果 A 到 B 既存在单射也存在满射, 也即 A 的结构既可以覆盖 B , 又可以被 B 包含, 它们的结构是否一定相同呢?

对于集合, 这个问题的答案是肯定的, 称为 Cantor-Bernstein 定理——听名字就能知道, 这个结论并没有看上去那么显然。

用更抽象的语言，双射就是一个集合之间的同构。

回到向量空间中，向量空间比起集合还多了运算，因此，刻画向量空间关联的方式虽然也是映射，但是需要保证对加法和数乘的相容性。

于是，我们将向量空间之间的同态（或称为线性映射）定义为满足 $f(a) + f(b) = f(a+b)$ 、 $f(\lambda a) = \lambda f(a)$ 的映射 f ，注意到，同态事实上是某种可交换性，加法、数乘与映射可以交换。若将映射后的元素看作映射前的元素的对应，它们在运算下也是无法区分的。

如果在元素的无法区分下更进一步希望空间无法区分，我们希望这个映射还是元素的一一对应，也即双射。由此，线性双射可称为线性同构（思考为何如此定义的映射能导致空间无法区分）。此外，是单射的线性映射称为单同态，对应结构的被包含；是满射的线性映射称为满同态，对应结构的覆盖。

列举出如下基本结论，大家可以自行通过定义验证：

1. 零向量在同态下变为零向量（利用 $0 + 0 = 0$ ）。
2. 同态的复合是同态（利用定义计算）。
3. 线性相关的向量组在同态下变为线性相关的向量组（即 $\alpha_1, \dots, \alpha_r$ 线性相关则 $f(\alpha_1), \dots, f(\alpha_r)$ 线性相关）。
4. 只要给出 U 的一组基到 V 的一个映射，就能给出 U 到 V 的一个同态，反之亦然（基的像可以确定同态），由此任何两向量空间存在同态，且构造同态只需说明基的像。
5. 若同态是单同态，线性无关的向量组变为线性无关的向量组，特别地，原空间一组基变为线性无关的向量组（反证，若结论不成立，0 会存在两个原像）。
6. 同态是单同态当且仅当零向量的原像只有零向量（若某元素有两不同原像，作差可得零向量的非零原像）。
7. 向量空间 U 到 V 存在单同态当且仅当 $\dim U \leq \dim V$ （利用之前的结论，并在条件满足时通过基构造同态）。
8. 若同态是满同态，原空间一组基的像包含新空间一组基（注意一组基可以生成原空间，若其像不能生成新空间则矛盾）。
9. 向量空间 U 到 V 存在单同态当且仅当 $\dim U \geq \dim V$ （利用之前的结论，并在条件满足时通过基构造同态，可将多的部分映射为 0）。
10. 向量空间 U 到 V 存在线性同构当且仅当 $\dim U = \dim V$ （利用之前的结论）。
11. 线性同构把一组基映射到一组基。
12. 线性同构的逆映射（由于其为双射，必然存在）也是线性的（考虑用基进行表示）。

* 事实上，同构的定义一般是双射、保持结构、且逆映射保持结构。由于线性映射与双射足够推出逆映射是线性的，无需第三条定义。此外，最后一件事就说明了线性同构存在即等价于我们之前说的同构。

3.2.3 第一同构定理

线性映射刻画了结构的传递，为了验证这点，我们要进行一个奇妙的操作：从线性映射出发，试着构造一个不损失信息的线性同构。

我们仍然从大家更为熟悉的集合开始，在集合中，上述的操作是：从一个映射出发，试着构造一个不损失信息的双射。

我们考虑一个最简单的映射 $f : \{1, 2, 3, 4, 5\} \rightarrow \{a, b, c, d\}$, 满足

$$f(1) = c, f(2) = d, f(3) = b, f(4) = c, f(5) = d$$

首先, 对于这个映射中根本没有“提到”的元素 a , 我们可以直接把它删掉, 这样 f (不妨记为 \tilde{f}) 就成为了 $\{1, 2, 3, 4, 5\} \rightarrow \{b, c, d\}$ 的满射。

接下来, 为了将 \tilde{f} 变为单射, 一个直观的想法是直接从 1、4 中选一个, 2、5 中选一个, 将剩下两个扔掉即可剩下单射。不过, 这样做事实上损失了另一个元素的信息, 因此, 我们将 \tilde{f} 定义为以集合为元素的映射, 也即 $\tilde{f} : \{\{1, 4\}, \{2, 5\}, \{3\}\} \rightarrow \{b, c, d\}$ 。

$$\tilde{f}(\{1, 4\}) = c, \quad \tilde{f}(\{2, 5\}) = d, \quad \tilde{f}(\{3\}) = b$$

* 事实上, 这么做是用 f 在集合 $\{1, 2, 3, 4, 5\}$ 上定义了一个等价关系, 两元素 x, y 等价当且仅当 $f(x) = f(y)$, 再划分出等价类后将等价类作为元素。更直观来看, 它是将像相同的元素打包成一个整体。

线性映射是一个映射, 自然也可以如此操作。但是, 我们希望得到的是一个线性同构, 这就对线性性有了很高的要求。仍然一步步进行操作, 设 f 是 U 到 V 的线性映射:

1. 如同集合, 我们要扔掉 V 中未必“提到”的元素, 事实上也就是只保留 f 的像集中的元素, 我们将它称为 $\text{Im } f$ 。为保证线性性, 我们希望 $\text{Im } f$ 是一个向量空间, 才能谈论线性同构。

证明: 利用定义, 对 $\text{Im } f$ 中的 a, b , 找到它们的原像 x, y , 则 $f(x + y) = a + b$ 、 $f(\lambda x) = \lambda a$, 从而均在 $\text{Im } f$ 中, 得证。

2. 下面, 为了将元素“打包”, 我们需要观察一个元素 $a \in \text{Im } f$ 的原像到底是什么。设 $f(x) = f(y) = a$, 则 $f(x - y) = 0$; 反之, 若 $f(z) = 0$ 、 $f(x) = a$, 则 $f(x + z) = a$ 。由此, 可将任何元素的原像与零向量的原像一一对应, 于是只需考虑零向量原像, 设其为 $\text{Ker } f$ 。

可以发现, 这里的一一对应过程其实就是之前说明齐次线性方程组与非齐次线性方程组的解的关系的过程。我们这里来更严谨地用双射说明。

证明: 设 $a \in \text{Im } f$ 的原像集合 $f^{-1}(a)$, 由于其非空, 设 $x \in f^{-1}(a)$ 。我们构造映射 $\varphi : f^{-1}(a) \rightarrow \text{Ker } f$, 使得 $\varphi(y) = y - x$ 。由于之前的推导, $x, y \in f^{-1}(a)$ 可知 $y - x \in \text{Ker } f$, 这的确是良好定义的映射。为验证其双射, 需要证明单射与满射:

- 若 $y \neq z$, 则 $\varphi(y) - \varphi(z) = (y - x) - (z - x) = y - z \neq 0$, 由此 $\varphi(y) \neq \varphi(z)$, 因此其为单射。
- 对任何 $y \in \text{Ker } f$, 有 $f(y + x) = f(y) + f(x) = f(x) = a$, 且 $\varphi^{-1}(y + x) = y$, 由此其原像非空, 因此映射为满射。

3. 根据刚才的证明过程, 我们已经发现了, 任给元素 $a \in \text{Im } f$, 若 $x \in f^{-1}(a)$, 有 $f^{-1}(a) = x + \text{Ker } f$ 。这个原像的形式非常类似我们之前定义的线性流形, 于是我们要观察 $\text{Ker } f$ 是否为向量空间, 答案是肯定的 (完全类似齐次线性方程组的解集)。

证明: 若 $f(a) = 0$ 、 $f(b) = 0$, 有 $f(a + b) = f(a) + f(b) = 0$ 、 $f(\lambda a) = \lambda f(a) = 0$, 从而得证。

4. 由此, 我们最终定义

$$\tilde{f}(x + \text{Ker } f) = f(x)$$

我们将所有 $x + \text{Ker } f$ 构成的集合 (去掉重复) 记为 $U/\text{Ker } f$, 则 f 为 $U/\text{Ker } f \rightarrow \text{Im } f$ 的映射。这是一个符合要求的双射。

证明:

- 良好定义性: 若 $x + \text{Ker } f = y + \text{Ker } f$ (集合意义下相等), 则根据 $0 \in \text{Ker } f$ (见上一部分的性质, 或利用其为向量空间) 可知 $x \in x + \text{Ker } f$ 。利用集合相等性, $x \in y + \text{Ker } f$, 必然有 $z \in \text{Ker } f$ 使得 $x = y + z$, 从而 $x - y = z \in \text{Ker } f$ 。于是 $f(y) = f(y + (x - y)) = f(x)$, 从而 $\tilde{f}(x + \text{Ker } f) = \tilde{f}(y + \text{Ker } f)$ 。这就说明了不会将同一个元素映射到两个元素, 这的确是映射。
- 单射: 若 $\tilde{f}(x + \text{Ker } f) = \tilde{f}(y + \text{Ker } f)$, 即 $f(x) = f(y)$, 有 $f(x - y) = 0$, $x - y \in \text{Ker } f$, 设其为 z 。对任何 $x + \text{Ker } f$ 中的元素, 设其写为 $x + w$, $w \in \text{Ker } f$, 有 $x + w = y + z + w = y + (z + w)$, 于是其在 $y + \text{Ker } f$ 中, 反之同理, 因此可得 $x + \text{Ker } f = y + \text{Ker } f$ 。
- * 这里证明的其实是单射的逆否命题, 由 $\tilde{f}(x + \text{Ker } f) = \tilde{f}(y + \text{Ker } f)$ 推出 $x + \text{Ker } f = y + \text{Ker } f$ 。
- 满射: 对任何 $a \in \text{Im } f$, 设 $f(x) = a$, 则 $\tilde{f}(x + \text{Ker } f) = a$, 于是其为满射。

5. 最后, 为了让它具有线性结构, 我们还要求它是一个线性映射, 这就必须把 $U / \text{Ker } f$ 看作某个“向量空间”。我们这么考虑: 设 $\text{Ker } f$ 的一组基 (已经证明了它是向量空间) 为 $\alpha_1, \dots, \alpha_m$, 扩充为 U 的一组基 $\alpha_1, \dots, \alpha_m, \beta_{m+1}, \dots, \beta_n$, 由于任何向量可以表出为

$$x = \sum_{i=1}^m \lambda_i \alpha_i + \sum_{i=m+1}^n \mu_i \beta_i$$

可知

$$f(x) = \sum_{i=1}^m \lambda_i f(\alpha_i) + \sum_{i=m+1}^n \mu_i f(\beta_i) = \sum_{i=m+1}^n \mu_i f(\beta_i)$$

设 β_i 生成的 U 的子空间为 W , 构造映射:

$$\psi : U / \text{Ker } f \rightarrow W, \quad \psi(x + \text{Ker } f) = (x + \text{Ker } f) \cap W$$

我们证明 $x + \text{Ker } f \cap W$ 一定有且仅有一个元素 (这就证明了良好定义性), 且其为双射。

证明:

- 存在性: 考虑 x 的上述表示, 由其为向量空间, $-\sum_{i=1}^m \lambda_i \alpha_i \in \text{Ker } f$, 因此 $x - \sum_{i=1}^m \lambda_i \alpha_i \in x + \text{Ker } f$, 而根据 W 的定义即知它属于 W , 从而它在 $(x + \text{Ker } f) \cap W$ 中。
- 唯一性: 若 $a, b \in (x + \text{Ker } f) \cap W$, 根据之前已证 $c = a - b \in \text{Ker } f$, 且根据 W 为向量空间可知 $c \in W$ 。由此 c 可在基 $\alpha_1, \dots, \alpha_m$ 下表出, 也可在 $\beta_{m+1}, \dots, \beta_n$ 下表出, 由线性无关性可知其只能为 0, 从而 $a = b$ 。
- 单射: 若 $\psi(x + \text{Ker } f) = \psi(y + \text{Ker } f)$, 说明 $(x + \text{Ker } f) \cap W = (y + \text{Ker } f) \cap W$, 设此元素为 a , 则 $a \in (x + \text{Ker } f) \cap (y + \text{Ker } f)$, 于是 $x - a \in \text{Ker } f$ 且 $y - a \in \text{Ker } f$, 有 $y - x = (y - a) - (x - a) \in \text{Ker } f$, 利用之前已证, 这就说明了 $x + \text{Ker } f = y + \text{Ker } f$ 。
- 满射: 由于 $w \in w + \text{Ker } f$, 即可知对任何 $w \in W$, $\psi(w + \text{Ker } f) = w$, 这就证明了满射。

6. 由此, 我们将 $U / \text{Ker } f$ 看作 W , 其上的运算即为 W 上的运算, 而这是普通的加法和数乘, 再利用 $x + \text{Ker } f$ 与 $y + \text{Ker } f$ 某元素和一定在 $x + y + \text{Ker } f$ 中, $x + \text{Ker } f$ 中某元素 λ 倍一定在 $\lambda x + \text{Ker } f$ 中 (可自行验证) 即有

$$(x + \text{Ker } f) + (y + \text{Ker } f) = x + y + \text{Ker } f$$

$$\lambda(x + \text{Ker } f) = \lambda x + \text{Ker } f$$

* 以上做法事实上是定义了 $(x + \text{Ker } f) + (y + \text{Ker } f) = \psi^{-1}(\psi(x + \text{Ker } f) + \psi(y + \text{Ker } f))$, 对数乘类似, 也可以脱离这种做法直接通过抽象线性空间定义。

7. 最后, 在这样的将 $U/\text{Ker } f$ 看成向量空间的看法下, \tilde{f} 是线性映射。

证明:

$$\begin{aligned}\tilde{f}(x + \text{Ker } f) + \tilde{f}(y + \text{Ker } f) &= f(x) + f(y) = f(x + y) = \tilde{f}(x + y + \text{Ker } f) \\ \lambda \tilde{f}(x + \text{Ker } f) &= \lambda f(x) = f(\lambda x) = \tilde{f}(\lambda x + \text{Ker } f)\end{aligned}$$

* 刚才的过程里基本包含了对抽象线性映射会使用到的技巧, 包括 Ker 、 Im 的性质和操作等。

* 在上述证明过程里已经蕴含了 $\dim \text{Ker } f + \dim \text{Im } f = \dim U$ 。

这个定理称为线性空间的第一同构定理。之后学到更多抽象代数结构时, 也会有更多对应的第一同构定理。

§3.3 矩阵视角的行列变换

3.3.1 整体思想

为什么突然从线性映射跳到了矩阵——因为这两者事实上是密切相关的, 也即线性变换一定可以用坐标表示成一个矩阵。在这个意义上, 矩阵乘法可以看作线性变换的复合, 从而利用映射的结合律自然有结合律。

简单来说, 假设 $f: U \rightarrow V$ 是 n 维空间 U 到 m 维空间 V 的线性变换, 并假设 U 的一组基 $\alpha_1, \dots, \alpha_n$ 、 V 的一组基 β_1, \dots, β_m 。 U 中任何元素可看作一个坐标 (这里用列向量表示) $(\lambda_1, \dots, \lambda_n)'$, V 中任何元素也可看作一个坐标 $(\mu_1, \dots, \mu_m)'$ 。

假设 e_j 为只有第 j 个元素为 1, 其余为 0 的 n 维向量, 设 $f(e_j) = (c_{1j}, c_{2j}, \dots, c_{mj})'$, 将所有 c_{ij} 拼成的矩阵 C (记为 $C = (c_{ij})$) 称为 f 在这两组基下的矩阵表示。可发现其为 $m \times n$ 矩阵, 这里 m 为 V 的维数, n 为 U 的维数。

设 $x = (\lambda_1, \dots, \lambda_n)'$, 则可直接由 $x = \sum_i \lambda_i e_i$ 证明 $f(x)$ 的第 i 个分量为

$$\sum_j c_{ij} \lambda_j$$

我们将这个结果定义为矩阵与向量的乘法, 即矩阵 $C = (c_{ij})$ 与向量 $\lambda = (\lambda_j)$ 的乘法为向量 $C\lambda$, 满足

$$(C\lambda)_i = \sum_j c_{ij} \lambda_j$$

另一方面, 若还有线性变换 $g: V \rightarrow W$, 对 V 中的 β_1, \dots, β_m , W 的一组基 $\gamma_1, \dots, \gamma_p$, 其矩阵表示应为一个 $p \times m$ 矩阵 (d_{ij}) 。于是有

$$g(\mu)_i = \sum_j d_{ij} \mu_j$$

考虑 $g(f(\lambda))$, 也即代入 $\mu = f(\lambda)$ 可知

$$g(f(\lambda))_i = \sum_j d_{ij} \mu_j = \sum_j d_{ij} \sum_k c_{jk} \lambda_k = \sum_k \sum_j d_{ij} c_{jk} \lambda_k$$

因此根据矩阵表示定义, $g \circ f$ (\circ 表示复合) 的矩阵表示为

$$F = (f_{ik}), \quad f_{ik} = \sum_j d_{ij} c_{jk}$$

将矩阵 F 定义为矩阵 C, D 的乘积, 记为 $F = DC$ 。

* 由此, 矩阵 $A_{m \times n}$ (这里用下标表示行列数, 之后也有这种用法) 也可看作线性映射 $\lambda \rightarrow A\lambda$, 这里 $\lambda \in \mathbb{R}^n$ 。

不过, 由于缺乏对矩阵的研究, 对线性映射的讨论暂时只能到此为止。在这部分, 我们要先放下向量空间相关的理论, 去研究这个之前研究线性方程组时出现的、具有行列的数组的性质——当然, 我们已经知道了线性映射可以看作矩阵乘法, 那么研究矩阵和研究向量空间之间的线性映射本质上是相同的。

由于目前并未讲太多矩阵相关, 我们只强调一件事, 即整体思想。很多时候, 看待矩阵可以直接作为整体, 而不再关心行列的具体数, 例如, 作业里, 只要知道 $B^3 = O$, 就有 $(I - B)(I + B + B^2) = I$, 又如本讲义第一章的行列式最后一道例题, 无需做任何具体的运算就能得到结果。

某种意义上, 这就像我们用符号指代了数形成代数一样, 用大写字母 A, B 将矩阵整体代换为了一些新的量, 并存在它们自己的运算规则与性质。

当然, 遇到更具体的问题时, 还是很可能要将矩阵写回分量的形式的。此外, 其实还有一种部分分量的写法, 也就是分块矩阵——不过考虑到本章发布时大概还没有讲到分块矩阵, 或只是浅浅提及, 这部分的内容留待下次习题课说明。

3.3.2 秩的矩阵论定义

上节习题课里跳过了这么一个结论: 秩等于最大非零子式的阶数。这是因为, 这个结论本质上是一个脱离空间的矩阵论写法, 理应放在矩阵的部分介绍。

这个结论的证明是简单的, 下设矩阵的秩为 r :

- 根据行秩定义, 从行向量中可取出极大线性无关组 A_1 , 而再根据 A_1 行秩为 r , 可知其列秩也为 r , 从而又可取出其列向量的极大线性无关组, 这是一个 $r \times r$ 的非零子式。
- 由于矩阵的任何 $r + 1$ 行线性相关, 必有一行能被其他行表出, 而对任何一个 $r + 1$ 阶子式, 由于其为原行缩短, 必然也有一行能被其他行表出, 从而行列式为 0。

* 这里可以给大家展示一下 Laplace 展开定理的最后荣光, 我们来证明矩阵存在任何 $k \leq r$ 阶的非零子式。当 $k < r$ 时, 我们考虑之前取出的 r 阶非零子式 R , 若其 k 阶子式均为 0, 根据 Laplace 展开定理可知行列式为 0, 矛盾, 从而得证。

注意到, 上述的结论完全可以无视线性无关、线性相关进行定义。并且, 它还有一个更本质的视角, 我们先给出如下定理: 任何 $m \times n$ 矩阵 A 都能通过行初等变换与列初等变换成为 [这称为 **Hermite 标准形**]

$$\begin{pmatrix} I_{r \times r} & O_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{pmatrix}$$

这里 $r = \text{rank } A$ 。

证明: 先通过行初等变换将其变为简化阶梯形矩阵, 此时主元个数即为 r 。接着, 将主元所在列交换到前 r 列, 矩阵成为了 (X 未知)

$$\begin{pmatrix} I_{r \times r} & X_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{pmatrix}$$

通过减去前 r 列的适当倍数可消去 X , 从而得证。

矩阵论的视角里, 初等变换是什么呢? 4.2 节给出了答案。一系列行初等变换可以看作一列初等矩阵进行左乘, 而列初等变换可以看作一列初等矩阵进行右乘。我们定义变换阵 [这个词是生造的, 之后会学到它的实际定义] 为能写成一列初等阵相乘的阵, 则上述结论可以表达为: 对任何矩阵 $A_{m \times n}$, 存在变换阵 $P_{m \times m}$ 与 $Q_{n \times n}$ 使得

$$PAQ = \begin{pmatrix} I_{r \times r} & O_{r \times (n-r)} \\ O_{(m-r) \times r} & O_{(m-r) \times (n-r)} \end{pmatrix}$$

* 之后我们把左右乘变换阵称为作行变换 P 、作列变换 Q 。

* 如果还记得之前所说的等价关系，秩相等构成一个等价关系，而上方的过程说明了对同阶矩阵，“秩相等”与“Hermite 标准形相同”是一致的。更进一步地，它们和“可以行列变换得到”一致——利用行列变换是可逆的，两个矩阵可以行列变换为同一个矩阵意味着它们可以行列变换得到。

为了展示上述结论的应用，我们来从矩阵论的视角解决之前题目中的一些秩相关的问题：

1. $\text{rank } A = \text{rank } A'$: 利用最大可逆子式的定义， A 与其转置对称位置的子式也为转置关系，而行列式在转置后不变，于是同时为 0 或不为 0，即得证。
2. $\text{rank } A_{m \times n} \leq \min(m, n)$: 通过子式阶数只能取到 $\min(m, n)$ 即得。
3. $\text{rank } A = 0$ 等价于 $A = O$: 右推左利用定义知结论，左推右利用任何非零数都是一阶可逆子式即得。
4. 子矩阵的秩不超过原矩阵：子矩阵的子式一定是原矩阵子式，从而得证。
5. 去掉全为 0 的行/列不影响秩：子式只要选中了这些行/列，一定为 0，从而非零子式不可能取到它们，即得证删去后非零子式相同。
6. $\text{rank} \begin{pmatrix} A & O \\ O & B \end{pmatrix} = \text{rank } A + \text{rank } B$: 见书 3.5 节例 8 即为矩阵角度的证明。

* 用向量角度，可验证其 A 对应行的极大线性无关组与 B 对应行的极大线性无关组线性无关，共同构成全矩阵极大线性无关组。

7. $\text{rank}(A, B) \leq \text{rank } A + \text{rank } B$: 假设 A 通过行变换 P 列变换 Q 可成为 Hermite 标准形，对所有行作行变换 P ，可发现 B 的部分也进行了行变换 P ，因此成为

$$(PA, PB)$$

再对 A 所在的列作列变换 Q ，即可得到 ($r = \text{rank } A$)

$$\begin{pmatrix} I_r & O & (PB)_1 \\ O & O & (PB)_2 \end{pmatrix}$$

这里 $(PB)_1$ 指其前 r 行， $(PB)_2$ 指其后 $m - r$ 行 (设行数为 m)。

直接通过列变换可消去 $(PB)_1$ ，于是利用上个结论可知 $\text{rank}(A + B) = \text{rank } A + \text{rank}(PB)_2$ 。

而由于 $(PB)_2$ 是 PB 的子矩阵，即有 $\text{rank}(PB)_2 \leq \text{rank}(PB) = \text{rank } B$ (行变换不影响秩)，由此也可知等号取到当且仅当 $\text{rank}(PB)_2 = \text{rank}(PB)$ 。

最后，我们来解决一个较难的问题：证明

$$\text{rank} \begin{pmatrix} A_{s \times n} & C_{s \times m} \\ O & B_{l \times m} \end{pmatrix} \geq \text{rank } A + \text{rank } B$$

且等号成立当且仅当存在 $X_{n \times m}, Y_{s \times l}$ 使得 $C = AX + YB$ 。

* 书上 3.5 节的版本并没有等号成立条件，因为简单的分析方法是无法证明的。

我们设 $r = \text{rank } A$ 、 $t = \text{rank } B$ ，且 A 在行变换 P_A 列变换 Q_A 下成为 Hermite 标准形，且 B 在行变换 P_B 列变换 Q_B 下成为 Hermite 标准形。

将左侧矩阵的前 s 行进行行变换 P_A ，类似之前可发现成为 $(P_A A, P_A C)$ ，再将前 n 列进行列变换 Q_A ，可发现下方的 O 并未产生变化，由此得到

$$\begin{pmatrix} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} & P_A C \\ O & B \end{pmatrix}$$

将其后 m 列进行列变换 Q_B , 同样类似之前可发现右上角成为 $P_A C Q_B$, 再将后 l 行进行行变换 P_B , 最终得到

$$\begin{pmatrix} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} & P_A C Q_B \\ O & \begin{pmatrix} I_t & O \\ O & O \end{pmatrix} \end{pmatrix}$$

设

$$P_A C Q_B = \begin{pmatrix} S_1 & S_2 \\ S_3 & S_4 \end{pmatrix}$$

其中 S_1 为 $r \times t$ 阶, S_4 为 $(s-r) \times (m-t)$ 阶, 则可以写为

$$\begin{pmatrix} I_r & O & S_1 & S_2 \\ O & O & S_3 & S_4 \\ O & O & I_t & O \\ O & O & O & O \end{pmatrix}$$

利用 I_r 列变换消去 S_1 、 S_2 , 利用 I_t 行变换消去 S_3 , 最终得到

$$\begin{pmatrix} I_r & O & O & O \\ O & O & O & S_4 \\ O & O & I_t & O \\ O & O & O & O \end{pmatrix}$$

去掉为 0 的行列、适当交换后利用上方结论 6 即可发现其秩为 $r + t + \text{rank } S_4$ 。

由此, 大于等于号成立, 且等号成立当且仅当 $\text{rank } S_4 = 0$, 根据结论 3 即 $S_4 = O$ 。下面说明这当且仅当 $C = AX + YB$ 。

考虑方程 $C = AX + YB$, 利用矩阵乘法分配律 $P_A C Q_B = P_A A X Q_B + P_A Y B Q_B$ 。由于行变换可逆, 我们假设按照 Q_A 反向进行列变换对应变换阵为 \tilde{Q}_B , 应有对任意 M

$$M Q_B \tilde{Q}_B = M$$

同样, 假设按 P_B 反向进行行变换对应变换阵为 \tilde{P}_B , 应有对任意 N

$$\tilde{P}_B P_B N = N$$

由此

$$P_A C Q_B = P_A A X Q_B + P_A Y B Q_B = P_A A Q_A \tilde{Q}_A X Q_B + P_A Y \tilde{P}_B P_B B Q_B$$

而由条件这即是

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \tilde{Q}_A X Q_B + P_A Y \tilde{P}_B \begin{pmatrix} I_n & O \\ O & O \end{pmatrix}$$

我们记 $\hat{X} = \tilde{Q}_A X Q_B$ 、 $\hat{Y} = P_A Y \tilde{P}_B$, 并设 (注意变换阵都是方阵, 相乘不会改变阶数)

$$\hat{X} = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}, \quad X_1 \in \mathbb{R}^{r \times t}, \quad X_4 \in \mathbb{R}^{(n-r) \times (m-t)}$$

$$\hat{Y} = \begin{pmatrix} Y_1 & Y_2 \\ Y_3 & Y_4 \end{pmatrix}, \quad Y_1 \in \mathbb{R}^{r \times t}, \quad Y_4 \in \mathbb{R}^{(s-r) \times (l-t)}$$

直接利用矩阵乘法定义计算可知 (计算阶数可发现符合之前设为 S_1, S_2, S_3, S_4 的阶数)

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \hat{X} + \hat{Y} \begin{pmatrix} I_n & O \\ O & O \end{pmatrix} = \begin{pmatrix} X_1 + Y_1 & X_2 \\ Y_3 & O \end{pmatrix} = \begin{pmatrix} S_1 & S_2 \\ S_3 & S_4 \end{pmatrix}$$

当 $S_4 \neq 0$ 时此方程无解，而当 $S_4 \neq 0$ 时，取 $X_1 = S_1$ 、 $Y_1 = O$ 、 $X_2 = S_2$ 、 $Y_3 = S_3$ ，其余任取，即可得到 \hat{X}, \hat{Y} 的一组解。

注意到 \hat{X}, \hat{Y} 是由 X, Y 进行一些行变换与列变换得到的，将这些变换反向进行即可得到 X, Y 的解。
* 学完矩阵的逆后，我们会重新规范表述上方的证明。

四 补充：通用技巧

* 本部分讲义对应第一次补充习题课，因此会尽量避免涉及前文内容，必须涉及处会有索引。

顾名思义，这次习题课的主题是给矩阵的诸多技巧性很强的题目一些共同的通用技巧。之所以要讲这些，是因为教材的编排并不尽如人意：期中考试的范围是到 4.3 节，而 4.2、4.3 的作业中已经出现了大量矩阵相关的技巧题。然而，较为普遍的矩阵论方法在 4.4、4.5 与 5.2 节中方才涉及，空间方法则更后，这就导致在拥有后续知识前几乎无法总结前述问题的技巧与思路。

因此，这节习题课将简述矩阵与空间视角中现在可以学会并用到的（尽量少的）基本工具，并且以它们解决当前绝大部分的问题，以展示它们的普遍性与强大之处。

§4.1 基本工具

4.1.1 限制映射

我们已经知道，一个线性映射 $\mathcal{A} : U \rightarrow V$ 的像定义为

$$\text{Im } \mathcal{A} = \{y \mid \exists x, y = \mathcal{A}(x)\} \subset V$$

核定义为

$$\text{Ker } \mathcal{A} = \{x \mid \mathcal{A}(x) = 0\} \subset U$$

它们都是向量空间，且用它们可以刻画一个线性映射的基本性质，而其中最重要的维数等式为

$$\dim \text{Im } \mathcal{A} + \dim \text{Ker } \mathcal{A} = \dim U$$

对任何矩阵 $A_{m \times n}$ ，其可以自然看作线性映射（考虑向量为列向量）

$$\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad \mathcal{A}(x) = Ax$$

此后，我们将直接以对应的花体字母表示矩阵对应的线性映射，并不再特殊说明，无歧义时可记 $\mathcal{A}(x)$ 为 $\mathcal{A}x$ 。

* 由此，矩阵乘积 AB 看作的线性映射即为 \mathcal{A} 复合 \mathcal{B} ，这是矩阵乘法的线性映射视角，也是某种意义上最本质的一个。矩阵乘法结合律事实上来自映射复合的结合律。

* 注意 \mathbb{R}^n 与 \mathbb{R}^m 一般不同，因此 $\text{Im } \mathcal{A}$ 与 $\text{Ker } \mathcal{A}$ 无法进行比较或运算，但当 $m = n$ （即 A 为方阵）时，它们便能够一起处理，如考虑两空间的交与和等。

但是，上述的性质分析只能针对整体，我们希望能够考察一个映射的局部性质，这就引出了限制映射的定义。先考虑对一般的映射 $f : X \rightarrow Y$ ：

- 对任何 $X_0 \subset X$ ，可以定义限制映射 $f_{X_0 \rightarrow Y}$ ，满足

$$\forall x \in X_0, \quad f_{X_0 \rightarrow Y}(x) = f(x)$$

- 对任何 $f(X) \subset Y_0 \subset Y$ ，可以定义限制映射 $f_{X \rightarrow Y_0}$ ，满足

$$\forall x \in X, \quad f_{X \rightarrow Y_0}(x) = f(x)$$

这两个定义是非常直观的，前者相当于“缩小定义域”形成的新映射，后者则相当于“缩小陪域”(由于值域并未改变，这样的定义是非常合理的)形成的新映射。由于它们都反映了原映射的部分性质，我们将它们称为限制映射。

更一般地，我们可以先缩小定义域，再缩小值域：对任何 $X_0 \subset X$ 与 $f(X_0) \subset Y_0 \subset Y$ ，定义限制映射 $f_{X_0 \rightarrow Y_0}$ 满足

$$\forall x \in X_0, \quad f_{X_0 \rightarrow Y_0}(x) = f(x)$$

它即成为了 $X_0 \rightarrow Y_0$ 的映射。

对于线性映射，我们将上述的映射改为线性映射，并将“包含”改为“子空间”，即得到了线性映射的限制映射。它可用于反映线性映射的部分性质。

4.1.2 可逆与逆

可逆阵相关的理论非常多，也相对复杂，我们仅陈述能用到的最简单的定义与结论：

- 可逆定义：若对方阵 A ，存在同阶方阵 B 使得 $AB = BA = I$ ，则称 A 可逆， B 称为 A 的逆，记作 $B = A^{-1}$ ，由定义可发现 $(A^{-1})^{-1} = A$ ；
- 行列变换可逆性：任何初等方阵均可逆 (考虑行/列变换的描述可以得到逆的具体表达式)。
- 乘积可逆性：当 A 、 B 均可逆时， AB 可逆，且 $(AB)^{-1} = B^{-1}A^{-1}$ 。

上述一个定义、两个性质 (下称逆的性质 1、2) 已经足够解决秩的相关问题。严谨起见，我们对它们给出证明：

1. 直接计算验证可得

$$P(j, i(k))^{-1} = P(j, i(-k))$$

$$P(i, j)^{-1} = P(i, j)$$

$$P(i(c))^{-1} = P(i(c^{-1}))$$

2. 利用矩阵乘法结合律有

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

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

利用性质 1、2 可发现，任何一系列行列变换形成的矩阵都是可逆的，之后即将用此得到相抵标准形的结论。

在线性映射的视角下，由于 $Ix = x$ ，可逆即代表存在某个映射与其左右复合均为恒等映射，也即 A 可逆等价于 \mathcal{A} 可逆。

我们用线性映射的视角证明一个之后可能用到的性质。

先给出单侧逆的定义：若矩阵 $A_{m \times n}$ 与 $B_{n \times m}$ 满足 $AB = I_m$ ，则 A 称为 B 的左逆， B 称为 A 的右逆。

单侧逆在方阵时事实上可以推出双侧：若方阵 A 有左逆或右逆 B ，则 $B = A^{-1}$ 。

只需对左逆说明即可，右逆同理，我们分为两部分进行证明：

1. \mathcal{A} 将一组基映射到一组基。

设 A 阶数为 n , 从线性映射出发, 这即代表 $\mathcal{B}\mathcal{A}$ 为恒等映射。考虑 \mathbb{R}^n 的一组基 e_1, \dots, e_n , 并记 $a_i = \mathcal{A}e_i$ 。若 a_n 线性相关, 也即存在不全为 0 的 $\lambda_1, \dots, \lambda_n$ 使得

$$\sum_i \lambda_i a_i = 0$$

从而

$$\mathcal{A}\left(\sum_i \lambda_i e_i\right) = 0$$

利用 \mathcal{B} 为线性映射可知 $\mathcal{B}(0) = 0$, 从而

$$\mathcal{B}\mathcal{A}\left(\sum_i \lambda_i e_i\right) = 0$$

但根据定义其应为 $\sum_i \lambda_i a_i$, 利用线性无关性知非零, 矛盾。

2. $\mathcal{A}\mathcal{B}$ 为恒等映射。

由前一部分证明, a_1, \dots, a_n 也为 \mathbb{R}^n 的一组基, 且根据 $\mathcal{B}\mathcal{A}e_i = e_i$ 可知 $e_i = \mathcal{B}a_i$ 。将 \mathbb{R}^n 中任何元素表示为

$$\sum_i \mu_i a_i$$

则利用线性性可知

$$\mathcal{A}\mathcal{B}\left(\sum_i \mu_i a_i\right) = \sum_i \mu_i \mathcal{A}\mathcal{B}a_i = \sum_i \mu_i \mathcal{A}e_i = \sum_i \mu_i a_i$$

由此可知 $\mathcal{A}\mathcal{B}$ 亦为恒等映射, 得证。

4.1.3 相抵标准形

本节只介绍一个结论与它的一个基本应用, 更多用法即在后面的题目中:

对任何矩阵 $A_{m \times n}$, 存在可逆阵 $P_{m \times m}$ 与 $Q_{n \times n}$ 使得 (这称为 A 的相抵标准形)

$$A = P \begin{pmatrix} I_r & O_{(m-r) \times r} \\ O_{r \times (m-r)} & O_{(m-r) \times (n-r)} \end{pmatrix} Q$$

且 $r = \text{rank } A$ 。

证明: 存在一系列行变换将 A 变为简化阶梯形, 设这些行变换视为矩阵乘法后相乘而得的矩阵为 \tilde{P} 。进一步地, 将主元所在列换到前 r 列, 并以此消去 $r+1$ 到 n 列, 即可得到 $\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$ 的形式, 设这些列变换视为矩阵乘法后相乘得到的矩阵为 \tilde{Q} , 则有

$$\tilde{P} A \tilde{Q} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$$

利用逆的性质 1、2, 一系列行/列变换形成的矩阵可逆, 从而等式两边同时左乘 \tilde{P}^{-1} 、右乘 \tilde{Q}^{-1} 得到

$$A = \tilde{P}^{-1} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \tilde{Q}^{-1}$$

令 $P = \tilde{P}^{-1}$, $Q = \tilde{Q}^{-1}$ 即是结论。

* 这里用到了重要的技巧: 通过同乘其逆, 可逆阵可以在等式两边消去。

我们以此证明 $n \times n$ 矩阵 A 可逆当且仅当 $\text{rank } A = n$ 。

右推左：此时分解即为 $A = PQ$ ，于是 $A^{-1} = Q^{-1}P^{-1}$ ，得证可逆。

左推右：若否，可知 $\det A = 0$ ，但利用 Binet-Cauchy 公式，若 A 可逆有 $\det A \det A^{-1} = \det I = 1$ ，矛盾。

§4.2 习题讲解

4.2.1 秩相关的三种处理

事实上，上述的基本工具都可以用于处理秩相关的题目，我们以下面这道题为例，介绍不同处理手段的特点：

问题：求证

$$\forall A_{n \times m}, B_{m \times n}, \quad \text{rank}(I_n - AB) = n - m + \text{rank}(I_m - BA)$$

1. 限制映射出发证明

要想用线性映射的办法，一定要将目标尽量向 Ker 或 Im 的维数去凑。本题的结论可以自然变形为

$$n - \text{rank}(I_n - AB) = m - \text{rank}(I_m - BA)$$

这就出现了明显的解空间形式，将其写为

$$\dim \text{Ker}(\mathcal{I} - \mathcal{A}\mathcal{B}) = \dim \text{Ker}(\mathcal{I} - \mathcal{B}\mathcal{A})$$

利用定义也即要证明（注意 $\mathcal{I}x = x$ ）

$$\dim\{x \mid x = \mathcal{A}\mathcal{B}x\} = \dim\{y \mid y = \mathcal{B}\mathcal{A}y\}$$

将左侧空间记为 W_1 ，右侧为 W_2 ，可发现 W_1 为 \mathbb{R}^n 子空间，而 W_2 为 \mathbb{R}^m 子空间，因此无法通过包含关系进行维数比较，只能构造映射。

注意到， $x = \mathcal{A}\mathcal{B}x$ 存在某种“逆映射”的形式。利用之前关于限制映射的讨论，由于 W_1 上任何 x 有 $x = \mathcal{A}\mathcal{B}x$ ，可得

$$\mathcal{A}\mathcal{B}_{W_1 \rightarrow \mathbb{R}^m} = \mathcal{I}_{W_1}$$

记 $\mathcal{B}_1 = \mathcal{B}_{W_1 \rightarrow \mathbb{R}^m}$ ，上式这虽然不足以说明 \mathcal{B}_1 可逆，但已经有足够强的性质了。事实上， \mathcal{B}_1 必须为单射才能保证存在这样的形式：若 $\mathcal{B}_1(x) = \mathcal{B}_1(x')$ ，可发现 $x = \mathcal{A}(\mathcal{B}_1(x)) = \mathcal{A}(\mathcal{B}_1(x')) = x'$ 。

另一方面，可发现当 $x \in W_1$ 时

$$\mathcal{B}_1x = \mathcal{B}x = \mathcal{B}(\mathcal{A}\mathcal{B}_1x) = (\mathcal{B}\mathcal{A})(\mathcal{B}_1x)$$

这即说明了 $\mathcal{B}_1x \in W_2$ ，也即 $\mathcal{B}_1(W_1) \subset W_2$ ，定义 $\mathcal{B}_{12} = \mathcal{B}_{W_1 \rightarrow W_2}$ ，由 $x \in W_1$ 时 $\mathcal{B}_{12}x = \mathcal{B}_1x = \mathcal{B}x$ 可知 \mathcal{B}_{12} 为单射。

下面我们证明一个简单的引理： W_1 到 W_2 存在单射 \mathcal{B}_{12} 意味着 $\dim W_1 \leq \dim W_2$ 。若否， W_1 的一组基 a_1, \dots, a_s 的像 $\mathcal{B}_{12}a_1, \dots, \mathcal{B}_{12}a_s$ 在 W_2 中必然线性相关，也即存在不全为 0 的 λ_i 使得

$$\sum_{i=1}^s \lambda_i \mathcal{B}_{12}a_i = 0$$

也即

$$\mathcal{B}_{12} \left(\sum_{i=1}^s \lambda_i a_i \right) = 0$$

但由于 a_i 为一组基，括号中非零，再结合 $\mathcal{B}_{12}0 = 0$ 即与单射矛盾。

完全类似地，可定义限制映射 $\mathcal{A}_{21} = \mathcal{A}_{W_2 \rightarrow W_1}$ ，并可证明其为单射，从而 $\dim W_1 \geq \dim W_2$ ，与上一部分结合得到 $\dim W_1 = \dim W_2$ 。

2. 行列变换技巧出发证明

考虑加边的技巧，可发现

$$\text{rank}(I_n - AB) + m = \text{rank} \begin{pmatrix} I_m & O \\ O & I_n - AB \end{pmatrix}$$

通过第一类初等列变换可将右上角的 O 变为任何矩阵，为方便消去，变换为

$$\text{rank} \begin{pmatrix} I_m & B \\ O & I_n - AB \end{pmatrix}$$

下面，设 A 的行向量组为 $\alpha_1^T, \dots, \alpha_n^T$ ，则可发现

$$AB = \begin{pmatrix} \alpha_1^T B \\ \vdots \\ \alpha_n^T B \end{pmatrix}$$

由此右下角为

$$\begin{pmatrix} e_1 - \alpha_1^T B \\ \vdots \\ e_n - \alpha_n^T B \end{pmatrix}$$

注意到 $\alpha_1^T B$ 可以看作 B 的行向量的线性组合，利用行变换消去 $-AB$ ，并对应计算左下角的 O 中产生的系数，最终可以得到

$$\text{rank}(I_n - AB) + m = \text{rank} \begin{pmatrix} I_m & B \\ A & I_n \end{pmatrix}$$

同理

$$\text{rank}(I_m - BA) + n = \text{rank} \begin{pmatrix} I_n & A \\ B & I_m \end{pmatrix}$$

将上方矩阵后 n 行交换为前 n 行、后 n 列交换为前 n 列即可得到下方矩阵，从而两者秩相同，得证。

* 在学习了分块矩阵的初等变换计算后，上述证明能写成非常简单的形式

$$\begin{aligned} \begin{pmatrix} O & I_n \\ I_m & O \end{pmatrix} \begin{pmatrix} I_m & O \\ A & I_n \end{pmatrix} \begin{pmatrix} I_m & O \\ O & I_n - AB \end{pmatrix} \begin{pmatrix} I_m & B \\ O & I_n \end{pmatrix} \begin{pmatrix} O & I_m \\ I_n & O \end{pmatrix} &= \begin{pmatrix} I_n & A \\ B & I_m \end{pmatrix} \\ \begin{pmatrix} I_n & B \\ O & I_m \end{pmatrix} \begin{pmatrix} I_n & O \\ O & I_m - BA \end{pmatrix} \begin{pmatrix} I_n & A \\ O & I_m \end{pmatrix} &= \begin{pmatrix} I_n & A \\ B & I_m \end{pmatrix} \end{aligned}$$

但此形式的来源事实上即为上方的行列变换过程，这样简洁的式子是不可能凭空产生的。

3. 相抵标准形出发证明

设 $A = P\Sigma Q$ ，这里 $\Sigma = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$ ，且 P, Q 可逆，则

$$\text{rank}(I_n - AB) = \text{rank}(I_n - P\Sigma QB)$$

$$\text{rank}(I_m - BA) = \text{rank}(I_m - BP\Sigma Q)$$

利用行列变换不影响秩，可知乘一系列行列变换阵的乘积不影响秩，而行列变换阵的逆仍然能写成一系列行列变换阵的乘积，从而为统一形式，先消去 P 再增添 P 可发现

$$\text{rank}(I_n - P\Sigma QB) = \text{rank}(P^{-1}(I_n - P\Sigma QB)P) = \text{rank}(I_n - \Sigma QBP)$$

同理

$$\text{rank}(I_m - BP\Sigma Q) = \text{rank}(Q(I_m - BP\Sigma Q)Q^{-1}) = \text{rank}(I_m - QBP\Sigma)$$

* 这两步中用到了逆的另一个重要应用，即通过转移将一些东西变为相同的形式，从而可看作整体。

由此可设 $QBP = B'$ ，直接计算可发现 $\Sigma B'$ 为 B' 前 r 行不变，并添加 $n-r$ 行 0， $B'\Sigma$ 为 B' 的前 r 列不变，并添加 $m-r$ 列 0，从而设（注意只要标注左上角、右下角的矩阵阶数，其他阶数可自然得到）

$$B' = \begin{pmatrix} X_{r \times r} & Y \\ Z & W_{(m-r) \times (n-r)} \end{pmatrix}$$

则有

$$\text{rank}(I_n - AB) = \text{rank} \begin{pmatrix} I_r - X & Y \\ O & I_{n-r} \end{pmatrix}$$

$$\text{rank}(I_m - BA) = \text{rank} \begin{pmatrix} I_r - X & O \\ Z & I_{m-r} \end{pmatrix}$$

再利用行变换消去 Y 、列变换消去 Z ，得到

$$\text{rank}(I_n - AB) = \text{rank} \begin{pmatrix} I_r - X & O \\ O & I_{n-r} \end{pmatrix} = n-r + \text{rank}(I_r - X)$$

$$\text{rank}(I_m - BA) = \text{rank} \begin{pmatrix} I_r - X & O \\ O & I_{m-r} \end{pmatrix} = m-r + \text{rank}(I_r - X)$$

作差即得结论。

三种处理各有特点与难点：线性映射相关的处理需要熟悉线性映射与像与核的性质，最为抽象化，但避免了复杂的计算；行列变换相关的处理较有技巧性且需要对行列变换非常熟悉（最好会一些分块行列变换），但过程最为简洁；相抵标准形的处理计算最为复杂、需要熟悉一定的分块结论（如 $\Sigma B'$ 与 $B'\Sigma$ 的结果），但相对思维难度最小，且容易得过程分。

个人在秩相关的题中较为推荐第三种做法，因为它利用计算规避了很多思维难度。其过程基本上是：设题中某矩阵 $A = P\Sigma Q$ ，并想办法消去 P 、 Q （如合并到其他变量中并视为整体），只保留 Σ ，再在 Σ 的情况下通过计算直接得到结果。我们再看两道更简单的习题作为例子：

1. 满秩分解：若 $\text{rank}(A_{m \times n}) = r$ ，则存在 $B_{m \times r}, C_{r \times n}$ 使得 $A = BC$ 。

证明：设相抵标准形 $A = P\Sigma Q$ ，取

$$B = P \begin{pmatrix} I_r \\ O_{(m-r) \times r} \end{pmatrix}, \quad C = \begin{pmatrix} I_r & O_{r \times (n-r)} \end{pmatrix} Q$$

可直接计算验证成立。

* 书上习题有 $r = 1$ 的情况。

2. 矩阵 $A_{m \times n}$ 有左逆当且仅当 A 列满秩，有右逆当且仅当 A 列满秩（定义见前一部分）。

证明：只说明左逆的情况，右逆同理。

设相抵标准形 $A = P\Sigma Q$ ，若 A 列满秩，即知

$$\Sigma = \begin{pmatrix} I_n \\ O_{(m-n) \times n} \end{pmatrix}$$

直接构造

$$B = Q^{-1} \begin{pmatrix} I_n & O_{n \times (m-n)} \end{pmatrix} P^{-1}$$

可验证成立。

* 左侧放 Q^{-1} , 右侧放 P^{-1} 是为了与 $P\Sigma Q$ 乘法消去, 这类的构造是重要的。

反之, 若 A 不列满秩, 可知 Σ 有一列 0, 计算可得无论 C 为何, $C\Sigma$ 一定有一列 0, 于是 $BP\Sigma$ 有一列 0, 其秩不可能为 n , 而由列变换不改变秩可知 $\text{rank}(BP\Sigma Q) = \text{rank}(BP\Sigma)$, 不可能为 I 。

这两题事实上也都可以通过线性映射的思路解决, 我们这里只介绍思路, 想练习线性映射的同学可以自行尝试:

1. 也即 $\dim \text{Im } \mathcal{A} = r$, 不妨设其将 a_1, \dots, a_r 映射为了 $\text{Im } \mathcal{A}$ 的一组基, 构造 $\mathcal{C} : \mathbb{R}^n \rightarrow \mathbb{R}^r$ 使得 $\text{Ker } \mathcal{C} = \text{Ker } \mathcal{A}$ 且 $\mathcal{C}(a_i) = e_i$ (思考: 为何可如此规定?), $\mathcal{B} : \mathbb{R}^r \rightarrow \mathbb{R}^m$ 使得 $\mathcal{B}(e_i) = \mathcal{A}(a_i)$, 则可验证 $\mathcal{A} = \mathcal{B}\mathcal{C}$ 。
2. 上一部分已经证明 \mathcal{A} 有左逆时其 \mathcal{A} 为单射, 而其为单射时通过一组基的像可直接构造出左逆, 从而有左逆等价于其为单射。考虑原空间标准基的像, 证明单射等价于列满秩即可。

4.2.2 方阵的特性

本节中, 我们来用向量空间和矩阵的思路解决方阵的问题。如之前所说, 当 A 为方阵时, 其 $\text{Im } \mathcal{A}$ 与 $\text{Ker } \mathcal{A}$ 可视为同一空间中的, 由此可以有更复杂的性质。

我们首先给出直和的定义: 若同一向量空间的两子空间 U, V 满足 $V \cap U = \{0\}$, 则 $U + V$ 可记为 $U \oplus V$, 称为直和。

问题: 证明以下三者相互等价:

- a. $A^2 = A$;
- b. 设 A 的相抵标准形为 Σ , 存在 P 使得 $A = P\Sigma P^{-1}$;
- c. $\mathbb{R}^n = \text{Ker}(\mathcal{I} - \mathcal{A}) \oplus \text{Ker } \mathcal{A}$ 。

证明: 由于第二个描述显然为矩阵视角的描述, 第三个描述为空间视角, 而第一个条件是整体性的, 因此从矩阵、空间出发都可行。我们从矩阵论出发证明前两个等价, 再从空间出发证明一三等价。

1. b 推 a

计算得 $\Sigma^2 = \Sigma$, 从而有

$$A^2 = P\Sigma P^{-1}P\Sigma P^{-1} = P\Sigma^2 P^{-1} = A$$

2. a 推 b

设相抵标准形 $A = P\Sigma Q$, 由 $A^2 = A$ 可知

$$P\Sigma Q P\Sigma Q = P\Sigma Q$$

从而

$$\Sigma Q P \Sigma = \Sigma$$

与上一部分相抵标准形出发的证明类似, 将 QP 看作整体, 并假设

$$QP = \begin{pmatrix} X_{r \times r} & Y \\ Z & W_{(n-r) \times (n-r)} \end{pmatrix}$$

则计算可知

$$\Sigma Q P \Sigma = \begin{pmatrix} X & O \\ O & O \end{pmatrix} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$$

即 $X = I_r$ 。

为了能够进行下一步化简，我们拆分出作为整体的 QP ，得到

$$A = P \Sigma Q P P^{-1} = P \begin{pmatrix} I_r & Y \\ O & O \end{pmatrix} P^{-1}$$

由此，只需要最终消去 Y 即可得到结果。利用列变换消去 Y ，可写出对应的列变换阵为

$$\begin{pmatrix} I_r & Y \\ O & O \end{pmatrix} \begin{pmatrix} I_r & -Y \\ O & I_{n-r} \end{pmatrix} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} = \Sigma$$

而注意到右乘 Σ 相当于保留前 r 列，即有

$$\begin{pmatrix} I_r & Y \\ O & I_{n-r} \end{pmatrix} \begin{pmatrix} I_r & Y \\ O & O \end{pmatrix} \begin{pmatrix} I_r & -Y \\ O & I_{n-r} \end{pmatrix} = \begin{pmatrix} I_r & -Y \\ O & I_{n-r} \end{pmatrix} \Sigma = \Sigma$$

计算验证可得（注意到这即为第一类初等变换阵的形式，但将 Y 看作了整体）

$$\begin{pmatrix} I_r & Y \\ O & I_{n-r} \end{pmatrix}^{-1} = \begin{pmatrix} I_r & -Y \\ O & I_{n-r} \end{pmatrix}$$

从而最终得到

$$\begin{pmatrix} I_r & Y \\ O & O \end{pmatrix} = \begin{pmatrix} I_r & -Y \\ O & I_{n-r} \end{pmatrix} \Sigma \begin{pmatrix} I_r & Y \\ O & I_{n-r} \end{pmatrix}$$

于是

$$A = P \begin{pmatrix} I_r & -Y \\ O & I_{n-r} \end{pmatrix} \Sigma \begin{pmatrix} I_r & Y \\ O & I_{n-r} \end{pmatrix} P^{-1}$$

设 $\tilde{P} = P \begin{pmatrix} I_r & -Y \\ O & I_{n-r} \end{pmatrix}$ ，利用逆的性质 2 可验证 $A = \tilde{P} \Sigma \tilde{P}^{-1}$ ，得证。

* 这里消去 Y 的步骤本质上利用了分块矩阵的初等变换，期中范围内基本不会出现这样的技巧，随便看看就好。

3. c 推 a

由于 $\text{Ker}(\mathcal{I} - \mathcal{A}) = \{x \mid Ax = x\}$ 、 $\text{Ker} \mathcal{A} = \{y \mid Ay = 0\}$ ，由条件可知 \mathbb{R}^n 中任何向量 u 可写为 $x + y$ ，满足 $Ax = x$ 、 $Ay = 0$ ，从而

$$A^2 u = A^2 x + A^2 y = x = Ax + Ay = Au$$

这即说明了 $\mathcal{A}^2 = \mathcal{A}$ ，与 $A^2 = A$ 等价（考虑取 u 为 \mathbb{R}^n 一组基）。

4. a 推 c

要证明直和，需要先证明交为 $\{0\}$ 。

* 千万别写两个线性空间交为空！

由 Ker 定义，交中元素应满足 $Ax = x$ 、 $Ax = 0$ ，从而 $x = 0$ ，得证。

接下来证明和为全空间，观察 c 推 a 的证明过程，记

$$x = Au, \quad y = u - Au$$

有 $x + y = u$, 从而只需证明 $x \in \text{Ker}(\mathcal{I} - \mathcal{A})$ 、 $y \in \text{Ker} \mathcal{A}$ 即可, 而

$$(\mathcal{I} - \mathcal{A})x = (\mathcal{I} - \mathcal{A})\mathcal{A}u = \mathcal{A}u - \mathcal{A}^2u = 0$$

$$\mathcal{A}y = \mathcal{A}(u - \mathcal{A}u) = \mathcal{A}u - \mathcal{A}^2u = 0$$

从而得证。

事实上还能给出第四个等价条件: 由于 $A^2 = A$ 等价于对一切 x 有 $A(Ax) = Ax$, 而所有 Ax 构成 $\text{Im} \mathcal{A}$, 利用限制映射的思路, 即可得到这等价于

$$\mathcal{A}_{\text{Im} \mathcal{A} \rightarrow \text{Im} \mathcal{A}} = \mathcal{I}_{\text{Im} \mathcal{A}}$$

我们将在下一题中详细叙述限制映射的思路。

问题: 证明以下三者相互等价:

- a. $\text{rank } A^2 = \text{rank } A$;
- b. $\mathcal{A}_{\text{Im} \mathcal{A} \rightarrow \text{Im} \mathcal{A}}$ 可逆;
- c. $\mathbb{R}^n = \text{Im} \mathcal{A} \oplus \text{Ker} \mathcal{A}$.

证明: 由于题目中只出现了整体与空间, 我们自然会想到利用线性映射进行计算。不过, 为了展现矩阵方法的普适性, 我们仍然介绍一个通过相抵标准形计算的思路。

1. a 推 b

利用线性空间的视角, $\text{rank } A^2 = \text{rank } A$ 即为 $\dim \text{Im} \mathcal{A}^2 = \dim \text{Im} \mathcal{A}$ 。利用 Im 的定义, 应有

$$\text{Im} \mathcal{A}^2 = \mathcal{A}(\text{Im} \mathcal{A}) \subset \text{Im} \mathcal{A}$$

而从维数相同即得

$$\text{Im} \mathcal{A}^2 = \text{Im} \mathcal{A}$$

由此可知 $\mathcal{A}(\text{Im} \mathcal{A}) = \text{Im} \mathcal{A}$, 从而 $\mathcal{A}_{\text{Im} \mathcal{A} \rightarrow \text{Im} \mathcal{A}}$ 为满射。

下面只需要说明, 相同维度向量空间 U, V 之间的满射 f 是可逆的。

考虑 V 的一组基 v_1, \dots, v_r , 由于 f 是 $U \rightarrow V$ 的满射, 一定存在 u_1, \dots, u_r 满足 $f(u_i) = v_i$ 。

若 u_1, \dots, u_r 线性相关, 存在非零 λ_i 使得 $\sum_i \lambda_i u_i = 0$, 从而

$$\sum_i \lambda_i v_i = \sum_i \lambda_i f(u_i) = f\left(\sum_i \lambda_i u_i\right) = f(0) = 0$$

与 $\{v_i\}$ 为一组基矛盾。因此 $\{u_i\}$ 必然线性无关, 由维数相同可知它们为 U 的一组基。

由于 f 将一组基映射到一组基, 与之前完全类似可验证 f 是单射, 从而其可逆, 得证。

2. b 推 c

注意到, 由于 $\dim \text{Im} \mathcal{A} + \dim \text{Ker} \mathcal{A} = n$, 可知

$$n = \dim(\text{Im} \mathcal{A} + \text{Ker} \mathcal{A}) + \dim(\text{Im} \mathcal{A} \cap \text{Ker} \mathcal{A})$$

从而只要证明了 $\text{Im} \mathcal{A} + \text{Ker} \mathcal{A} = \mathbb{R}^n$, 其交必然为 $\{0\}$, 是直和。

利用上一题 a 推 c 的思路, 对 $u \in \mathbb{R}^n$, 由于 $\mathcal{A}_{\text{Im} \mathcal{A} \rightarrow \text{Im} \mathcal{A}}$ 可逆, 存在唯一 $x \in \text{Im} \mathcal{A}$ 使得 $\mathcal{A}x = \mathcal{A}u$, 记 $y = u - x$, 则

$$\mathcal{A}y = \mathcal{A}u - \mathcal{A}x = 0$$

从而 $y \in \text{Ker } \mathcal{A}$, 而假设中已知 $x \in \text{Im } \mathcal{A}$, 即得到 $\mathbb{R}^n = \text{Im } \mathcal{A} + \text{Ker } \mathcal{A}$ 。

* 若没有 b, 直接证明 a 推 c, 一个比较自然的想法是考虑 $\text{Im } \mathcal{A}$ 与 $\text{Ker } \mathcal{A}$ 各一组基后证明线性无关, 这本质上还是会说明限制映射是满射。

3. c 推 a

与 a 推 b 类似, 只需证明 $\mathcal{A}(\text{Im } \mathcal{A}) = \text{Im } \mathcal{A}$ 。

对任何 $x \in \text{Im } \mathcal{A}$, 设 $x = \mathcal{A}y$, 由条件可知存在 z, w 使得 $z \in \text{Im } \mathcal{A}, w \in \text{Ker } \mathcal{A}$, $y = z + w$, 于是

$$\mathcal{A}z = \mathcal{A}(y - w) = x - 0 = x$$

即可得到任何 $\text{Im } \mathcal{A}$ 中元素有 $\text{Im } \mathcal{A}$ 中原像, 从而 $\mathcal{A}_{\text{Im } \mathcal{A} \rightarrow \text{Im } \mathcal{A}}$ 是满射, 得证。

4. a 推 c (矩阵方法)

设相抵标准形 $A = P\Sigma Q$, 则与上题 a 推 b 完全相同, 设 $QP = \begin{pmatrix} X_{r \times r} & Y \\ Z & W_{(n-r) \times (n-r)} \end{pmatrix}$ 算得

$$A^2 = P\Sigma QP\Sigma Q = P \begin{pmatrix} X & O \\ O & O \end{pmatrix} Q$$

由于行列变换不影响秩, P, Q 均为行列变换, 可知

$$\text{rank } A^2 = \text{rank } X = r$$

根据之前的证明, r 阶方阵 X 秩为 r 意味着 X^{-1} 存在。

类似写出

$$A = P \begin{pmatrix} X & Y \\ O & O \end{pmatrix} P^{-1}$$

下面计算 $\text{Im } \mathcal{A}$ 与 $\text{Ker } \mathcal{A}$ 。

若 $\mathcal{A}x = 0$, 可得

$$P \begin{pmatrix} X & Y \\ O & O \end{pmatrix} P^{-1}x = 0$$

两侧同时左乘 P^{-1} , 并设 $y = P^{-1}x = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, 其中 y_1 为 r 行, 得到

$$\begin{pmatrix} X & Y \\ O & O \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 0$$

直接计算可发现这代表着 $Xy_1 + Yy_2 = 0$, 于是 y_2 可任取, 对应的 $y_1 = -X^{-1}Yy_2$, 最终得到

$$\text{Ker } \mathcal{A} = \left\{ P \begin{pmatrix} -X^{-1}Yy_2 \\ y_2 \end{pmatrix}, \quad y_2 \in \mathbb{R}^{n-r} \right\}$$

对于 $\text{Im } \mathcal{A}$, 其为

$$\left\{ P \begin{pmatrix} X & Y \\ O & O \end{pmatrix} P^{-1}x, \quad x \in \mathbb{R}^n \right\}$$

设 $y = P^{-1}x$, 则 $x = Py$, 由此 x 与 y 一一对应, 可改写为

$$\left\{ P \begin{pmatrix} X & Y \\ O & O \end{pmatrix} y, \quad y \in \mathbb{R}^n \right\}$$

与之前类似分为 y_1, y_2 , 可发现其为

$$\left\{ P \begin{pmatrix} Xy_1 + Yy_2 \\ 0 \end{pmatrix}, \quad y \in \mathbb{R}^n \right\}$$

由于对任何 $t \in \mathbb{R}^r$, 取 $y_1 = X^{-1}r, y_2 = 0$ 即可得到 $Xy_1 + Yy_2 = t$, 也即上方的分量实际上可任取, 得到

$$\text{Im } \mathcal{A} = \left\{ P \begin{pmatrix} t \\ 0 \end{pmatrix}, \quad t \in \mathbb{R}^r \right\}$$

只要证明 $\mathbb{R}^n = \text{Ker } \mathcal{A} + \text{Im } \mathcal{A}$ 即可, 而对任何 $u \in \mathbb{R}^n$, 设 $v = P^{-1}u = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, 则取

$$p = P \begin{pmatrix} -X^{-1}Yv_2 \\ v_2 \end{pmatrix}$$

$$q = P \begin{pmatrix} v_1 + X^{-1}Yv_2 \\ 0 \end{pmatrix}$$

根据形式可知 $p \in \text{Ker } \mathcal{A}, q \in \text{Im } \mathcal{A}$, 且 $p + q = Pv = u$, 从而得证。

* 虽然相比线性空间方法, 这里的计算显著更加复杂, 但其实总体思维难度是更低的, 顺序计算即可, 最后的构造也相对自然。这里用到的重要技巧: 可逆矩阵作为映射也是可逆的, 从而可以方便地进行代换。

* 事实上, 类比前一部分的证明, 可知存在可逆矩阵 P 使得

$$A = P \begin{pmatrix} X & O \\ O & O \end{pmatrix} P^{-1}$$

最后, 我们来看两道构造性的题目, 观察如何从线性映射基与矩阵标准形出发构造。

问题: 证明对任何矩阵 $A_{m \times n}$, 存在 $B_{n \times m}$ 使得 $ABA = A$ 。

证明:

1. 线性映射角度

也即找 \mathcal{B} 使得 $\mathcal{AB}(\mathcal{Ax}) = (\mathcal{Ax})$ 对任何 x 成立, 而这即代表 (注意根据定义可知 $\text{Im } \mathcal{AB} \subset \text{Im } \mathcal{A}$, 从而可以如此限制)

$$(\mathcal{AB})_{\text{Im } \mathcal{A} \rightarrow \text{Im } \mathcal{A}} = \mathcal{I}_{\text{Im } \mathcal{A}}$$

于是, 设 $\text{Im } \mathcal{A}$ 一组基为 x_1, \dots, x_r , 取 y_1, \dots, y_r 使得 $\mathcal{A}y_i = x_i$, 由线性无关性可以构造 \mathcal{B} 使得 $\mathcal{B}x_i = y_i$ (将 $\{x_i\}$ 扩充为 $\text{Im } \mathcal{A}$ 一组基, \mathcal{B} 在其他基上的像任取, 在 x_i 上的像指定为 y_i), 从而即可验证 $\mathcal{AB}A = \mathcal{A}$ 。

* 注意基可以确定线性映射, 因此构造映射只须指定基的像。

* 若令 \mathcal{B} 将其他基都映射到 0, 可发现 $\mathcal{B}\mathcal{AB} = \mathcal{B}$ 也成立。

2. 矩阵角度

设相抵标准形 $A = P\Sigma Q$, 考虑 $B = Q^{-1}Y P^{-1}$ 的形式, 若原命题满足, 可发现 $\Sigma Y \Sigma = \Sigma$, 根据之前经历的计算, 只要取 Y 左上角为 I_r 即可。

* 若取 $Y = \Sigma$, 可发现 $BAB = B$ 也成立。

* 满足 $ABA = A$ 、 $BAB = B$ 的 B 称为 A 的广义逆。若还有 AB 、 BA 均对称 (或对复矩阵 Hermite)，则称为 **Moore-Penrose 广义逆**，MP 广义逆对任何矩阵存在唯一。

问题：证明若 $A_{m \times n}, B_{n \times p}$ 满足 $\text{rank}(AB) = \text{rank } A$ ，则存在 $C_{p \times n}$ 使得 $ABC = A$ 。

证明：

1. 线性映射角度

由 $\dim \text{Im } AB = \dim \text{Im } A$ 与 $\text{Im } AB \subset \text{Im } A$ 可知 $\text{Im } AB = \text{Im } A$ 。

而利用定义 $\text{Im } AB = A(\text{Im } B)$ ，这也即是说，对 $\text{Im } A$ 一组基 a_1, \dots, a_r ，存在 $b_1, \dots, b_r \in \text{Im } B$ 使得 $Ab_i = a_i$ 。

设 c_1, \dots, c_r 使得 $Bc_i = b_i$ 。与 $\text{rank } A^2 = \text{rank } A$ 推 $A_{\text{Im } A \rightarrow \text{Im } A}$ 可逆相似论证，可以从 $\{a_i\}$ 线性无关得到 $\{b_i\}$ 线性无关，于是可以构造 C 使得 $Cb_i = c_i$ ，由此计算得

$$ABCb_i = Ab_i$$

下面证明，一定可以把 $\{b_i\}$ 扩充成 \mathbb{R}^n 一组基使得增添的 $\alpha_1, \dots, \alpha_{n-r}$ 构成 $\text{Ker } A$ 一组基，从而只要定义 $C\alpha_j = 0$ 即可计算验证 $ABC = A$ 。

设 b_1, \dots, b_r 生成的子空间为 S ，则由线性无关性 $\dim S = r$ ，而 $\dim \text{Ker } A = n - r$ ，从而只要说明 $S + \text{Ker } A = \mathbb{R}^n$ ，即有两者交为 $\{0\}$ ，能进行如上扩充。另一方面，对任何 $u \in \mathbb{R}^n$ ，由 b_1, \dots, b_r 映为 $\text{Im } A$ 一组基可取 $b \in S$ 使得 $Au = Ab$ ，从而 $A(u - b) = 0$ ， $u - b \in \text{Ker } A$ ，这就得到了证明。

* 若两子空间交为 $\{0\}$ ，和为全空间，则称它们互为补空间。上面的证明说明了 S 与 $\text{Ker } A$ 互补。

2. 矩阵角度

设相抵标准形 $A = P\Sigma Q$ ，并设 $B = Q^{-1}XP^{-1}$ ，则利用行列变换不改变秩可知

$$\text{rank}(AB) = \text{rank}(\Sigma X) = r$$

而利用之前计算， ΣX 为 X 前 r 行不变，之后变成 0，这即说明了 X 的前 r 行是线性无关的，设

$$X = \begin{pmatrix} Y_{r \times p} \\ Z_{(n-r) \times p} \end{pmatrix}$$

即有 Y 行满秩。根据上一部分证明的结论，存在右逆 $W_{p \times r}$ 使得 $YW = I_r$ ，于是构造

$$C = P \begin{pmatrix} W & O_{p \times (n-r)} \end{pmatrix} Q$$

可计算验证成立。

* 上述矩阵与线性映射的方法似乎很有关联性。大致来说，相似标准形似乎与取基有关，之后的学习中将学到更明确的关联。一定记住，只要掌握一条路径，就基本可以处理出所有题目。而考虑到与转置相关的题目难以通过线性映射进行处理，仍然较推荐掌握矩阵操作的方法。

4.2.3 无穷维的线性算子

* 本节为拓展内容。

我们之前的线性空间讨论都在有限维之中，而有限维的讨论一定可以归于矩阵技巧，因此线性映射似乎不存在特殊的效果。不过，在无穷维，情况就有所改变了。由于我们并未学到严谨的线性空间定义，当前只作简单说明。我们的目标是求解微分方程

$$y''(x) + y'(x) - 2y(x) = e^x$$

考虑映射 $D : y(x) \rightarrow y'(x)$, 则它是一个函数到函数的映射 (可以称为算子), 且具有线性性:

$$D(\lambda y(x) + \mu z(x)) = \lambda y'(x) + \mu z'(x)$$

类似矩阵的运算 (也与函数的加法、数乘相似), 我们可以定义线性算子的加法、数乘、乘法:

$$(D_1 + D_2)(y(x)) = D_1 y(x) + D_2 y(x)$$

$$(\lambda D_1)y(x) = \lambda D_1 y(x)$$

$$(D_1 D_2)(y(x)) = D_1(D_2(y(x)))$$

这时, 原方程即可以化为

$$(D^2 + D - 2I)y(x) = e^x$$

这里 $Iy = y$ 代表恒等算子。

它似乎形式上成为了一个非齐次线性方程组, 也的确有非齐次方程组解的性质: 其通解可以写为一个特解加上对应的齐次线性方程组 $(D^2 + D - 2I)y(x) = 0$ 的通解, 证明与线性方程组情况完全相同。

更具体的计算是数分内容, 此处略过, 不过, 利用定义可发现

$$D(f(x)e^{ax}) = ((D + aI)f(x))e^{ax}$$

由此取 $a = 1$ 可发现 $f(x) = \frac{1}{3}x$ 对应的 $\frac{1}{3}xe^x$ 是原方程特解, 再计算齐次线性方程通解为 $C_1 e^x + C_2 e^{-2x}$ 可得原方程通解。

五 期中总结

§5.1 习题解答

1. 习题 4.1-7

设 $K = A - \lambda I_n$, 利用本节例 9 可知 K^p 当 $p \geq n$ 时为 0, 否则当且仅当 $j - i = p$ 时 K^p 的第 i 行第 j 列为 0。

利用二项式定理与 I 同任何矩阵可交换, 直接展开, 注意 K^n 后的项为 0, 可知

$$A^m = (\lambda I + K)^m = \sum_{p=0}^{\min(n-1, m)} C_m^p \lambda^{m-p} K^p$$

2. 习题 4.1-15

设 $C = AB$, 各元素为 c_{ij} , 设第 p 行为 0, 利用定义可知

$$c_{pj} = \sum_{k=1}^n a_{pk} b_{kj} = 0$$

由于所有元素非负, 可知 a_{pk} 、 b_{kj} 中至少有一个为 0, 而这对任何 k, j 成立。

若 b 的某行全为 0, 则结论成立, 否则对每个 k , 存在 j 使得 $b_{kj} \neq 0$, 从而 $a_{pk} = 0$, 这即代表 A 的第 p 行为 0。

3. 习题 4.2-8

直接取 X 左上角为 1, 其余为 0 可验证成立。

* 构造方式可以从本讲义第四章中 $B\Sigma$ 与 ΣB 的乘积结论中联想到, 建议熟悉 $\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$ 与其他矩阵乘法的结果, 因为它们和初等方阵的乘积可以形成任何矩阵 (相抵标准形), 由此任何矩阵乘法可以拆分为这些部分的计算。

4. 习题 4.2-5

设 $A_{m \times n}$, 利用转置定义直接计算 (用下标表示分量)

$$(AA^T)_{ij} = \sum_{k=1}^n a_{ik}a_{jk} = 0$$

当 $i = j$ 时, 这即意味着所有 a_{ik} 平方后对 k 求和为 0, 也即所有 a_{ik} 为 0。利用 i 可任取即得结论。

5. 习题 4.3-14

空间视角: 当 $\text{rank}(AB) = \text{rank } B$ 时, 有

$$\dim \text{Im } (AB) = \dim \text{Im } B$$

从而

$$m - \dim \text{Im } (AB) = m - \dim \text{Im } B$$

即

$$\dim \text{Ker } (AB) = \dim \text{Ker } B$$

而利用定义可发现

$$\text{Ker } B \subset \text{Ker } (AB)$$

于是即有 $\text{Ker } AB = \text{Ker } B$, 这即是结论。

矩阵视角: 与空间视角完全类似得到 $ABx = 0$ 与 $Bx = 0$ 的解空间维数相同, 而 $Bx = 0$ 的解一定为 $ABx = 0$ 的解, 因此后者包含于前者, 再由维数相同知相等。

6. 习题 4.3-15

空间视角: 与习题 4.3-14 相同, 两侧被 r 减后只需验证

$$\text{Ker } (ABC) = \text{Ker } (BC)$$

而利用定义可发现

$$\text{Ker } (ABC) = \{x \mid ABCx = 0\} = \{x \mid Cx \in \text{Ker } (AB)\}$$

$$\text{Ker } (BC) = \{x \mid BCx = 0\} = \{x \mid Cx \in \text{Ker } B\}$$

于是利用习题 4.3-14 得到的结论即知 $\text{Ker } (ABC) = \text{Ker } (BC)$ 。

矩阵视角: 由于 $ABx = 0$ 与 $Bx = 0$ 的解相同, 令 $x = Cy$ 可由定义验证 $ABCy$ 与 BCy 的解相同, 从而解空间维数相同, 再利用解空间维数定理可知秩相同。

* 由于 AB 与 B 的定义域相同, 一般考虑包含在定义域中的 Ker , 而 AB 与 A 陪域相同, 一般考虑包含在陪域中的 Im 。

7. 习题 4.3-5

* 注意本讲义最前的记号说明。Hermite 矩阵的概念很重要, 对应实数中的对称阵。

利用 $(AB)^T = B^T A^T$, 根据共轭的定义可发现转置与共轭可交换, 从而

$$(BB^H)^T = (\overline{B}^T)^T B^T = \overline{B}B^T$$

而利用 $\overline{ab} = \overline{a}\overline{b}$ 与 $\overline{a+b} = \overline{a} + \overline{b}$, 即可发现 $\overline{AB} = \overline{A}\overline{B}$, 从而注意到共轭两次得到自身可发现

$$(BB^H)^H = \overline{\overline{B}B^T} = BB^H$$

对 $B^H B$ 完全同理。

8. 习题 4.3-6

与本节例 3 类似, 只需证明 $B^H Bx = 0$ 与 $Bx = 0$ 同解即可。根据定义后者的解一定是前者的解, 而对前者的解 y , 左乘 y^H 有 (注意 $(AB)^H = B^H A^H$ 可类似习题 4.3-5 证明成立)

$$y^H B^H B y = (B y)^H (B y) = 0$$

对于复向量 z , 可发现 $z^H z$ 为每个位置模长的平方和, 从而只能 $B y = 0$, 即得证 $\text{rank}(B^H B) = \text{rank } B$ 。对上式代入 B^H 并利用两次共轭转置回到自身 (分别考虑共轭与转置) 可知

$$\text{rank}(B B^H) = \text{rank } B^H = \text{rank } B$$

从而得证。

* 这里事实上还需要证明 $\text{rank } B^H = \text{rank } B$, 由于已有 $\text{rank } B^T = \text{rank } B$, 还需证 $\text{rank } \bar{B} = \text{rank } B$ 。通过将 B 拆分出实部与虚部可验证向量组取共轭后线性相关/线性无关性不变, 因此得证。

9. 习题 4.3-17

直接计算

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2 + bc & ab + bd \\ ca + dc & d^2 + bc \end{pmatrix} = O$$

由右上、左下元素可知, 或 $a + d = 0$, 或 b, c 均为 0, 分类讨论。当 $b = c = 0$ 时, 即可知 $a^2 = d^2 = 0$, 从而 $a = d = 0$, $a + d = 0$ 也成立, 因此只需考虑 $a + d = 0$ 的情况。此时, 利用 $d^2 = a^2$, 四个方程化为 $a^2 + bc = 0$, 因此平方等于 O 当且仅当

$$\begin{cases} d = -a \\ bc = -a^2 \end{cases}$$

即 $a = 0$ 时 bc 至少有一个为 0, 或 $a \neq 0$ 时 $b \neq 0$, $d = -a$, $c = -\frac{a^2}{b}$ 。

10. 补充题三-2

见本讲义 3.2.1。

11. 习题 4.3-4

仍与本节例 9 类似, 只需说明 $AA^T Ax = 0$ 与 $Ax = 0$ 同解。根据定义后者的解一定是前者的解, 而对前者的解 y , 左乘 $y^T A^T$ 可得到

$$y^T A^T A A^T A y = 0$$

计算得即

$$(A^T A y)^T (A^T A y) = 0$$

从而完全类似得证。

* 也可以利用相抵标准形进行计算。设 $A = P \Sigma Q$, 则计算后消去两侧的行列变换阵可发现

$$\text{rank}(AA^T A) = \Sigma Q Q^T \Sigma P^T P \Sigma$$

它的左上角 r 阶子方阵为 $Q Q^T$ 左上角 r 阶子方阵乘 $P^T P$ 左上角 r 阶子方阵, 分别利用 $\text{rank } AA^T = \text{rank } A$ 与 $\text{rank } A^T A = \text{rank } A$ 可知两者秩都为 r , 从而乘积秩为 r , 得证。

不过, 这个做法需要的前置知识较多, 如乘满秩方阵不改变秩, 与一些分块计算基本知识。事实上, 用相抵标准形处理转置问题虽然不像线性映射一样完全不可行, 但也绝对谈不上简洁。真正较为一般的方法是奇异值分解, 之后可能会学到。

* 也可利用 $\text{rank } A^T A = \text{rank } A$ 得到

$$\text{rank}(A^T A A^T A) = \text{rank}((A^T A)^T A^T A) = \text{rank}(A^T A) = \text{rank } A$$

再利用乘积秩不等式证明。

§5.2 拾遗

5.2.1 分块矩阵

事实上，我们之前已经多次将矩阵写为分块的形式了（并且主要是分为四块）。它的作用是，可以将分块矩阵作为某个“元素”进行整体的乘法。我们只介绍一个最基本的定理，它足以完成之前的所有推导：

$$\begin{pmatrix} A_{a \times b} & B_{a \times d} \\ C_{c \times b} & D_{c \times d} \end{pmatrix} \begin{pmatrix} X_{b \times p} & Y_{b \times q} \\ Z_{d \times p} & W_{d \times q} \end{pmatrix} = \begin{pmatrix} AX + BZ & AY + BW \\ CX + DZ & CY + DW \end{pmatrix}$$

事实上，实际使用这个定理时，完全无需对维数进行复杂的验证，因为只要右侧每个乘法都可以进行，分块矩阵的乘积一定可以看作元素进行乘积。

* 注意矩阵乘法不可交换，等式右侧不能交换乘法顺序。

利用这个结论就可以验证一些计算，例如

$$\begin{pmatrix} I_a & Y \\ O & I_b \end{pmatrix}^{-1} = \begin{pmatrix} I_a & -Y \\ O & I_b \end{pmatrix}$$

或 4.2.1 节第一题行列变换出发的巧妙证明

$$\begin{pmatrix} O & I_n \\ I_m & O \end{pmatrix} \begin{pmatrix} I_m & O \\ A & I_n \end{pmatrix} \begin{pmatrix} I_m & O \\ O & I_n - AB \end{pmatrix} \begin{pmatrix} I_m & B \\ O & I_n \end{pmatrix} \begin{pmatrix} O & I_m \\ I_n & O \end{pmatrix} = \begin{pmatrix} I_n & A \\ B & I_m \end{pmatrix}$$

$$\begin{pmatrix} I_n & B \\ O & I_m \end{pmatrix} \begin{pmatrix} I_n & O \\ O & I_m - BA \end{pmatrix} \begin{pmatrix} I_n & A \\ O & I_m \end{pmatrix} = \begin{pmatrix} I_n & A \\ B & I_m \end{pmatrix}$$

上述几个算式留给读者自行验证。拥有了分块变换技巧后，相抵标准形出发的证明才真正完整。

* 由此，看完讲义的 4.1.2、4.1.3 与本节后，利用相抵标准形直接计算就可以成为一个备选的方法，无论如何都能通过合理的过程混到一些分。

5.2.2 行列式技巧 II

比起上次自成一节，行列式技巧 II 只剩下一小节了，这是因为事实上此处只有一个需要记忆的公式，即 Binet-Cauchy 公式的方阵版本：

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

要想用好这个公式，必须对矩阵乘法的累加定义十分熟悉：

$$(AB)_{ij} = \sum_k a_{ik} b_{kj}$$

这个定义也即相当于在 i, j 之间“插入”了符号 k ，并对 k 求和。事实上这个操作对于多个矩阵连乘也成立，大家可以自行验证：

$$(ABC)_{ij} = \sum_{k,l} a_{ik} b_{kl} c_{lj}$$

由此，当我们看到求和符号时，就可以考虑这样进行拆分了。考察如下例子：给定 x_1, \dots, x_n ，并且假设 n 阶行列式的第 i 行第 j 列为 $x_1^{i+j-1} + x_2^{i+j-1} + \dots + x_n^{i+j-1}$ ，求行列式。

设其为 D , 有

$$(D)_{ij} = \sum_{k=1}^n x_k^{i+j-1}$$

看作 $a_{ik}b_{kj}$ 的形式, 可发现能取 $a_{ik} = x_k^i$ 、 $b_{kj} = x_k^{j-1}$ 。矩阵 A 在第 k 列提出 x_k 后成为 Vandermonde 阵, 而矩阵 B 直接为 Vandermonde 阵的转置, 由此行列式即为

$$\det D = \det A \det B = \prod_{k=1}^n x_k \prod_{i < j} (x_j - x_i)^2$$

* 一般来说, 即使给的矩阵是以省略号写出的, 只要每项有 n 个数求和, 都可用 $(D)_{ij} = \sum_k f(i, j, k)$ 的形式写出来, 再观察能否拆分为矩阵乘法。

* 需要拆分时往往都和 Vandermonde 行列式有很大关系, 因此须记熟其结论。

事实上, 行列式除了可以看作有向体积外, 还可以成为矩阵对应的线性变换的有向体积放大率。由此, $\det(AB) = \det A \det B$ 在线性映射的视角下是自然的: 复合的放大率自然是放大两次的结果。

* 很多优美的公式都存在一个自然的视角, 更多行列式相关的几何可以参考 3Blue1Brown 的线性代数的本质系列视频, 但个人并不推荐过于依赖几何直观——代数必须严谨。例如, 如果要用上面的有向体积进行刻画, 我们至少要严谨定义体积、证明行列式确为体积、证明对所有立体体积放大率一致, 这些又成为了细致的步骤。

* 到这里行列式计算的相关内容就基本告一段落了。然而, 在行列式结合方程后, 还会出现更有意思的结果, 这就是矩阵的特征值。在学习特征值相关知识后, 还会存在一个非常少用但的确可用的行列式计算方法, 如果这学期能讲到的话会在那时说明。

§5.3 知识与技巧整理

* 以节为单位, 梳理重要的定理与技巧, 并只介绍基本练习题, 更综合的内容将在考前复习课介绍。直接应用算法的简单计算题不在这里列出, 但总得练练。

5.3.1 线性方程组与行列式

1.1 矩阵消元法

定义: 三类初等行变换、矩阵/方阵/零矩阵/元/矩阵的相等、齐次/非齐次线性方程组、矩阵表示

* 主元一类的概念大概看看即可, 在之后利用矩阵描述后很难再出现了。

算法: 将矩阵初等行变换化为简化阶梯形矩阵

1.2 解的判别

算法: 通过简化阶梯形矩阵确定解的形式

定理: 齐次线性方程组方程个数小于未知数个数时必有解

练习: 1.2.2 例 4 (分类讨论技巧, 每次出现未知数在主元/需要被除时进行讨论)

1.3 数域

定义: 数域 (绝大部分时候考虑 \mathbb{R} 即可, 一般的 \mathbb{K} 可完全类似)

2.1 n 元排列

定义: n 元排列、逆序数

算法: 通过需要的对换 (注意无需相邻) 次数计算排列奇偶性

2.2 行列式完全展开

定义: n 阶行列式 (完全展开)、上/下三角阵 (下三角类似上三角定义)

算法: 完全展开式计算行列式

练习: 习题 2.2.4 (符号判断非常重要)、习题 2.3.4(1) (非零元素少的矩阵可尝试直接展开)、应用小天地例 2 (求和次序交换技巧, 很多需要求和展开的场合都会用到)

2.3 行列式性质

定义: 转置

算法: 变换为上三角阵计算行列式 (对一般行列式的最快计算方法)

定理: 转置不影响行列式、三种初等行变换对行列式的影响

练习: 习题 2.3.4(3) (尽可能消去容易消去的部分)、2.3.2 例 3 (拆分技巧)、习题 2.3.2(1) (另一种思路为加边技巧, 详见本讲义 1.3.2)

2.4 行列式按行展开

定义: 余子式、代数余子式

算法: 按行展开计算行列式

定理: 错位乘积求和结果为 0 (书上定理 3、4)、范德蒙德行列式 (记忆形式与结果)

练习: 2.4.2 例 6 (本质与完全展开有类似处, 但存额外的构造递推策略)

2.5 Cramer 法则

定理: Cramer 法则 (被之后线性空间对解的讨论涵盖, 基本无需单独记忆/练习)

2.6 行列式多行展开

定理: $\det \begin{pmatrix} A & O \\ C & B \end{pmatrix} = \det A \det B$ 、Laplace 定理 (优先级很低, 几乎所有题目都能规避)

* 为什么说是“几乎”: 如习题 4.3-13 就不太可规避。

5.3.2 向量空间

3.1 向量空间概念

定义: 向量空间、加法、数乘、线性表出、线性组合、子空间、生成 (生成的概念非常重要)

定理: 线性方程组与线性表出的等价性 (书上命题 1)

3.2 线性相关/线性无关

定义: 线性相关、线性无关

算法: 是否线性相关的判定 (利用线性方程组)

定理: 线性相关与线性无关向量组的一大堆性质

* 比较重要的是 (3) 线性方程组视角、(4) n 个向量是否相关与行列式关系 (利用 Cramer 法则) 与 (7) 线性表出视角, 其他结论也时常用到。不过, 这些性质比起记忆, 更多应该能够自己从定义推出。

练习: 3.2.2 例 2 (结论重要)、3.2.2 例 6 (注意严谨的两方向说明)

3.3 向量组的秩

定义: 向量组等价、极大线性无关组、向量组秩

算法: 极大线性无关组的找法 (3.3.2 例 8, 此题结论与对应的算法都非常重要)

定理: 极大线性无关组个数相等 (秩存在)、表出向量组的秩不超过原向量组

* 注意秩相同未必等价。

练习: 3.3.2 例 11 (方程组视角)、习题 3.3.6 (注意利用定义与基本结论的推导)、习题 3.3.10 (寻找自己熟悉的路径证明)

3.4 向量空间基与维数

定义: 基、维数、行空间、列空间

* 实质上就是看作无穷多个向量的向量组的极大线性无关组与秩。

定理: 基扩充定理 (习题 3.4.5)

* 本节无特殊的练习, 技巧基本可以在之前找到, 一些比较有用的结论可见本讲义 2.4.3。

3.5 矩阵的秩

定义: 行秩、列秩、秩、行/列满秩、满秩

定理: 行列秩相等 (注意证明过程如何拆分简化问题)、秩的非零子式定义

* 事实上秩的非零子式定义应在第四章中结合矩阵论证明。

练习: 3.5.2 例 7 (标准的回归向量组说明思路)、3.5.2 例 8 (结论重要)、3.5.2 例 9 (结论重要, 非零子式定义的应用, 矩阵论出发的更强结论见本讲义 3.3.2)

* 在复习到此节时可以不急做题, 矩阵复习结束后再回头看能较为清晰, 这里也并不列举作业题中的练习。

3.6 线性方程组有解条件

定理: 线性方程组有解判别定理

* 其实会建议先看 3.7 对齐次线性方程组的研究, 不过这节确实相对简单。

3.7 齐次线性方程组的解空间

定义: 基础解系

算法: 齐次线性方程组的解 (至少需要学会算法, 相对清晰的推导过程可见本讲义 2.5.1, 其中包含线性表出问题的重要技巧向量组的方程转化为系数方程)

* 本章除了行列秩相等与此定理外, 其他定理应做到能独立写出完整证明。

定理: 解空间维数定义 (极常用, 注意基础解系等价于解空间的基)

练习: 习题 3.7.5 (一维向量空间等价于任何两元素成比例, 注意转化为线性方程组问题)

3.8 非齐次线性方程组的解集结构

定义: 导出组

定理: 非齐次线性方程组的解集结构

练习: 习题 3.8.3 (解集结构的另一种视角, 实质上是凸性)、习题 3.8.5 (利用定义将问题转化为向量组相关)

5.3.3 矩阵

4.1 矩阵运算

定义: 矩阵加法、数乘、乘法、单位阵、可交换、矩阵的多项式

* 在矩阵理论中, 需要熟练记忆 $A^0 = I$ 的定义, 矩阵多项式的常数项视为乘 I 。

定理: 矩阵运算的基本性质

* 注意线性方程组的矩阵乘法形式, 它蕴含了右乘向量 x 相当于对 A 的列向量线性组合。

练习: 4.1.2 例 9 (归纳计算, 结论有用但在期中范围没用)、4.1.2 例 12 (结论重要)、习题 4.1.13 (整体思想计算)

4.2 特殊矩阵

定义: 对角矩阵、基本矩阵、三角矩阵、初等矩阵、对称/斜对称矩阵

定理: 上述相关的很多结论, 尤其三角阵乘积性质与行列变换看作初等矩阵乘积。

练习: 4.2.2 例 9 (注意结论与求和操作)、习题 4.2.5 (元素出发的计算)、习题 4.2.9 (整体出发的计算)

* 分块的好处: 介于元素与整体之间, 刻画部分性质

4.3 矩阵乘积的秩与行列式

定理: $\text{rank}(AB) \leq \min(\text{rank } A, \text{rank } B)$, Binet-Cauchy 公式

* 与 Laplace 定理类似, Binet-Cauchy 公式的记忆优先级也非常低, 一般只要会方阵版本与 $s > n$ 时为 0 即可。

练习: 4.3.2 例 1 (极大线性无关组与运算结合)、4.3.2 例 3 (结论重要, 回归方程组视角)、4.3.2 例 6 (注意求和符号转为矩阵乘法的行列式计算)、4.3.2 例 10 (Cauchy 不等式, 结论非常重要, 但目前用到不多)

* 方程组视角其实本质就是线性映射视角。

4.4 可逆矩阵

不在期中范围内, 推荐掌握内容见本讲义 4.1.2, 是矩阵论技巧的一部分。

4.5 分块矩阵

不在期中范围内, 推荐掌握内容见本讲义 5.2.1, 是矩阵论技巧的一部分。

4.7 线性映射

不在期中范围内, 推荐掌握内容: 线性映射、核空间、像空间定义, 矩阵 A 视为线性映射 $x \rightarrow Ax$ 的像与核计算 (解空间维数定理的线性映射版本)。

进阶内容为本讲义 4.1.1 (使用限制映射证明最好先叙述一下定义, 毕竟不在书上), 是向量空间技巧的一部分。

5.2 相抵标准形

不在期中范围内, 推荐掌握内容见本讲义 4.1.3, 是矩阵论技巧的一部分。

* 理论来说, 掌握矩阵论技巧与向量空间技巧是等价的, 学明白其中一条路即可。

六 补充：综合复习

§6.1 复习题

1. (1) 设 n 阶行列式 D 的第 i 行第 j 列为 d_{ij} , 且 $|i - j| = 2$ 时 $d_{ij} = x$, 否则为 0, 计算此行列式;
 (2) 设 n 阶行列式 D 的第 i 行第 j 列为 d_{ij} , 且 $d_{ij} = (a_i + b_j)^{n-1}$, 计算此行列式;
 (3) 设 n 阶行列式 D 的第 i 行第 j 列为 d_{ij} , 且 d_{ij} 当 $i = j$ 时为 $1 + \cos(\theta_i - \varphi_j)$, 否则为 $\cos(\theta_i - \varphi_j)$, 计算此行列式。
2. (1) 若 \mathbb{R}^n 中的向量组 $\alpha_1, \dots, \alpha_s$ 线性无关, 向量组 $\alpha_1 - \lambda\alpha_2, \alpha_2 - \lambda\alpha_3, \dots, \alpha_s - \lambda\alpha_1$ 线性相关, 求实数 λ ;
 (2) 若向量组 $\alpha_1, \dots, \alpha_s$ 线性无关, 且有表出

$$\forall i = 1, \dots, t, \quad \beta_i = \sum_{j=1}^s c_{ij} \alpha_j$$
 求证 β_1, \dots, β_t 线性无关当且仅当 c_{ij} 拼成的矩阵 C 行满秩。
3. (1) 设向量空间 W 为向量空间 V 子空间, 证明 $\dim W \leq \dim V$, 等号成立当且仅当 $W = V$;
 (2) 对向量空间 $W \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$, 设 $U \subset \mathbb{R}^{m+n}$ 为前 m 个分量在 W 中、后 n 个分量在 V 中的向量构成的集合, 求证其构成向量空间, 并求其维数;
 (3) 设 W 一组基为 $\alpha_1, \dots, \alpha_r$, V 一组基为 β_1, \dots, β_s (均至少有一个), 构造上间的 U 的一组基, 使得这组基中任何向量的前 m 个分量不全为 0, 后 n 个分量不全为 0。
4. 设 $\alpha_1, \dots, \alpha_r$ 是向量空间 $V \subset \mathbb{R}^n$ 的一组基, 证明 V 中有无穷多个向量 β 满足至少有 $r - 1$ 个分量为 0。
5. 已知线性方程组 $Ax = b$ 有解, 证明 A 的第 k 个列向量不能被其他列向量线性表出的充要条件是任何解 x 第 k 个分量相同。
6. 设 $A \in \mathbb{R}^{m \times n}$, 对线性方程组 $Ax = 0$, 设其解空间为 W , A 的行空间为 U , 证明对任何向量 $y \in \mathbb{R}^n$, 存在唯一 $x \in W$ 与 $z \in U$ 使得 $y = x + z$ 。
7. (1) 若线性方程组 $Ax = 0$ 的解都是方程 $b^T x = 0$ 的解, 证明 b^T 可被 A 的行向量线性表出;
 (2) 若线性方程组 $Ax = 0$ 与 $Bx = 0$ 的解集相同, 证明存在矩阵 C 使得 $B = CA$;
 (3) 若线性方程组 $Ax = 0$ 与 $Bx = 0$ 的解集相同, 在 A, B 行数相同时, 证明 A 可通过初等行变换变换为 B ;
 * 提示: 取出极大线性无关组后不断利用替换定理。
 (4) 若线性方程组 $Ax = p$ 与 $Bx = q$ 的解集相同且非空, 在 A, B 行数相同时, 证明增广矩阵 (A, p) 可以通过初等行变换变换为 (B, q) 。
 * 这即证明了解集非空时第一章中化为的简化阶梯形能够唯一确定解集。
8. 称 m 个数域 \mathbb{K} 上的 n 阶方阵 A_1, \dots, A_m “满秩相关”, 若存在不全为 0 的 $\lambda_1, \dots, \lambda_m \in \mathbb{K}$ 使得 $\lambda_1 A_1 + \dots + \lambda_m A_m$ 不满秩。
 - (1) 证明任何 $n + 1$ 个 \mathbb{K} 上的 n 阶方阵满秩相关;
 - (2) 取 $\mathbb{K} = \mathbb{Q}$, 对 $n > 1$, 找两个不满秩相关的 n 阶方阵;
 - (3) 取 $\mathbb{K} = \mathbb{C}$, 证明任何两个 n 阶方阵满秩相关;

- (4) 取 $\mathbb{K} = \mathbb{R}$, 证明任何两个 $2k+1$ 阶方阵满秩相关。
注: \mathbb{R} 上 n 阶满秩无关方阵的最大个数称为 Radon-Hurwitz 数。
9. 设 $\text{tr}(A)$ 为方阵 A 对角线元素的和, 称为方阵 A 的迹。对矩阵 $A \in \mathbb{R}^{m \times n}$ 与 $B \in \mathbb{R}^{n \times m}$, 求证 $\text{tr}(AB) = \text{tr}(BA)$, 并证明不存在方阵 A, B 使得 $AB - BA = I$ 。
10. (1) 若 $x \in \mathbb{R}^{s \times 1}, y \in \mathbb{R}^{n \times 1}, A \in \mathbb{R}^{s \times n}$, 证明 $x^T A y \neq 0$ 时 $\text{rank}(A y x^T A) = 1$, 其逆命题是否成立?
(2) 证明或否定, 对同阶方阵 A, B 有 $\text{rank}(AB) = \text{rank}(BA)$;
(3) 证明或否定, 对行数相同的矩阵 A, B , 当 A 的任何列不能被 B 的列线性表出、 B 的任何列不能被 A 的列线性表出时有 $\text{rank}(A B) = \text{rank } A + \text{rank } B$;
(4) 证明
- $$\text{rank}(A^2 - AB + B^2) \leq \text{rank}(A) + \text{rank}(B)$$
- (5) 证明 $A^2 - 3A + 2I = O$ 当且仅当 $\text{rank}(A - 2I) + \text{rank}(A - I) = n$ 。

§6.2 解答

1. (1) 考虑先按第一行、再按第一列、再按剩下的第一行、再按剩下的第一列展开 (每次展开的行/列中均只有一个元素非零, 注意代数余子式的符号), 可发现 $D_n = x^4 D_{n-4}$, 从而考虑 1、2、3、4 阶的情况可知当 n 为 4 的倍数时 $D_n = x^n$, 否则为 0。
(2) 将次方展开写成求和的形式有

$$d_{ij} = \sum_{k=1}^n C_{n-1}^{k-1} a_i^{k-1} b_j^{n-k}$$

从而设 $a_{ik} = C_{n-1}^{k-1} a_i^{k-1}$ 、 $b_{kj} = b_j^{n-k}$, 分别拼成方阵 AB , 应有结果为

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

注意到 A 的第 k 列可以提出 C_{n-1}^{k-1} , 设 $c_{ik} = a_i^{k-1}$, 拼成方阵 C , 有

$$\det A = \prod_{k=1}^n C_{n-1}^{k-1} \det C$$

利用 Vandermonde 行列式结论可知

$$\det C = \prod_{i < j} (a_j - a_i)$$

而对 $\det B$, 可发现 B 将所有行反向排列后为 Vandermonde 行列式, 而将 n 行从正向交换为反向可以通过 $n(n-1)/2$ 次相邻交换得到, 从而

$$\det B = (-1)^{n(n-1)/2} \prod_{i < j} (b_j - b_i)$$

最终得到原行列式为

$$(-1)^{n(n-1)/2} \prod_{k=1}^n C_{n-1}^{k-1} \prod_{i < j} (a_j - a_i) \prod_{i < j} (b_j - b_i)$$

(3) 直接展开可知

$$\cos(\theta_i - \varphi_j) = \cos \theta_i \cos \varphi_j + \sin \theta_i \sin \varphi_j$$

我们设所有 $\cos \theta_i$ 拼成列向量 Θ_1 , 所有 $\sin \theta_i$ 拼成列向量 Θ_2 , 所有 $\cos \varphi_i$ 拼成列向量 Φ_1 , 所有 $\sin \varphi_i$ 拼成列向量 Φ_2 , 则可发现原行列式化为了

$$\det \left(I + \begin{pmatrix} \Theta_1 & \Theta_2 \end{pmatrix} \begin{pmatrix} \Phi_1^T \\ \Phi_2^T \end{pmatrix} \right)$$

由此, 利用 4.5 节的公式 $\det(I + AB) = \det(I + BA)$, 可得原行列式等于

$$\det \left(I_2 + \begin{pmatrix} \Phi_1^T \\ \Phi_2^T \end{pmatrix} \begin{pmatrix} \Theta_1 & \Theta_2 \end{pmatrix} \right) = \det \begin{pmatrix} 1 + \Phi_1^T \Theta_1 & \Phi_1^T \Theta_2 \\ \Phi_2^T \Theta_1 & 1 + \Phi_2^T \Theta_2 \end{pmatrix}$$

也即最终为

$$(1 + \Phi_1^T \Theta_1)(1 + \Phi_2^T \Theta_2) - \Phi_1^T \Theta_2 \Phi_2^T \Theta_1$$

* 放这个题在这里主要是建议大家记一下这个形式简单的公式, 一些行列式可以用它进行简便计算, 例如往年题中出现的, 主对角线为 a 、其他为 1 的行列式。公式的证明可见本讲义 4.2.1 节第一个问题的行列变换技巧出发的证明 (注意行列变换对行列式的影响)。

2. (1) 利用线性无关定义考虑以 $\lambda_1, \dots, \lambda_s$ 为未知量的方程

$$\lambda_1(\alpha_1 - \lambda\alpha_2) + \dots + \lambda_s(\alpha_s - \lambda\alpha_1) = 0$$

此方程组可整理为

$$(\lambda_1 - \lambda\lambda_s)\alpha_1 + (\lambda_2 - \lambda\lambda_1)\alpha_2 + \dots + (\lambda_s - \lambda\lambda_{s-1})\alpha_s = 0$$

由于 $\alpha_1, \dots, \alpha_s$ 线性无关, 上式成立当且仅当

$$\lambda_1 - \lambda\lambda_s = \lambda_2 - \lambda\lambda_1 = \dots = \lambda_s - \lambda\lambda_{s-1} = 0$$

此方程组可化为 $\lambda_1 = \lambda^s \lambda_1$ 与

$$\forall k = 1, \dots, s, \quad \lambda_k = \lambda^{k-1} \lambda_1$$

由此, 若 $\lambda^s = 1$, 即 s 为奇数时 $\lambda = 1$ 或 s 为偶数时 $\lambda = \pm 1$, 取 $\lambda_1 = 1$ 即得到了一组非零解, 它们线性相关; 否则, 必须 λ_1 为 0, 可推出 λ_i 全为 0, 因此它们线性无关。

综合以上讨论即知 s 为奇数时 $\lambda = 1$, s 为偶数时 $\lambda = \pm 1$ 。

(2) 利用线性无关定义考虑以 $\lambda_1, \dots, \lambda_t$ 为未知量的方程

$$\sum_{i=1}^t \lambda_i \beta_i = 0$$

展开即得其为

$$\sum_{i=1}^t \lambda_i \sum_{j=1}^s c_{ij} \alpha_j = 0$$

利用乘法分配律可将其整理为

$$\sum_{j=1}^s \left(\sum_{i=1}^t \lambda_i c_{ij} \right) \alpha_j = 0$$

由于 $\alpha_1, \dots, \alpha_s$ 线性无关, 上式成立当且仅当

$$\forall j = 1, \dots, s, \quad \sum_{i=1}^t \lambda_i c_{ij} = 0$$

将 λ_i 拼成列向量 λ , 将线性方程组写为矩阵的形式即

$$C^T \lambda = 0$$

利用解空间维数定理, λ 解空间维数为 $t - \text{rank } C^T = t - \text{rank } C$, 而只存在零解也即解空间维数为 0, 当且仅当 $\text{rank } C = t$, 这也就是行满秩。

3. * 这题是纯我自己出的, 目的是让大家不要只记得操作, 忘了向量空间相关的基本定义。

(1) 根据基的定义, 设 W 一组基为 $\alpha_1, \dots, \alpha_r$, V 一组基为 β_1, \dots, β_s , 则由 $W \subset V$ 可知 $\alpha_1, \dots, \alpha_r$ 可被 β_1, \dots, β_s 线性表出, 又由它们线性无关可知 $r \leq s$, 从而得证。

若 $W = V$, 利用定义 $\dim W = \dim V$; 若 $\dim W = \dim V$, 设 W 一组基为 $\alpha_1, \dots, \alpha_r$, 它们是 V 中等于维数的线性无关组, 因此也是 V 的一组基, 即得结论。

* 也可从基扩充定理的角度说明, W 一组基可以扩充为 V 一组基, 从而得到证明。

(2) 由于 $U \subset \mathbb{R}^{m+n}$, 其中的加法与数乘是向量的加法、数乘, 无需验证性质, 只需证明 U 中两个向量的和与某个向量的数乘都在 U 中即可。设 U 中向量为 (u_1, u_2) , 这里 u_1 含 m 个分量, u_2 含 n 个分量, 则

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2)$$

$$\lambda(u_1, u_2) = (\lambda u_1, \lambda u_2)$$

由 U 的定义可知 $a_1, b_1, u_1 \in W$ 且 $a_2, b_2, u_2 \in V$, 从而 $a_1 + b_1, \lambda u_1 \in W$ 且 $a_2 + b_2, \lambda u_2 \in V$, 这就得到了证明。

设 W 一组基为 $\alpha_1, \dots, \alpha_r$, V 一组基为 β_1, \dots, β_s , 考虑

$$(\alpha_1, 0), \dots, (\alpha_r, 0), (0, \beta_1), \dots, (0, \beta_s)$$

对任何 $(u, v) \in U$, 设

$$u = \sum_{i=1}^r \lambda_i \alpha_i, \quad v = \sum_{j=1}^s \mu_j \beta_j$$

则有

$$(u, v) = \sum_{i=1}^r \lambda_i (\alpha_i, 0) + \sum_{j=1}^s \mu_j (0, \beta_j)$$

从而 U 中任何向量可被它们表出。另一方面, 若

$$\sum_{i=1}^r \lambda_i (\alpha_i, 0) + \sum_{j=1}^s \mu_j (0, \beta_j) = 0$$

可发现 $\sum_{i=1}^r \lambda_i \alpha_i = 0$ 且 $\sum_{j=1}^s \mu_j \beta_j = 0$, 从而所有 λ_i 与 μ_j 必须全为 0, 即证明了它们线性无关。由于它们线性无关且可表出 U 中任何向量, 由定义它们是 U 的一组基, 于是 $\dim U = \dim W + \dim V$ 。

(3) 根据定义, U 中任何与

$$(\alpha_1, 0), \dots, (\alpha_r, 0), (0, \beta_1), \dots, (0, \beta_s)$$

等价的向量组都构成一组基。我们将第一个向量加到后 s 个, 得到

$$(\alpha_1, 0), \dots, (\alpha_r, 0), (\alpha_1, \beta_1), \dots, (\alpha_1, \beta_s)$$

再把最后一个向量加到前 r 个, 得到

$$(2\alpha_1, \beta_s), \dots, (\alpha_r + \alpha_1, \beta_s), (\alpha_1, \beta_1), \dots, (\alpha_1, \beta_s)$$

由于这里只进行了初等变换，变换前后向量组等价，因此变换后构成一组基。由线性无关性， α_1 、 $2\alpha_1$ 、 $\alpha_2 + \alpha_1, \dots, \alpha_r + \alpha_1$ 均不可能为 0，同理 β_1, \dots, β_s 不可能为 0，这就证明了这组基符合要求。

* 构造方法有很多，不过基本思路都来自初等变换。

4. 由于 V 是一个向量空间，只要存在一个至少 $r - 1$ 个分量是 0 的非零向量，考虑其数乘 $\lambda \in \mathbb{R}$ 即可得到无穷多个。

将 $\alpha_1, \dots, \alpha_r$ 作为列拼成矩阵 A ，由线性无关性可知 $\text{rank } A = r$ ，从而其存在一个 r 阶非零子式，设第 i_1, \dots, i_r 行构成非零子式 R 。

利用 Cramer 法则， $Rx = e_1$ 必然存在唯一解，这里 e_1 指第一个分量是 1，其他为 0 的向量，设解为 x_1, \dots, x_r ，由于 A 每列是 R 每列的延伸，计算即可发现

$$x_1\alpha_1 + x_2\alpha_2 + \dots + x_r\alpha_r$$

的第 i_1 个分量为 1，第 i_2, i_3, \dots, i_r 个分量为 0，这就找到了至少 $r - 1$ 个分量是 0 的非零向量，结论成立。

* 这可能是这套复习题里思路最莫名其妙的一个，但确实出自往年题。这题用到的秩的非零子式定义、Cramer 法则刻画存在性，都是考得不多但值得注意的知识点。如果说有什么办法能想出这题，大概是“存在性”问题除了构造（而从条件出发没有明显的构造方式）外，就只能想到 Cramer 法则了，然后再向此方向去靠拢：由于需要找 $r - 1$ 个分量是 0 的非零向量，可以想到考虑某个 r 阶非零子式。

* 另解：从齐次线性方程组出发考虑。将 $\alpha_1, \dots, \alpha_r$ 作为列拼成矩阵 A ，所有使得 Ax 的前 $r - 1$ 个分量为 0 的 x 。这是一个齐次线性方程组，且有 $r - 1$ 个方程、 r 个未知数，因此一定有非零解，而由于 A 的列线性无关， x 非零可知

$$Ax = x_1\alpha_1 + \dots + x_r\alpha_r \neq 0$$

这就找到了至少 $r - 1$ 个分量是 0 的非零向量。

* 又另解：从行变换不影响行空间出发考虑，将 $\alpha_1, \dots, \alpha_r$ 作为行拼成矩阵 A ，并将 A 初等行变换为简化阶梯形，则变换后所有行必然仍在 V 中。此时，由于秩为 r ，最后一行必然非零，且根据阶梯形的要求最后一行前 $r - 1$ 个分量为 0，这就找到了至少 $r - 1$ 个分量是 0 的非零向量。

5. 充分性：若任何解 x 的第 k 个分量相同，根据非齐次线性方程组解集结构，可知 $Ay = 0$ 的任何解 y 满足第 k 个分量为 0。

设 $A \in \mathbb{R}^{m \times n}$ ，若其第 k 个列向量可被其他列向量表出，有

$$\alpha_k = \sum_{i \neq k} \lambda_i \alpha_i$$

考虑 $y = (\lambda_1, \dots, \lambda_{k-1}, -1, \lambda_{k+1}, \dots, \lambda_n)^T$ ，即可验证

$$Ay = \sum_{i \neq k} \lambda_i \alpha_i - \alpha_k = 0$$

与 y 第 k 个分量为 0 矛盾。

必要性：与充分性同理，只需证明 $Ay = 0$ 的任何解 y 满足第 k 个分量为 0。若解 $y = (y_1, \dots, y_n)^T$ 满足 $y_k \neq 0$ ，有

$$-y_k \alpha_k = \sum_{i \neq k} y_i \alpha_i$$

也即

$$\alpha_k = \sum_{i \neq k} -\frac{y_i}{y_k} \alpha_i$$

也即 α_k 能被其他列向量表出，矛盾。

* 注意线性方程组用线性组合语言叙述的结果： Ax 就是 A 的列向量的某个线性组合。

6. 设 A 的行向量为 $a_1^T, a_2^T, \dots, a_m^T$ ，这里 a_m 为 n 维列向量，则线性方程组可以写为

$$a_1^T x = a_2^T x = \dots = a_m^T x = 0$$

先证明 $W \cap U = \{0\}$ 。对任何 $x \in W \cap U$ ，其满足上式成立，且能写为

$$x = \sum_{i=1}^m \lambda_i a_i$$

但由 $x \in W$ 可知

$$0 = \sum_{i=1}^m \lambda_i a_i^T x = x^T x$$

由于 $x^T x$ 即为其各分量平方和， x 只能为零向量。

记 $r = \text{rank } A$ 。利用行空间维数为 r 与解空间维数为 $n-r$ ，设 U 的一组基为 $\alpha_1, \dots, \alpha_r$ ， W 的一组基为 $\beta_1, \dots, \beta_{n-r}$ ，下面说明它们构成 \mathbb{R}^n 一组基。由于其个数等于维数，只需验证线性无关性，而若有

$$\sum_{i=1}^r \lambda_i \alpha_i + \sum_{j=1}^{n-r} \mu_j \beta_j = 0$$

可知

$$\sum_{i=1}^r \lambda_i \alpha_i = -\sum_{j=1}^{n-r} \mu_j \beta_j$$

由于左侧在 U 中，右侧在 W 中，根据之前已证必有

$$\sum_{i=1}^r \lambda_i \alpha_i = \sum_{j=1}^{n-r} \mu_j \beta_j = 0$$

但 $\{\alpha_i\}$ 与 $\{\mu_j\}$ 是线性无关的，即得到所有 λ_i 与 μ_j 均为 0，得证。

最后，由于 \mathbb{R}^n 任何向量 y 可被这组基唯一表出，设其为

$$\sum_{i=1}^r \lambda'_i \alpha_i + \sum_{j=1}^{n-r} \mu'_j \beta_j$$

记

$$x = \sum_{j=1}^{n-r} \mu'_j \beta_j, \quad z = \sum_{i=1}^r \lambda'_i \alpha_i$$

即可验证 $y = x + z$ 符合要求。若还有 $y = x_0 + z_0$ ，则 $x - x_0 = z_0 - z$ ，但 $x - x_0 \in W$ ， $z_0 - z \in U$ ，于是 $x - x_0 = z_0 - z = 0$ ，这就得到了表出唯一。

* 这里用到了线性方程组从行空间出发的叙述方式，事实上是内积的语言，将在之后深入学习。当问题中出现转置相关时（对于线性方程组自然的讨论是 A 的列向量，因此讨论行向量涉及转置），往往与某种配方有关，书上 $\text{rank}(A^T A) = \text{rank } A$ 的证明也是如此。

* 此题事实上代表了行空间与解空间具有某种“互补性”，因此可以反过来从给出的基础解系构造线性方程组：若 x_1, \dots, x_r 为基础解系，将 x_1^T, \dots, x_r^T 作为矩阵 B 的行，解方程 $By = 0$ ，将其基础解系记为 a_1, \dots, a_{n-r} ，并将 a_1^T, \dots, a_{n-r}^T 拼成矩阵 A ，则 $Ax = 0$ 就是以 x_1, \dots, x_r 为基础解系的线性方程组。

7. (1) 由条件可知添加方程 $b^T x = 0$ 不会影响解集，因此通过解空间维数定理有

$$\text{rank } A = \text{rank} \begin{pmatrix} A \\ b^T \end{pmatrix}$$

考虑 A 的行向量的一个极大无关组，记为 $a_{i_1}^T, \dots, a_{i_r}^T$ ，若 b^T 不能被它们表出，则根据书上结论知添加 b^T 后它们仍然线性无关，从而 $\begin{pmatrix} A \\ b^T \end{pmatrix}$ 的秩至少为 $\text{rank } A + 1$ ，矛盾。

(2) 由条件可知给 $Ax = 0$ 添加方程 $Bx = 0$ 不会影响解集，因此通过解空间维数定理有

$$\text{rank } A = \text{rank} \begin{pmatrix} A \\ B \end{pmatrix}$$

与 (1) 完全类似考虑 B 的每一行，可发现 B 的每一行均可被 A 的行线性表出。设 A 的行为 a_1^T, \dots, a_s^T ， B 的行为 b_1^T, \dots, b_t^T ，由此知存在 c_{ij} 使得

$$b_i^T = \sum_{j=1}^s c_{ij} a_j^T$$

这也即说明对任何 k 有

$$b_{ik} = \sum_{j=1}^s c_{ij} a_{jk}$$

将 c_{ij} 拼成矩阵 C ，上述式子即代表 $B = CA$ 。

* 注意矩阵乘法的线性组合叙述方式，左乘相当于对行线性组合，右乘相当于对列线性组合。

(3) 由 (2) 可知 B 的行向量组可被 A 的行向量组表出，同理 A 的行向量组可被 B 的行向量组表出，因此两者等价。我们只需证明等价的向量组可以通过初等变换得到，下考虑等价的向量组 $\alpha_1, \dots, \alpha_n$ 与 β_1, \dots, β_n ，我们设法将 $\alpha_1, \dots, \alpha_n$ 变换为 β_1, \dots, β_n 。

由于两者等价，秩必然相同，不妨设 $\alpha_1, \dots, \alpha_r$ 与 β_1, \dots, β_r 分别构成两向量组的极大无关组。根据极大无关组的性质， $\alpha_{r+1}, \dots, \alpha_n$ 可被 $\alpha_1, \dots, \alpha_r$ 表出，利用第一类初等变换（将之后的每个向量减去 $\alpha_1, \dots, \alpha_r$ 的适当倍数）可将它们均消成零向量，这样就剩余了 $n - r$ 个零向量。

利用向量组等价性， β_1 可被 $\alpha_1, \dots, \alpha_r$ 表出，设

$$\beta_1 = \sum_{i=1}^r \lambda_i \alpha_i$$

由于其构成极大无关组的一部分，不可能为零向量，因此一定存在某个 α_k 前的 $\lambda_k \neq 0$ 。我们将 α_k 乘 λ_k 后加上其他 α_i 的适当倍数，即将 α_k 初等变换为了 β_1 ，且其他 α_i 不变。利用行秩等于列秩的证明过程中已证的，初等变换不影响等价性， $\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+1}, \dots, \alpha_r, \beta_1$ 仍然与 β_1, \dots, β_r 等价。

于是，存在 μ_i 使得

$$\beta_2 = \sum_{i \neq k} \mu_i \alpha_i + \mu_k \beta_1$$

若所有 $i \neq k$ 的 $\mu_i = 0$ ，可得 β_2 能被 β_1 表出，与 β_1, β_2 的线性无关性矛盾，从而即得有某个 α_l 前的系数不为 0，乘 μ_l 后加上其他 α_i 与 β_1 的适当倍数，即将原向量组初等变换为了 β_1, β_2 与 $i \notin \{k, l\}$ 的 α_i 。

重复此过程，利用 $\{\beta_i\}$ 的线性无关性，每次剩余的 α_i 前的系数都不可能全为 0，因此总能进行替换，最终可以将 $\alpha_1, \dots, \alpha_r$ 初等变换为 β_1, \dots, β_r 。

最后，由于 $\beta_{r+1}, \dots, \beta_n$ 可被 β_1, \dots, β_r 表出，进行第一类初等变换可把之前剩余的 $n - r$ 个零向量加为 $\beta_{r+1}, \dots, \beta_n$ ，这就得到了结论。

* 其实这一问利用相抵标准形与可逆来说明会非常简单，下半学期将会具体介绍。

* 一般地，在 A, B 行数不同时，只要允许初等行变换加/减零向量，即可将 A 变换为 B ：将 A 的行加上等同于 B 的行数的零向量，再在下方变换出 B ，然后用 B 消去 A ，将 A 部分的零行去除。由于两者行向量组等价，这是可以实现的。

- (4) 先证明 $Ay = 0$ 与 $By = 0$ 的解集相同。取出 $Ax = p$ 与 $Bx = q$ 的任何一个公共解 x_0 ，利用非齐次线性方程组的解集性质可知 $Ay = 0$ 的解集是 $Ax = p$ 的解集减 x_0 ， $By = 0$ 的解集是 $By = q$ 的解集减 x_0 ，从而两者相同。

由此， A 可以通过初等行变换变换为 B ，对 (A, p) 进行相同的初等行变换后，设结果为 (B, p_0) ，即说明了 $Ax = p$ 的解集与 $Bx = p_0$ 相同。若 $p_0 \neq q$ ，对 $Bx = p_0$ 的任何解，不可能同时满足 $Bx = q$ ，矛盾，因此 $p_0 = q$ ，这就直接得到了结论。

* 一般地，在 A, B 行数不同时，允许初等行变换加/减零向量后，唯一可能让 (A, p) 无法进行相同的初等行变换的情况是变换中 A 的某一行成为了 0，但此时 p 对应分量不为 0。然而，这种情况还原回线性方程组可发现 $Ax = p$ 无解，与解集非空矛盾，因此不可能出现，由此仍能得到结论。

8. * 基本结论：方阵的特殊性在于满秩等价于行列式非零，因此这实际上是一个行列式问题。下面默认应用了此结论。

- (1) 由于 $n+1$ 个 \mathbb{K} 上的 n 阶向量线性相关，对任何 n 阶方阵 A_1, \dots, A_{n+1} ，存在非零的 $\lambda_1, \dots, \lambda_{n+1}$ 使得

$$\sum_{i=1}^{n+1} \lambda_i A_i$$

的第一行为 0，于是其不可能满秩。

* 看起来非常简单但是鬼知道怎么想到的的题目，出自往年真题。

- (2) 考虑

$$A_1 = I_n, \quad (A_2)_{ij} = \begin{cases} 1 & j - i = 1 \\ 2 & i = n, j = 1 \\ 0 & \text{otherwise.} \end{cases}$$

直接利用完全展开式计算行列式可发现

$$\det(\lambda A_1 + \mu A_2) = \lambda^n + (-1)^{n-1} 2\mu^n = \lambda^n - 2(-\mu)^n$$

若存在非零有理数 λ, μ 使得其为 0，即说明了 $\sqrt[n]{2}$ 是有理数，这是不可能的。

* 如何想到这样的构造？接下来两问的证明过程事实上才是想法的来源，也即通过没有有理根的多项式进行构造。

- (3) 对 n 阶方阵 P, Q ，若 $\det P = 0$ ，则 $1P + 0Q$ 不满秩，已经满秩相关，由此只需研究 $\det P \neq 0$ 的情况。考虑看作 λ 多项式的

$$f(\lambda) = \det(\lambda P + Q)$$

利用完全展开式，设 P, Q 的分量为 p_{ij}, q_{ij} ，则

$$f(\lambda) = \sum_{j_1 \dots j_n} (-1)^{\tau(j_1 \dots j_n)} \prod_{k=1}^n (\lambda p_{kj_k} + q_{kj_k})$$

利用此形式可看出 $f(\lambda)$ 是 λ 的至多 n 次多项式，且首项 λ^n 前的系数必须从乘积中每项都选择了 λp_{kj_k} ，从而为

$$\sum_{j_1 \dots j_n} (-1)^{\tau(j_1 \dots j_n)} \prod_{k=1}^n p_{kj_k} = \det P$$

由于已经架设了其非零, $f(\lambda)$ 是 λ 的 n 次多项式, 因此根据代数学基本定理必有复数根, 这就说明了存在 λ 使得 $\lambda P + Q$ 不满秩, 得证。

* 主要注意完全展开式的应用, 这里由于具有对称性, 元素并不特别, 因此想到利用完全展开式说明。

- (4) 与上问完全类似, 只需考虑 $\det P \neq 0$ 的情况, 可发现 $f(\lambda) = \det(\lambda P + Q)$ 是首项系数为 $\det P$ 的 $2k + 1$ 次多项式。由于 $f(\lambda)$ 次数为奇数, 在 $\lambda \rightarrow \pm\infty$ 时其正负性不同, 因此利用介值定理一定存在实根, 这就说明了存在 λ 使得 $\lambda P + Q$ 不满秩, 得证。

9. 直接计算可发现

$$\text{tr}(AB) = \sum_{i=1}^m \left(\sum_{k=1}^n a_{ik} b_{ki} \right) = \sum_{k=1}^n \left(\sum_{i=1}^m b_{ki} a_{ik} \right) = \text{tr} BA$$

另一方面, 对同阶方阵可直接验证 $\text{tr}(P - Q) = \text{tr}(P) - \text{tr}(Q)$, 于是

$$\text{tr}(AB - BA) = 0$$

但利用定义 $\text{tr}(I_n) = n$, 从而左右不可能相等。

10. (1) 由于 $x^T A y \neq 0$, 其平方不为 0, 也即

$$x^T A y x^T A y \neq 0$$

由于这是一个数, 有

$$\text{rank}(x^T A y x^T A y) = 1$$

重新看作矩阵乘法, 利用乘积的秩不超过原矩阵的秩可发现 (注意 y 列数为 1, 秩不可能超过 1)

$$\text{rank}(x^T A y x^T A y) \leq \text{rank}(A y x^T A) \leq \text{rank } y \leq 1$$

因此必须全部取等, 这就得到了结论。

其逆命题不成立: 考虑

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad x = y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

- (2) 结论是错误的, 考虑

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

则可验证 $AB = O$ 但 $BA \neq O$, 于是两者秩不同。

* 构造还是依托于熟悉 $\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$ 的乘法结果。

- (3) 结论是错误的, 考虑

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

* 这个问题的本质是交空间, 从空间取基的角度或许更容易想到构造。

- (4) 利用 $\text{rank}(X + Y) \leq \text{rank } X + \text{rank } Y$ 与 $\text{rank } XY \leq \text{rank } X$ 即有

$$\text{rank}(A^2 - AB + B^2) \leq \text{rank}(A(A - B)) + \text{rank}(B^2) \leq \text{rank } A + \text{rank } B$$

* 秩不等式有时是简单的脑筋急转弯, 所以优先尝试整体的处理, 处理不过来再采用更高级的方法。

(5) 只需证明

$$\text{rank}(A - 2I) + \text{rank}(A - I) = \text{rank}(A^2 - 3A + 2I) + n$$

利用秩的等式也即

$$\text{rank} \begin{pmatrix} A - 2I & O \\ O & A - I \end{pmatrix} = \text{rank} \begin{pmatrix} A^2 - 3A + 2I & O \\ O & I \end{pmatrix}$$

左侧通过前 n 行加上后 n 行可变为

$$\text{rank} \begin{pmatrix} A - 2I & A - I \\ O & A - I \end{pmatrix}$$

再用后 n 列减前 n 列即成为

$$\text{rank} \begin{pmatrix} A - 2I & I \\ O & A - I \end{pmatrix}$$

用 I 作行变换消去后 n 行的 $A - I$, 可发现此时左侧变为 (可将 $A - 2I, A - I$ 看作整体 B, C 进行行列变换, 4.3 节证明 Binet-Cauchy 公式时有类似操作)

$$\text{rank} \begin{pmatrix} A - 2I & I \\ -(A - I)(A - 2I) & O \end{pmatrix}$$

利用 I 消去前 n 列的 $A - 2I$ 即得到其变为

$$\text{rank} \begin{pmatrix} O & I \\ -(A - I)(A - 2I) & O \end{pmatrix}$$

可发现 $(A - I)(A - 2I) = A^2 - 3A + 2I$, 给后 n 行添加负号并交换行列可得结论。

* 另解: 空间思路, 原等式可视为

$$(n - \text{rank}(A - 2I)) + (n - \text{rank}(A - I)) = n - \text{rank}(A^2 - 3A + 2I)$$

也即 $Ax = 2x$ 的解空间维数与 $Ay = y$ 的解空间维数之和为 $(A^2 - 3A + 2I)z = 0$ 的解空间维数。

设三个空间分别为 W_2, W_1 与 W , 代入可验证 W_2, W_1 都是 W 的子空间, 且若 $Ax = 2x, Ax = x$, 作差有 $x = 0$, 因此 $W_2 \cap W_1 = \{0\}$ 。

另一方面, 设 $(A^2 - 3A + 2I)z = 0$, 则可发现 $(A - 2I)(A - I)z = 0$, 于是 $(A - I)z \in W_2$, 同理, 由于 $(A - I)(A - 2I)z = 0$ (注意同一个矩阵的多项式可交换), 有 $(A - 2I)z \in W_1$, 而 $z = (A - I)z - (A - 2I)z$, 因此 z 可写为 W_1 与 W_2 各一元素的和。

考虑 W_1 一组基 $\alpha_1, \dots, \alpha_t$ 与 W_2 一组基 β_1, \dots, β_s , 由前述推导可知它们都在 W 中, 且能表出 W 中任何元素。与复习题 6 相同可从 $W_2 \cap W_1 = \{0\}$ 推出 $\alpha_1, \dots, \alpha_t, \beta_1, \dots, \beta_s$ 线性无关, 这就说明了它们构成 W 的一组基, 即有维数结论。

七 可逆矩阵

§7.1 期中考试

7.1.1 试题

- 计算行列式:

$$(a) \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix};$$

$$(b) \det D_{n \times n}, \quad d_{ij} = \begin{cases} 5 & i = j \\ 3 & i = j - 1 \\ 2 & i = j + 1 \\ 0 & |i - j| > 1 \end{cases};$$

$$(c) \det D_{n \times n}, \quad d_{ij} = \frac{1 - a_i^n b_j^n}{1 - a_i b_j}.$$

2. 设

$$\alpha_1 = \begin{pmatrix} 1 \\ 3 \\ -5 \\ -9 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 2 \\ -1 \\ -3 \\ -4 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} -3 \\ 5 \\ 1 \\ -1 \end{pmatrix}, \quad \alpha_4 = \begin{pmatrix} -4 \\ 6 \\ 2 \\ 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} -5 \\ -1 \\ 11 \\ 17 \end{pmatrix}$$

(a) 求 $\alpha_1, \dots, \alpha_4$ 的一个极大线性无关组;

(b) 判断 β 是否可以被 $\alpha_1, \dots, \alpha_4$ 线性表出, 若可以表出, 给出所有的表出方式。

3. 求 λ 的值使得矩阵

$$\begin{pmatrix} 3 & 1 & 1 & 4 \\ \lambda & 4 & 10 & 1 \\ 1 & 7 & 17 & 3 \\ 2 & 2 & 4 & 3 \end{pmatrix}$$

的秩最小。

4. 设矩阵 $A \in M_{3 \times 4}(\mathbb{K})$, 且齐次线性方程组 $Ax = 0$ 的解空间由向量 $(1, 1, 1, 0)^T$ 和 $(-2, -1, 0, 1)^T$ 生成。

(a) 求 A 对应的简化阶梯形矩阵;

(b) 求以下空间的维数: A 的列空间 $C(A)$ 、 A^T 的列空间 $C(A^T)$ 、齐次方程组 $A^T x = 0$ 的解空间 $N(A^T)$;

(c) 上一问中哪些空间的基可以确定? 并写出相应空间的一组基。

5. 设 $A \in M_{m \times n}(\mathbb{K})$ 且 $\text{rank } A = r$, 若 A 的前 r 行线性无关、前 r 列也线性无关, 证明它们相交处的 r 阶子式非零。

6. 设 $A \in M_{m \times n}(\mathbb{K})$ 、 $B \in M_{r \times n}(\mathbb{K})$, 且齐次线性方程组 $Ax = 0$ 与 $Bx = 0$ 同解, 证明 A 与 B 的行向量组等价。

7. 设 $A \in M_{r \times n}(\mathbb{K})$ 、 $\alpha \in M_{1 \times r}(\mathbb{K})$ 、 $\beta \in M_{n \times 1}(\mathbb{K})$ 满足 $\alpha A \beta = (a)$ 且 $a \neq 0$, 令

$$B = A - a^{-1}(A\beta\alpha A)$$

并记 $N(A)$ 、 $N(B)$ 分别为齐次线性方程组 $Ax = 0$ 与 $Bx = 0$ 的解空间, 证明 $N(B)$ 可由 $N(A)$ 和 β 生成, 并进而证明 $\text{rank } B = \text{rank } A - 1$ 。

7.1.2 解答

1. (a) 直接行列变换或展开计算得结果为 -16 。
- (b) 设 n 阶情况为 D_n , 先按首行展开、再按首列展开可以发现

$$D_n = 5D_{n-1} - 6D_{n-2}$$

且根据定义有

$$D_1 = 5, \quad D_2 = 19$$

使用待定系数法可发现配凑

$$D_n - 3D_{n-1} = 2(D_{n-1} - 3D_{n-2}), \quad D_n - 2D_{n-1} = 3(D_{n-1} - 2D_{n-2})$$

分别当作等比数列可求解出 $D_n - 3D_{n-1}$ 与 $D_n - 2D_{n-1}$, 消去 D_{n-1} 即得到

$$D_n = 3^{n+1} - 2^{n+1}$$

- (c) 反向利用等比数列求和公式可发现展开

$$d_{ij} = \sum_{k=1}^n a_i^{k-1} b_j^{k-1}$$

于是令 $a_{ik} = a_i^{k-1}$ 、 $b_{kj} = b_j^{k-1}$, 分别拼成矩阵 A, B , 即有

$$D = AB, \quad \det D = \det A \det B$$

由于 A, B 均为范德蒙德阵 (与其转置), 直接利用范德蒙德行列式可知结果为

$$\prod_{i < j} (a_j - a_i)(b_j - b_i)$$

2. (a) 将它们拼成矩阵, 利用行变换变换为简化阶梯形, 即可得到它们中的任两个向量均构成极大线性无关组。
- (b) 这即为非齐次线性方程组问题, 设

$$\beta = x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3 + x_4\alpha_4$$

可直接计算得到通解为

$$\forall s, t \in \mathbb{K}, \quad x_1 = -1 - s - \frac{8}{7}t, \quad x_2 = -2 + 2s + \frac{18}{7}t, \quad x_3 = s, \quad x_4 = t$$

* 由于变元和特解不同, 可能存在不同正确写法。

3. * 有不同的消元路径与计算方法, 结果正确即可。

作列变换可发现第三行能被 1、4 两行消去, 进一步作行变换可发现第三列能被 1、4 两列消去, 由此原矩阵的秩等于

$$\text{rank} \begin{pmatrix} 3 & 1 & 4 \\ \lambda & 4 & 1 \\ 2 & 2 & 3 \end{pmatrix}$$

可发现 2、3 两列线性无关, 从而此矩阵秩至少为 2, 当且仅当第一列能被另两列表出时秩为 2, 而

$$3 = \frac{1}{5}(4 \times 4 - 1), \quad 2 = \frac{1}{5}(4 \times 3 - 2)$$

由此当且仅当 $\lambda = 0$ 时成立。

4. (a) 将解写为自由变量的形式 (参考 3.7 节的最终形式), 即

$$\begin{cases} x_1 = x_3 - 2x_4 \\ x_2 = x_3 + x_4 \end{cases}$$

于是对应的简化行阶梯形矩阵为

$$\begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

(b) 由条件可知 $\dim N(A) = 2$, 利用解空间维数定理可知 $\dim C(A) = 2$, 利用行秩等于列秩可知 $\dim C(A^T) = 2$, 再次利用解空间维数定理可知 $\dim N(A^T) = 1$ 。

(c) 由于 A 可以且仅可以任意作初等行变换, 其行空间 (即 $C(A^T)$) 的基是确定的, 即为 (a) 中的前两行。

其列空间的基会被行变换改变, 因此不确定, 而 $N(A^T)$ 由 $C(A)$ 决定, 会被初等行变换改变 (可举例观察), 因此也无法确定。

5. * 经典错误: 认为若干向量线性无关则它们的缩短组线性无关。

由于 A 的前 r 行构成极大线性无关组, 它们可线性表出其他行, 因此可利用初等行变换将其他行消去。

由于初等行变换不改变秩, 变换后的矩阵仍然是秩为 r 的。由于初等行变换不改变列的线性无关性, 前 r 列仍然线性无关, 由此它们可构成所有列的极大线性无关组, 同理进行初等列变换将其他列消去。

如此变换后, A 除了左上角的 $r \times r$ 子式外均为 0, 且秩仍为 r , 即可得到左上角为非零子式, 得证。

6. 由于 $Ax = 0$ 与 $Bx = 0$ 同解, 其亦与

$$\begin{cases} Ax = 0 \\ Bx = 0 \end{cases} \iff \begin{pmatrix} A \\ B \end{pmatrix} x = 0$$

同解。由解空间相同可知维数相同, 从而有

$$\text{rank } A = \text{rank } \begin{pmatrix} A \\ B \end{pmatrix}$$

考虑行秩。不妨设 A 的行的一个极大线性无关组为 a_1^T, \dots, a_r^T , 若 B 的某一行 b_i^T 无法被它们线性表出, 则可发现 $a_1^T, \dots, a_r^T, b_i^T$ 线性无关, 右侧秩至少为 $r+1$, 矛盾, 因此 B 的行可被 A 的极大线性无关组表出, 也即可被 A 的行表出。

同理 A 的行可被 B 的行表出, 得证等价。

* 从秩相同推出行可被表出的必要过程是这题核心, 不能省略。

7. 分为五部分证明:

- $N(A) \subset N(B)$

当 $Ax = 0$ 时, 直接代入验证可知 $Bx = Ax - a^{-1}A\beta\alpha(Ax) = 0$ 。

- $\beta \in N(B)$

直接计算得 $B\beta = A\beta - a^{-1}A\beta(\alpha A\beta) = A\beta - A\beta = 0$ 。

- $\beta \notin N(A)$

若否, 从 $A\beta = 0$ 可推出 $\alpha A\beta = 0$, 矛盾。

- $N(B)$ 可由 $N(A)$ 与 β 生成

提供两种做法:

- 从秩的角度考虑, 由于 $\text{rank}(a^{-1}A\beta\alpha A) \leq \text{rank } \alpha$ 可知其为 0 或 1, 从而有

$$\text{rank } B \geq \text{rank } A - \text{rank}(a^{-1}A\beta\alpha A) \geq \text{rank } A - 1$$

利用解空间维数定理可知

$$\dim N(B) \leq \dim N(A) + 1$$

但根据上述分析, 考虑 $N(A)$ 一组基与 β , 由于 $\beta \notin N(A)$ 可知它们线性无关, 因此这是 $N(B)$ 中的 $\dim N(A) + 1$ 个线性无关的向量, 由此只能 $\dim N(B) = \dim N(A) + 1$, 这些向量构成一组基。

- 从方程的角度考虑, $y \in N(B)$ 可推出

$$Ay - a^{-1}A\beta\alpha Ay = 0$$

注意到 αAy 是 1×1 矩阵, 可看作数乘, 即得到

$$A(y - a^{-1}(\alpha Ay)\beta) = 0$$

也即

$$y - a^{-1}(\alpha Ay)\beta \in N(A)$$

从而 y 可写为 $N(A)$ 中某向量加 β 某倍数的形式, 得证。

* 这部分的分值很高, 且大部分同学都没有正确说明。

- $\text{rank } B = \text{rank } A - 1$

与上一部分秩的角度完全相同, 从 $N(B)$ 由 $N(A)$ 、 β 生成与 $\beta \notin N(A)$ 可知

$$\dim N(B) = \dim N(A) + 1$$

再从解空间维数定理得到最终结论。

* 必须要提到解空间维数定理或相应的公式。

§7.2 习题解答

1. 习题 4.4-4

可发现 $A(A^2 - 2A + 3I) = I$, 于是 $A^{-1} = A^2 - 2A + 3I$ 。

2. 习题 4.4-9

利用本节性质 4 与例 5, 将下三角矩阵取转置得到上三角矩阵, 从而有结论。

3. 5.2 节例 8

见例题过程。

4. 习题 4.4-8(1)

将其记为 $AX = B$, 计算可发现 $\det A \neq 0$, 从而 A 可逆, 同左乘 A^{-1} 并计算可得

$$X = A^{-1}B = \frac{1}{7} \begin{pmatrix} 13 & -2 \\ 10 & -13 \\ 18 & -1 \end{pmatrix}$$

5. 习题 4.1-10(3)

直接方法: 此即为求解 $AX = XA$, 而展开可发现其事实上为线性方程组, 从而可利用线性方程组消元法得到结果。

更快的方法: 由于

$$A = 2I + H, \quad H = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

可发现

$$AX = XA \iff (2I + H)X = X(2I + H) \iff 2X + HX = 2X + XH \iff HX = XH$$

直接求解 $HX = XH$ 更为简单, 最终得到结果为

$$\begin{pmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{pmatrix}, \quad a, b, c \in \mathbb{K}$$

6. 习题 3.2-5

直接利用定义或由 3.2 节例 2 计算行列式为 -15 得到线性无关。

§7.3 可逆矩阵

7.3.1 基本性质

首先, 为什么要定义矩阵的可逆性?

对于一个映射 f , 我们可以谈论它的逆映射, 也即满足 $g \circ f$ 与 $f \circ g$ 均为恒等映射的映射 g 。逆映射的存在事实上说明了映射 f 是一个一一对应, 也即某种意义下保持了信息的无损。可逆矩阵的定义是映射的想法自然得来的: 矩阵 A 可逆当且仅当映射 $\mathcal{A}: x \rightarrow Ax$ 为可逆映射, 我们将在下次习题课中进行更详尽的分析。

从映射的角度, 我们可以得出如下的定义 (同样详见下次习题课): 若对矩阵 $A_{m \times n}$, 存在矩阵 $B_{n \times m}$ 使得 $AB = I_m$ 、 $BA = I_n$, 则称矩阵 A 可逆, 并将 B 称为 A 的逆, 记作 $B = A^{-1}$ 。

下面将从此定义出发得到一些结论, 我们先声明一个引理 (可以直接按分量展开计算证明): 若矩阵 B 的列向量为 $\beta_1, \beta_2, \dots, \beta_r$, 且 A, B 可乘, 则

$$AB = (A\beta_1 \ A\beta_2 \ \dots \ A\beta_r)$$

1. 可逆矩阵一定是方阵。

否则不妨设 $m < n$, 有 $\text{rank}(BA) \leq \text{rank } A \leq m < n = \text{rank}(I_n)$, 矛盾。

2. 可逆矩阵必行列式非零, 即满秩。

后两者等价性之前已证。由 $1 = \det I = \det(AB) = \det A \det B$ 可知 $\det A \neq 0$, 从而可逆矩阵行列式非零。

3. 可逆矩阵的逆唯一。

若 A 的逆为 B, C , 则

$$A(B - C) = AB - AC = O$$

而利用引理, 这意味着 A 乘 $B - C$ 的任何一列都是零向量, 由于 A 满秩, 根据解空间维数定理 (或 Cramer 法则) 可知这只能 $B - C$ 任何一列都全 0, 也即 $B = C$ 。

* 事实上, 有了唯一性后 $B = A^{-1}$ 的记号方才严谨。

4. 若对方阵 A, B 有 $AB = I$, 则 $BA = I$, 即 $B = A^{-1}$ 。

由 $AB = I$ 可知 $\det(AB) = 1$, 从而 $\det A \neq 0$, A 满秩。而 $A(BA - I) = ABA - A = IA - A = O$, 与上个性质类似从 A 满秩与引理可推出 $BA - I = O$, 即得证。

5. 满秩矩阵必可逆。

设 e_i 为第 i 个分量为 1, 其他为 0 的向量, 利用 Cramer 法则可知 $Ax = e_i$ 一定有解, 将解记为 β_i 。将所有列向量 β_i 排成一行得到矩阵 B , 利用引理可验证 $AB = I$, 再由性质 4 即得 $B = A^{-1}$ 。

6. 一些简单计算性结论 (下假设 A, B 可逆, $a \neq 0$):

- $I^{-1} = I$;
- $(A^{-1})^{-1} = A$;
- $(AB)^{-1} = B^{-1}A^{-1}$;
- $(aA)^{-1} = \frac{1}{a}A^{-1}$;
- $(A^T)^{-1} = (A^{-1})^T$ 。

利用性质 4 直接计算可得结论, 注意此处第三条结论可以推广为任意多个方阵乘积, 乘积的逆即为逆的逆序乘积。

7. 初等方阵均可逆, 且逆仍为相同类型的初等方阵。

不妨看作行变换, 逆是容易通过行变换的含义构造的:

- 第 i 行乘 k 倍的行变换的逆是第 i 行乘 $\frac{1}{k}$ 倍;
- 第 i, j 行交换的逆是第 i, j 行交换, 即其逆为自身;
- 第 i 行加第 j 行的 k 倍的逆是第 i 行加第 j 行的 $-k$ 倍。

再写出对应的初等变换阵计算验证即可。

8. 有限个初等方阵的乘积可逆。

利用性质 6、7 可直接得到。

9. 对任何矩阵 $A_{m \times n}$, 设 $r = \text{rank } A$, 则存在可逆矩阵 $P_{m \times m}, Q_{n \times n}$ 使得 (P, Q) 中的矩阵称为 A 的相抵标准形)

$$A = P \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Q$$

见本讲义 4.1.3。

10. 可逆矩阵能写为有限个初等方阵乘积。

对可逆矩阵, 由于其满秩, 上式的分解变为了 $A = PQ$, 而根据本讲义 4.1.3 的证明过程, P, Q 均为有限个初等方阵的乘积, 从而得证。

7.3.2 伴随方阵与摄动法

上述的 10 条性质中, 除了第 9 条证明过程相对复杂, 其他都是直接的。不过, 我们显然需要回答一个基本的问题: 给定一个方阵, 如何设计一个算法判定其是否可逆, 并在可逆时计算它的逆?

对于前者, 由于利用性质 2、5 可以直接计算行列式, 问题已经解决。对于后者, 性质 5 的证明过程中事实上也给出了构造: 只要将所有 $Ax = e_i$ 的解拼为一行即可。

设 $B = A^{-1}$, 其第 i 行第 j 列为 b_{ij} , 则由上方讨论, b_{ij} 是 $Ax = e_j$ 的解的第 i 个分量。利用 Cramer 法则, 有

$$b_{ij} = \frac{1}{\det A} \det B_i^{(j)}$$

其中 $B_i^{(j)}$ 代表以 e_j 替换 A 的第 i 列得到的行列式。将此行列式按照第 i 列展开, 由于只有一个分量为 1, 即可发现 $B_i^{(j)} = A_{ji}$, 也即

$$b_{ij} = \frac{1}{\det A} A_{ji}$$

这里 A_{ji} 为 A 对应位置的代数余子式。

由于解中有一个自然的数乘 $\frac{1}{\det A}$, 我们将满足 $c_{ij} = A_{ji}$ 的矩阵 C 记作 A^* , 称为伴随方阵, 则有

$$A^{-1} = \frac{1}{\det A} A^*$$

根据我们的定义, 在 A 可逆时有 $A^*A = AA^* = (\det A)I$ 。不过, 伴随方阵不止对可逆方阵可以定义, 我们需要知道 A 不可逆时是否具有这样的性质, 而完成这件事的常用方法是摄动法 (这个词来自物理, 意为微小扰动), 我们先介绍一个完整的标准摄动法过程, 下方的讨论都在 \mathbb{R} 上进行:

- 对任何 n 阶方阵 A , 至多有 n 个 λ 使得 $\lambda I + A$ 不可逆。

考虑完全展开可知 $\det(\lambda I + A)$ 是 λ 的 n 次多项式, 且首项系数为 1 (可见期中复习题 8(3) 证明过程), 由于 $n > 1$, 根据代数学基本定理可知其必有 n 个复根 (含重数), 从而至多有 n 个实根。

- 对任何 n 阶方阵 A , 存在 δ 使得 $\lambda \in (-\delta, 0) \cup (0, \delta)$ 时 $\lambda I + A$ 可逆。

由于其至多有 n 个实根, 其非零实根至多 n 个, 它们的绝对值一定能找到最小值, 取 δ 比最小值小即可。

- $(\lambda I + A)^*(\lambda I + A)$ 、 $(\lambda I + A)(\lambda I + A)^*$ 与 $\det(\lambda I + A)I$ 的每个位置都是对 λ 连续的函数。

由于它们的每个位置都是通过加法、减法、乘法得到的, 而 $\lambda I + A$ 的每个位置是 λ 的多项式, 从而它们的每个位置也是 λ 的多项式, 因此是连续函数。

- 对任何 n 阶方阵 A 有 $A^*A = AA^* = (\det A)I$ 。

取第二部分证明中的 δ , 则对 $(-\delta, 0) \cup (0, \delta)$ 中的任何 λ , 由可逆性知有

$$(\lambda I + A)^*(\lambda I + A) = (\lambda I + A)(\lambda I + A)^* = (\det(\lambda I + A))I$$

令 $\lambda \rightarrow 0$, 利用连续性可知 0 处三者相等仍然成立, 从而得证。

不过, 在上方的讨论中, 我们已经可以发现, 在谈论可逆矩阵相关的问题时, 数域的选取成为了一个巨大的难题。在只涉及加、减、乘、除时, 任何数域 \mathbb{K} 都可以做完全类似的处理, 但只要涉及多项式的根, 不同数域会有完全不同的结果。这个难题的影响将在之后讨论相似时彻底爆发, 而至少在此处, 我们可以通过一个权宜之计来解决:

- 由于任何数域上的加减乘除都与 \mathbb{C} 上完全一致, 我们可以将任何数域上的矩阵看成 \mathbb{C} 上的矩阵。此时, 上方证明的第一步仍然成立, 第二步也仍然可以找到 (只找 \mathbb{R} 上的 λ)。对于第三步, 这里需要将每个位置的实部、虚部分开看, 但由于只涉及了加、减、乘, 实部和虚部必然都是 λ 的多项式, 因此仍然为对 λ 连续的函数。最后, 第四步证明考虑实部、虚部分别的极限, 仍然可以成立。
- 但是, 这样的证明依托在 \mathbb{C} 具有连续的结构。事实上矩阵的定义可以在一般的域 (抽象代数概念) 而非数域上, 这时域的结构就未必连续了。不过, 那时, 类似的证明仍然可以完成: 将 $\lambda I + A$ 的每个位置看作 A 的多项式, 并考虑域上的有理函数域, 其在有理函数域中一定可逆, 从而可证明左右的每个位置在作为 λ 的多项式时是相等的, 于是它们代入 $\lambda = 0$ 能得到相同的值, 这就完成了证明。

我们举另一个摄动法的经典例子：证明 $(AB)^* = B^*A^*$ 。为说明方便，这里仍然考虑 \mathbb{R} 上，在 \mathbb{C} 中时拆分实虚部可完全类似证明。

- 对可逆阵 A, B 有 $(AB)^* = B^*A^*$ 。

直接计算可知（利用性质 6）

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

- 对任何 n 阶方阵 A, B ，存在 δ 使得 $\lambda \in (-\delta, 0) \cup (0, \delta)$ 时 $\lambda I + A$ 与 $\lambda I + B$ 均可逆。

与上方类似，至多有 $2n$ 个 λ 使得 $\lambda I + A$ 或 $\lambda I + B$ 不可逆，取它们中非零的绝对值最小值，令 δ 比其小即可。

- $((\lambda I + A)(\lambda I + B))^*$ 与 $(\lambda I + B)^*(\lambda I + A)^*$ 的每个位置都是对 λ 连续的函数。

由于运算只涉及了加、减、乘，同样可以得到每个位置都是 λ 的多项式。

- 对任何方阵 A, B 有 $(AB)^* = B^*A^*$ 。

由上方取定 δ 并令 $\lambda \rightarrow 0$ 即可。

不过，当涉及不连续时，无法通过摄动法解决，例如如下经典结论，对 n 阶方阵 A 有：

1. $\text{rank } A = n \iff \text{rank } A^* = n$ 。

当 A 可逆时， A^* 为 A^{-1} 的非零倍数，从而秩为 n 。

反之，若 A^* 可逆、 A 不可逆，有 $A^*A = (\det A)I = O$ ，但从 A^* 可逆类似之前性质证明知只能 $A = O$ ，但此时 $A^* = O$ ，矛盾。

2. $\text{rank } A < n - 1 \iff \text{rank } A^* = 0$ 。

由于 A^* 为 A 的全部 $n - 1$ 阶子式加符号拼成的矩阵，当 $\text{rank } A < n - 1$ 时 $A^* = O$ 。

反之，若 $A^* = O$ ，可知其全部 $n - 1$ 阶子式非零，而 $\det A$ 按第一行展开可发现亦为 0，从而非零子式至多 $n - 2$ 阶， $\text{rank } A < n - 1$ 。

* 这里并不能通过“摄动法”得出 $\text{rank } A^*$ 仍为 n 的荒谬结论，本质是因为秩作为函数不连续。

3. $\text{rank } A = n - 1 \iff \text{rank } A^* = 1$ 。

利用本讲义 3.1 已证的习题 3.7-5， $\text{rank } A = n - 1$ 时 A^* 任何两行成比例，也即极大线性无关组至多只有一行，于是 $\text{rank } A^* \leq 1$ ，而由于 A 存在 $n - 1$ 阶的非零子式， $A^* \neq O$ ，于是只有 $\text{rank } A^* = 1$ 。

当 $\text{rank } A^* = 1$ 时，由于前两种情况 $\text{rank } A < n - 1$ 与 $\text{rank } A = n$ 已经排除，只能 $\text{rank } A = n - 1$ 。

7.3.3 消去与拆分

可逆阵的一大作用是进行消去，具体来说，对可逆阵 A 有

$$AX = AY \implies X = Y$$

$$XA = YA \implies X = Y$$

第一式可以通过等式两边同时左乘 A^{-1} 得到，第二式可以通过矩阵两边同时右乘 A^{-1} 得到。务必注意，由于矩阵乘法的不可交换性，左乘右乘必须区分清楚。

利用消去律, 同样可以非常容易得到线性方程组的解。也即, 对可逆阵 A , $Ax = b$ 当且仅当 $x = A^{-1}b$ 。

另一个可逆的常用方法是拆分, 也即利用 $P^{-1}P = I$, 可写出诸如 $AB = AP^{-1}PB$ 这样“凭空造出”想要的 PB 的结构。为了介绍这样做好处, 我们需要先证明如下重要结论: $A_{n \times n}$ 可逆当且仅当 $\mathcal{A}: \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}$, $\mathcal{A}(X) = AX$ 为可逆映射。

- 当: 反证。若 A 不可逆, 利用解空间维数定理可知 $Ax = 0$ 有非零解 $x \in \mathbb{R}^{n \times 1}$ 。将 m 个 x 排成一行成为 X , 利用本讲义 7.3.1 中的引理可知 $AX = O$, 且直接计算有 $AO = O$, 于是 \mathcal{A} 不为单射, 不可逆, 矛盾。
- 仅当: 利用消去律可知 $AX = AY$ 时 $X = Y$, 从而为单射, 而对任何 $Y \in \mathbb{R}^{m \times n}$ 有 $A(A^{-1}Y) = Y$, 从而为满射。

由此, 对可逆矩阵 A , 我们可以放心地做 $y = Ax$ 或 $Y = AX$ 这样的换元, 因为若换元前结论对所有 X 成立, 根据双射性, 换元后结论应对所有 Y 成立, 反之亦然。

回顾本讲义 4.2.2 的那些例题, 至此, 我们已经几乎完全掌握了可逆如何作为工具进行运用。可以发现, 4.2.2 节中的所有证明几乎都与秩密切相关, 因此之前介绍的性质 9 (相抵标准形) 事实上可以成为可逆矩阵解决一般矩阵问题的桥梁。正是因为一般矩阵可以在乘可逆矩阵后化为充分简单的形式, 我们才能利用消去与拆分解决问题。

——不过直到现在我们还没有解释到底为什么那个东西称为相抵标准形。到底什么是相抵, 标准形又体现在何处呢?

§7.4 相抵与相似

7.4.1 相抵标准形

早在本讲义 1.2.2, 我们就以简化阶梯形矩阵解释了何为等价类与标准形。简单来说, 等价类是基于某种等价关系的分类方法, 而标准形 (或代表元) 则对应其中最简单的一个。

用代表元讨论问题的好处显而易见: 在“线性方程组可同解变形得到”对应的等价类中, 从系数矩阵看出原方程组的解是几乎不可能的, 但从简化阶梯形系数矩阵看出原方程组的解则是直接的, 因此, 讨论线性方程组解相关问题时, 只要先找到简化阶梯形矩阵对应的性质, 再想办法化归为简化阶梯形即可。

在此之前, 我们考虑的复杂问题往往是关于秩的, 而若条件只有秩相关, 我们希望能通过如下三个步骤证明:

1. 在所有秩相同的矩阵中找到最易于使用的一个 Σ ;
2. 证明原问题对 Σ 成立;
3. 将一般问题化归为 Σ 的问题。

由此, 我们直接给出相抵的定义: 两个 $m \times n$ 矩阵相抵当且仅当它们秩相等。以此划分等价类后, 根据之前的性质 9 即可得到两个矩阵相抵当且仅当它们具有相同的相抵标准形。

由于相抵标准形的形式足够简单, 它的确可以作为第一步中的“最易于使用”。但是, 如果只进行这样的定义, 我们并不知道第三步的“化归”如何进行。也即, 我们需要把具有相同不变量的矩阵联系上。由此, 我们需要给出等价定义: 两个矩阵 A, B 相抵当且仅当存在可逆阵 P, Q 使得 $A = PBQ$ 。

- 仅当: 设 A, B 的相抵标准形为 Σ , 则存在可逆阵 P_A, P_B, Q_A, Q_B 使得 $A = P_A \Sigma Q_A, B = P_B \Sigma Q_B$, 从而 $B = P_B P_A^{-1} A Q_A^{-1} Q_B$ 。
- 当: 若 $A = PBQ$, 利用乘积秩关系可知 $\text{rank } A \leq \text{rank } B$, 而 $B = P^{-1} A Q^{-1}$, 于是 $\text{rank } B \leq \text{rank } A$, 即得证。

* 证明过程即是利用可逆阵的消去律任意移动。我们事实上也证明了重要性质：乘可逆阵不改变秩。

下面给出一个用以上过程解决问题的实例：证明 $\text{rank } AB = \text{rank } B$ 时存在 C 使得 $CAB = B$ 。

- 先证明当 $A = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$ 时结论成立。此时，计算可知 AB 即为将 B 前 r 行以外的行改成 0，因此 $\text{rank } AB = \text{rank } B$ 当且仅当 B 的前 r 行有行向量的极大线性无关组。

根据极大线性无关组的性质， B 的前 r 行可以表出 B 的任何行向量。而计算可发现 CB 即为对 B 的每一行进行线性组合，从而即可利用表出方式构造 C ，具体可见期中复习题 7(2) 的解答。

- 再证明对任何 A 可化为上述情况。设 A 的相抵标准形为 Σ ，存在可逆阵 P, Q 使得 $A = P\Sigma Q$ ，则条件可改为

$$\text{rank } P\Sigma Q = \text{rank } B$$

利用乘可逆阵不影响秩，左侧左乘 P^{-1} 、右侧左乘 Q (这是为了凑出相同形式) 即得到

$$\text{rank}(\Sigma(QB)) = \text{rank}(QB)$$

由此，根据已经证明的情况，存在 C_0 使得 $C_0\Sigma Q = QB$ ，由此即有

$$Q^{-1}C_0\Sigma Q = B$$

再拆分出 P 得到

$$(Q^{-1}C_0P^{-1})(P\Sigma Q)B = B$$

令 $C = Q^{-1}C_0P^{-1}$ 即可。

* 一般来说，由于我们无法提前知道如何进行化归是合理的，上面的第二步往往会先于第一步进行，也即先化归出较简单形式再对此形式证明，而非先证明简单形式再设法化归。更多的例子见本讲义第四章的矩阵技巧相关。

7.4.2 相似的动机

在刚才的例子中我们已经发现，若研究关于某个不变量的问题，可以考虑以其定义的等价类中最易于处理的元素，也即标准形。然而，更多的问题中，往往难以直接看出不变量，因此需要先进行化归的尝试。我们考虑两个问题：

- (可交换性问题) 对 n 阶方阵 A ，求所有 B 使得 $AB = BA$ 。

出于之前的讨论，我们可以尝试考虑相抵下的结果。假设 $A = P\Sigma Q$ ，其中 Σ 为相抵标准形， P, Q 可逆，则上述方程可以化为

$$P\Sigma Q = BP\Sigma Q$$

利用消去律可以合并为

$$\Sigma(QBQ^{-1}) = (P^{-1}BP)\Sigma$$

然而，由于 QBQ^{-1} 和 $P^{-1}BP$ 未必相等——注意矩阵乘法不能交换——上式是无法化简的。不过，可以发现若 $Q = P^{-1}$ ，则 $QBQ^{-1} = P^{-1}BP$ ，问题就变为了简单的形式。

由此，我们考虑 $A = PA_0P^{-1}$ ，完全类似计算可知原方程化为

$$A_0(P^{-1}BP) = (P^{-1}BP)A_0$$

反之, 若 $A_0 B_0 = B_0 A_0$, 由于 $A_0 = P^{-1}AP$, 可计算得到

$$A(PB_0P^{-1}) = (PB_0P^{-1})A$$

利用乘可逆阵为可逆映射, 可发现与 A_0 可交换的矩阵集合同与 A 可交换的矩阵集合存在一一对应, 映射即为

$$B_0 \rightarrow PB_0P^{-1}$$

于是若 A_0 的形式相对简单, 可以完全解决与 A_0 可交换的矩阵, 则原方程的解即为

$$\{PB_0P^{-1} \mid A_0 B_0 = B_0 A_0\}$$

2. (多项式问题) 解矩阵方程 $A^n = I$ 。

同样先考虑相抵标准形, 设分解为 $A = P\Sigma Q$, 有

$$A^n = P\Sigma Q P\Sigma Q P\Sigma Q \cdots P\Sigma Q$$

此形式并没有很好的刻画, 但若 $Q = P^{-1}$, 中间的部分即可全部消去。

由此, 我们考虑 $A = PA_0P^{-1}$, 直接计算可知

$$A^n = P^{-1}A_0^n P = I$$

进一步消去即得到 $A_0^n = I$ 。

这启发我们, 如果存在可逆的 P 使得 $A = PA_0P^{-1}$ 是一个等价关系, 我们就可以进行分类。由于每一类或都满足 $A^n = I$, 或都不满足, 我们只需验证所有代表元是否满足 $A^n = I$, 即可得到一般的 $A^n = I$ 的解。

出于以上观察, 我们定义 n 阶方阵 A 与 B 相似当且仅当存在可逆的 P 使得 $A = PBP^{-1}$ 。

7.4.3 相似的基本结论

与刚引入逆时类似, 我们希望用一系列基本的结论去研究相似的性质, 下面默认写出的矩阵均为 n 阶方阵, 并设 f 为多项式:

1. (自反性/反身性) A 与自身相似。

由 $I^{-1} = I$ 可知 $A = I^{-1}AI$ 。

2. (对称性) A 与 B 相似则 B 与 A 相似。

设 $A = PBP^{-1}$, 则消去可知 $B = P^{-1}AP$, 而 $P = (P^{-1})^{-1}$, 由此将 P^{-1} 看作整体可发现相似。

3. (传递性) A 与 B 相似、 B 与 C 相似则 A 与 C 相似。

设 $A = PBP^{-1}$ 、 $B = QCQ^{-1}$, 则 $A = PQCQ^{-1}P^{-1}$ 。由可逆性质 $Q^{-1}P^{-1} = (PQ)^{-1}$, 从而将 PQ 看作整体可发现相似。

4. 相似是一个等价关系, 从而可以分出等价类。

上面三条性质即为等价关系的定义。

5. A 与 B 相似则 $\text{rank } A = \text{rank } B$ 。

由乘可逆阵不改变秩可得到结论。

6. A 与 B 相似则 $\text{tr } A = \text{tr } B$ 。(定义见期中复习题 9)

设 $A = PBP^{-1}$, 由期中复习题 9 已证的 $\text{tr}(AB) = \text{tr}(BA)$ 可知 $\text{tr}(PBP^{-1}) = \text{tr}(P^{-1}PB) = \text{tr } B$ 。

7. A 与 B 相似则 $\det A = \det B$ 。

设 $A = PBP^{-1}$, 由 Binet-Cauchy 公式有 $\det A = \det P \det B (\det P^{-1}) = \det B \det(PP^{-1}) = \det B$ 。

8. 对任何 a , 与 aI 相似的矩阵只有 aI 。

直接计算可得结论, 这本质上是由于 aI 与任何矩阵可交换。

9. 即使 $\text{rank } A = \text{rank } B$ 、 $\text{tr } A = \text{tr } B$ 、 $\det A = \det B$, A 与 B 也未必相似。

考虑 $A = I$, $B = I + \mu E_{12}$, 这里 E_{12} 表示第一行第二列为 1, 其他为 0 的矩阵。计算可发现两矩阵均满秩、迹为 n 且行列式为 1, 但由性质 8 可知不相似。

这意味着, 以上三个量仍然不足以刻画相似下的分类——但它们已经是我们学过的所有关于矩阵的量了……这个悲伤的事实表明, 矩阵在相似下的分类可能会非常复杂。

10. 设 $A = PBP^{-1}$, 则 $f(A) = Pf(B)P^{-1}$ 。

在上一节中我们已经计算了 $A^n = PB^nP^{-1}$, 而直接计算可知 $aA = P(aB)P^{-1}$, 从而对单项式已经成立。另一方面, 若 $X = PYP^{-1}$ 、 $Z = PWP^{-1}$, 利用两次乘法分配律可知 $X+Z = P(Y+W)P^{-1}$, 于是对单项式进行若干次求和可知结论对多项式成立。

由此, 矩阵的多项式相关问题可以通过相似进行刻画。更进一步地, 如果相似标准形真的存在, 我们希望它易于计算多项式, 这样任何矩阵的多项式都方便计算了。

* 冷知识, 相似标准形的多项式计算已经在之前的作业里出现过了, 这意味着它确实简单到了刚学矩阵乘法的人也能计算出结果的程度。

11. A 与 B 相似则 $f(A)$ 与 $f(B)$ 的 rank 、 tr 、 \det 均相同。

由性质 10 可知 $f(A)$ 与 $f(B)$ 相似, 从而得证。特别地, 考虑 $\text{rank} = 0$ 的情况可发现 $f(A) = O$ 当且仅当 $f(B) = O$, 由此某种意义上矩阵多项式的“根”是若干个相似下的等价类。

12. 即使 $\text{rank } A = \text{rank } B$ 、 $\text{tr } A = \text{tr } B$ 、 $\det A = \det B$, 且 $f(A) = O$ 当且仅当 $f(B) = O$, A 与 B 也未必相似。

这个例子较为复杂, 这里不加验证地写出结果: $A = I + E_{12} + E_{34}$ 、 $B = I + E_{12}$ 。

至此, 相似的特性似乎已经非常复杂了, 但还有更致命的一件事在。我们先给出两个方阵相似的更严谨版本定义: 数域 \mathbb{K} 上的 n 阶方阵 A 与 B 相似当且仅当存在可逆的 \mathbb{K} 上方阵 P 使得 $A = PBP^{-1}$ 。

由此, 会自然产生一个问题: 若对两个 \mathbb{K} 上的 n 阶方阵 A, B , 存在可逆的 \mathbb{C} 上方阵 P 使得 $A = PBP^{-1}$, 它们是否一定相似? 也即, 因为 $\mathbb{K} \subset \mathbb{C}$, \mathbb{K} 上的方阵自然可以看作复方阵, 但看作复方阵时的相似性与原本的相似性究竟是否一致呢?

对于相抵关系, 这个问题是容易解决的。我们之前事实上默认了一个结论: 一个方阵看作任何数域上时相抵标准形不变。这是由于我们进行的行列变换只需要加、减、乘、除, 而这些操作是可以保证在同一个数域中进行的。比起它更早的结论是, 一个向量组看作任何数域上时线性相关性不变, 这也可以直接通过线性方程组的解一定在系数所在的数域中 (仍然因为只进行了加减乘除) 而得到。

但对于相似关系, 似乎并没有一个直接的方法能判断这件事。幸运的是, 这件事的确是成立的, 也即在任何数域上的相似与 \mathbb{C} 上的相似等价。不幸则比幸运更多: 首先, 这个结论的证明需要用到非常后面的知识 (不变因子), 并没有简单的办法; 其次, 虽然不同数域上的相似确实等价, 但相似标准形并不相同, 这是因为 \mathbb{C} 上的相似标准形可能出现虚数, 从而未必能在更小的数域中写出。

为了规避这样的问题，我们此后谈论相似时默认考虑复方阵，有的结论对更一般的数域上也成立（例如之前列出的所有性质），希望大家能在阅读的过程中仔细分辨。

最后，有的同学可能注意到了，在给出性质 11 后，性质 12 并没有否定 $\text{rank } f(A) = \text{rank } f(B)$ 、 $\text{tr } f(A) = \text{tr } f(B)$ 、 $\det f(A) = \det f(B)$ 恒成立时 A 与 B 的相似性。事实上，可以给出一个终极结论（右推左已经在性质 11 中）：

$$A, B \text{ 相似当且仅当对任何多项式 } f \text{ 有 } \text{rank } f(A) = \text{rank } f(B)。$$

不仅如此，通过此结论还可以给出相似标准形的一个计算方法。不过，这个结论的证明实在是太过复杂了——也并不在这学期的范围内，因此我们必须回到“寻找不变量”上一点点进行研究。

7.4.4 特征系统

在之前我们已经提到，无论是秩、行列式还是迹，对于刻画相似下的不变性都过于弱了。不过，由于 $f(A)$ 与 $f(B)$ 的秩、行列式与迹仍然相同，我们可以用矩阵多项式定义一些性质。

出于空间角度的考虑（详见下次习题课），我们可以考察如下的矩阵方程（称为 A 的特征方程）：

$$Ax = \lambda x$$

由于它可以改写为 $(\lambda I - A)x = 0$ ，利用解空间维数定理可知其有非零解当且仅当 $\det(\lambda I - A) = 0$ 。记 $\varphi_A(\lambda) = \det(\lambda I - A)$ ，类似本讲义 7.3.2 中摄动法的第一步，可知 $\varphi_A(\lambda)$ 是 λ 的 n 次多项式，且首项系数为 1，称为特征多项式。

利用代数学基本定理，其可以分解为（代数学基本定理并不在这门课的范围内，因此直接相信这个结论就好）

$$\varphi_A(\lambda) = (\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_k)^{m_k}$$

其中 $\lambda_1, \dots, \lambda_k$ 是互不相同的复数，它们构成 $\varphi_A(\lambda)$ 的全部根，称为 A 的特征值，指数 m_i 即根的重数，称为对应特征值的代数重数。

* 利用代数学基本定理，所有不同特征值代数重数和为矩阵阶数 n 。

这样，特征方程有解当且仅当 λ 为 A 的某个特征值 λ_i ，将解称为属于该特征值的特征向量，解空间称为该特征值的特征子空间，解空间的维数称为该特征值的几何重数。

由于我们已经证明过 $\det f(A) = \det f(B)$ ，即可知 $\det(\lambda I - A) = \det(\lambda I - B)$ ，于是相似矩阵的特征值与特征多项式相同，从而进一步有相似矩阵对应特征值代数重数相同。另一方面，对特征值 λ_i ，利用解空间维数定理可知 A 的特征值 λ_i 几何重数为 $n - \text{rank}(\lambda_i I - A)$ ，同理 B 的特征值 λ_i 几何重数为 $n - \text{rank}(\lambda_i I - B)$ ，利用已证的 $\text{rank } f(A) = \text{rank } f(B)$ 可知相似矩阵对应特征值几何重数相同。

对于特征子空间的关系，设 $A = PBP^{-1}$ ，若 $Ax = \lambda x$ ，利用消去律（注意数乘可以任意换位）有

$$PBP^{-1}x = \lambda x \iff B(P^{-1}x) = \lambda(P^{-1}x)$$

由于 P^{-1} 可逆， $x \rightarrow P^{-1}x$ 是一一对应，也即相似矩阵相同特征值的特征子空间存在一一对应（而此映射由定义可验证是线性的，也即相似矩阵相同特征值的特征子空间存在自然的线性同构，这将在下一次习题课中详细说明）。

八 补充：可逆的空间视角

§8.1 逆矩阵的计算

本部分是补充上次习题课中未涉及的可逆矩阵相关的一些操作。

8.1.1 伴随方阵 vs 初等变换

在上次习题课，我们给出了可逆矩阵的逆的一个理论计算方法：

$$A^{-1} = \frac{1}{\det A} A^*$$

不过，这个方法需要计算 n^2 个行列式，复杂度是过于高的。为了寻找更简单的方法，我们希望能有一个类似行列式计算时的直接行列变换的思路。

受到相抵标准形的启发，我们知道， A 可以由一系列初等方阵组合而成，而将这些初等方阵逐个逆向操作即可得到 A 的逆，而这恰恰对应将 A 变换为 I 的行列变换。由此，只要将 A 行列变换为 I ，同时对 I 施以相同的行列变换，即能得到 A^{-1} 。

由于初等方阵都可以看作行变换，而对 A, I 同时进行相同的行变换也即对 $(A \ I)$ 进行行变换，我们可以得到算法：对 $(A \ I)$ 进行行变换得到 $(I \ B)$ ，则 $B = A^{-1}$ 。用矩阵乘法表示也即

$$A^{-1} (A \ I) = (I \ A^{-1})$$

正如行列式的行列变换，此方法是计算矩阵求逆的一般方法，对于给定阶数的矩阵也有着所有算法中较低的复杂度。

* 我们并没有给出将 $(A \ I)$ 行变换为 $(I \ B)$ 的算法，因为这事实上就是行变换得到简化阶梯形矩阵的算法。至于为何简化阶梯形矩阵一定是如此形式，注意到消元的过程不会改变前 r 列的线性无关性，根据主元的性质即可得到。

考虑习题 4.5-10(2)，即计算上三角部分为 1、其他为 0 的矩阵 B 的逆。

将此矩阵化为 I 只需从第一行开始每行减下一行，将完全相同的操作作用到 I 上即得到

$$\begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \\ & & & & 1 \end{pmatrix}$$

从而这就是 B 的逆。

* 对习题 4.5-10(3)，只需将“每行减下一行”的操作再重复一遍即可。

理论上来说，正如完全展开式的作用，对于矩阵元素完全对称的场合，伴随方阵可能具有更好的效果。不过，由于此时基本都是运用整体思想的情况，我们将在下一部分进行介绍。

8.1.2 整体思想

整体思想，也即不拆分元素进行整体的考虑。这个思想本身是简单的，但将其运用好也需要一些技巧，我们直接以例题说明：

1. 若方阵 A 的第一行全为 1，证明其所有代数余子式之和为 $\det A$ 。

记 γ 为所有元素均为 1 的列向量， A 的第一行全为 1 可以得到

$$e_1^T A = \gamma^T$$

设 A 的伴随方阵为 A^* ，计算可发现 A 所有代数余子式之和即为

$$\gamma^T A^* \gamma$$

由此利用 $AA^* = (\det A)I$ 计算即得

$$\gamma^T A^* \gamma = (e_1^T A) A^* \gamma = e_1^T (\det A) I \gamma = (\det A) e_1^T \gamma = \det A$$

2. 若 n 阶方阵 A 的每行和全为 1, 证明其所有代数余子式之和为 $n \det A$ 。

记 γ 为所有元素均为 1 的列向量, A 的每行和全为 1 可以得到

$$\gamma^T A = \gamma^T$$

设 A 的伴随方阵为 A^* , 计算可发现 A 所有代数余子式之和即为

$$\gamma^T A^* \gamma$$

由此利用 $AA^* = (\det A)I$ 计算即得

$$\gamma^T A^* \gamma = (\gamma^T A) A^* \gamma = \gamma^T (\det A) I \gamma = (\det A) \gamma^T \gamma = n \det A$$

3. 对 $A_{m \times n}, B_{n \times m}$, 若 $I_m - AB$ 可逆, 则 $I_n - BA$ 可逆, 并求其逆。

* 一个想法: 利用等比数列求和可知 $|x| < 1$ 时 $(1-x)^{-1} = 1 + x + x^2 + \dots$, 从而可以写出

$$(I - BA)^{-1} = I + BA + BABA + \dots = I + B(I + AB + ABAB + \dots)A = I + B(I - AB)^{-1}A$$

毫无疑问, 这个过程是不严谨的, 因为我们并没有考虑级数的收敛性问题, 但它至少可以用来猜测结果。

验证结果是简单的, 利用乘法分配律有

$$\begin{aligned} (I - BA)(I + B(I - AB)^{-1}A) &= I - BA + B(I - AB)^{-1}A - BAB(I - AB)^{-1}A \\ &= I - BA + B(I - AB)(I - AB)^{-1}A \\ &= I - BA + BA = I \end{aligned}$$

4. 对同阶可逆方阵 A, D 与和它们同阶的方阵 B , 若 $B^T AB + D^{-1}$ 可逆, 证明

$$(A + BDB^T)^{-1} = A^{-1} - A^{-1}B(B^T A^{-1}B + D^{-1})^{-1}B^T A^{-1}$$

直接计算

$$\begin{aligned} (A + BDB^T)(A^{-1} - A^{-1}B(B^T A^{-1}B + D^{-1})^{-1}B^T A^{-1}) \\ = I - B(B^T A^{-1}B + D^{-1})^{-1}B^T A^{-1} + BDB^T A^{-1} - BDB^T A^{-1}B(B^T A^{-1}B + D^{-1})^{-1}B^T A^{-1} \end{aligned}$$

为消去, 合并含逆的两项得到

$$I + BDB^T A^{-1} - B(I + DB^T A^{-1}B)(B^T A^{-1}B + D^{-1})^{-1}B^T A^{-1}$$

可以发现此形式与逆中已经非常像了, 而利用消去律可在左侧提出 D , 从而其为

$$\begin{aligned} I + BDB^T A^{-1} - BD(D^{-1} + B^T A^{-1}B)(B^T A^{-1}B + D^{-1})^{-1}B^T A^{-1} \\ = I + BDB^T A^{-1} - BDB^T A^{-1} = I \end{aligned}$$

5. 若 $A^m = O$, 求以下矩阵的逆:

$$I + \sum_{k=1}^{m-1} \frac{1}{k!} A^k$$

由于 $A^m = O$, 求逆的目标是 A 的多项式 $f(A)$, 只要找到多项式 $g(x)$ 使得

$$f(x)g(x) = 1 + x^m h(x)$$

其中 h 为多项式，即有

$$f(A)g(A) = I + A^m h(A) = I$$

于是 $g(A)$ 即为 $f(A)$ 的逆。

为了找出符合上述形式的多项式，由于 $A^m = O$ ，只需找至多 $m - 1$ 次的多项式即可。设

$$g(x) = \sum_{k=0}^{m-1} a_k x^k$$

而由于 (设 $0! = 1$)

$$f(x) = \sum_{k=0}^{m-1} \frac{1}{k!} x^k$$

有

$$a_0 = 1, \quad \forall k = 1, \dots, m-1, \quad \sum_{s=0}^k \frac{1}{s!} a_{k-s} = 0$$

算几项寻找规律后可直接归纳得到

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

总结来说，整体思想也即尽量将条件表述为矩阵乘法，并利用整体关系式配凑。

8.1.3 分块行列变换

对于能够进行整体配凑的题目，使用整体方法当然是简单的。但是，更多情况下，将其完全视为整体无法解决问题，而展开为元素则过于复杂。由此，我们需要寻求一个整体与局部之间的方法，也即分块矩阵。具体的基本结论见本讲义 5.2.1，我们将其进行简单的推广，总结为“看着能乘就能乘”。

具体来说，如果两个矩阵都写成了分块的矩阵，一个分为了 $a \times b$ 块，另一个分为了 $c \times d$ 块，为了让它们“看着像”可以进行元素的乘法，必须 $b = c$ ；此外，在 $b = c$ 时，我们可以形式上地写出每块看作元素的矩阵乘法结果，为了让结果合理，其中涉及的所有块的乘法必须是可以进行的 ($a = b = c = 2$ 时的具体阶数要求见本讲义 5.2.1，这也是我们运用最普遍的情况)。

我们断言，只要以上两个条件满足，分块矩阵的确可以看作元素进行乘法，但务必注意，块与元素不同之处在于左右乘不可交换。

* 此结论证明见教材 4.5 节，需要一些非常复杂的计算，完全可以跳过。

利用这样的结论，我们马上可以写出分块的“初等方阵”，这里以 2×2 为例，并以行变换 (左乘) 作为解释：

1. 第一类初等变换，对元素即将某行的倍数加到另一行，其分块形式为

$$\begin{pmatrix} I_a & Y \\ O & I_b \end{pmatrix}, \quad \begin{pmatrix} I_a & O \\ Z & I_b \end{pmatrix}$$

它们作为行变换的意思分别是，将后 b 行左乘 Y 后加到前 a 行、将前 a 行左乘 Z 后加到后 b 行，且计算验证可知逆分别为

$$\begin{pmatrix} I_a & -Y \\ O & I_b \end{pmatrix}, \quad \begin{pmatrix} I_a & O \\ -Z & I_b \end{pmatrix}$$

2. 第二类初等变换，对元素即交换两行，其分块形式为

$$\begin{pmatrix} O & I_a \\ I_b & O \end{pmatrix}$$

它作为行变换的意思是，将后 a 行与前 b 行交换，计算验证可知逆为

$$\begin{pmatrix} O & I_b \\ I_a & O \end{pmatrix}$$

3. 第三类初等变换，对元素即某行乘非零倍数，其分块形式为

$$\begin{pmatrix} P_{a \times a} & O \\ O & I_b \end{pmatrix}, \quad \begin{pmatrix} I_a & O \\ O & Q_{b \times b} \end{pmatrix}$$

它们作为行变换的意思分别是，前 a 行左乘矩阵 P 、后 b 行左乘矩阵 Q 。为使结果可逆，计算行列式可发现这等价于 P 或 Q 可逆，计算验证可知逆分别为

$$\begin{pmatrix} P^{-1} & O \\ O & I_b \end{pmatrix}, \quad \begin{pmatrix} I_a & O \\ O & Q^{-1} \end{pmatrix}$$

用分块行列变换解决秩问题的例子可以参考本讲义 3.3.2 最后较难的问题，利用它们应当可以严谨写出过程了，留给大家自己完成。为了展示其与可逆结合的应用，我们仍然研究几道例题：

1. 若 $\lambda I - A$ 可逆，且 $\text{rank } A = r$ ，在 r 很小时，给出一个快速判断是否可逆并计算逆的方法。

* 习题 4.4-10(1)(4)(5) 均是这类问题，(4)(5) 中 $r = 1$ ，(1) 中 $r = 2$ 。

利用满秩分解（见本讲义 4.2.1 后方习题或 4.3 节例题），可将 A 分解为 $n \times r$ 矩阵 B 与 $r \times n$ 矩阵 C 的乘积。由此问题变为研究

$$\lambda I - BC$$

在 4.5 节中已经证明 $\det(I - XY) = \det(I - YX)$ ，从而 λ 非零时

$$\det(\lambda I - BC) = \lambda^n \det(I - \lambda^{-1}BC) = \lambda^n \det(I - C\lambda^{-1}B) = \lambda^{n-r} \det(\lambda I - CB)$$

利用连续性可知 $\lambda = 0$ 时上式仍成立。由于 $r \leq n$ ，可逆当且仅当 $\det(\lambda I - CB) \neq 0$ ，且 $r < n$ 时 $\lambda \neq 0$ 。

下面假设上述条件成立，计算对应的逆。直接使用上节例题 3 的结论自然是一个办法，除此之外，我们也介绍一种从行列变换的视角出发解决的方法。当 $r = n$ 、 $\lambda = 0$ 时的情况即为直接计算 A 的逆，下设 $r < n$ 且 $\lambda \neq 0$ 。

类似行列式时的证明，我们可以想到在 $\lambda I_n - BC$ 的左上角增添一个 I_r ，并设法转移。利用分块计算可知

$$\begin{pmatrix} I_r & \\ & \lambda I_n - BC \end{pmatrix}^{-1} = \begin{pmatrix} I_r & \\ & (\lambda I_n - BC)^{-1} \end{pmatrix}$$

于是只要计算整体的逆即可。

为了进行消去，我们先将前 r 列右乘 C 倍（请务必注意左乘行变换右乘列变换）加到后 n 列，得到

$$\begin{pmatrix} I_r & \\ & \lambda I_n - BC \end{pmatrix} \begin{pmatrix} I_r & C \\ & I_n \end{pmatrix} = \begin{pmatrix} I_r & C \\ & \lambda I_n - BC \end{pmatrix}$$

前 r 行左乘 B 倍加到后 n 行即可消去右下角（这些操作某种意义上都可以将原方阵看成 2×2 矩阵得到），得到

$$\begin{pmatrix} I_r & \\ B & I_n \end{pmatrix} \begin{pmatrix} I_r & C \\ & \lambda I_n - BC \end{pmatrix} = \begin{pmatrix} I_r & C \\ B & \lambda I_n \end{pmatrix}$$

由于 $\lambda \neq 0$ ，右下角可逆，可以用右下角消去其他部分，也即前 r 行减去后 n 行左乘 $\lambda^{-1}C$ 倍，前 r 列减去后 r 列右乘 $\lambda^{-1}B$ 倍，得到

$$\begin{pmatrix} I_r & -\lambda^{-1}C \\ & I_n \end{pmatrix} \begin{pmatrix} I_r & C \\ B & \lambda I_n \end{pmatrix} \begin{pmatrix} I_r & \\ -\lambda^{-1}B & I_n \end{pmatrix} = \begin{pmatrix} I_r - \lambda^{-1}CB & \\ & \lambda I_n \end{pmatrix}$$

由于已知 $\lambda I_r - CB$ 可逆，且 λ 非零，利用上次习题课可逆矩阵性质 6 可知可知

$$(I_r - \lambda^{-1}CB)^{-1} = \lambda(\lambda I_r - CB)^{-1}$$

于是有

$$\begin{pmatrix} I_r - \lambda^{-1}CB & \\ & \lambda I_n \end{pmatrix}^{-1} = \begin{pmatrix} \lambda(\lambda I_r - CB)^{-1} & \\ & \lambda^{-1}I_n \end{pmatrix}$$

而整理刚才的变换过程得到

$$\begin{pmatrix} I_r & -\lambda^{-1}C \\ & I_n \end{pmatrix} \begin{pmatrix} I_r & \\ B & I_n \end{pmatrix} \begin{pmatrix} I_r & \\ & \lambda I_n - BC \end{pmatrix} \begin{pmatrix} I_r & C \\ & I_n \end{pmatrix} \begin{pmatrix} I_r & \\ -\lambda^{-1}B & I_n \end{pmatrix} = \begin{pmatrix} I_r - \lambda^{-1}CB & \\ & \lambda I_n \end{pmatrix}$$

利用 $(XY)^{-1} = Y^{-1}X^{-1}$ ，与分块初等变换的逆矩阵，两边求逆，记 $Z = (\lambda I_n - BC)^{-1}$ 得到

$$\begin{pmatrix} I_r & \\ \lambda^{-1}B & I_n \end{pmatrix} \begin{pmatrix} I_r & -C \\ & I_n \end{pmatrix} \begin{pmatrix} I_r & \\ & Z \end{pmatrix} \begin{pmatrix} I_r & \\ -B & I_n \end{pmatrix} \begin{pmatrix} I_r & \lambda^{-1}C \\ & I_n \end{pmatrix} = \begin{pmatrix} \lambda(\lambda I_r - CB)^{-1} & \\ & \lambda^{-1}I_n \end{pmatrix}$$

进一步利用消去可知

$$\begin{pmatrix} I_r & \\ & Z \end{pmatrix} = \begin{pmatrix} I_r & C \\ & I_n \end{pmatrix} \begin{pmatrix} I_r & \\ -\lambda^{-1}B & I_n \end{pmatrix} \begin{pmatrix} \lambda(\lambda I_r - CB)^{-1} & \\ & \lambda^{-1}I_n \end{pmatrix} \begin{pmatrix} I_r & -\lambda^{-1}C \\ & I_n \end{pmatrix} \begin{pmatrix} I_r & \\ B & I_n \end{pmatrix}$$

直接计算右侧并对比右下角块的值可发现

$$Z = \frac{1}{\lambda}(I_n + B(\lambda I_r - CB)^{-1}C)$$

* 此处，我们通过分块行列变换分解为乘积并进行求逆。事实上，对于分块矩阵，仍可以通过整体行列变换直接算出逆，将在下题中介绍。

2. 若 A 可逆， D 为方阵，求以下矩阵可逆的充要条件，并在可逆时求其逆：

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

我们尝试进行行列变换计算，先在右侧增加一个 I ，并对应分块为

$$\begin{pmatrix} A & B & I & O \\ C & D & O & I \end{pmatrix}$$

由于只能进行行变换，我们将第二行的块减去第一行的块的 CA^{-1} 倍（左乘）以消去 C ，得到

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

若 $\det(D - CA^{-1}B) = 0$ ，则原矩阵进行行变换后 $D - CA^{-1}B$ 对应的行线性相关，从而不可能可逆；若 $\det(D - CA^{-1}B) \neq 0$ ，设其为 G ，将第一行的块减去第二行的块的 BG^{-1} 倍即可得到

$$\begin{pmatrix} A & O & I + BG^{-1}CA^{-1} & -BG^{-1} \\ O & G & -CA^{-1} & I \end{pmatrix}$$

最后，第一行的块左乘 A^{-1} 、第二行的块左乘 G^{-1} ，得到

$$\begin{pmatrix} I & O & A^{-1} + A^{-1}BG^{-1}CA^{-1} & -A^{-1}BG^{-1} \\ O & I & -G^{-1}CA^{-1} & G^{-1} \end{pmatrix}$$

于是原矩阵可逆当且仅当 G 可逆，且可逆时的逆为

$$\begin{pmatrix} A^{-1} + A^{-1}BG^{-1}CA^{-1} & -A^{-1}BG^{-1} \\ -G^{-1}CA^{-1} & G^{-1} \end{pmatrix}$$

* 这里的行列变换与逆的形式称为 **Schur 公式**，而 G 称为 A 的 Schur 补，用行列变换阵的形式可写为

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

3. 证明 $\text{rank}(ABC) + \text{rank } B \geq \text{rank}(AB) + \text{rank}(BC)$ 。

利用

$$\text{rank } X + \text{rank } Y = \text{rank} \begin{pmatrix} X & O \\ O & Y \end{pmatrix}$$

只需证

$$\text{rank} \begin{pmatrix} ABC & O \\ O & B \end{pmatrix} \geq \text{rank} \begin{pmatrix} AB & O \\ O & BC \end{pmatrix}$$

为将前者行列变换到后者，可以将第一行的块加上第二行的块的 A 倍（左乘），即可得到一个 AB ，也即问题变为证

$$\text{rank} \begin{pmatrix} ABC & AB \\ O & B \end{pmatrix} \geq \text{rank} \begin{pmatrix} AB & O \\ O & BC \end{pmatrix}$$

注意到，只要将第一列的块减去第二列的块的 C 倍（右乘），就可以消去 ABC ，只保留 B ，问题变为证

$$\text{rank} \begin{pmatrix} O & AB \\ -BC & B \end{pmatrix} \geq \text{rank} \begin{pmatrix} AB & O \\ O & BC \end{pmatrix}$$

给左侧第一列乘 $-I$ ，并交换行列，可得只需证明

$$\text{rank} \begin{pmatrix} AB & O \\ B & BC \end{pmatrix} \geq \text{rank} \begin{pmatrix} AB & O \\ O & BC \end{pmatrix}$$

由 3.5 节习题 11 即得到结论。

* 教材 3.5 节的例 8、例 9 与习题 11 常用于证明秩不等式。本题结论称为 **Frobenius 秩不等式**。

* 将 B 取为 I 的 Frobenius 秩不等式称为 **Sylvester 秩不等式**，最常用的情况即

$$\text{rank}(A_{m \times n} B_{n \times p}) = O \implies \text{rank } A + \text{rank } B \leq n$$

§8.2 线性映射与相抵

本部分中，我们希望继续本讲义 3.2.2 对线性映射的讨论，并且彻底解决线性映射和矩阵的关系。为方便之后的研究，我们将所有数域上的矩阵都看成复矩阵。

8.2.1 线性映射的矩阵表示

直接验证可以发现，对 $A_{m \times n}$ ，映射 $\mathcal{A} : \mathbb{C}^n \rightarrow \mathbb{C}^m$ ， $\mathcal{A}(x) = Ax$ 是一个线性映射。在之前的讨论中，我们已经大量利用了这个线性映射与矩阵性质的关系，这里列举最重要的两条：

$$\dim \text{Im } \mathcal{A} = \text{rank } A, \quad \dim \text{Ker } \mathcal{A} = n - \text{rank } A$$

更多的相关定义与性质可参考本讲义第三章。

我们下面证明一个非常重要的性质：对任何线性映射 $f : \mathbb{C}^n \rightarrow \mathbb{C}^m$ ，一定存在矩阵 $F_{m \times n}$ 使得 $f(x) = Fx$ ，从而 $x \rightarrow Ax$ 就是全部的线性映射。

为了证明此结论，我们需要先思考 F 的元素是如何确定的。直观来看，取 $x = e_i$ ，则 Fx 就是 F 的第 i 列。由此，反向进行操作，设 F 的第 i 列是 $f(e_i)$ ，下面证明对任何向量 x 有 $f(x) = Fx$ 。

对任何向量 x ，由于 e_1, \dots, e_n 构成 \mathbb{C}^n 的一组基，可设其为

$$x = \sum_{i=1}^n x_i e_i, \quad x_i \in \mathbb{C}$$

利用线性映射的线性性即有

$$f(x) = \sum_{i=1}^n x_i f(e_i)$$

直接计算可发现这就是 Fx 的表达式。由于考虑 Fe_i 可发现不同 F 对应的映射不同，我们即证明了对 $\mathbb{C}^n \rightarrow \mathbb{C}^m$ 的线性映射 f ，存在唯一矩阵 $F_{m \times n}$ 使得 $f(x) = Fx$ 。

因此，我们可以得到， \mathbb{C}^n 到 \mathbb{C}^m 的线性映射与 $m \times n$ 矩阵存在一个一一对应，这就完全刻画了线性映射。然而，对于两个更一般的向量空间 U 与 V ，能否完全刻画其间的线性映射呢？

出于对向量空间刻画的基本需求，我们设 $\dim U = n$ ，一组基为 $S = \{\alpha_1, \dots, \alpha_n\}$ ； $\dim V = m$ ，一组基为 $S' = \{\beta_1, \dots, \beta_m\}$ 。某种意义上，给定基的 k 维向量空间就可以看作 \mathbb{C}^k ：对向量空间 W 与其一组基 $T = \{\gamma_1, \dots, \gamma_k\}$ ，任何 $x \in W$ 存在唯一表示

$$x = \sum_{i=1}^k x_i \gamma_i$$

将列向量 $(x_1, \dots, x_k)^T$ 称为 x 在基 T 下的坐标，记为 x_T （这不是一个通用记号，这里引入是为了方便书写）。

我们从坐标的定义出发进行推导：

1. 坐标映射 $\pi_T(x) = x_T$ 是 $W \rightarrow \mathbb{C}^k$ 的映射。

利用基的性质，任何向量 x 一定在基 T 下可以写出对应的表达式，且表达式是唯一的，由此确实构成映射。

2. $\pi_T(x)$ 是 $W \rightarrow \mathbb{C}^k$ 的线性映射。

直接验证 $\pi_T(\lambda x) = \lambda x_T$ 与 $\pi_T(x + y) = x_T + y_T$ 即可。

3. $\pi_T(x)$ 是 $W \rightarrow \mathbb{C}^k$ 的单射。

由于其为线性映射，为单射仅需证明 $\text{Ker } \pi_T = \{0\}$ ，而 $\pi_T(x) = 0$ 意味着 $x_T = 0$ ，根据定义可发现有 $x = 0$ 。

4. $\pi_T(x)$ 是 $W \rightarrow \mathbb{C}^k$ 的满射。

利用向量空间封闭性，对任何 \mathbb{C}^k 中向量，其作为坐标可确定 W 中元素，从而为满射。

* 由此， π_T 构成 $W \rightarrow \mathbb{C}^k$ 的双射，或称线性同构。

5. 对不同的基 T, T' ，映射 π_T 与 $\pi_{T'}$ 不同。

对两组不同的基，由个数相等不可能 $T \subset T'$ ，因此一定有某个向量 α 在 T 中但不在 T' 中，由此 $\pi_T(\alpha)$ 为某个 e_i ，但 $\pi_{T'}(\alpha)$ 不为某个 e_i ，于是它们不同。

6. 对 $U \rightarrow V$ 的线性映射 f ， $x \rightarrow \pi_{S'}(f(\pi_S^{-1}(x)))$ 是 $\mathbb{C}^n \rightarrow \mathbb{C}^m$ 的线性映射。

由本讲义 3.2.2 的结论，线性同构的逆亦为线性同构，于是 π_S^{-1} 是 $\mathbb{C}^n \rightarrow U$ 的线性映射、 f 是 $U \rightarrow V$ 的线性映射、 $\pi_{S'}$ 是 $V \rightarrow \mathbb{C}^m$ 的线性映射，由线性映射的复合是线性映射知结论。

7. 对 $U \rightarrow V$ 的线性映射 f , 存在唯一矩阵 $F_{S,S'} \in \mathbb{C}^{m \times n}$ 使得 $\pi_{S'}(f(\pi_S^{-1}(x))) = F_{S,S'}x$ 。

由于 $x \rightarrow \pi_{S'}(f(\pi_S^{-1}(x)))$ 是 $\mathbb{C}^n \rightarrow \mathbb{C}^m$ 的线性映射, 本节开头已证其一定对应某个矩阵的乘法。

8. 对任何矩阵 $F \in \mathbb{C}^{m \times n}$, 映射 $x \rightarrow \pi_{S'}^{-1}(F\pi_S(x))$ 是 $U \rightarrow V$ 的线性映射。

由于 $x \rightarrow Fx$ 是 $\mathbb{C}^n \rightarrow \mathbb{C}^m$ 的线性映射, π_S 是 $U \rightarrow \mathbb{C}^n$ 的线性映射, $\pi_{S'}^{-1}$ 是 $\mathbb{C}^m \rightarrow V$ 的线性映射, 复合可知结论。

9. 给定基 S, S' 后, $U \rightarrow V$ 的线性映射 f 与 $m \times n$ 矩阵 $F_{S,S'}$ 一一对应。

前两个性质已经给出了线性映射对应到矩阵与矩阵对应到线性映射的方式, 而若矩阵 F 对应的线性映射为 $\pi_{S'}^{-1}(F\pi_S(x))$, 其对应的矩阵为 (利用映射结合律)

$$\pi_{S'}(\pi_{S'}^{-1}(F\pi_S(\pi_S^{-1}(x)))) = Fx$$

反之类似可验证, 于是上述两种对应是互逆的, 这就说明了构成一一对应。

10. $y = f(x)$ 当且仅当 $y_{S'} = F_{S,S'}x_S$ 。

利用 $x_S = \pi_S(x)$, $y_{S'} = F_{S,S'}x_S$ 等价于 $\pi_{S'}(y) = F_{S,S'}\pi_S(x)$, 而两边同时作用 $\pi_{S'}^{-1}$ 可得其等价于 $y = \pi_{S'}^{-1}(F_{S,S'}\pi_S(x))$, 根据性质 8、9 可知这即等价于 $y = f(x)$ 。

上方定义的 $F_{S,S'}$ 即称为线性映射 f 在基 S, S' 下的矩阵表示。上述过程事实上表明了, 矩阵表示也即看作坐标时对应的矩阵乘法, 于是, 只要是 n 维空间到 m 维空间的线性映射, 均可看作矩阵乘法。

矩阵表示的一个等价定义是: 设 (注意右侧的下标与矩阵乘法时通常出现的 ij 对 j 不同, 是 ij 对 i)

$$f(\alpha_j) = \sum_{i=1}^m f_{ij} \beta_i$$

将 f_{ij} 拼成的矩阵 F 定义为 f 在基 S, S' 下的矩阵表示。这个定义看似比上面复杂的映射操作要简洁, 但对矩阵表示意味着可看作矩阵乘法失去了直观。它与我们之前定义的等价性可由性质 10 计算验证。

8.2.2 作为基变换的可逆阵

虽然上述的讨论已经完全刻画了两个向量空间之间的线性映射, 但仍有一个不足之处: 线性映射的矩阵表示是依赖基的, 会随着基的变化而变化。

很自然地, 我们希望能够知道, 线性映射在怎样的基下拥有简单的表示。为了研究这一性质, 我们必须刻画基之间的关系。同样先给出定义: 对空间 U 的两组基 $S = \{\alpha_1, \dots, \alpha_n\}$ 与 $T = \{\beta_1, \dots, \beta_n\}$, 设

$$\beta_j = \sum_{i=1}^n p_{ij} \alpha_i$$

则称所有 p_{ij} 拼成的 n 阶矩阵 P 为 S 到 T 的过渡矩阵。

* 仍需注意右侧的下标是 ij 对 i 。

由于任何 U 中向量在基 S 下存在唯一表示, 此定义确实是合理的, 且根据 3.2 节例 2 可知 P^T 可逆 (由于下标影响, 例 2 中实际的矩阵为 P^T), 从而由逆的基本性质可知 P 可逆。反之, 给定 S 与可逆矩阵 P , 可以构造出 T , 同样由 3.2 节例 2 与 P^T 可逆可知 T 也为一组基。

上述的讨论给出了可逆矩阵的重要性质: 可以作为基之间的过渡矩阵。事实上, 若 P 是 S 到 T 的过渡矩阵, 则 P^{-1} 是 T 到 S 的过渡矩阵, 此外, $x_T = P^{-1}x_S$ 。这俩个性质同样体现了可逆的必要性。

为证明第一个性质, 我们设 T 到 S 的过渡矩阵为 Q , 对应元素 q_{ij} , 有

$$\alpha_j = \sum_{k=1}^n q_{kj} \beta_k = \sum_{k=1}^n q_{kj} \sum_{i=1}^n p_{ik} \alpha_i = \sum_{i=1}^n \left(\sum_{k=1}^n p_{ik} q_{kj} \right) \alpha_i$$

利用基的线性无关性，当且仅当 $i = j$ 时 $\sum_{k=1}^n p_{ik}q_{kj} = 1$ ，否则其为 0，而这即为矩阵乘法的分量表示，从而 $PQ = I$ ，得证。

为证明第二个性质，设

$$x = \sum_{j=1}^n x_j \beta_j$$

有

$$x = \sum_{j=1}^n x_j \sum_{i=1}^n p_{ij} \alpha_i = \sum_{i=1}^n \left(\sum_{j=1}^n p_{ij} x_j \right) \alpha_i$$

利用矩阵乘法的分量表示即知 $x_S = Px_T$ ，从而 $x_T = P^{-1}x_S$ 。

由此，假设 S, T 为 U 的两组基， S', T' 为 V 的两组基，若已知 f 的矩阵表示 $F_{S,S'}$ ，我们希望能从过渡矩阵出发得到矩阵表示 $F_{T,T'}$ 。

假设 S 到 T 的过渡矩阵为 P ， S' 到 T' 的过渡矩阵为 Q ，利用性质 10 有

$$y_{S'} = F_{S,S'} x_S \iff y = f(x) \iff y_{T'} = F_{T,T'} x_T$$

而上方证明了 $y_{T'} = Q^{-1}y_{S'}$ 、 $x_T = P^{-1}x_S$ ，从而

$$y_{S'} = F_{S,S'} x_S \iff Q^{-1}y_{S'} = F_{T,T'} P^{-1} x_S$$

将右侧利用消去进一步改写为 $y_{S'} = QF_{T,T'} P^{-1} x_S$ ，由于等价性对任何 $x_S, y_{S'}$ 成立，必有 $QF_{T,T'} P^{-1} x_S = F_{S,S'} x_S$ 对一切 x_S 成立，因此 $QF_{T,T'} P^{-1} = F_{S,S'}$ ，消去得到

$$F_{T,T'} = Q^{-1} F_{S,S'} P$$

从而可知线性映射在不同基下的矩阵表示相抵。

反之，由于从任何过渡矩阵出发都可以构造出对应的基，所有与 $F_{S,S'}$ 相抵的矩阵都是 f 在某组基下的矩阵表示。

最后，我们回顾本讲义第七章的相抵标准形结论，这意味着，对任何线性映射 $f: U \rightarrow V$ ，若 $\dim U = n$ 、 $\dim V = m$ ，存在 U 的一组基 $S = \{\alpha_1, \dots, \alpha_n\}$ 与 V 的一组基 $S' = \{\beta_1, \dots, \beta_m\}$ 使得

$$F_{S,S'} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$$

利用矩阵表示的等价定义，这意味着

$$\forall i \in \{1, \dots, r\}, \quad f(\alpha_i) = \beta_i$$

$$\forall i \in \{r+1, \dots, n\}, \quad f(\alpha_i) = 0$$

这就是相抵标准形对线性映射的意义：选取基使得线性映射的矩阵表示最为简洁。

* 由于相抵不改变秩，我们可以定义一个线性映射的秩 $\text{rank } f$ 为其某个矩阵表示的秩 $\text{rank } F_{S,S'}$ ，其含义见下节。

8.2.3 不可区分的线性映射

* 本节为附加内容，仅为相抵提供另一种理解方式。

在本讲义 3.2.2 中，我们提到了向量空间不可区分当且仅当同构，也即维数相等。那么，两个线性映射在何时不可区分呢？

考虑 $f: U \rightarrow V$ 与 $f': U' \rightarrow V'$ 。若 U 与 U' 不同构或 V 与 V' 不同构，它们自然是可以区分的。否则，若存在线性同构 $\varphi: U \rightarrow U'$ 与线性同构 $\psi: V' \rightarrow V$ 使得

$$f = \psi \circ f' \circ \varphi$$

两个映射即不可区分——这意味着，只要对 U, U', V, V' 选取合适的基，能使两映射形式上相同。更进一步地，由于我们已经给出了矩阵表示与坐标映射的定义，形式上相同也即具有相同的矩阵表示。

* 利用矩阵表示与坐标映射的关系，由于同构将一组基映射到一组基，可验证 $f = \psi \circ f' \circ \varphi$ 确实与存在相同矩阵表示等价。

根据上节，具有相同的矩阵表示等价于矩阵表示的相抵标准形相同，也即秩相同。再由上节性质 10 知

$$\text{rank } F_{S,S'} = \dim \text{Im } F_{S,S'} = \dim \{y_{S'} \mid y_S = F_{S,S'}x_S\} = \dim \{\pi_{S'}(y) \mid y = f(x)\} = \dim \pi_{S'}(\text{Im } f)$$

由于 $\pi_{S'}$ 为同构，可验证其将 $\text{Im } f$ 一组基映射为 $\pi_{S'}(\text{Im } f)$ 一组基，于是矩阵表示的秩即为 $\dim \text{Im } f$ 。

综合上述讨论， $f: U \rightarrow V$ 与 $f': U' \rightarrow V'$ 不可区分当且仅当

$$\dim U = \dim U', \quad \dim V = \dim V', \quad \dim \text{Im } f = \dim \text{Im } f'$$

* 思考：从相抵标准形出发构造对应的 ψ 与 φ 。

§8.3 线性变换与相似

8.3.1 线性变换

线性变换的定义非常简单：一个空间 U 到自身的线性映射 $\rho: U \rightarrow U$ 称为线性变换。

如此定义的关键作用是，线性变换可以和自身复合，我们定义线性变换的次方 ρ^n 代表 ρ 和自身复合 n 次，加法为 $(\rho_1 + \rho_2)(x) = \rho_1(x) + \rho_2(x)$ ，数乘为 $(a\rho)(x) = a\rho(x)$ ，则可以谈论线性变换的多项式，其中常数项定义为 $a\mathcal{I}$ ， \mathcal{I} 为恒等变换 $x \rightarrow x$ 。

* 这里加法和数乘的定义类比函数的加法，是非常自然的。

由于线性变换也是线性映射，我们之前关于线性映射的结论可以直接用在线性变换上，不过，由于其定义域与陪域是在同一个空间，我们自然希望谈论它们的时候用的是同一组基。考虑线性变换 $\rho: U \rightarrow U$ (称为 U 上的线性变换) 与 U 的一组基 $S = \{\alpha_1, \dots, \alpha_n\}$ ，直接从线性映射的讨论得到如下结论 (这里采用复合的写法简化书写)：

1. 存在唯一矩阵 $P_S \in \mathbb{C}^{n \times n}$ 使得 $\pi_S \circ \rho \circ \pi_S^{-1}(x) = P_S x$ 。
2. 对任何矩阵 $P \in \mathbb{C}^{n \times n}$ ，映射 $x \rightarrow \pi_S^{-1}(P\pi_S(x))$ 是 U 上的线性变换。
3. 给定基 S 后， U 上的线性变换 ρ 与 n 阶方阵 P_S 一一对应。
4. $y = \rho(x)$ 当且仅当 $y_S = P_S x_S$ 。
5. $\text{rank } P_S = \dim \text{Im } \rho$ 。

可以发现，在这个视角下，方阵即对应着线性变换。不过，出于方阵的特殊性，我们还有着更多的结论：

5. ρ 是同构当且仅当 P_S 可逆。

由于已知其为线性映射， ρ 是同构等价于 ρ 可逆。本讲义第七章中已证明映射 $\phi: x_S \rightarrow P_S x_S$ 可逆当且仅当 P_S 可逆，而 $\rho = \pi_S^{-1} \circ \phi \circ \pi_S$ ，若 ϕ 可逆，利用可逆映射复合可逆知 ρ 可逆。反之，若 ρ 可逆，由 $\phi = \pi_S \circ \rho \circ \pi_S^{-1}$ 同理知 ϕ 可逆。

6. 若 f 为多项式, $f(\rho)$ 在基 S 下的矩阵表示为 $f(P_S)$ 。

定义 $\phi: x_S \rightarrow P_S x_S$, 由于 $\rho = \pi_S^{-1} \circ \phi \circ \pi_S$, 利用映射结合律相消有 $\rho^n = \pi_S^{-1} \circ \phi^n \circ \pi_S$, 进一步利用它们都为线性映射, 线性组合得到 $f(\rho) = \pi_S^{-1} \circ f(\phi) \circ \pi_S$ 。由于 $f(\rho)$ 的矩阵表示对应映射为 $\pi_S \circ f(\rho) \circ \pi_S^{-1} = f(\phi)$, 直接计算 $f(\phi)$ 可发现对应矩阵是 $f(P_S)$ 。

* 注意到, 这里事实上出现了一个与相似差不多的结构, 而证明过程事实上也与相似矩阵的多项式相似非常接近。这种 $a^{-1}ba$ 类型的结构称为共轭, 在 a, b 可交换时是平凡的, 只能为 b , 但对不交换的系统具有一些非常本质的意义, 无论是矩阵乘法还是映射复合都是如此。

7. 设基 S 到 T 的过渡矩阵为 Q , 则 $P_T = Q^{-1}P_S Q$ 。

由于 P_T 即为看作线性映射的 $P_{T,T}$, 利用上一部分结论即得。

8. 对任何可逆矩阵 Q , $Q^{-1}P_S Q$ 可以作为 ρ 在某组基下的矩阵表示。

由于从任何可逆矩阵 Q 与基 S 出发可以构造基 T , 此结论成立。

由此, 线性变换在不同基下的矩阵相似, 且所有与某个矩阵表示相似的矩阵都可以看作其的矩阵表示。从而, 相似也具有了一个直观的意义: 选取基使得线性变换的矩阵表示最为简洁。

* 仿照上一部分的最后一节, 可以说明两个线性变换不可区分当且仅当对应的空间同构, 且各自的某个矩阵表示相似。

* 所有相似下的不变量都可以定义在线性变换上, 例如 rank、det 与 tr, 定义即为其某个矩阵表示的 rank、det 与 tr。

8.3.2 复向量中的特征系统

我们希望知道, 除了 rank、det 与 tr, 与能定义在线性映射上的 Ker、Im, 还有什么性质是对于一个线性变换可以研究的。

对于一个变换 f , 我们往往会考察其不动点, 也即满足 $f(x) = x$ 的点。对于 U 上的线性变换 ρ (设 U 的一组基为 S), 这个问题即成为了方程

$$\rho(x) = x$$

由于这即为 $(\rho - I)x = 0$, 根据之前线性变换多项式的矩阵表示, 这当且仅当 $(P_S - I)x_S = 0$ 。利用 Cramer 法则, 此方程组有解当且仅当 $\det(P_S - I) = 0$, 而只要解出了符合要求的 x_S , 作 $\pi_S^{-1}(x_S)$ 即能得到所有的 x 。

更进一步地, 由于向量空间中可以定义数乘, 我们关系 $\rho(x)$ 的“广义不动点问题”, 也即只改变方向、未改变大小的点, 这就得到了线性映射的特征方程:

$$\rho(x) = \lambda x$$

1. 计根的重数时, 恰有 n 个 λ (计重根) 使得此方程有非零解, 它们称为线性映射的特征值, 对应的 λ 重数称为特征值的代数重数。

见本讲义 7.4.4, 由于相似不影响特征多项式, 这里任取某个矩阵表示考虑特征多项式即可。

* 再次提醒所有不同特征值代数重数和为 n 。

2. 所有 n 个特征值的和为 $\text{tr}(A)$ 。

利用韦达定理, 由于 $\varphi(A)(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i)$, 其 $n-1$ 次项为 $-\sum_i \lambda_i$, 因此只需证明 $\det(\lambda I - A)$ 的 $n-1$ 次项为 $-\text{tr } A$ 。

考虑完全展开式, 主对角线外的项至少包含两个主对角线外的元素, 从而次数至多 $n-2$, 不影响 $n-1$ 次项。由此, $n-1$ 次项即为主对角线乘积 $\prod_i (\lambda - a_{ii})$ 的 $n-1$ 次项, 再次利用韦达定理即可得到它是 $-\text{tr } A$ 。

3. 所有 n 个特征值的乘积为 $\det(A)$ 。

利用韦达定理可知乘积为 $\det(\lambda I - A)$ 的常数项乘 $(-1)^n$ ，完全类似期中复习题 8(2) 中的说明可知常数项为 $(-1)^n \det A$ ，再乘 $(-1)^n$ 可抵消，从而得证。

同样利用韦达定理，所有特征值乘积即为 $\varphi_A(\lambda)$ 的常数项

4. 对每个特征值 λ ，此方程的解构成线性空间，称为其对应的**特征子空间**，其中的**非零向量**称为此特征值的**特征向量**，维数称为此特征值的**几何重数**。

5. 不同特征值的特征子空间交为 $\{0\}$ 。

否则，存在非零向量 x 使得 $\rho(x) = \lambda_1 x = \lambda_2 x$ ，但 $\lambda_1 \neq \lambda_2$ ，矛盾。

6. 所有不同特征值的特征子空间和是直和。

* 这里多个空间求和可看作逐个相加，和是直和意为每一步都是直和。

假设这些特征子空间为 W_1, \dots, W_k ，进行归纳。结论 5 证明了两个子空间的和是直和。

若某步并非直和，意味着 $W_1 \oplus \dots \oplus W_i$ 与 W_{i+1} 有非零交，设这个元素为 x ，其表示为

$$x_1 + \dots + x_i, \quad x_i \in W_i$$

也即存在互不相同的 $\lambda_1, \dots, \lambda_i, \lambda_{i+1}$ ，使得

$$\lambda_{i+1}x = \rho(x) = \rho(x_1 + \dots + x_i) = \lambda_1 x_1 + \dots + \lambda_i x_i$$

于是

$$(\lambda_{i+1} - \lambda_1)x_1 + \dots + (\lambda_{i+1} - \lambda_i)x_i = 0$$

但是，由于系数均非零且 x_j 不全为 0，考虑使得 j 最大的非零 x_j ，其可被 x_1, \dots, x_{j-1} 表出，与之前每步和为直和矛盾。

7. 任何特征值几何重数不超过代数重数。

考虑特征值 μ 与对应的特征子空间 W ，设 W 一组基 $\alpha_1, \dots, \alpha_k$ ，并扩充为全空间一组基

$$\alpha_1, \dots, \alpha_k, \quad \beta_1, \dots, \beta_{n-k}$$

由于

$$\rho(\alpha_1) = \mu\alpha_1, \quad \dots, \quad \rho(\alpha_k) = \mu\alpha_k$$

根据矩阵表示的等价定义， ρ 在这组基下的矩阵表示为

$$\begin{pmatrix} \mu I_k & X \\ O & Y \end{pmatrix}$$

而直接计算特征多项式可发现其行列式为 $\det((\lambda - \mu)I_k) \det(\lambda I - Y) = (\lambda - \mu)^k \det(\lambda I - Y)$ ，从而 μ 的代数重数至少为 k 。

这样，我们就从空间视角对特征系统给出了明确的意义：它事实上代表一个线性变换的“广义不动点”的集合性质。

8.3.3 可对角化 I

最后，让我们来聊一聊对角化的相关问题。

若一个矩阵可以相似为对角阵，则称其为可对角化的矩阵。从线性映射 ρ 的角度来说，若能选取一组基 $\alpha_1, \dots, \alpha_n$ 使得其的矩阵表示为 $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ ，则利用矩阵表示的等价定义可知这即代表

$$\rho(\alpha_1) = \lambda_1 \alpha_1, \quad \rho(\alpha_2) = \lambda_2 \alpha_2, \quad \dots, \quad \rho(\alpha_n) = \lambda_n \alpha_n$$

这意味着， λ_i 都是 ρ 的特征值，从而可对角化等价于存在一组基均为 ρ 的特征向量（由于可对角化本质是某些相似等价类的性质，相似不改变可对角化性，于是可定义在线性变换上）。本节中，我们先不关心怎样的矩阵是可对角化的，专注于可对角化矩阵的好处。下假设 $\rho: U \rightarrow U$ 是可对角化的：

1. ρ 任何特征值的几何重数等于代数重数。

设 ρ 的所有特征子空间为 W_1, \dots, W_k ，由于每个特征向量都属于某个特征子空间，而可对角化时任何矩阵向量可拆分为特征向量的线性组合，即得

$$U = W_1 + \dots + W_k$$

但是，由于和的维数不超过维数的和、 $\dim W_i$ 不超过 λ_i 的代数重数，通过所有不同特征值代数重数和为 n 可知上述两个“不超过”必须取等，从而得证。

2. 给定 $x = \sum_i \mu_i \alpha_i$ ，则 $f(\rho)(x) = \sum_i \mu_i f(\lambda_i) \alpha_i$ ，从而 $f(\rho)$ 也是可对角化的。

利用 $f(\rho)$ 定义直接计算验证即可。由此， $f(\rho)\alpha_i = f(\lambda_i)\alpha_i$ ，而 $\{\alpha_i\}$ 构成 U 一组基，因此可对角化。

3. 若 ρ 是同构，给定 $x = \sum_i \mu_i \alpha_i$ ，则 $\rho^{-1}(x) = \sum_i \mu_i \lambda_i^{-1} \alpha_i$ ，从而 ρ^{-1} 也是可对角化的。

由于 ρ 是同构， $\ker \rho = \{0\}$ ，也即所有 λ_i 均非零，于是 $x \rightarrow \sum_i \mu_i \lambda_i^{-1} \alpha_i$ 确实是良好定义的映射，计算验证可发现 ρ 与此映射复合为 \mathcal{I} ，因此其确实为 ρ 的逆。可对角化的理由与上个结论相同。

4. 线性变换 ϕ 与 ρ 可交换当且仅当对任何 ρ 的特征子空间 W_i 有 $\phi(W_i) \subset W_i$ 。

由结论 1 过程可知 $U = W_1 + \dots + W_k$ ，再由上节结论 6 得

$$U = W_1 \oplus \dots \oplus W_k$$

类似两空间直和的情况，归纳可知 U 中任何向量 x 存在唯一分解 $x_1 + \dots + x_k$ 使得 $x_i \in W_i$ 。由此，利用线性性， $\rho \circ \phi = \phi \circ \rho$ 当且仅当对 $x_i \in W_i$ 有 $\rho(\phi(x_i)) = \phi(\rho(x_i))$ 。

设 W_i 对应特征值 λ_i ，若 $\phi(W_i) \subset W_i$ ，即有

$$\rho(\phi(x_i)) = \lambda_i \phi(x_i) = \phi(\lambda_i x_i) = \phi(\rho(x_i))$$

将上式反向观察即可发现 $\phi(x_i) \in W_i$ ，从而得证。

最后，我们给出上面四个结论的矩阵论版本，设 A 是一个可对角化的方阵，且存在可逆阵 P 使得

$$A = P \text{diag}(\lambda_1, \dots, \lambda_n) P^{-1}$$

1. A 任何特征值的几何重数等于代数重数。

与上方第一个结论的说明方式完全相同。

2. $f(A) = P \text{diag}(f(\lambda_1), \dots, f(\lambda_n)) P^{-1}$ 。

利用相似的多项式性质，直接计算对角阵的多项式发现相当于每个位置作多项式，从而得证。

3. A 可逆时 $A^{-1} = P(\lambda_1^{-1}, \dots, \lambda_n^{-1})P^{-1}$ 。

由 A 可逆可知不存在零特征值，因此右侧矩阵良好定义。直接计算可验证右侧矩阵乘 A 为 I ，从而得证。

4. B 与 A 可交换当且仅当

$$B = PB_0P^{-1}$$

其中 B_0 满足当且仅当 $\lambda_i = \lambda_j$ 时 $(B_0)_{ij}$ 可任取，否则为 0。

直接计算 $AB = BA$ 并消去两边的 P 与 P^{-1} 得到对任何 i, j 有 $\lambda_i(B_0)_{ij} = \lambda_j(B_0)_{ij}$ ，从而得证。

也即，对可对角化的矩阵，多项式、逆与交换性问题都是容易解决的，相当于对每个特征值操作。

九 可逆阵的技巧

* 本次习题课讲义的补充内容部分均以习题形式组织，具体的知识背景、相关定义请参考第七章与第八章。

§9.1 习题解答

1. 习题 5.5-2

- (1) 特征值 $1 + i\sqrt{3}$ 对应特征向量 $(i, 1)^T$ ，特征值 $1 - i\sqrt{3}$ 对应特征向量 $(-i, 1)^T$ 。看成实数域上矩阵无特征值。
- (2) 特征值 i 对应特征向量 $(-1 + 2i, 1 - i, 2)^T$ ，特征值 $-i$ 对应特征向量 $(-1 - 2i, 1 + i, 2)^T$ ，特征值 1 对应特征向量 $(-2, 1, 1)^T$ 。看成实数域上矩阵特征值只有 1，对应特征向量不变。

2. 习题 5.5-9

当 $\lambda \neq 0$ 时，由 4.5 节命题 2

$$\det(\lambda I - AB) = \lambda^n \det(I - (\lambda^{-1}A)B) = \lambda^n \det(I - B(\lambda^{-1}A)) = \det(\lambda I - BA)$$

而左右均为 λ 的多项式，它们在 0 以外的点相同即说明它们相同（或利用 Binet-Cauchy 公式单独讨论 0 的情况）。

3. 习题 5.5-16

利用满秩分解（见 4.3 节例题或本讲义 4.2.1 后方习题）可设 $A = \alpha\beta^T$ ，其中 α, β 为列向量，且由 $A \neq O$ 可知 α 与 β 均为非零向量。

由本讲义 8.1.3 第一题直接计算可知（注意 $\beta^T\alpha$ 是 1×1 矩阵，行列式为自身）

$$\det(\lambda I - A) = \lambda^{n-1}(\lambda - \beta^T\alpha)$$

若 $\beta^T\alpha = 0$ ，有 $A^2 = \alpha(\beta^T\alpha)\beta^T = O$ ，矛盾，因此根据特征值重数定义得证。

4. 习题 5.5-7

设 A 的特征多项式为 $(\lambda_1, \dots, \lambda_n)$ 为计重数的全部特征值

$$\det(\lambda I - A) = \prod_{i=1}^n (\lambda - \lambda_i)$$

由 $k \neq 0$ 直接计算可得

$$\det(\lambda I - kA) = k^n \det((k^{-1}\lambda)I - A) = k^n \varphi_A(k^{-1}\lambda) = k^n \prod_{i=1}^n (k^{-1}\lambda - \lambda_i) = \prod_{i=1}^n (\lambda - k\lambda_i)$$

从而 A 的特征值 λ_i 与 kA 的特征值 $k\lambda_i$ 一一对应，即得证。

5. 习题 5.5-14

若 A 可逆, 即特征值均非零, 利用习题 5.5-7 与本节例 6 可知 A^* 的特征值为

$$\frac{\det A}{\lambda_1}, \dots, \frac{\det A}{\lambda_n}$$

再由本讲义 8.3.2 结论 3 可化为

$$\prod_{i \neq 1} \lambda_i, \prod_{i \neq 2} \lambda_i, \dots, \prod_{i \neq n} \lambda_i$$

* 此时利用摄动法与特征值、 A^* 对 A 的连续性可直接得到 A^* 不可逆时结论同上。不过, 特征值的连续性并不显然 (在考试中尽量别用, 可以用来猜结论), 我们这里还是采用较复杂的讨论。

若 A 不可逆, 其有零特征值, 不妨设 $\lambda_1 = 0$, 利用本讲义 7.3.2 的结论可知 $\text{rank } A^* \leq 1$, 即 0 的几何重数至少为 $n - 1$, 于是 0 的代数重数至少为 $n - 1$, 也即 A^* 至少有 $n - 1$ 个特征值为 0。

根据本讲义 8.3.2 的结论, $\text{tr}(A^*)$ 为 A^* 特征值的和, 也即唯一一个可能非零的特征值。另一方面直接利用完全展开计算 (或由书上结论) 可发现 $\text{tr}(A^*) = \sum_i A_{ii}$ 是 $\det(\lambda I - A)$ 的一次项乘 $(-1)^{n-1}$ 。由于

$$\det(\lambda I - A) = \lambda \prod_{i=2}^n (\lambda - \lambda_i)$$

其一次项为 $(-1)^{n-1} \lambda_2 \dots \lambda_n$, 也即其最后一个特征值为 $\lambda_2 \dots \lambda_n$, 与可逆时结论相同。

6. 习题 5.6-4

由上三角阵的特征值结论可知 A 的特征值 1 代数重数 r , 特征值 -1 代数重数 $n - r$ 。直接利用行列变换消去 B (类似相抵标准形证明过程) 可知 $\text{rank}(A - I) = n - r$ 、 $\text{rank}(A + I) = r$, 从而得证可对角化。

7. 习题 4.4-13

直接配凑可得

$$(I - A)(I - B) = I - A - B + AB = I$$

于是 $I - A$ 与 $I - B$ 互逆, 从而

$$I = (I - B)(I - A) = I - B - A + BA$$

这即得到 $A + B = BA$, 因此 $AB = BA$ 。

8. 习题 4.5-19

记向量 $\alpha = (a_1, \dots, a_n)^T$, $\beta = (b_1, \dots, b_n)^T$, 则原行列式为 $I + \alpha\beta^T$, 由此利用 4.5 节命题 2 (注意 $\beta^T\alpha$ 是 1×1 矩阵, 行列式为自身)

$$\det(I + \alpha\beta^T) = 1 + \beta^T\alpha = 1 + \sum_{i=1}^n a_i b_i$$

9. 习题 4.5-21

将其看作矩阵

$$\begin{pmatrix} I_{n-1} & B \\ C & D \end{pmatrix}$$

利用习题 4.5-14 (证明见本讲义中 Schur 公式) 可知

$$\det A = \det(D - CB)$$

设 $F = D - CB$, 其各元素为 f_{ij} , 可发现

$$f_{ij} = \begin{cases} (1-n)a & i = j < n \\ a & i = j = n \\ -nb & i = j - 1 \\ n & i = n, j = 1 \end{cases}$$

直接利用完全展开可发现只有两项, 从而计算得结果为

$$(1-n)^{n-1}a^n + n^n b^{n-1}$$

10. 习题 4.6-2

直接利用分配律计算可发现

$$\beta^T \sum_{i=1}^s \lambda_i \alpha_i = \sum_{i=1}^s \lambda_i^s \beta^T \alpha_i = 0$$

11. 习题 4.6-5

直接利用正交化可得结果为

$$\beta_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \\ 0 \\ 0 \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} \sqrt{6}/6 \\ -\sqrt{6}/6 \\ \sqrt{6}/3 \\ 0 \end{pmatrix}, \quad \beta_3 = \begin{pmatrix} \sqrt{3}/6 \\ -\sqrt{3}/6 \\ -\sqrt{3}/6 \\ -\sqrt{3}/2 \end{pmatrix}$$

12. 习题 4.6-14

直接计算可得结果为

$$\arccos\left(-\frac{1}{\sqrt{602}}\right)$$

§9.2 矩阵论技巧

9.2.1 更复杂的分块

1. 若 $A^2 = O$, 求证存在可逆矩阵 P 使得

$$A = P^{-1} \begin{pmatrix} O & I_r \\ O & O \end{pmatrix} P$$

设 A 相抵标准形为 $R\Sigma Q$, 由 $A^2 = O$ 通过消去可得 $\Sigma Q R \Sigma = O$, 设 QR 分块为

$$\begin{pmatrix} X_{r \times r} & Y \\ Z & W_{(n-r) \times (n-r)} \end{pmatrix}$$

计算可得 $X = O$, 从而计算可得

$$A = R\Sigma(QR)R^{-1} = R \begin{pmatrix} O & Y \\ O & O \end{pmatrix} R^{-1}$$

由于乘可逆阵不改变秩, 应有 $\text{rank } Y = r$, 即其行满秩 (由 Y 为 $r \times (n-r)$ 阶矩阵可得 $r \leq n-r$, 即 $2r \leq n$), 于是其与 $(O_{r \times (n-2r)} I_r)$ 相抵, 设

$$Y = P_0 \begin{pmatrix} O & I_r \end{pmatrix} Q_0$$

则利用第三类初等变换的分块形式可发现

$$A = R \begin{pmatrix} P_0 & \\ & I_{n-r} \end{pmatrix} \begin{pmatrix} O & I_r \\ O & O \end{pmatrix} \begin{pmatrix} I_r & \\ & Q_0 \end{pmatrix} R^{-1}$$

为配凑逆进一步计算可发现

$$A = R \begin{pmatrix} P_0 & \\ & I_{n-r} \end{pmatrix} \begin{pmatrix} I_r & \\ & Q_0^{-1} \end{pmatrix} \begin{pmatrix} O & I_r \\ O & O \end{pmatrix} \begin{pmatrix} I_r & \\ & Q_0 \end{pmatrix} \begin{pmatrix} P_0^{-1} & \\ & I_{n-r} \end{pmatrix} R^{-1}$$

由此取

$$P = \begin{pmatrix} I_r & \\ & Q_0 \end{pmatrix} \begin{pmatrix} P_0^{-1} & \\ & I_{n-r} \end{pmatrix} R^{-1}$$

可利用分块初等变换阵的逆验证结论成立。

2. 若 $A^3 = A$, 求证存在可逆矩阵 P 使得

$$A = P^{-1} \begin{pmatrix} I_a & O & O \\ O & -I_b & O \\ O & O & O \end{pmatrix} P$$

设 A 相抵标准形为 $R\Sigma Q$, 由 $A^3 = A$ 通过消去可得

$$\Sigma Q R \Sigma Q R \Sigma = \Sigma$$

设 QR 分块为

$$\begin{pmatrix} X_{r \times r} & Y \\ Z & W_{(n-r) \times (n-r)} \end{pmatrix}$$

计算可得 $X^2 = I$, 由此类似本讲义 4.2.2 第一问 (更详细的证明见之后的 9.2.3) 可证得存在 P_0 使得

$$X = P_0 \begin{pmatrix} I_a & O \\ O & -I_b \end{pmatrix} P_0^{-1}$$

而直接计算可得

$$A = R\Sigma(QR)R^{-1} = R \begin{pmatrix} X & Y \\ O & O \end{pmatrix} R^{-1}$$

由 $X^2 = I$ 知 X 可逆, 从而可利用之前的 Schur 公式消去 Y , 并进一步通过初等变换阵的逆配凑得到

$$A = R \begin{pmatrix} X & O \\ O & O \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ O & I \end{pmatrix} R^{-1} = R \begin{pmatrix} I & -X^{-1}Y \\ O & I \end{pmatrix} \begin{pmatrix} X & O \\ O & O \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ O & I \end{pmatrix} R^{-1}$$

再利用 X 的表达式计算可得

$$A = R \begin{pmatrix} I & -X^{-1}Y \\ O & I \end{pmatrix} \begin{pmatrix} P_0 & O \\ O & I \end{pmatrix} \begin{pmatrix} I_a & O & O \\ O & -I_b & O \\ O & O & O \end{pmatrix} \begin{pmatrix} P_0^{-1} & O \\ O & I \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ O & I \end{pmatrix} R^{-1}$$

由此取

$$P = \begin{pmatrix} P_0^{-1} & O \\ O & I \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ O & I \end{pmatrix} R^{-1}$$

可利用分块初等变换阵的逆验证结论成立。

* 这两题中多次应用了由于矩阵某些部分为 0, 作初等变换后配凑发现另一侧乘上对应的逆无影响, 这是相似配凑的常用技巧。

9.2.2 多项式相关

1. 若 $\lambda_1, \dots, \lambda_n$ 是 A 的全部特征值 (含重数), 证明对多项式 g 有

$$\det(g(A)) = \prod_{i=1}^n g(\lambda_i)$$

设 $g(x) = \prod_{k=1}^m (x - \mu_k)$, 则利用矩阵多项式的可交换性可得

$$g(A) = \prod_{k=1}^m (A - \mu_k I)$$

由 Binet-Cauchy 公式得到

$$\det(g(A)) = \prod_{k=1}^m \det(A - \mu_k I)$$

利用定义, $A - \mu_k I$ 的特征多项式为 $\det((\lambda + \mu_k)I - A)$, 由此其全部特征值为 $\lambda_1 - \mu_k, \dots, \lambda_n - \mu_k$, 再利用行列式等于特征值乘积可得

$$\det(g(A)) = \prod_{k=1}^m \prod_{i=1}^n (\lambda_i - \mu_k)$$

先对 k 作乘积即得到

$$\det(g(A)) = \prod_{i=1}^n g(\lambda_i)$$

2. 证明

$$x^k - y^k = \prod_{j=0}^{k-1} (x - \omega^j y), \quad \omega = e^{2\pi i/k}$$

先考虑 $y = 1$ 时。由 ω 的表达式可验证 $\omega^k = e^{2\pi i} = 1$, 由此 ω 是 $x^k - 1$ 的根, 而同理计算得 $\omega^0, \omega^1, \dots, \omega^{k-1}$ 都是 $x^k - 1$ 的根。

此外, 对 $0 \leq t < s \leq k - 1$, 直接计算可得 (最后不等号是由于其对应小于 2π 的某角度)

$$\omega^{s-t} = e^{2(s-t)\pi i/k} \neq 1$$

从而这些根互不相同。根据代数学基本定理, $x^k - 1$ 与 $\prod_{j=0}^{k-1} (x - \omega^j)$ 只相差倍数, 再由它们首项系数相同可知相等。

对一般的 y , 令 $t = x/y$, 左侧即可写为 $y^k(t^k - 1)$, 写出分解后再将 y^k 乘到分解式中即可。

3. 已知 A 的特征多项式为 $\varphi_A(\lambda)$, 求 A^k 的特征多项式。

由定义可知要求的多项式

$$\varphi_{A^k}(\lambda) = \det(\lambda I - A^k) = (-1)^n \det(A^k - \lambda I)$$

* 此处作相反数是为了保持 A 部分的独立, 方便之后化为 A 的特征多项式。

将其看作 (这里对复数任取某 k 次方根即可, 不影响之后讨论) $\det(A^k - (\lambda^{1/k} A)^k)$, 利用上题可写出分解

$$\varphi_{A^k}(\lambda) = (-1)^n \det \left(\prod_{j=0}^{k-1} (A - \omega^j \lambda^{1/k} I) \right) = (-1)^n \prod_{j=0}^{k-1} \det(A - \omega^j \lambda^{1/k} I), \quad \omega = e^{2\pi i/k}$$

为将其化为 A 的特征多项式形式, 再在每个因式中提出 $(-1)^n$ 即得

$$\varphi_{A^k}(\lambda) = (-1)^{kn+n} \prod_{j=0}^{k-1} \det(\omega^j \lambda^{1/k} I - A) = (-1)^{kn+n} \prod_{j=0}^{k-1} \varphi_A(\omega^j \lambda^{1/k})$$

* 若之前不作相反数, 这里提出的系数将会更加复杂, 不过可通过精细计算化出相同的结论。

4. 证明 A^k 的全部特征值为 A 的全部特征值对应作 k 次方。

即要证明

$$\det(\lambda I - A^k) = \prod_{t=1}^n (\lambda - \lambda_t^k)$$

类似上题写出分解

$$\det(\lambda I - A^k) = \prod_{j=0}^{k-1} \det(\lambda^{1/k} I - \omega^j A)$$

利用习题 5.5-7 可知 $\omega^j A$ 特征值为 $\omega^j \lambda_1, \dots, \omega^j \lambda_n$ ，于是可知行列式分解为 (将 $\lambda^{1/k}$ 看作整体)

$$\prod_{j=0}^{k-1} \prod_{t=1}^n (\lambda^{1/k} - \omega^j \lambda_t)$$

更换乘积次序，先对 j 作乘积，再次利用分解式即得到这是

$$\prod_{t=1}^n (\lambda - \lambda_t^k)$$

* 稍复杂的多项式题目往往需要熟练的代数变换与单位根 (即 $e^{2\pi i/k}$) 相关的技巧，一个取巧的方法是利用之后介绍的相似三角化，但这也并不能保证解决所有问题。

9.2.3 可对角化 II

1. 若 A 可对角化，求证存在无重根的多项式 f 使得 $f(A) = O$ 。

设 A 的所有不同特征值为 $\lambda_1, \dots, \lambda_k$ ，其对角化为 $A = P^{-1}DP$ ，则设 $f(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)$ ，利用多项式的相似计算可知

$$f(A) = P^{-1}f(D)P$$

由于 D 的某个位置一定等于某个 λ_i ，且

$$f(D) = \prod_{i=1}^k (D - \lambda_i I)$$

D 的某个位置一定在某个因子中为 0，由此直接计算可得 $f(D) = O$ ，即得证 $f(A) = O$ 。

2. 若存在无重根的多项式 f 使得 $f(A) = O$ ，求证 A 可对角化。

回顾 $\text{diag}(A_1, A_2, \dots, A_k)$ 表示将 A_1, \dots, A_k 排在主对角线上的分块对角阵。设

$$f(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i)$$

且 $\lambda_1, \dots, \lambda_k$ 互不相同，下面证明存在 P 使得

$$A = P^{-1} \text{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \dots, \lambda_k I_{n_k})P$$

对 k 归纳。当 $k = 1$ 时，直接得到 $A = \lambda_1 I$ ，符合要求，下面设对 $1, \dots, k-1$ 均成立，考虑 k 的情况。

* 终极目标：化为已解决的情况。为此，我们需要不断简化命题。接下来设 $\lambda_k = 0$ 是因为我们可以方便地通过“平移”将情况化到 $\lambda_k = 0$ 时，而 $\lambda_k = 0$ 的情况思路来自之前分析 $A^3 = A$ 时，详见后续证明。

先证明 $\lambda_k = 0$ 时情况成立。此时, $f(\lambda)$ 没有零次项, 可设其为

$$f(\lambda) = \prod_{i=1}^k (\lambda - \lambda_i) = \lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_1\lambda$$

则有

$$A^k + a_{k-1}A^{k-1} + \cdots + a_1A = O$$

设 A 相抵标准形为 $R\Sigma Q$, 代入上式得到

$$(R\Sigma Q)^k + a_{k-1}(R\Sigma Q)^{k-1} + \cdots + a_1R\Sigma Q = O$$

注意每项最左侧都为 P 、最右侧都为 Q , 可消去, 且计算可发现消去后原式化为 (可直接展开计算说明 $R^{-1}(R\Sigma Q)^k Q^{-1} = (\Sigma Q R)^{k-1} \Sigma$)

$$(\Sigma Q R)^{k-1} \Sigma + a_{k-1}(\Sigma Q R)^{k-2} \Sigma + \cdots + a_1 \Sigma = O$$

设 QR 分块为

$$\begin{pmatrix} X_{r \times r} & Y \\ Z & W_{(n-r) \times (n-r)} \end{pmatrix}$$

计算可得

$$(\Sigma Q R)^t \Sigma = \begin{pmatrix} X^t & O \\ O & O \end{pmatrix}$$

于是可进一步化为

$$X^{k-1} + a_{k-1}X^{k-2} + \cdots + a_1I = O$$

这是一个以 $\lambda_1, \dots, \lambda_{k-1}$ 为根的 $k-1$ 次多项式, 于是根据归纳假设有

$$X = P_0^{-1} \operatorname{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \dots, \lambda_{k-1} I_{n_{k-1}}) P_0$$

此外, 由于 f 无重根, $\lambda_1, \dots, \lambda_{k-1}$ 中不会有 0, 因此 X 可逆。

直接计算可得

$$A = R\Sigma(QR)R^{-1} = R \begin{pmatrix} X & Y \\ O & O \end{pmatrix} R^{-1}$$

与本讲义 9.2.1 第二题完全相同可得

$$A = R \begin{pmatrix} I & -X^{-1}Y \\ O & I \end{pmatrix} \begin{pmatrix} X & O \\ O & O \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ O & I \end{pmatrix} R^{-1}$$

完全相同继续计算得

$$A = R \begin{pmatrix} I & -X^{-1}Y \\ O & I \end{pmatrix} \begin{pmatrix} P_0 & O \\ O & I \end{pmatrix} \operatorname{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \dots, \lambda_{k-1} I_{n_{k-1}}, O) \begin{pmatrix} P_0^{-1} & O \\ O & I \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ O & I \end{pmatrix} R^{-1}$$

而由于 $\lambda_k = 0$, 取

$$P = \begin{pmatrix} P_0^{-1} & O \\ O & I \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ O & I \end{pmatrix} R^{-1}$$

这已经是符合要求的形式。

下面证明一般的情况。对任何 A , 令 $B = A - \lambda_k I$, 则直接代入 $f(A) = O$ 可发现

$$\prod_{i=1}^{i-1} (B - (\lambda_i - \lambda_k)I)B = O$$

由 $\lambda_1, \dots, \lambda_k$ 互不相同, 它们同减 λ_k 后仍互不相同, 这即符合 $\lambda_k = 0$ 的情况, 于是存在 P 使得

$$B = P^{-1} \operatorname{diag}((\lambda_1 - \lambda_k)I_{n_1}, \dots, (\lambda_{k-1} - \lambda_k)I_{n_{k-1}}, O)P$$

还原回 A , 由于可交换性 $\lambda_k I = P^{-1}(\lambda_k I)P$, 利用乘法分配律最终得到

$$A = P^{-1}(\operatorname{diag}((\lambda_1 - \lambda_k)I_{n_1}, \dots, (\lambda_{k-1} - \lambda_k)I_{n_{k-1}}, O) + \lambda_k I)P$$

设 O 为 n_k 阶, 即有最终结论

$$A = P^{-1} \operatorname{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \dots, \lambda_k I_{n_k})P$$

* 这本质是最小多项式相关的结论, 将在之后介绍。

§9.3 相似的特殊结论

* 本部分的结论大都非常重要, 之后也会大量用到, 因此建议大家至少熟悉这里的所有结论。

9.3.1 相似三角化

1. 证明任何复方阵与某上三角阵相似。

核心思路: 特征方程 $A\alpha = \lambda\alpha$ 用矩阵乘法表述 (将 λ 看作一阶方阵) 事实上是 $A\alpha = \alpha\lambda$, 具有相似的形式。

对阶数归纳, 一阶时直接成立, 若对 $n-1$ 阶成立, n 阶时选出 A 的某特征值 λ 与特征向量 α , 并由 $\alpha \neq 0$ 利用基扩充构造一个以 α 为第一列的可逆阵, 记为 P , 计算可发现 AP 第一列应为 $\lambda\alpha$, 进一步计算有 (这里 x, X 为未知向量/矩阵)

$$AP = P \begin{pmatrix} \lambda & x \\ 0 & X \end{pmatrix}$$

利用归纳假设, 设 $X = QUQ^{-1}$, U 为上三角阵, 则计算可发现

$$A = P \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} \lambda & xQ \\ 0 & U \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q^{-1} \end{pmatrix} P^{-1}$$

记

$$R = \begin{pmatrix} 1 & 0 \\ 0 & Q^{-1} \end{pmatrix} P^{-1}$$

则 RAR^{-1} 已经为上三角阵。

* 利用相似不改变特征多项式, 此上三角阵的对角元即为其特征值, 且由于算法中每一步取的特征值是任意的, 对角元可按任意顺序排列。

2. 若 $\lambda_1, \dots, \lambda_n$ 是 A 的全部特征值, 对多项式 f , 求证 $f(A)$ 的全部特征值为 $f(\lambda_1), \dots, f(\lambda_n)$ 。

由于 $A = P^{-1}UP$, U 是对角元为 $\lambda_1, \dots, \lambda_n$ 的上三角阵, 有 $f(A) = P^{-1}f(U)P$, 而直接计算可发现 $f(U)$ 是对角元为 $f(\lambda_1), \dots, f(\lambda_n)$ 的上三角阵。由于 $f(A)$ 与 $f(U)$ 相似, 二者特征值相同, 而上三角阵特征值为对角元, 从而得证。

9.3.2 置换相似

- 称每行每列恰有一个元素为 1, 其他为 0 的矩阵为 **置换阵**, 证明置换阵可逆, 并计算其逆。

利用完全展开式即得置换阵行列式为 ± 1 , 从而可逆。

设其非零元素为 $a_{1m_1}, a_{2m_2}, \dots, a_{nm_n}$, 考虑其左乘一个方阵, 即相当于把第 m_1 行交换到第一行, 第 m_2 行交换到第二行……由此从变换的角度其逆应为 $a_{m_11}, a_{m_22}, \dots, a_{m_nn}$ 非零的置换阵, 而这恰为原矩阵转置。下面验证此结论。

记 δ_{ab} 当 $a = b$ 时为 1, 否则为 0, 则置换阵可以写为 $a_{ij} = \delta_{jm_i}$ 。其转置即 $(A^T)_{ij} = a_{ji} = \delta_{im_j}$ 。

直接计算发现

$$(AA^T)_{ij} = \sum_{k=1}^n \delta_{km_i} \delta_{km_j}$$

当 $i = j$ 时, $m_i = m_j$, 这 n 项求和中有一项为 1, 从而为 1; 否则由每列只有一个元素为 1 可知 $m_i \neq m_j$, 从而为 0。这就已经得到了单位阵的形式。

- 若 K 为置换阵, 设其第 i 行的非零元素为 m_i , 对任何方阵 A 计算 $K^{-1}AK$, 并以此说明相似对角化中特征值可按任意顺序排列。

直接计算 (这里若觉得过程不清晰可以举具体例子观察)

$$(KAK^{-1})_{ij} = \sum_{kl} \delta_{m_i k} a_{kl} \delta_{l m_j} = a_{m_i m_j}$$

由此, 对于对角阵 D , 用这样的 K 置换相似后, 即将其第 i 个对角元变为第 m_i 个对角元, 非对角仍保持 0 ($i \neq j$ 时 $m_i \neq m_j$), 这就足以任意指定顺序了。

9.3.3 Caylay-Hamilton 定理

本节记 $\varphi_A(\lambda) = \det(\lambda I - A)$ 为 A 的特征多项式, 下方的第一个习题即为 Caylay-Hamilton 定理。

- 证明 $\varphi_A(A) = O$ 。

由于特征多项式不会随 A 看作的数域改变, 将 A 看作复方阵不影响结论。利用相似三角化, 设 $A = P^{-1}UP$, U 为上三角阵, 则 $\varphi_A(A) = P^{-1}\varphi_A(U)P$ 。为证 $\varphi_A(A) = O$, 只需证明对于对角元为 $\lambda_1, \dots, \lambda_n$ 的上三角阵 U 有

$$\prod_{i=1}^n (\lambda_i I - U) = O$$

对 U 的阶数归纳, 利用

$$\begin{pmatrix} 0 & x \\ \mathbf{0} & X_{(n-1) \times (n-1)} \end{pmatrix} \begin{pmatrix} \lambda_1 & y \\ 0 & O \end{pmatrix} = O$$

即可得到结论。

- 证明若 A 可逆, 则存在多项式 f 使得 $A^{-1} = f(A)$, 且 f 的次数小于 n 。

若 A 可逆, 其无零特征值, 也即 $\varphi_A(\lambda)$ 所有根都非零, 因此其常数项非零。设其为

$$\lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$$

则利用 Caylay-Hamilton 定理

$$A^n + a_{n-1}A^{n-1} + \dots + a_1A + a_0I = O$$

直接计算可发现

$$A^{-1} = -\frac{1}{a_0}(A^{n-1} + a_{n-1}A^{n-2} + \dots + a_1I)$$

3. 证明若 A 的多项式可逆, 则逆是 A 的多项式, 且可使其次数小于 n 。

由于 A 的多项式的逆是 A 的多项式的多项式, 其也必然为 A 的多项式。此外, 如上题设 φ_A , 有

$$A^n = -a_{n-1}A^{n-1} - \cdots - a_1A - a_0I$$

因此 A 的任何多项式大于等于 n 次的项都可以不断代入降低次数, 最终得到其等于某小于 n 次的多项式。

4. 计算循环方阵 A 的逆 (假设的确可逆), 也即 n 阶方阵 A 的元素满足

$$a_{ij} = b_k \quad j - i = k \text{ 或 } i - j = n - k$$

设置换方阵 K 满足 $a_{12} = a_{23} = \cdots = a_{n-1,n} = 1$, $a_{n1} = 1$, 其余为 0, 则计算可发现 K^s 恰好满足 $j = i = s$ 或 $i - j = n - s$ 时为 1, 否则为 0, 由此

$$A = \sum_{k=0}^{n-1} b_k K^k$$

利用上题可设

$$A^{-1} = \sum_{k=0}^{n-1} c_k K^k$$

此外, 计算可发现 $K^n = I$, 于是 AA^{-1} 的零次项事实上相当于多项式乘积的 0 与 n 次项之和, 一次项事实上相当于多项式乘积的 1 与 $n+1$ 次项之和, 由此可得到方程组。

对于一般情况, 直接求解此方程组的相对复杂, 需要利用单位根相关的较多知识, 本质是离散 Fourier 变换, 因此我们需要另一种思路。记

$$f(\lambda) = \sum_{k=0}^{n-1} b_k \lambda^k$$

直接计算可发现 K 的特征多项式是 $\lambda^n - 1$, 于是其特征值为 $1, \omega, \dots, \omega^{n-1}$, 这里 $\omega = e^{2\pi i/n}$ 。

* 由此也可直接得到 A 的特征多项式。

由其特征值互不相同, 特征多项式无重根, 根据 Caylay-Hamilton 定理, 它是可对角化的。事实上, 求解其对每个特征值 ω^j 的特征向量 (具体形式仍然需要一些单位根的性质), 并拼成矩阵 P , 可直接验证

$$P^{-1}KP = \text{diag}(1, \omega, \dots, \omega^{n-1})$$

由此有

$$P^{-1}AP = P^{-1}f(K)P = f(P^{-1}KP) = \text{diag}(f(1), f(\omega), \dots, f(\omega^{n-1}))$$

从而

$$A = P \text{diag}(f(1), f(\omega), \dots, f(\omega^{n-1}))P^{-1}$$

于是直接计算得可逆时逆为

$$A^{-1} = P \text{diag}(f(1)^{-1}, f(\omega)^{-1}, \dots, f(\omega^{n-1})^{-1})P^{-1}$$

9.3.4 一般数域上的相似

对 \mathbb{K} 上 n 阶方阵 A , 若有 $\lambda \in \mathbb{K}$ 与 $\alpha \in \mathbb{K}^n$ 使得 $A\alpha = \lambda\alpha$, 则称 λ 为 A 的特征值, α 为 A 的特征向量。

1. 证明 λ 为 A 的特征值当且仅当 $\lambda \in \mathbb{K}$ 为 A 看作 \mathbb{C} 上矩阵的特征值。

由于 $A\alpha = \lambda\alpha$ 解存在仍然等价于 $\det(\lambda I - A) = 0$, 而此特征多项式与数域无关, 因此其解 λ 一定在 \mathbb{C} 上也为解, 这就说明了 \mathbb{K} 上特征值当且仅当在 \mathbb{K} 中且为 \mathbb{C} 上的特征值。

2. 若 A 作为 \mathbb{C} 上方阵可对角化, 且特征值均在 \mathbb{K} 中, 证明 A 作为 \mathbb{K} 上方阵可对角化。

利用本讲义 9.2.3 的等价定义, 由条件可知存在无重根且根都在 \mathbb{K} 中 (从而系数都在 \mathbb{K} 中) 的多项式 f 使得 $f(A) = O$ 。

观察本讲义 9.2.3 第二题证明。由于相抵标准形与数域无关, R, X, Y 均为 \mathbb{K} 中的矩阵, 由此每一步构造的 P 比起 P_0 多乘的部分都是在 \mathbb{K} 中的 (注意逆是在同一个数域中的), 以此归纳即可构造出 \mathbb{K} 中的相似对角化。

* 事实上从其他可对角化等价定义中也可证明此结论, 留给读者思考。

3. 若 A 作为 \mathbb{K} 上方阵可对角化, 证明 A 作为 \mathbb{C} 上方阵可对角化, 且特征值均在 \mathbb{K} 中。

由于存在 \mathbb{K} 上可逆方阵 P 使得 $P^{-1}AP$ 为对角阵, 其对角元必然也在 \mathbb{K} 中, 由此特征值在 \mathbb{K} 中。由于 P 也可看作 \mathbb{C} 上的可逆方阵, A 在 \mathbb{C} 上亦可对角化。

§9.4 空间技巧

* 本节需要熟悉本讲义 8.3.1、8.3.2 与之前空间相关的基本结论。由于需要的前置较多, 一般更推荐练习矩阵技巧。不过, 由于空间视角蕴含一些更本质的思想, 这里也进行介绍。在熟悉矩阵表示的语言后, 有不少题目也的确用空间叙述更加简便。

9.4.1 选取基的艺术

1. 若 $A^2 = O$, 求证存在可逆矩阵 P 使得

$$A = P^{-1} \begin{pmatrix} O & I_r \\ O & O \end{pmatrix} P$$

由此, 由线性变换矩阵表示与坐标的联系, 利用直接的计算结论

$$\begin{pmatrix} O & I_r \\ O & O \end{pmatrix} e_i = \begin{cases} 0 & i \leq n-r \\ e_{i-n+r} & i > n-r \end{cases}$$

A 能写为上述形式当且仅当存在一组基 $\alpha_1, \dots, \alpha_n$ 使得

$$A\alpha_1 = 0, \dots, A\alpha_{n-r} = 0$$

$$A\alpha_{n-r+1} = \alpha_1, \dots, A\alpha_n = \alpha_r$$

下面我们设法找到这组基。由 $A^2 = O$ 可知 $\mathcal{A}(\text{Im } \mathcal{A}) = \{0\}$, 也即 $\text{Im } \mathcal{A} \subset \text{Ker } \mathcal{A}$, 且根据解空间维数定理 $\text{Im } \mathcal{A}$ 为 r 维、 $\text{Ker } \mathcal{A}$ 为 $n-r$ 维。由此, 取 $\alpha_1, \dots, \alpha_r$ 为 $\text{Im } \mathcal{A}$ 一组基, 并将其扩充成 $\text{ker } \mathcal{A}$ 的一组基 $\alpha_1, \dots, \alpha_{n-r}$, 则

$$A\alpha_1 = 0, \dots, A\alpha_{n-r} = 0$$

已经满足。

为使第二组条件成立, 由于 $\alpha_1, \dots, \alpha_r$ 在 $\text{Im } \mathcal{A}$ 中, 可设它们的某个原像为 $\alpha_{n-r+1}, \dots, \alpha_n$, 则第二组条件也得到了满足, 只需说明选出的 $\alpha_1, \dots, \alpha_n$ 线性无关即可。假设

$$\sum_{i=1}^n \lambda_i \alpha_i = 0$$

首先, 这可以推出

$$A \sum_{i=1}^n \lambda_i \alpha_i = \sum_{i=1}^n \lambda_i A \alpha_i = 0$$

而根据两组条件, 这即代表

$$\sum_{i=1}^r \lambda_{i+n-r} \alpha_i = 0$$

由于 $\alpha_1, \dots, \alpha_r$ 为 $\text{Im } \mathcal{A}$ 一组基, 只能 $\lambda_{n-r+1}, \dots, \lambda_n$ 全为 0。此时, 原式化为

$$\sum_{i=1}^{n-r} \lambda_i \alpha_i$$

但 $\alpha_1, \dots, \alpha_{n-r}$ 为 $\text{Ker } \mathcal{A}$ 一组基, 从而必须 $\lambda_1, \dots, \lambda_{n-r}$ 也全为 0, 线性无关性得证, 于是这组基即符合要求。

* 这里过程略显复杂是因为写得较详细, 实际上是比之前的矩阵做法简单得多的。

2. 若 $A^3 = A$, 求证存在可逆矩阵 P 使得

$$A = P^{-1} \begin{pmatrix} I_a & O & O \\ O & -I_b & O \\ O & O & O \end{pmatrix} P$$

仍然直接计算可知 A 能写为上述形式当且仅当存在一组基 $\alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b, \gamma_1, \dots, \gamma_{n-a-b}$ 使得

$$A\alpha_1 = \alpha_1, \dots, A\alpha_a = \alpha_a$$

$$A\beta_1 = -\beta_1, \dots, A\beta_b = -\beta_b$$

$$A\gamma_1 = 0, \dots, A\gamma_{n-a-b} = 0$$

这些式子事实上意味着 $\alpha_i \in \text{Ker}(\mathcal{I} - \mathcal{A})$ 、 $\beta_j \in \text{Ker}(\mathcal{I} + \mathcal{A})$ 、 $\gamma_k \in \text{Ker } \mathcal{A}$, 而只要证明

$$\mathbb{C}^n = \text{Ker}(\mathcal{I} - \mathcal{A}) + \text{Ker}(\mathcal{I} + \mathcal{A}) + \text{Ker } \mathcal{A}$$

利用和空间定义即可在右侧三空间中取出一些基成为全空间的基 (具体来说, 考虑每个空间的一组基, 由定义它们可以生成全空间, 于是它们的极大线性无关组构成全空间一组基)。

* 事实上, 由于这些空间都是特征子空间 ($\text{Ker } \mathcal{A} = \text{Ker}(-\mathcal{A})$ 为 0 的特征子空间), 根据本讲义第八章结论, 右侧的和为直和。

我们分两步证明:

- $\mathbb{C}^n = \text{Ker } \mathcal{A} \oplus \text{Ker}(\mathcal{A}^2 - \mathcal{I})$

先说明它们交为 $\{0\}$ 。它们交中的元素应满足 $Ax = 0$ 且 $A^2x - x = 0$, 而第一式代入第二式即得 $x = A(Ax) = 0$, 于是得证。

将 $A^3 = A$ 写为 $A(A^2 - I) = O$, 利用 Sylvester 秩不等式 (见本讲义 8.1.3) 可得

$$\text{rank } A + \text{rank}(A^2 - I) \leq \text{rank } O + \text{rank } I = n$$

于是由解空间维数定理

$$\dim \text{Ker } \mathcal{A} + \dim \text{Ker}(\mathcal{A}^2 - \mathcal{I}) \geq n + n - n = n$$

然而, 由于两者和为直和, 根据和空间维数可知

$$\dim(\text{Ker } \mathcal{A} \oplus \text{Ker}(\mathcal{A}^2 - \mathcal{I})) = \dim \text{Ker } \mathcal{A} + \dim \text{Ker}(\mathcal{A}^2 - \mathcal{I}) \geq n$$

从而其只能为全空间, 得证。

- $\text{Ker}(\mathcal{A}^2 - \mathcal{I}) = \text{Ker}(\mathcal{I} - \mathcal{A}) \oplus \text{Ker}(\mathcal{I} + \mathcal{A})$

由于 $A^2 - I = (A + I)(A - I) = (A - I)(A + I)$, 无论是 $(A - I)x = 0$ 还是 $(A + I)x = 0$ 都可推出 $(A^2 - I)x = 0$, 因此右侧两空间确实包含于左侧。此外, 利用特征子空间和为直和可知右侧确实为直和。

记 $V = \text{Ker}(\mathcal{A}^2 - \mathcal{I})$, 且记 $\dim V = n_0$ 。由于若 $x \in V$, 有

$$(A^2 - I)(Ax) = A((A^2 - I)x) = 0$$

于是可知 $Ax \in V$, 因此 \mathcal{A} 可看作 $V \rightarrow V$ 的线性变换 \mathcal{A}_0 , 设其某个矩阵表示为 A_0 。记 $V \rightarrow V$ 的恒等映射为 \mathcal{I}_0 。

* 利用限制映射的语言, $\mathcal{A}_0 = \mathcal{A}|_{V \rightarrow V}$, 注意 A_0 为 $n_0 \times n_0$ 阶方阵。

由于对 V 中任何 x 有 $A^2x = x$, 可知 $(\mathcal{A}_0^2 - \mathcal{I}_0)x = 0$, 于是 $\mathcal{A}_0^2 = \mathcal{I}$, 由此对矩阵表示也有 $A_0^2 = I_0$ 。此外, 由于右侧两空间在 V 中, 利用定义应有

$$\text{Ker}(\mathcal{I} - \mathcal{A}) = \text{Ker}(\mathcal{I}_0 - \mathcal{A}_0), \quad \text{Ker}(\mathcal{I} + \mathcal{A}) = \text{Ker}(\mathcal{I}_0 + \mathcal{A}_0)$$

为了证明两空间和为 V , 我们再次进行维数计算, 利用 $A_0^2 - I = O$, 由 Sylvester 秩不等式可得

$$\text{rank}(I_0 - A_0) + \text{rank}(I_0 + A_0) \leq n_0$$

从而完全类似第一步证明得到

$$\dim(\text{Ker}(\mathcal{I}_0 - \mathcal{A}_0) \oplus \text{Ker}(\mathcal{I}_0 + \mathcal{A}_0)) \geq n_0$$

但此空间又在 V 中, 因此只能为 V , 得证。

* 这里事实上也可以由 Sylvester 秩不等式直接说明, 见之后证明根子空间的和是直和时的做法。

* 大家或许注意到了两步证明的相似性。事实上, 它们在进行类似的化归操作。第二步证明虽然抽象, 但是揭示了映射语言的一种优势: 可以利用限制映射统一不同阶数的情况。

* 从本题空间角度与矩阵角度的联系或许也可以看出, 分块与限制映射本质是统一的, 具体来说, 限制映射的矩阵表示可以在适当的基下看成原矩阵的子矩阵, 之后将详细说明。

* 此题的做法也可以推广证明本讲义 9.2.3 的结论, 这里不再赘述。

9.4.2 特征子空间与根子空间

1. 若 n 阶方阵 A, B 满足 $\text{rank } A + \text{rank } B < n$, 求证它们有公共特征向量。

类似上节说明, $\text{Ker } \mathcal{A}$ 为 A 的零特征值对应的特征子空间, $\text{Ker } \mathcal{B}$ 为 B 的零特征值对应的特征子空间, 而条件可利用解空间维数定理改写为

$$\dim \text{Ker } \mathcal{A} + \dim \text{Ker } \mathcal{B} > n + n - n = n$$

若两者交只有 $\{0\}$, 则它们直和的维数将超过 n , 与全空间维数为 n 矛盾, 因此必然有非零交, 而根据特征子空间定义这就是 A, B 的公共特征向量。

2. 对特征值 λ , 考虑空间 $\text{Ker}(\lambda\mathcal{I} - \mathcal{A})^k$, 将满足 $\text{Ker}(\lambda\mathcal{I} - \mathcal{A})^k = \text{Ker}(\lambda\mathcal{I} - \mathcal{A})^{k+1}$ 的最小正整数 k 对应的 $\text{Ker}(\lambda\mathcal{I} - \mathcal{A})^k$ 称为其根子空间, 证明上述定义合理, 且记 $r_t = \text{Ker}(\lambda\mathcal{I} - \mathcal{A})^t$, 则 r_t 在 $1, \dots, k$ 严格单调增加, 而对任何 $j \geq k$ 有 $r_j = r_k$ 。

由于 $\text{Ker}(\lambda\mathcal{I} - \mathcal{A})^k \subset \text{Ker}(\lambda\mathcal{I} - \mathcal{A})^{k+1}$, 空间维数不减, 而最高维数为全空间维数, 因此这样的 k 一定存在。

类似习题 4.3-16, 利用 Frobenius 秩不等式可通过维数计算证明, 只要

$$\text{Ker}(\lambda\mathcal{I} - \mathcal{A})^k = \text{Ker}(\lambda\mathcal{I} - \mathcal{A})^{k+1}$$

即有

$$\text{Ker}(\lambda\mathcal{I} - \mathcal{A})^k = \text{Ker}(\lambda\mathcal{I} - \mathcal{A})^{k+m}$$

对任何正整数 m 成立。

3. 证明 λ 对应的根子空间维数等于 λ 的代数重数。

设 $A = P^{-1}UP$, U 为上三角阵, 且由于可任意排练顺序, 不妨设其前 m 个对角元为 λ , 之后均不为 λ 。利用多项式相似结论有

$$(\lambda\mathcal{I} - A)^k = P^{-1}(\lambda\mathcal{I} - U)^k P$$

记 $U_0 = \lambda\mathcal{I} - U$, 其为前 m 个对角元为 0, 其余非零的上三角阵, 将其分块为

$$\begin{pmatrix} X_{m \times m} & Y \\ O & W_{(n-m) \times (n-m)} \end{pmatrix}$$

直接计算发现, U_0^m 的左上角为 X^m , 可计算得其为 O , 于是其只有 $n - m$ 个非零列, $\text{rank } U_0^m$ 至多为 $n - m$ 。另一方面, 右下角为对角元非零的上三角阵, 因此构成 $n - m$ 阶非零子式, 于是 $\text{rank } U_0^m = n - m$, 也即 $\text{Ker } U_0^m = m$ 。

对任何高于 m 的次方, 上述推理仍然成立, 于是此后维数恒为 m , 根据上题, 维数不变时即为根子空间维数, 从而得证。

* 与代数重数相关的题目往往归于相似三角化, 因为一般并没有更好的刻画代数重数的方法。

4. 证明所有不同特征值的根子空间的和是直和。

我们证明一个更一般的结论: 若多项式 f, g 没有公共根, 则

$$\text{Ker } f(\mathcal{A}) \oplus \text{Ker } g(\mathcal{A}) = \text{Ker}(f(\mathcal{A})g(\mathcal{A}))$$

由此根据根子空间定义可以归纳得到结果。

先说明左侧交为 $\{0\}$ 。这里需要利用多项式的裴蜀定理 (可通过辗转相除证明), 也即, 由于 f, g 无公共根, 它们的最大公因式为 1, 因此存在多项式 u, v 使得

$$f(\lambda)u(\lambda) + g(\lambda)v(\lambda) = 1$$

由此, 设 x 在左侧交集中, 有

$$x = (u(\mathcal{A})f(\mathcal{A}) + v(\mathcal{A})g(\mathcal{A}))x = 0 + 0 = 0$$

再说明和包含在右侧中。由于 $f(\mathcal{A})g(\mathcal{A}) = g(\mathcal{A})f(\mathcal{A})$, 左侧的两个空间确实为右侧的子空间, 从而

$$\text{Ker } f(\mathcal{A}) \oplus \text{Ker } g(\mathcal{A}) \subset \text{Ker}(f(\mathcal{A})g(\mathcal{A}))$$

利用和空间维数公式, 这也可以推出

$$\dim \text{Ker } f(\mathcal{A}) + \dim \text{Ker } g(\mathcal{A}) \leq \dim \text{Ker}(f(\mathcal{A})g(\mathcal{A}))$$

最后, 利用 Sylvester 秩不等式有

$$\text{rank}(f(\mathcal{A})g(\mathcal{A})) \geq \text{rank } f(\mathcal{A}) + \text{rank } g(\mathcal{A}) - n$$

于是

$$\dim \text{Ker}(f(\mathcal{A})g(\mathcal{A})) \leq \dim \text{Ker } f(\mathcal{A}) + \dim \text{Ker } g(\mathcal{A})$$

这与之前的不等式结合可得到左右维数相等, 又由包含关系知等号成立。

* 通过结论 3、4, 计算维数可以得到所有根子空间直和为全空间, 这称为根子空间分解。

9.4.3 可对角化 III

1. 证明 A 可对角化当且仅当其所有特征子空间直和为全空间。

所有特征子空间的和是直和见 8.3.2 节, 也可利用上节证明的 f, g 无公共根则

$$\text{Ker } f(\mathcal{A}) \oplus \text{Ker } g(\mathcal{A}) = \text{Ker } (f(\mathcal{A})g(\mathcal{A}))$$

进行归纳, 这里只进行等价性的证明。由本讲义 8.3.3, 可对角化等价于存在一组基使得它们均为 A 的特征向量。

若所有特征子空间直和为全空间, 取每个特征子空间的一组基, 利用和空间定义可知它们可以线性组合出 \mathbb{R}^n 中任何向量。此外, 利用和空间维数定理可知这些基的总个数为所有特征子空间维数和, 即为 n , 由此它们构成 \mathbb{R}^n 一组基, 从而 A 可对角化。

若 A 可对角化, 取出符合要求的一组基, 由于每个都是 A 的特征向量, 它们必然都在所有特征子空间的和空间中。由此, 所有特征子空间的和包含全空间, 从而只能为全空间, 得证。

2. 证明 A 可对角化当且仅当其所有特征值代数重数等于几何重数。

可对角化阵的特征值代数重数等于几何重数证明见本讲义 8.3.3, 这里只证明另一边。

若所有特征值代数重数等于几何重数, 由于所有特征子空间和是直和, 利用和空间维数定理可得所有特征子空间直和维数即为所有特征值代数重数和, 而根据代数重数定义这即为 n , 由此和空间必然为全空间。

3. 证明 A 可对角化当且仅当其任何特征子空间是根子空间。

根据定义, 每个特征子空间包含在对应的根子空间中, 而所有根子空间和为全空间, 由此, 若每个特征子空间都为根子空间, 则它们和为全空间。反之, 只要某个特征子空间真包含于根子空间, 所有特征子空间维数和不足所有根子空间维数和 n , 因此不可能和为全空间。

4. 证明上三角阵可对角化当且仅当

$$i < j, \quad a_{ii} = a_{jj} \implies a_{ij} = 0$$

对于上三角阵, 利用代数重数定义可知特征值 λ 的代数重数等于其在对角元出现的次数。设特征值 λ 出现的位置为 i_1, \dots, i_r , 由于相似上三角化可任意指定对角元顺序, 可将此上三角阵相似为 λ 出现在 $1, \dots, r$ 的上三角阵, 且此过程不改变可对角化性。

* 上述顺序调整也可直接通过置换相似构造得到。

我们下面证明, λ 的几何重数为 r 当且仅当 $a_{ij} = 0$ 对 $i < j \leq r$ 成立, 这就可以通过上一个结论得到证明。

若 $a_{ij} = 0$ 对 $i < j \leq r$ 成立, 直接计算可发现 e_1, \dots, e_r 满足 $Ae_i = \lambda e_i$, 由它们线性无关, 几何重数至少为 r , 再由代数重数为 r 与几何重数不超过代数重数可知几何重数为 r 。

若几何重数为 r , 设 $B = \lambda I - A$, 其为前 r 个对角元为 0 的上三角阵, 且非对角元有 $b_{ij} = -a_{ij}$ 。

将 $Bx = 0$ 展开, 利用上三角性与 b_{11} 到 b_{rr} 为 0 可得

$$\forall 1 \leq i \leq r, \quad \sum_{j=i+1}^n b_{ij}x_j = 0$$

$$\forall r+1 \leq i \leq n, \quad \sum_{j=i}^n b_{ij}x_j = 0$$

* 此结论和相似三角化结合, 利用归纳计算也可证明 A 可对角化当且仅当存在无重根的多项式 f 使得 $f(A) = O$ 。

十 正交方阵

* 为了便于几何理解, 本章与下一章的所有讨论中, 我们均默认谈论的向量空间为 \mathbb{R}^n 或其子空间, 称为实向量空间。但事实上, 将两章中出现的所有转置替换为共轭转置 (由此对称阵变为了 Hermite 阵, 正交阵变为了酉方阵), 我们往往可以得到 \mathbb{C}^n 或其子空间 (称为复向量空间) 中类似的结论。

§10.1 习题解答

1. 习题 4.6-6

直接由 Schmidt 正交化算法计算可得

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{5}{2\sqrt{11}} & \frac{1}{\sqrt{66}} \\ -\frac{1}{2} & -\frac{1}{2\sqrt{11}} & \sqrt{\frac{2}{33}} \\ -\frac{1}{2} & \frac{3}{2\sqrt{11}} & \frac{5}{\sqrt{66}} \\ \frac{1}{2} & -\frac{3}{2\sqrt{11}} & \sqrt{\frac{6}{11}} \end{pmatrix}, \quad R = \begin{pmatrix} 2 & -\frac{7}{2} & \frac{19}{2} \\ 0 & \frac{3\sqrt{11}}{2} & -\frac{1}{2\sqrt{11}} \\ 0 & 0 & 4\sqrt{\frac{6}{11}} \end{pmatrix}$$

2. 求方程组 $Ax = b$ 的最小二乘解, 其中增广矩阵 (A, b) 分别为

$$\begin{pmatrix} 1 & 1 & 5 \\ 1 & -1 & 2 \\ 2 & -1 & 7 \end{pmatrix}, \quad \begin{pmatrix} 2 & -3 & 0 & 8 \\ 2 & -1 & 1 & 12 \\ 4 & -5 & 1 & 15 \\ 2 & 0 & 4 & 1 \end{pmatrix}$$

根据教材 4.6 节最小二乘的相关结论, 只需求解方程组 $A^T Ax = A^T b$ 即可, 两题中解都是唯一的, 直接计算得解分别为

$$\begin{pmatrix} \frac{55}{14} \\ \frac{9}{7} \end{pmatrix}, \quad \begin{pmatrix} \frac{1993}{182} \\ \frac{61}{13} \\ -\frac{68}{13} \end{pmatrix}$$

* 教材中虽然给出了 QR 分解计算 $A^T Ax = A^T b$ 解的办法, 但事实上日常计算中直接求解线性方程组往往比涉及正交的算法快很多, QR 分解更适合计算机使用。

3. 求 $(1, 1, 0, 0)$ 与 $(1, 1, 1, 2)$ 生成的空间 U 中使得 $\|u - (1, 2, 3, 4)\|$ 最小的 u 。

由于空间中的任何向量都可以写成 $\lambda(1, 1, 0, 0) + \mu(1, 1, 1, 2)$, 构造

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

问题即成为了方程组 $Ax = b$ 的最小二乘解问题, 求解 $A^T Ax = A^T b$ 得到唯一解

$$\lambda = -\frac{7}{10}, \quad \mu = \frac{11}{5}$$

再代入计算即得

$$u = \left(\frac{3}{2}, \frac{3}{2}, \frac{11}{5}, \frac{22}{5} \right)$$

4. 习题 5.7-1(2)

直接由教材中所给的特征向量正交化算法计算得到可取

$$T = \begin{pmatrix} \frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ \frac{1}{3} & 0 & \frac{2\sqrt{2}}{3} \\ \frac{1}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \end{pmatrix}$$

此时 $T^{-1}AT = \text{diag}(6, -3, -3)$ 。

* 这里 T 中特征值 6 对应的列唯一, 另两列可能有不同答案。

5. 习题 5.7-3

基本思路见本讲义 9.3.1 第一题。由于 $\lambda \in \mathbb{K}$, $A\alpha = \lambda\alpha$ 看成系数在 \mathbb{K} 中的齐次线性方程组, 且系数行列式为 0, 因此一定存在 \mathbb{K} 中的非零向量 α 作为解, 由此归纳过程可以进行。

6. 习题 5.7-7

若 $A\alpha = \lambda\alpha$, 两侧同取共轭转置得到 $\alpha^H A^H = \bar{\lambda}\alpha^H$ 。由条件 $A^H = A$, 于是 $\alpha^H A\alpha = \bar{\lambda}\alpha^H$ 。第一式同左乘 α^H , 第二式同右乘 α 得到

$$\lambda\alpha^H\alpha = \alpha^H A\alpha = \bar{\lambda}\alpha^H\alpha$$

由于 $\alpha^H\alpha$ 为其所有元素模长平方和, α 为非零向量可知 $\lambda = \bar{\lambda}$, 得证。

7. 习题 5.6-6

* 这题最好不要不设 $a_1 \neq 0$, 因为不同位置非零时 P 的形式存在区别。

设 $a_i \neq 0$, 特征方程

$$A^T A\alpha = \lambda\alpha$$

可以改写为

$$(A\alpha)A^T = \lambda\alpha$$

由此分类讨论: 当 $\lambda = 0$ 时, $A\alpha = 0$ 且 α 非零就能保证 α 为特征向量。根据解空间维数定理, $\text{rank } A = 1$ 可知 0 的特征子空间维数 $n-1$; 当 $\lambda \neq 0$ 时, 必须 α 为 A^T 的倍数, 且可验证 $\alpha = A^T$ 时

$$(AA^T)A^T = \lambda A^T$$

于是 $\lambda = AA^T$, 对应的特征子空间维数为 1。

综合以上两种情况, 特征子空间维数和为 n , 因此可对角化, 取每个特征子空间的一组基拼成矩阵即为 P 。特征值 AA^T 的特征子空间中取 $\alpha = A^T$ 即可, 特征值 0 的特征子空间中, 由 $a_i = 0$ 可取所有

$$\alpha_j = a_i e_j - a_j e_i, \quad j = 1, \dots, i-1, i+1, \dots, n$$

作为一组基 (这里 e_j 指第 j 个分量为 1, 其余为 0 的向量, 可直接验证它们都满足 $A\alpha_j = 0$ 且线性无关), 取矩阵

$$P = (\alpha, \alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n)$$

即为所求, 相似对角化结果为 $\text{diag}(AA^T, 0, \dots, 0)$ 。

8. 习题 5.7-9

设 A 的正交相似对角化为 $P^T \text{diag}(\lambda_1, \dots, \lambda_n)P$, 取

$$B = P^T \text{diag}(\sqrt[3]{\lambda_1}, \dots, \sqrt[3]{\lambda_n})P$$

可直接计算转置验证 B 实对称, 且 $B^3 = A$ 。

9. 补充题五-11

由条件可设 $B = P^{-1}AP$, 由 (可参考本讲义 7.3.2) $(XY)^* = Y^*X^*$ 得

$$B^* = P^*A^*(P^{-1})^*$$

由 P 可逆有

$$P^* = (\det P)P^{-1}$$

$$(P^{-1})^* = \det(P^{-1})(P^{-1})^{-1} = (\det P)^{-1}P$$

从而提出常数即得

$$B^* = P^{-1}A^*P$$

得证相似。

10. 习题 6.1-4(2)

与教材过程类似, 最终可取

$$z_1 = x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3, \quad z_2 = x_1, \quad z_3 = \frac{1}{2}x_2 - \frac{1}{2}x_3$$

得到

$$f(x_1, x_2, x_3) = z_1^2 - z_2^2 - z_3^2$$

* 结果形式可能会有不同, 但一定是一项正两项负。

11. 习题 6.1-5

利用正交相似对角化, 由相似不改变秩与对角阵秩为非零对角元个数可知秩为 r 恰好对应 r 个非零特征值, 于是设 $\text{rank } A = r$, 且 $A^T = A$, 存在正交阵 P 使得

$$A = P^T \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)P$$

其中 $\lambda_i \neq 0$ 。

由此有

$$A = \sum_{i=1}^r P^T \text{diag}(\lambda_i e_i)P$$

由相似不改变秩可知求和中每个秩都为 1, 得证。

12. 习题 6.2-5

由条件可设存在可逆矩阵 P 使得

$$P^T AP = \text{diag}(I_s, -I_s, O)$$

且 $s > 0$ 。

将 P 按列分块为 (p_1, \dots, p_n) , 并将右侧对角阵记为 D , 利用分块矩阵知识可以直接计算发现

$$p_i^T A p_j = d_{ij}$$

于是, 取 $\alpha_1 = p_1$ 即得 $\alpha_1^T A \alpha_1 = 1 > 0$, 取 $\alpha_2 = p_{s+1}$ 即得 $\alpha_2^T A \alpha_2 = -1 < 0$ 。

取 $\alpha_3 = p_1 + p_{s+1}$, 由于 $p_1^T A p_{s+1} = 0$ 即有

$$\alpha_3^T A \alpha_3 = p_1^T A p_1 + p_{s+1}^T A p_{s+1} = 1 - 1 = 0$$

* 不能取 $\alpha_3 = p_{2s+1}$, 因为可能有 $2s = n$ 。

§10.2 内积

10.2.1 保内积的线性映射

内积的引入源自一个非常自然的想法：我们希望刻画 \mathbb{R}^n 中的长度与角度。与大家在高中学过的 \mathbb{R}^3 中情况相近，只要定义了两个 \mathbb{R}^n 中列向量的内积（这里将一阶方阵 $x^T y$ 视为一个数）

$$(x, y) = x^T y$$

就能由勾股定理写出一个向量的长度（记为 $\|x\|$ ，也称为模长）为

$$\|x\| = \sqrt{(x, x)}$$

两个向量的夹角为（利用 4.3.2 例 10 的 Cauchy 不等式可说明其定义良好）

$$\arccos \frac{(x, y)}{\|x\| \|y\|}$$

由于几何并不是这门课的重点，我们省略过于细节的几何结论，回到对代数的关注。

对于任何实向量空间 $V \subset \mathbb{R}^n$ ，我们当然可以与 \mathbb{R}^n 完全相同定义其中两个向量的内积。不过，这就自然产生了一个问题：两个赋予内积的实向量空间（我们将其称为欧氏空间）何时是不可区分的？

* 这个问题的重要性在于，所有经典几何（如平面几何）基本都与长度和角度相关，因此，如果两个欧氏空间不可区分，我们就可以研究相同的经典几何。例如，所有与 \mathbb{R}^2 不可区分的欧氏空间上，都具有完全相同的平面几何结论。

首先，对两个实向量空间来说，它们不可区分（也即存在线性双射）当且仅当维数相同。由于内积是在线性结构上附加了一层结构，两个欧氏空间不可区分至少也需要维数相同。反之，从几何直观上来看，任何一个二维的实向量空间都应该是一个“平面”，于是理应不可区分，由此可以想象欧氏空间不可区分的条件仍然是维数相同。

为了给不可区分一个严谨的定义，仿照保持线性结构的线性映射，我们定义如下的保持线性与内积结构的正交映射：若线性映射 $f : U \rightarrow V$ 满足 $(u, v) = (f(u), f(v))$ （注意左右代表的是不同空间中的内积），则称其为正交映射。

若两个空间之间存在正交双射，它们即是不可区分的。

* 事实上更严谨的定义是“存在正交双射使得其逆也是正交双射”，不过任何正交双射的逆一定是正交双射，因此可以省略后一部分，证明留给读者思考。

根据之前的几何直观，我们希望证明，只要两个欧氏空间的维数相同，就一定存在正交的双射。我们设 $\dim U = \dim V = r$ ，为构造出这样的双射，必须研究内积的基本性质。从内积表达式可直接得出以下三条性质：

- 双线性性，即 $(\lambda x_1 + \mu x_2, y) = \lambda(x_1, y) + \mu(x_2, y)$ 且 $(x, \lambda y_1 + \mu y_2) = \lambda(x, y_1) + \mu(x, y_2)$ ，对两个分量都是线性映射，这在其作为线性空间之上的结构时是必要的，代表与线性结构的相容；
- 对称性，即 $(x, y) = (y, x)$ ；
- 正定性，即 $(x, x) \geq 0$ ，且等号成立当且仅当 $x = 0$ 。

正如有了线性性后只要确定一组基的像就能确定线性映射的像，有了双线性性后，只要确定一组基的内积就能通过坐标确定任何两个向量的内积。具体来说，对 U 的一组基 $\alpha_1, \dots, \alpha_r$ ，设

$$x = \sum_{i=1}^r \lambda_i \alpha_i, \quad y = \sum_{i=1}^r \mu_i \alpha_i$$

则有

$$(x, y) = \sum_{i=1}^r \lambda_i (\alpha_i, y) = \sum_{i=1}^r \sum_{j=1}^r \lambda_i \mu_j (\alpha_i, \alpha_j)$$

对于欧氏空间 U 的一组基 $\alpha_1, \dots, \alpha_r$, 我们定义满足 $g_{ij} = (\alpha_i, \alpha_j)$ 的矩阵 G 为它们的内积矩阵。不是所有方阵都可以成为内积矩阵: 从内积的对称性可以得到 $g_{ij} = g_{ji}$, 而从正定性与基不包含零向量可知 $g_{ii} > 0$ 。

非常自然地, 我们想问, 在允许任意选取基的情况下, 内积矩阵能简化到何种程度。由对称性与对角元大于 0 的要求, 最理想的情况是内积矩阵即为 I , 此时的一组基称为标准正交基。

* 标准在此处指 $(\alpha_i, \alpha_i) = 1$, 而正交指 $i \neq j$ 时 $(\alpha_i, \alpha_j) = 0$ 。一般地, 我们将一个模长为 1 的向量称为标准的, 两个内积为 0 的向量称为正交的。

一个非常令人开心的事实是, 任何欧氏空间都存在标准正交基, 我们将在之后的两小节以不同方式研究这个结论。在本小节的最后, 我们用此结论证明同一维数的欧氏空间不可区分, 由此, 可将所有 r 维欧氏空间视为 \mathbb{R}^r , 只需在 \mathbb{R}^r 中讨论即可。容易给出 \mathbb{R}^r 的一组标准正交基: e_1, \dots, e_r 就满足这样的性质。

设 U 的一组标准正交基为 $\alpha_1, \dots, \alpha_r$, V 的一组标准正交基为 β_1, \dots, β_r , 构造线性映射 f 满足

$$f(\alpha_i) = \beta_i, \quad i = 1, \dots, r$$

由于一组基的像已经确定, 这个线性映射是唯一确定的, 且由于其将一组基映射到了一组基, 其确实为线性双射。下面说明其保持内积。对任何

$$x = \sum_{i=1}^r \lambda_i \alpha_i, \quad y = \sum_{i=1}^r \mu_i \alpha_i$$

有

$$(x, y) = \sum_{i=1}^r \lambda_i (\alpha_i, y) = \sum_{i=1}^r \sum_{j=1}^r \lambda_i \mu_j (\alpha_i, \alpha_j) = \sum_{i=1}^r \lambda_i \mu_i$$

而利用 f 线性性有

$$(f(x), f(y)) = \sum_{i=1}^r \lambda_i (f(\alpha_i), f(y)) = \sum_{i=1}^r \sum_{j=1}^r \lambda_i \mu_j (f(\alpha_i), f(\alpha_j)) = \sum_{i=1}^r \sum_{j=1}^r \lambda_i \mu_j (\beta_i, \beta_j) = \sum_{i=1}^r \lambda_i \mu_i$$

由此就证明了 f 是符合要求的映射。

10.2.2 正交化

现在, 我们试图给任何 r 维实向量空间 U 找到一组标准正交基 f_1, \dots, f_r 。

对于一般的 r 维实向量空间, 我们能做的只有找到其的一组基 $\alpha_1, \dots, \alpha_r$ 。如果能将这组基进行线性组合, 得到一组标准正交基, 我们就得到了证明。

在教材中, 给出了如下的正交化过程 (一般称为 Schmidt 正交化): 对 $i = 1, \dots, r$, 执行

$$\beta_i = \alpha_i - \sum_{j=1}^{i-1} \frac{(\beta_j, \alpha_i)}{(\beta_j, \beta_j)} \beta_j$$

这样就得到了相互正交的一组基 β_1, \dots, β_r , 从而再取

$$f_i = \frac{1}{\|\beta_i\|} \beta_i$$

即可得到结果 (可验证相互正交的向量乘倍数后仍然相互正交, 再直接计算验证 $(f_i, f_i) = 1$)。

由于前 $r-1$ 步即为对 α_1 到 α_{r-1} 的操作, 这个结论可以通过归纳证明: 一个向量时结论自然成立, 若 $\beta_1, \dots, \beta_{r-1}$ 已经为与 $\alpha_1, \dots, \alpha_{r-1}$ 等价的正交向量组, 直接由定义与归纳假设计算可知 $i < r$ 时

$(\beta_r, \beta_i) = 0$, 且由于 β_r 可由 $\alpha_r, \beta_1, \dots, \beta_{r-1}$ 表出、 α_r 可由 $\beta_r, \beta_1, \dots, \beta_{r-1}$ 表出, 再利用 $\beta_1, \dots, \beta_{r-1}$ 可由 $\alpha_1, \dots, \alpha_{r-1}$ 表出即得到等价性 (更具体的计算细节可以参考教材, 注意作除法时都要验证分母非零)。

Schmidt 正交化的计算过程本身是相对复杂的, 为了让其看上去稍微自然一些, 我们必须解释它几何角度的来源。

考虑两个不共线的向量 (由空间几何知识, 两个向量一定共面) α_1, α_2 , 为了从 α_2 中得到与 α_1 正交的向量, 在物理中常用的方式是利用平行四边形法则进行分解, 从而得到平行分量与垂直分量。

对更多个向量 $\alpha_1, \dots, \alpha_r$, 我们可以完全类似操作: 由于它们的线性无关性, $\alpha_1, \dots, \alpha_{r-1}$ 应生成一个 $r-1$ 维的空间, 而 α_r 不在其中。由此, 可以将 α_r 分解为与其平行 (也即在其中)、垂直 (也即正交) 的分量, 且由于 α_r 不在其中, 垂直分量一定非零。

将上方的介绍用数学语言表达, 也即作分解

$$\alpha_r = \lambda_1 \alpha_1 + \dots + \lambda_{r-1} \alpha_{r-1} + \beta_r$$

且 $(\alpha_1, \beta_r) = \dots = (\alpha_{r-1}, \beta_r) = 0$ 。这样, 前 $r-1$ 项就构成了平行分量, 最后一项就构成了垂直分量。

* 这里用到了习题 4.6-2 结论: 若 β 与 $\alpha_1, \dots, \alpha_{r-1}$ 正交, 则其与它们的任意线性组合正交。

为了求解 λ_i , 我们将

$$\beta_r = \alpha_r - \lambda_1 \alpha_1 - \dots - \lambda_{r-1} \alpha_{r-1}$$

代入 $(\beta_r, \alpha_j) = 0$, 利用内积双线性性得到

$$\forall j = 1, \dots, r-1, \quad (\alpha_r, \alpha_j) = \sum_{i=1}^{r-1} (\alpha_i, \alpha_j) \lambda_i$$

注意到未知数与方程个数均为 $r-1$, 由 Cramer 法则, 此方程组存在唯一解当且仅当系数矩阵可逆, 而此处系数矩阵即为 $\alpha_1, \dots, \alpha_{r-1}$ 对应的内积矩阵, 其一般情况的可逆性将在下一部分说明。

我们实际遇到的情况是, 由于 Schmidt 正交化的过程, $\alpha_1, \dots, \alpha_{r-1}$ 在这一步已经等价为了相互正交的 $\beta_1, \dots, \beta_{r-1}$, 而此时由正交, 方程组即成为

$$\forall j = 1, \dots, r-1, \quad (\alpha_r, \beta_j) = (\beta_j, \beta_j) \lambda_j$$

从此式中得到 λ 表达式后代回

$$\beta_r = \alpha_r - \lambda_1 \beta_1 - \dots - \lambda_{r-1} \beta_{r-1}$$

就是 Schmidt 正交化的表达式。

最后, Schmidt 正交化过程有一个非常重要的特点: α_i 可由 f_1, \dots, f_i 表出, 无需之后的 f_{i+1}, \dots, f_r 。由

$$\beta_i = \alpha_i - \sum_{j=1}^{i-1} \frac{(\beta_j, \alpha_i)}{(\beta_j, \beta_j)} \beta_j$$

可得

$$\alpha_i = \beta_i + \sum_{j=1}^{i-1} \frac{(\beta_j, \alpha_i)}{(\beta_j, \beta_j)} \beta_j$$

再利用 f_j 定义代入计算可知

$$\alpha_i = \|\beta_i\| f_i + \sum_{j=1}^{i-1} (f_j, \alpha_i) f_j$$

记 $r_{ii} = \|\beta_i\|$ (由此根据 β_i 构成一组基可知其必然为正)、 r_{ij} 在 $i < j$ 时为 (f_i, α_j) , 则有

$$\alpha_i = \sum_{j=1}^i r_{ji} f_j$$

将所有 α_i 拼成矩阵 A 、所有 f_i 拼成矩阵 Q 。由于 r_{ij} 只在 $i \geq j$ 时有定义，我们定义 $i < j$ 时其为 0 使其成为 r 阶上三角方阵，则利用分块矩阵计算有

$$A = QR$$

由于 α_i 构成一组基，上述的分解准确来说是：任何一个列满秩矩阵都可以分解成 QR ，其中 Q 的列均为单位向量且相互正交，而 R 为对角元均正的（可逆）上三角阵。这称为列满秩矩阵的 **QR 分解**。

10.2.3 合同

上述的算法虽然能给出一个漂亮的结论，但过程实际上是不具有对称性的，而是在进行逐个的操作。

所谓对称性是指，若我们已知 U 的一组基 $\alpha_1, \dots, \alpha_r$ ，它们的内积矩阵为 G ，希望能直接构造出另一组基 β_1, \dots, β_r 使得其内积矩阵为 I 。

设两组基的过渡矩阵（见本讲义 8.2.2）为 P ，也即

$$\beta_j = \sum_{i=1}^r p_{ij} \alpha_i$$

直接计算有

$$(\beta_s, \beta_t) = \sum_{i=1}^r p_{is} (\alpha_i, \beta_t) = \sum_{i=1}^r \sum_{j=1}^r p_{is} p_{jt} (\alpha_i, \alpha_j) = \sum_{i=1}^r \sum_{j=1}^r p_{is} g_{ij} p_{jt}$$

由于左侧拼成矩阵为 I ，而 p_{is} 为 P^T 的第 i 行第 s 列元素，将上述求和写成矩阵乘法即为

$$I = P^T G P$$

定义 A 与 B 合同当且仅当存在可逆阵 P 使得 $P^T A P = B$ ，可验证其为等价关系：

1. I 可逆，因此 A 合同于 A ；
2. 直接计算验证可知 P^T 的逆为 $(P^{-1})^T$ ，由此 A 与 B 合同时 $B = P^{-T} A P^{-1}$ 与 A 合同；
3. 由 $(PQ)^T = Q^T P^T$ 可得 A 与 B 合同、 B 与 C 合同时 A 与 C 合同。

* 合同也称为相合。

由此，任何欧氏空间存在标准正交基可以等价写为，任何内积矩阵与 I 合同。

关于合同的更多性质将在下一次习题课讨论，我们这里解决最后一个问题：到底怎样的矩阵可以成为内积矩阵？

根据定义，一组基 $\alpha_1, \dots, \alpha_r$ 对应的内积矩阵满足（这里 $(\alpha_i)_k$ 代表 α_i 第 k 个分量）

$$g_{ij} = \alpha_i^T \alpha_j = \sum_{k=1}^n (\alpha_i)_k (\alpha_j)_k$$

由此，将 $\alpha_1, \dots, \alpha_r$ 排成一行，设对应矩阵为 A ，有 $a_{ij} = (\alpha_j)_i$ ，于是写出矩阵乘法

$$G = A^T A$$

由于 α_i 构成一组基， A 应当列满秩，反之，只要 A 列满秩，其列均为某向量空间的一组基。因此，有 G 为内积矩阵当且仅当存在列满秩矩阵 A 使得 $G = A^T A$ 。

* 事实上，内积矩阵与我们下次习题课要介绍的正定方阵完全等价。

§10.3 正交相似

10.3.1 正交方阵

回到保内积的线性映射的问题。既然我们已经证明了任何 n 维欧氏空间都与 \mathbb{R}^n 等价, 只需要确定 $\mathbb{R}^n \rightarrow \mathbb{R}^m$ 的正交映射, 即可得到一般的结果。

由本讲义 8.2.1, 对 $\mathbb{R}^n \rightarrow \mathbb{R}^m$ 的线性映射, 存在矩阵 $A_{m \times n}$ 使得其为 $x \rightarrow Ax$, 于是保内积可由矩阵写为

$$(x, y) = (Ax, Ay)$$

也即

$$\forall x, y \in \mathbb{R}^n, \quad x^T y = x^T A^T A y$$

类似本讲义 8.2.1 对唯一性的证明, 上式对任何 y 成立可得到对任何 x 有 $x^T = x^T A^T A$, 同取转置即 $A^T A x = x$, 再由其对任何 x 成立可知 $A^T A = I_n$ 。

于是, $\mathbb{R}^n \rightarrow \mathbb{R}^m$ 的正交映射等价于矩阵 A 使得 $A^T A = I_n$ 。利用 $\text{rank}(A^T A) = \text{rank } A$, 这可以推出 $\text{rank } A = n$, 于是至少需要 $m \geq n$ 方才存在正交映射。

由于正交双射是更为重要的, 我们定义满足 $A^T A = I$ 的方阵 A 为正交 (方) 阵。利用逆的唯一性, 这可以直接推出 $A^T = A^{-1}$, 从而 $AA^T = I$ 。

另一个重要的观察时, 由上节对内积矩阵的讨论, $A^T A = I$ 恰好与其列向量组标准正交等价。因此, 正交阵也可以理解为 \mathbb{R}^n 的一组标准正交基拼成的方阵。

* 由于方阵行列的对偶性, 考虑转置或直接验证可得正交阵亦等价于 \mathbb{R}^n 的一组标准正交基的转置 (也即行向量标准正交) 拼成的方阵。

除了上面的几个等价定义外, 我们还需要一些基本性质:

- 所有置换阵 (见本讲义 9.3.2) 为正交阵, 特别地, I 为正交阵。

在本讲义 9.3.2 第一题中已证明置换阵的转置即为其逆。

- 若 A 为正交阵, 则 A^{-1} 为正交阵。

由于 $(A^{-1})^T = (A^T)^{-1} = (A^{-1})^{-1}$ 即得证。

- 若 A, B 为同阶正交阵, 则 AB 为正交阵。

利用 $(AB)^T AB = B^T A^T AB = B^T B = I$ 得证。

- A 的特征值模长均为 1。

若 $A\alpha = \lambda\alpha$, 作共轭转置可得

$$\alpha^H A^T = \bar{\lambda}\alpha^H$$

相乘得到

$$\alpha^H A^T A \alpha = \bar{\lambda}\lambda\alpha^H \alpha$$

由正交性与复数基本性质

$$\alpha^H \alpha = |\lambda|^2 \alpha^H \alpha$$

对非零复向量 α , 可发现 $\alpha^H \alpha$ 为每个位置模长 (指作为复数的模长) 平方和, 因此非零, 从而 $|\lambda|^2 = 1$, 于是 $|\lambda| = 1$, 得证。

* 共轭转置技巧在教材与作业题中也都已经出现了。由于不确定 λ 与 α 为实, 不能直接作转置, 而应转置后取共轭。由于 A 为实方阵, 其共轭转置与转置相等。

5. 彼此正交的非零向量组线性无关。

设 $\alpha_1, \dots, \alpha_n$ 满足题述性质。若否, 有

$$\alpha_i = \sum_{j \neq i} \lambda_j \alpha_j$$

利用习题 4.6-2, α_i 与 α_i 正交, 因此 $(\alpha_i, \alpha_i) = 0$, 与其非零矛盾。

6. 给定 \mathbb{R}^n 中彼此正交的单位 (即模长为 1) 向量 $\alpha_1, \dots, \alpha_r$, 存在正交阵 A 以它们为前 r 列。

由上个结论可知 $\alpha_1, \dots, \alpha_r$ 线性无关, 从而可扩充为 \mathbb{R}^n 一组基。对这组基执行正交化过程 (按 α_i , i 从小到大的顺序), 可发现前 r 个向量不变, 即得到了包含 $\alpha_1, \dots, \alpha_r$ 的一组标准正交基。将它们拼成 A 的列即得到结果。

有了正交矩阵后, 也就自然有了正交相似的定义: 若方阵 A, B 满足存在正交阵 P 使得 $A = P^{-1}BP$ (也可写为 $A = P^TBP$), 则称它们正交相似。

* 请读者自行验证, 利用前三个性质, 正交相似的确构成等价关系, 于是谈论正交相似是有意义的。

正交相似拥有一个相似不具有的性质, 即, 若 A 与 B 正交相似, 可以推出 $f(A, A^T)$ 与 $f(B, B^T)$ 也正交相似, 这里 f 为某个二元多项式, 且未必满足交换性。

例如, 取 $f(x, y) = xyx$, 利用 P 的正交性有

$$f(A, A^T) = AA^T A = P^T BP (P^T BP)^T P^T BP = P^T B P P^T B^T P P^T BP = P^T B B^T BP = P^T f(B, B^T) P$$

但若只知道 A, B 相似, $AA^T A$ 与 $BB^T B$ 是未必相似的, 这就是正交相似的额外作用。

10.3.2 实对称阵

很遗憾, 对于一般的方阵, 正交相似是一个过于复杂的操作, 也并没有正交相似标准形的结论。不过, 对于实对称阵, 结论将变得非常优美。

* 事实上对满足 $A^T A = AA^T$ 的实方阵 (称为规范方阵) 都有好看的正交相似标准形结果。

我们将满足 $A^T = A$ 的方阵称为对称阵, 而若其限制在实方阵范围内, 则称为实对称阵。由于当 A 对称时

$$(P^T AP)^T = P^T A^T P = P^T AP$$

实对称阵在正交相似后一定得到的还是实对称阵, 因此可以在实对称阵的范围内谈论正交相似。

直接给出三个结论:

1. 实对称阵的特征值都是实数;
2. 实对称阵一定可以正交相似对角化;
3. 两个实对角阵正交相似当且仅当它们相差排列。

有了这两个结论, 实对称矩阵在正交相似下的情况事实上和相似完全一致, 也即, 两个实对称阵正交相似当且仅当它们相似 (而由于它们都可对角化, 这又等价于特征值可一一对应)。下面我们来证明这三个结论:

1. 若 $A\alpha = \lambda\alpha$, 作共轭转置可得 (注意 A 为实方阵, 共轭转置即为转置)

$$\alpha^H A^T = \bar{\lambda} \alpha^H$$

第一式同左乘 α^H , 第二式同右乘 α , 利用 $A^T = A$ 有

$$\bar{\lambda} \alpha^H \alpha = \alpha^H A \alpha = \lambda \alpha^H \alpha$$

类似上节可知 $\alpha^H \alpha$ 非零, 于是 $\lambda = \bar{\lambda}$, 即其为实数。

2. 证明过程仿照本讲义 9.3.1 的相似三角化。

对阶数归纳, 一阶时直接成立, 若对 $n-1$ 阶成立, n 阶时选出 A 的某实特征值 λ 。由于 $Ax = \lambda x$ 为实线性方程组, 且系数矩阵秩至多 $n-1$, 其有实数解 x , 取 $\alpha = \frac{1}{\|x\|}x$, 可发现其为 λ 的实特征向量, 且模长为 1。

由上一小节结论, 存在以 α 为第一列的正交阵, 记为 P , 计算可发现 AP 第一列应为 $\lambda\alpha$, 进一步计算有 (这里 x, X 为未知向量/矩阵)

$$AP = P \begin{pmatrix} \lambda & x \\ 0 & X \end{pmatrix}$$

由 P 正交性, 这可以推出

$$P^T AP = \begin{pmatrix} \lambda & x \\ 0 & X \end{pmatrix}$$

而根据之前计算, $P^T AP$ 为实对称阵, 由此 $x = 0$, 且 X 为实对称阵。

利用归纳假设, 设 $X = Q^T D Q$, D 为对角阵、 Q 为正交阵, 则计算可发现

$$A = P \begin{pmatrix} 1 & 0 \\ 0 & Q^T \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix} P^T$$

记

$$R = \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix} P^T$$

则利用分块矩阵可计算知 R 为正交阵, 且 RAR^T 为对角阵。

3. 若两个实对角阵正交相似, 它们必须相似, 由此只能相差排列。反之, 若它们相差排列, 利用本讲义 9.3.2 第二题的结论即可得到它们正交相似, 从而得证。

实对称阵的正交相似对角化结果也可以称为正交相似标准形, 可以解决绝大部分的实对称阵相关问题。最后, 我们再给出一个转置/共轭转置技巧的例子, 证明实对称阵不同特征值对应的实特征向量正交。

假设 A 实对称, λ, μ 为其特征值, 已证明它们为实数。若 $A\alpha = \lambda\alpha$ 、 $A\beta = \mu\beta$, 且 α, β 为实向量, 第一式同时左乘 β^T 有

$$\beta^T A \alpha = \lambda \beta^T \alpha$$

第二式转置后右乘 α 有

$$\beta^T A^T \alpha = \beta^T A \alpha = \mu \beta^T \alpha$$

由于 $\lambda \neq \mu$, 即有 $\beta^T \alpha = 0$ 。

* 实对称阵正交相似另一个较重要的结论为教材 6.1.2 例 14, 证明需要熟悉本讲义 8.3.3 出现过的与对角阵可交换的矩阵形式。

§10.4 正交相抵

* 这部分在我们上课范围内, 但并不在考试范围, 不过仍然建议大家作为正交阵知识的应用来学习。奇异值分解的结论事实上非常重要。

10.4.1 反射变换

顾名思义, $A_{m \times n}$ 与 $B_{m \times n}$ 正交相抵当且仅当存在正交阵 $P_{m \times m}$ 与 $Q_{n \times n}$ 使得

$$A = PBQ$$

与正交相似类似，可以验证正交相抵也是一种等价关系。此外，正交相抵下的标准形易于求得，且有非常重要的意义。

为了寻找任意矩阵的正交相抵标准形，我们回顾相抵标准形的证明过程：通过一系列的初等方阵不断左右乘，直到化为了简单的形式。

若想类似得到正交相抵标准形，第二类初等方阵仍然可以使用，这是因为它们都为置换阵，于是都正交阵。这意味着，计算过程中，我们可以任意交换行或列。但是，仅有交换行列显然不足以得出标准形。为了能进行消元，我们需要介绍一类初等正交阵，称为 **Householder 阵**。

对任何单位向量 u ，定义其对应的 Householder 阵为

$$H = I - 2uu^T$$

乍一看，这个形式似乎十分奇怪。但事实上，这就是我们在平面或空间几何中已经非常熟悉的反射变换。

设 u 扩充为的一组标准正交基（见上节的性质）为 u, f_1, \dots, f_{n-1} ，且

$$x = \lambda u + \lambda_1 f_1 + \dots + \lambda_{n-1} f_{n-1}$$

则利用标准正交性可以直接计算发现 $\lambda = (x, u) = u^T x$ 。而

$$Hx = x - 2uu^T x = x - 2\lambda u = -\lambda u + \lambda_1 f_1 + \dots + \lambda_{n-1} f_{n-1}$$

因此， H 恰好对应 u 方向分量加负号，垂直分量不变的反射变换，也即关于与 u 垂直平面的反射。

我们对其的正交性进行一个简单的验证，利用 u 为单位向量， $u^T u = 1$ ，计算可得

$$H^T H = (I - 2uu^T)(I - 2uu^T) = I - 4uu^T + 2uu^T uu^T = I - 2uu^T = H$$

从上方过程事实上还可以看出， H 实对称，且 $H^2 = I$ 。

就像初等变换阵具有的变换性质，Householder 阵存在一个非常重要的操作：对任何 n 维向量 x, y 满足 $\|x\| = \|y\|$ 且 $x \neq y$ ，存在 Householder 阵 H 使得 $Hx = y$ 。

由于需要在 x, y 连线中垂面进行反射，其法向量即 $y - x$ ，从而利用反射的想法可直接构造 $u = \frac{y-x}{\|y-x\|}$ ，其对应的

$$H = I - 2 \frac{(y-x)(y-x)^T}{(y-x)^T(y-x)}$$

计算得到

$$Hx = x - 2 \frac{(y-x)^T x}{(y-x)^T(y-x)} (y-x) = -2 \frac{(y-x)^T x}{(y-x)^T(y-x)} y + \left(1 + 2 \frac{(y-x)^T x}{(y-x)^T(y-x)}\right) x$$

由 $x^T x = y^T y$ 有 $2(y-x)^T x = 2y^T x - 2x^T x = 2y^T x - x^T x - y^T y = -(y-x)^T(y-x)$ ，从而得证。

我们对奇异值分解的证明将大量依赖此操作进行构建。这里给出一个更好用的特殊化版本：对任何 n 维向量 x ，存在正交阵 H 使得 $Hx = \|x\|e_1$ 。

当 $x \neq \|x\|e_1$ 时，由于两者模长相等，可取 Householder 阵符合要求，否则直接取 I 即可。

* 事实上，利用上三角正交阵只能是每个对角元 ± 1 的对角阵，归纳可证任何一个 n 阶正交阵都可以分解为至多 n 个 Householder 阵的乘积，这与可逆阵分解为初等方阵乘积是相似的，这里省略详细证明，感兴趣的读者可以自行搜索进阶知识，或仿照下一小节奇异值分解的证明进行操作。

10.4.2 奇异值分解

至此，我们已经完成了准备，下面就将对正交相抵标准形进行探究。先给出结论：对任何实矩阵 $A_{m \times n}$ ，存在正交阵 $P_{m \times m}$ 、 $Q_{n \times n}$ 使得

$$A = P \begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix} Q$$

这里 $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, 且 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, 它们称为 A 的奇异值, 由 A 唯一确定。

* 事实上, 利用乘可逆阵不改变秩可知 $r = \text{rank } A$ 。下面记 $\text{rank } A = r$ 。

我们先证明奇异值分解的存在性, 再证明唯一性。证明过程分四步, 事实上与相抵标准形存在类似 (注意可以任意交换行列以简化证明过程), 但需要一个细节性操作, 即第二部分证明中的引理:

1. 更强的阶梯形矩阵

以下不妨设 $m \geq n$, 否则取转置即可。

设 A 的第一列为 α_1 , 构造正交阵 P_1 使得 $P_1\alpha_1 = \|\alpha_1\|e_1$, 则可发现此时的 P_1A 第一列为 $\|\alpha_1\|e_1$ 。考虑现在的 A 去掉第一列后剩下的矩阵, 设此时其第一行为 β_1^T , 构造正交阵 Q_1 使得 $Q_1\beta_1 = \|\beta_1\|e_1$, 于是, 计算有此时 (这里 * 表示未知元素)

$$P_1A \begin{pmatrix} 1 & 0 \\ 0 & Q_1 \end{pmatrix} = \begin{pmatrix} \|\alpha_1\| & \|\beta_1\| & 0 & \dots & 0 \\ 0 & * & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & * & * & \dots & * \end{pmatrix}$$

再将 A 第二列的其余元素收缩到 a_{22} 、将 A 第二行的其余元素收缩到 a_{23} , 不断重复, 最终可将 A 变换为 (这里利用了 $m \geq n$ 的假设, 未写出的元素为 0)

$$B = \begin{pmatrix} b_{11} & b_{12} & & & \\ & \ddots & \ddots & & \\ & & b_{n-1,n-1} & b_{n-1,n} & \\ & & & b_{nn} & \\ 0 & 0 & \dots & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & & 0 \end{pmatrix}$$

* 这样只有对角线和其上的斜线可能非零的矩阵称为 **Hessenberg** 阵。

2. 二阶引理

我们希望证明, 对给定的 a, b, c , 存在二阶正交阵 A, B 使得

$$A \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} B = \begin{pmatrix} d & 0 \\ 0 & e \end{pmatrix}$$

且 $d \geq 0, e \geq 0$ 。

由于乘可逆不改变秩, d, e 全不为 0, 又由于 $\text{diag}(\pm 1, \pm 1)$ 为正交阵, d, e 的正负可以任意指定, 只需正交相抵对角化即可。计算可发现

$$\begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

为二阶正交阵, 设 A 对应 θ , B 对应 φ , 直接计算可得要求等价于

$$a \cos \theta \sin \varphi + b \sin \theta \sin \varphi + c \sin \theta \cos \varphi = 0$$

$$-a \sin \theta \cos \varphi + b \cos \theta \cos \varphi - c \cos \theta \sin \varphi = 0$$

两式相加、相减, 设 $\alpha = \theta + \varphi, \beta = \theta - \varphi$ 有

$$-a \sin \beta + b \cos \beta + c \sin \beta = 0$$

$$a \sin \alpha + b \cos \alpha + c \sin \alpha = 0$$

无论对何种 a, b, c , 都可以找到符合比例的 α, β , 进而得到 θ, φ , 从而得证。

3. 正交相抵标准形

引理中计算了将二阶下三角阵正交相抵对角化的方式, 同理, 二阶上三角阵也可正交相抵对角化。将它们看成行变换, 也即, 给定两行、两列, 可以对它们进行初等行列变换使得交叉处的三角阵成为对角元均正的对角阵。

类似二阶引理进行操作, 通过左右复合旋转进行对角化 (这里的操作细节比较复杂, 大致思路是, 计算可发现 $B^T B$ 只有 $|i - j| \leq 1$ 的位置可能非零, 由此可利用 $n - 1$ 个对两行/列的旋转构造正交阵 Q 使得 $Q^T B^T B Q$ 为对角阵, 此时左乘合适的旋转 P 即使得 $P B Q$ 完成了对角化), 设此时的矩阵为 C , 由引理 C 只有 c_{11}, \dots, c_{nn} 可能非零, 且它们均非负。

最后, 由于正交相抵也是相抵, 不改变秩, c_{11} 到 c_{nn} 中非零元素恰好为 r 个, 将它们按照大小顺序交换到前 r 个对角元即可 (具体来说, 通过交换 i, j 两行、交换 i, j 两列, 即可实现 i, j 两个对角元的交换)。

4. 唯一性

由于

$$A^T A = Q^T \begin{pmatrix} \Sigma^2 & O \\ O & O \end{pmatrix} Q$$

所有 r 个奇异值的平方与 $n - r$ 个 0 构成 $A^T A$ 的全部特征值。由于 A 确定时 $A^T A$ 所有特征值是确定的, 而不可能有两个不同非负实数平方是同一个数, 奇异值可由 A 唯一确定。

最后, 我们给出一些简单的性质与算法:

- 奇异值最常用的算法是计算 $A^T A$ 或 AA^T 的非零特征值平方根。这个算法的正确性与上方第四部分可完全相同证明。
- 利用 $A^T A$ 的正交相似对角化, 奇异值分解事实上有更简单的证明方式。具体来说, 由 $\text{rank}(A^T A) = \text{rank } A$ 与 $A^T A$ 可对角化, 其零特征值必然有 $n - r$ 个。设 Q 是使得

$$Q(A^T A)Q^T = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$$

的正交阵, 并记 AQ^T 的每列为 α_i , 计算可发现 $\alpha_1, \dots, \alpha_r$ 已经正交, 且均非零。将它们单位化后扩充成标准正交基即得到了 P 。

* 这里“扩充成标准正交基”指先将 $\alpha_1, \dots, \alpha_r$ 扩充为一组基 $\alpha_1, \dots, \alpha_n$, 再进行 Schmidt 正交化, 之后作为列向量拼成矩阵 P 。

* 此证明方法是奇异值分解的常用计算方法, 而上方的过程则更接近计算机计算奇异值分解的过程。若 $A^T A$ 的阶数比 AA^T 的阶数更大, 也可以令 $B = A^T$, 计算 B 的奇异值分解 (这样涉及到需要正交相似对角化的矩阵阶数更低), 然后作转置得到 A 的。

- 奇异值分解可以处理一些转置相关的问题, 例如由它直接计算就可以得到 $\text{rank}(AA^T A) = \text{rank } A$ 。
- 由于特征值关于矩阵是连续的, 奇异值关于矩阵也是连续的, 因此其比秩更适合计算机操作。
- 奇异值可以看作秩的推广: 若一个矩阵可逆, 其最大奇异值除以最小奇异值可以刻画它有“多接近不可逆阵”, 而不可逆阵又称为奇异阵, 这就是其得名的原因。
- 也有将奇异值分解的标准形式写为 $A = P \text{diag}(\Sigma, O)Q^T$ 的, 一般来说右侧写 Q 或 Q^T 均可, 但要标注清楚。

10.4.3 矩阵等价关系再探

到此处，我们已经学过了五种矩阵之间的等价关系：相抵、合同、相似、正交相抵、正交相似。若等价过程与某个操作“可交换”，我们就可以称作等价关系保此操作。例如，由于

$$PAQ + PBQ = P(A + B)Q, \quad P(\lambda A)Q = \lambda(PAQ)$$

可知相抵保加法与数乘（即可称为保线性组合，意为将矩阵也看作向量）。

而由

$$P^{-1}APP^{-1}BP = P^{-1}(AB)P$$

可知相似保乘法。

综合之前进行过的分析可以发现：

- 相抵保线性组合；
- 合同保线性组合与转置；
- 相似保线性组合与乘法、求逆；
- 正交相似保线性组合与乘法、求逆、转置。

由于我们考虑的均为线性结构，考虑的等价关系自然都是保线性组合的。而合同、相似、正交相似则是在一步步增加条件，从而保住原矩阵更多的性质，可进行更多的运算。

对于正交相抵，虽然表面上其只保线性组合，但其与转置乘积即可得到正交相似的形式。由此，在诸如 Moore-Penrose 广义逆等问题中，利用它可以得到非常简洁的结果。此外，由于其可以定义在一般的矩阵上，可以看作对方阵特征值的某种推广，至少可以用于刻画线性映射前后的长度变化。

十一 补充：正定方阵

§11.1 合同下的标准化

在本讲义 10.2.3 中，我们已经引入了实方阵合同的定义。正如之前针对每一个等价关系研究的，我们希望能够解决合同下的等价类问题。

11.1.1 惯性指数

先考虑实对称方阵在合同下的标准形——直接计算可验证实对称方阵合同后仍为实对称方阵，因此这样的讨论是有意义的。在上次习题课，我们已经得到了任何实对称方阵的正交相似对角化结果，也即存在正交阵 P 使得 $A = P^T DP$ ，其中 $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ 为对角阵。

由于对角元可按任何顺序排列，我们不妨设前 a 个对角元为正、随后 b 个对角元为负，剩余 $n - a - b$ 个为 0（从而 $\text{rank } A = a + b$ ），设

$$Q = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_a}, \sqrt{-\lambda_{a+1}}, \dots, \sqrt{-\lambda_{a+b}}, 1, \dots, 1)$$

则由于对角阵的转置仍为自身，直接计算可得到 $D = Q^T \text{diag}(I_a, -I_b, O)Q$ ，从而有

$$A = (QP)^T \text{diag}(I_a, -I_b, O)(QP)$$

由定义可知 Q 的对角元都非零，因此可逆，这就导出了 A 合同于 $\text{diag}(I_a, -I_b, O)$ 。

为了确认这为 A 在合同下的标准形，我们还需要说明不同的 a 与 b 对应的 $\text{diag}(I_a, -I_b, O)$ 无法相互合同。

若 n 阶方阵 $\text{diag}(I_a, -I_b, O)$ 与 $\text{diag}(I_c, -I_d, O)$ 合同, 由于秩相等, 至少可以得到 $a + b = c + d$, 于是只需证明 $a = c$ 。若否, 不妨设 $a < c$, 取矩阵 P 使得

$$P^T \text{diag}(I_a, -I_b, O)P = \text{diag}(I_c, -I_d, O)$$

为了推出矛盾, 我们需要寻找一个合适的分块。设

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}$$

其中 $P_{11} \in \mathbb{R}^{a \times c}$ 、 $P_{22} \in \mathbb{R}^{b \times d}$ (由此左侧作分块乘法的结果恰好对应右侧的分块), 考虑左上角可计算得到

$$I_c = P_{11}^T P_{11} - P_{21}^T P_{21}$$

由于 $a < c$, 利用解空间维数定理可知 $P_{11}x = 0$ 存非零解, 记为 z , 上式两边同时左乘 z^T 、右乘 z 得到

$$z^T z = -z^T P_{21}^T P_{21} z$$

但是, 左侧为 z 所有元素平方和, 为正数, 右侧为 $P_{21}z$ 所有元素平方和的相反数, 因此非正, 矛盾。

* 教材中, 任何一个合同的对角阵都称为标准形, 而 $\text{diag}(I_a, -I_b, O)$ 称为规范形。为了避免歧义, 我们不再引入教材中标准形的定义, 并同样称此形式为规范形。

在这部分证明中, 我们会发现, 实对称矩阵的合同比起相抵, 额外要求了某种正负性的一致。将合同规范形 $\text{diag}(I_a, I_b, O)$ 中的 a 称为正惯性指数, b 称为负惯性指数, 则根据上方讨论可知, 实对称方阵的正负惯性指数唯一确定, 为其正/负特征值的个数。此外, 两同阶实对称方阵合同等价于正负惯性指数相等。

* 由于 a, b 是满足 $a + b \leq n$ 的任何非负整数, 固定 a 从 0 到 n , 对 b 的种类数求和可知 n 阶方阵的合同规范形共有 $(n+1) + \dots + 1 = \frac{(n+1)(n+2)}{2}$ 种。

* 在讨论合同相关的问题时, 往往会涉及主子式相关的内容, 而对合同后主子式的计算又依赖一般形式 (不正方阵形式) 的 Binet-Cauchy 公式, 比起期中前更需要大家的熟练度。

11.1.2 成对初等变换

虽然我们的证明过程看起来非常整体性, 也相对简洁, 但这样的证明并不具有可推广性, 因为它依赖一个过强的性质: 实对称方阵可以正交相似对角化。例如, 由于对斜对称方阵 (即 $A^T = -A$ 的实方阵) A , 有

$$(P^T AP)^T = P^T A^T P = -P^T AP$$

可知斜对称方阵合同后仍为斜对称方阵, 考虑其合同标准形是有意义的。

但是, 由于我们并没有斜对称方阵的正交相似结论 (其实有, 但我们不学), 当然也就无法简单得到结果。因此, 我们需要回到类似相抵标准形证明时的做法, 通过操作将其尽可能化为简单形式。

由于任何可逆阵 P 都可以分解成一系列初等变换阵乘积, 我们可以将 $P^T AP$ 看作, 每次对 A 施加某初等列变换, 并施加其转置对应的初等行变换。这样的变换方式称为成对初等变换。

通过直接的计算, 三类初等行列变换可以对应三类成对初等变换:

1. 将 A 的第 i 行的 k 倍加到第 j 行, 并将 A 的第 i 列的 k 倍加到第 j 列;
2. 将 A 的第 i, j 两行交换, 并将 A 的第 i, j 两列交换;
3. 将 A 的第 i 行乘 k 倍 ($k \neq 0$), 并将 A 的第 i 列乘 k 倍。

下面，我们将通过这些成对初等变换直接操作出斜对称阵的合同规范形

$$\text{diag}(K, \dots, K, 0, \dots, 0), \quad K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

由于 A 是斜对称方阵，其对角元素应全为 0。若 A 全为零，则已经符合规范形形式。否则，设 $a_{ij} \neq 0$ ，且 $i < j$ (由 $a_{ij} = -a_{ji}$ 可如此设)，将 $i, 1$ 两行/列交换，将 $j, 2$ 两行/列交换，可使 $a_{12} \neq 0$ 。

将第一行/列乘 $\frac{1}{a_{12}}$ 倍，则变换后 $a_{12} = 1$ ，利用斜对称性 $a_{21} = -1$ 。

考虑第三行/列减去第二行/列的 a_{13} 倍，利用斜对称性可知消去后 $a_{13} = a_{31} = 0$ ，同理，可将 a_{13} 至 a_{1n} 都消去，此时 a_{31} 至 a_{n1} 也被消去；考虑第三行/列减第一行/列的 a_{32} 倍，利用斜对称性可知消去后 $a_{32} = a_{23} = 0$ ，同理，可将 a_{32} 至 a_{n2} 都消去，此时 a_{23} 至 a_{n2} 也被消去。

* 这里建议读者举一个具体的例子操作验证为何上方的消去过程每步不会影响已经消去的部分。教材上的证明采用了分块的书写方法，本质上与上述操作是一致的。

由此， A 的前两行、前两列中，只有 $a_{12} = 1, a_{21} = -1$ ，其他元素全为 0，也即成为了 $\text{diag}(K, B)$ 的形式。由于 A 的斜对称性，其后 $n - 2$ 行、 $n - 2$ 列构成的子矩阵 B 也是斜对称阵，从而只要其不为 O ，就可以进行上述操作，直到化为规范形。

斜对称方阵规范形的唯一性是容易证明的：合同不改变秩，而规范形的秩等于其中 K 的个数乘 2，由规范形唯一确定。

* 由此可以看出斜对称方阵的秩一定为偶数。

* 当然，实对称阵也可以直接通过成对初等变换得到合同规范形，具体操作可参考教材，或仿照下一小节。对于一般的实方阵，合同也有一些结论，如任何不是斜对称阵的实方阵都可以合同于上三角阵。

11.1.3 复对称方阵

所谓复对称方阵，是指满足 $A^T = A$ 的复方阵，与实对称方阵类似，复对称方阵在合同下 (此时 P^TAP 的 P 为复方阵) 仍然为复对称方阵，因此谈论合同是有意义的，下面证明其规范形与相抵标准形相同，为 $\text{diag}(I_r, O)$ 。自然，由合同不改变秩可直接得到唯一性。

* 一个有关简化书写的附注：当且仅当 A 为方阵时，可以将 A 的相抵标准形写为 $\text{diag}(I_r, O)$ 。否则，若 A 列满秩/行满秩，其相抵标准形并不是 $\text{diag}(I_r, O)$ 的形式。

* 与实对称方阵不同，复对称方阵的相似性质是难以确定的，这是由于复方阵中应用共轭转置代替转置才能得到类似的结论。由此，得到其合同规范形必须通过行列变换操作。

若 A 全为零，则其已符合规范形形式。否则，设 $a_{ij} \neq 0$ ，考虑 a_{ii} 。若 $a_{ii} \neq 0$ ，直接进行下一步，否则可将第 i 行/列加第 j 行/列的一倍，此时 a_{ii} 即变成了 $a_{ij} + a_{ji} = 2a_{ij} \neq 0$ 。

将 $1, i$ 两行/列交换，可使 $a_{11} \neq 0$ 。由复数的性质，存在非零复数 t 使得 $t^2 = a_{11}$ (考虑模长变为平方根，辐角除以 2)，将第一行/列除以 t 即使得 $a_{11} = 1$ 。

第二行/列减第一行/列乘 a_{12} 后，可将 a_{12} 消为 0，此时由对称性 $a_{21} = 0$ 。同理，进一步消元能使 $a_{13} = \dots = a_{1n} = a_{31} = \dots = a_{n1} = 0$ ，于是 A 成为了 $\text{diag}(1, B)$ 的形式。由于 A 的对称性，其后 $n - 1$ 行、 $n - 1$ 列构成的子矩阵 B 也是复对称阵，从而只要其不为 O ，就可以进行上述操作，直到化为规范形。

本节中，我们用不同的方法讨论了三类不同方阵的合同规范形。值得注意的是，对不同类型的方阵，合同规范形指的是不同的形式，务必不要混淆。大部分情况下，谈论合同规范形是针对实对称方阵 (这是因为另两类方阵合同标准形与相抵标准形完全对应，无法刻画更多性质)，而对于一般的方阵，并没有合同规范形的定义。

另一个值得注意的情况是，上述的讨论说明了 $\text{diag}(I_a, -I_b, O)$ 与 $\text{diag}(I_{a+b}, O)$ 在看作复方阵时是合同的，但看作实方阵时不合同。这也就意味着，合同结果与所在数域相关，在谈论时我们必须指定谈论的数域，这与相抵、相似都不同（虽然我们尚未证明相似与数域无关，但这确实是成立的）。

那么，就谈论实对称阵而言，既然正交相似标准形已经足够好用，为什么我们还要谈论合同？

一个不错的视角是，合同所给出的信息是介于相抵和相似之间的。它包含的信息比秩要多，但对特征值又只关心其正负。可以想到，在主要关心正负的不等式类问题中，合同往往更加方便，而且由于它比正交相似弱，存在一些相似不具有的效果（如下面会介绍的有关同时合同对角化的结论）。

§11.2 正定方阵

经过之前的分析，我们对于实/复对称阵的讨论已经基本完善了。不过，直到现在，我们也没有说明过为何对称阵是一类重要的方阵。本节中，我们将说明对称与齐次二次多项式唯一对应，从而将中学学过的二次不等式结论推广到 n 元。

11.2.1 二次型

我们称系数 $q_{ij} \in \mathbb{K}$ 的 n 元齐次二次多项式

$$Q(x) = \sum_{1 \leq i \leq j \leq n} q_{ij} x_i x_j$$

为一个 \mathbb{K} 上的二次型。大部分情况下，我们只关心实二次型与复二次型。

记

$$a_{ij} = \begin{cases} q_{ij} & i = j \\ \frac{1}{2}q_{ij} & i \neq j \end{cases}$$

则 A 为对称阵，且直接计算可发现

$$Q(x) = x^T A x$$

反之，给定对称阵 A ，同样可计算得 $x^T A x$ 为一个二次型，且由于位置与系数的对应，对称阵与二次型一一对应。

由此，研究对称阵问题也可以看成是在研究二次型的问题——实对称方阵对应实二次型，复对称方阵则对应复二次型。若我们给出了 A 的某个合同对角化 $A = P^T D P$ ，有

$$Q(x) = x^T P^T D P x = (Px)^T D (Px)$$

设 $y = Px$ （由 P 可逆，这称为一个非退化线性替换），由 D 是对角阵计算即有

$$Q(x) = y^T D y = \sum_{i=1}^n d_{ii} y_i^2$$

也就是说，合同对角化事实上是将二次型换元为了平方和的形式。根据已经证明的结论，我们知道对实二次型可以换元使得 d_{ii} 为 ± 1 或 0，对复二次型可以换元使得 d_{ii} 为 1 或 0，这也称为对应的二次型的规范形。

另一方面，如果取 P 为正交阵，对应的 $y = Px$ 称为正交替换，于是，实二次型一定可以通过正交替换成为一系列平方的和或差。

* 有的时候，以二次型描述的问题会比起矩阵更加清晰，见教材 6.2.2 的例题。

11.2.2 正定性

对于一个实二次型，我们自然可以类似一元的二次函数，关心它的正负性。若 $Q(x) \geq 0$ 恒成立，称它是半正定的，若还有 $Q(x) = 0$ 当且仅当 $x = 0$ (即 $x \neq 0$ 时 $Q(x) > 0$)，则称它是正定的。同理，将上方的 \geq 改为 \leq ，可以得到负定与半负定的定义。由于二次型与对称阵的一一对应，上述的定义也可以针对对称阵给出。

为了说明合同变换的作用，我们先证明，若一个实对称阵是正定/半正定/负定/半负定/不定的，其合同后仍然满足这样的性质。设 $B = P^T AP$, P 可逆，则直接计算有

$$x^T B x = (Px)^T A (Px)$$

由于 P 可逆，根据之前对线性映射的讨论， x 到 Px 为双射，且 0 映射到 0，由此 x 取遍所有向量/非零向量时 Px 也取遍所有向量/非零向量，反之亦然。这就得到了证明。

下面，我们将给出一组等价定义，说明正定的性质。

对于 n 阶实对称阵 A ，以下命题相互等价：

1. (正定定义) 对任何 $x \neq 0$ 有 $x^T Ax > 0$;
2. A 特征值均为正实数;
3. A 正惯性指数为 n ;
4. 存在可逆实对称阵 P 使得 $A = P^2$;
5. 存在唯一正定阵 P 使得 $A = P^2$;
6. 存在可逆阵 P 使得 $A = P^T P$;
7. 存在列满秩阵 P 使得 $A = P^T P$;
8. A 的主子式都是正数;
9. A 的顺序主子式 (即前 r 行前 r 列交成的主子式) 都是正数。

我们设法找到一条相互推出的链：

- 2、3 等价

上节讨论中已经证明正惯性指数即为正特征值个数。

- 2 推 4

设 A 的正交相似对角化为 $Q^T \text{diag}(\lambda_1, \dots, \lambda_n) Q$ ，取

$$P = Q^T \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}) Q$$

可直接计算转置验证 P 实对称，且 $P^2 = A$ 。

- 4 推 6

实对称阵的转置即为自身，且 2 推 4 证明中构造的 P 可逆，从而即有 $A = P^2 = P^T P$ 。

- 6 推 7

可逆阵是列满秩的，从而性质 7 也成立。

• 7 推 1

取出性质 7 的 P 后, 由于 $x^T Ax = (Px)^T (Px)$, 可知其大于等于 0, 且等于 0 等价于 $Px = 0$ 。利用解空间维数定理, 对列满秩的 P 有 $Px = 0$ 只有零解, 从而成立。

• 1 推 2

设 A 的正交相似对角化为 $P^T DP$, 由上一小节计算, 记 $y = Px$ 可知

$$x^T Ax = \sum_{i=1}^n d_{ii} y_i^2$$

若某个 $d_{ii} \leq 0$, 取 $x = P^T e_i$, 则 $y = e_i$, 计算可知 $x^T Ax = d_{ii} \leq 0$, 矛盾。

• 2 推 5

回顾 2 推 4 证明中的构造

$$P = Q^T \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}) Q$$

由于我们上方已经证明了 1、2、4、6、7 等价, 由 P 符合性质 2 可知正定, 只需证明唯一性。

若存在另一个正定阵 Q 使得 $A = P^2 = Q^2$, 由平方的特征值为特征值的平方、 P, Q 的特征值均为正数, 至少可以得到 P 与 Q 特征值相同, 而由它们均实对称可知它们正交相似。

设 $T^T PT = Q$, 其中 T 为正交阵。此时 $P^2 = Q^2$ 可化为

$$P^2 T = T P^2$$

而我们要证 $P = Q$, 也即

$$PT = TP$$

于是, 只要证明任何与 P^2 可交换的矩阵都与 P 可交换即可。

设 $P = S^T DS$, 其中 S 为正交阵, 则 $P^2 = S^T D^2 S$ 。本讲义 8.3.3 的最后一题证明了与 P^2 可交换的矩阵一定能写为 $S^T B_0 S$, 其中 B_0 满足当且仅当 $d_{ii}^2 = d_{jj}^2$ 时 $(B_0)_{ij}$ 可任取, 否则为 0。但对两个正数, $d_{ii}^2 = d_{jj}^2$ 等价于 $d_{ii} = d_{jj}$, 从而 B_0 只在 $d_{ii} = d_{jj}$ 的元素非零, 这就说明了它与 P 可交换, 得证。

* 需要熟悉与对角阵可交换的矩阵表达式。

• 5 推 4

由于正定阵是可逆的, 符合性质 4 要求。

• 1 推 8

对第 m_1, \dots, m_r 行/列交成的子矩阵 C , 对任何 r 维非零向量 y , 考虑 n 维向量

$$x = \sum_{i=1}^r y_i e_{m_i} \neq 0$$

分块或直接展开计算可发现

$$y^T C y = x^T A x > 0$$

从而 C 正定。由性质 1、4 的等价性可知存在可逆阵 P_C 使得 $\det C = (\det P_C)^2 > 0$, 得证。

• 8 推 9

顺序主子式是主子式, 从而性质 9 成立。

• 9 推 1

对阶数归纳。 $n = 1$ 时成立，假设 $n = k$ 时成立，考虑 $n = k + 1$ 时，由对称性矩阵 A 可分块为

$$\begin{pmatrix} B_{k \times k} & \alpha \\ \alpha^T & a \end{pmatrix}$$

由 $n = k$ 时成立知 B 正定，且还有 $\det A > 0$ 。

由 B 特征值均正可知其可逆，从而可写出 Schur 公式

$$\begin{pmatrix} I & \\ -\alpha^T B^{-1} & 1 \end{pmatrix} \begin{pmatrix} B & \alpha \\ \alpha^T & a \end{pmatrix} \begin{pmatrix} I & -B^{-1}\alpha \\ & 1 \end{pmatrix} = \begin{pmatrix} B & 0 \\ 0 & a - \alpha^T B^{-1}\alpha \end{pmatrix}$$

式子左侧 A 左右乘的两个矩阵恰好互为转置（利用正交相似对角化可说明 B 对称则 B^{-1} 对称），于是 A 合同于 $\text{diag}(B, a - \alpha^T B^{-1}\alpha)$ ，且由 A 左右乘的两个矩阵行列式为 1 可知

$$\det A = \det \text{diag}(B, a - \alpha^T B^{-1}\alpha) = (a - \alpha^T B^{-1}\alpha) \det B$$

由 B 正定与 $\det A > 0$ 即得到 $a - \alpha^T B^{-1}\alpha > 0$ 。由此，考虑特征值利用性质 2 可直接验证 $\text{diag}(B, a - \alpha^T B^{-1}\alpha)$ 正定，从而 A 正定，得证。

从性质 8 的证明过程可以看出，若 A 正定，则其所有主子矩阵（即主子式对应的子矩阵）都正定，对半正定阵可类似证明此结论成立。上面的这些性质可以解决绝大部分关于正定阵的问题。此外，性质 9 证明中出现的对称阵的 Schur 公式本质是一种成对分块行列变换，这样的操作在一些细致的证明中时常会用到。

* 根据我们之前对内积矩阵的讨论，一个方阵能成为内积矩阵当且仅当它正定。

半正定阵也有类似的结论。对于 n 阶实对称阵 A ，以下命题相互等价：

1. (半正定定义) 对任何 x 有 $x^T Ax \geq 0$;
2. A 特征值均为非负实数;
3. A 负惯性指数为 0;
4. 存在实对称阵 P 使得 $A = P^2$;
5. 存在唯一半正定阵 P 使得 $A = P^2$;
6. 存在实方阵 P 使得 $A = P^T P$;
7. 存在实矩阵 P 使得 $A = P^T P$;
8. A 的主子式都非负。

性质 1 到 7 的等价性、1 推 8 于可与正定时几乎完全相同证明。值得注意的是，所有顺序主子式都非负不能说明半正定，如考虑三阶实方阵 $\text{diag}(1, 0, -1)$ ，其顺序主子式分别为 1、0、0，但它是不定的。

我们下面来证明 8 推 2。利用特征多项式 $\varphi_A(\lambda) = \det(\lambda I - A)$ 的定义，利用完全展开式可以发现

$$\varphi_A(\lambda) = \lambda^n + \sum_{k=1}^n (-1)^k \sigma_k \lambda^{n-k}$$

这里 σ_k 为所有 k 阶主子式的和，由性质 8 成立可知 $\sigma_k \geq 0$ 。

当 $\lambda < 0$ 时，有

$$\varphi_A(\lambda) = (-1)^n \left((-\lambda)^n + \sum_{k=1}^n \sigma_k (-\lambda)^{n-k} \right)$$

于是其恒与 $(-1)^n$ 同号, 从而非零, 这即说明 $\varphi_A(\lambda)$ 没有小于 0 的根, 即所有特征值非负 (由实对称方阵所有特征值为实数), 得证。

* 对于负定/半负定, 其等价于数乘 -1 后为正定/半正定, 因此上方的所有性质都有对应的版本, 这里不再赘述。

11.2.3 不等式问题

我们最后介绍两个合同对角化相关的引理, 与一些正定阵解决的不等式问题。

1. 合同对角化引理 1: 若 A 为正定阵, B 为同阶实对称阵, 存在可逆阵 P 使得 $P^TAP = I$ 、且 P^TBP 为对角阵。

由 A 正定, 其与 I 合同, 从而存在可逆阵 Q 使得 $Q^TAQ = I$ 。由于 Q^TBQ 仍对称, 设其正交相似对角化为

$$T^T(Q^TBQ)T = D$$

其中 T 正交, D 对角, 则有

$$(TQ)^T B (TQ) = D$$

$$(TQ)^T A (TQ) = T^T IT = I$$

再由可逆阵乘积可逆知令 $P = TQ$ 即可。

2. 合同对角化引理 2: 若 A, B 为同阶半正定阵, 存在可逆阵 P 使得 P^TAP, P^TBP 均为对角阵。

由 $x^T(A+B)x = x^TAx + x^TBx \geq 0$ 可知 $A+B$ 半正定, 从而存在可逆阵 Q 使得 $Q^T(A+B)Q = \text{diag}(I_r, O)$, 于是可设

$$Q^TAQ = \begin{pmatrix} A_1 & A_2 \\ A_2^T & A_3 \end{pmatrix}$$

有

$$Q^TBQ = \begin{pmatrix} I_r - A_1 & -A_2 \\ -A_2^T & -A_3 \end{pmatrix}$$

利用半正定阵的主子矩阵半正定与合同不改变半正定性, A_3 与 $-A_3$ 均半正定, 考虑惯性指数可知只能为 $A_3 = O$ 。

设 $Q^TAQ = G$, 其元素为 g_{ij} 。对 A_2 部分的任何一个元素 g_{ij} , 根据形式有 $i < j$ 且 $g_{jj} = 0$ 。考虑第 i, j 行/列相交形成的主子式, 其行列式为 $-g_{ij}^2$, 由半正定性知其大于等于 0, 从而只有 $g_{ij} = 0$, 由此得到 $A_2 = O$, 至此, 我们有

$$Q^TAQ = \begin{pmatrix} A_1 & O \\ O & O \end{pmatrix}$$

$$Q^TBQ = \begin{pmatrix} I_r - A_1 & O \\ O & O \end{pmatrix}$$

* 注意 $r = \text{rank}(A+B)$, 我们事实上得到了对半正定阵 A, B 有 $\text{rank } A \leq \text{rank}(A+B)$ 。

由于 A_1 对称, 设其正交相似对角化为 $A_1 = U^T DU$, 其中 U 为正交阵、 D 为对角阵, 则利用 $U(I - A_1)U^T = UU^T - D = I_r - D$ 有

$$\begin{pmatrix} U & O \\ O & I \end{pmatrix} Q^TAQ \begin{pmatrix} U^T & O \\ O & I \end{pmatrix} = \text{diag}(D, O)$$

$$\begin{pmatrix} U & O \\ O & I \end{pmatrix} Q^TBQ \begin{pmatrix} U^T & O \\ O & I \end{pmatrix} = \text{diag}(I_r - D, O)$$

验证可逆性可得这已经为同时相合对角化。

3. 证明若 A, B 均正定，则 AB 的所有特征值为正数。

设 $A = P^T P$, P 可逆, 有

$$P^{-T}(AB)P^T = PBP^T$$

由于 PBP^T 与 B 合同, 其仍正定, 因此特征值均为正, 而相似不改变特征值, 从而得证。

* 拆分合同与相似两步的技巧有时会有奇效。

4. 若 A 半正定, 对正整数 m , 证明存在唯一半正定阵 X 使得 $X^m = A$, 记为 $X = A^{1/m}$ 。

* 由此半正定阵的正有理数次方可以定义, 通过极限可定义其任何次方。

与正定阵性质 2 推 5 的证明非常类似。设 A 正交相似对角化为

$$Q^T \operatorname{diag}(\lambda_1, \dots, \lambda_n) Q$$

由于 $\lambda_i \geq 0$, 可取

$$X = Q^T \operatorname{diag}(\lambda_1^{1/m}, \dots, \lambda_n^{1/m}) Q$$

即符合要求。

为证明唯一性, 由于非负实数上 $x \rightarrow x^m$ 为双射, 类似可得与 X^m 可交换的矩阵都与 X 可交换, 从而若 $Y^m = X^m$, 且 Y 亦半正定, 利用特征值一一对应可知正交相似, 设 $T^T X T = Y$, 其中 T 正交, 则由于

$$X^m T = T X^m$$

可推出

$$X T = T X$$

从而得证唯一。

5. 若 A, B 均半正定, 证明 $\det(A + B) \geq \det A + \det B$ 。

利用合同对角化引理 2, 设可逆矩阵 P 使得 $P^T AP, P^T BP$ 均为对角阵, 要证的式子同左乘 $\det P^T$ 、右乘 $\det P$ (它们都非零) 可得

$$\det(P^T AP + P^T BP) \geq \det(P^T AP) + \det(P^T BP)$$

设 $P^T AP = \operatorname{diag}(a_1, \dots, a_n)$ 、 $P^T BP = \operatorname{diag}(b_1, \dots, b_n)$, 即化为要证在 a_i, b_i 非负时

$$\prod_{i=1}^n (a_i + b_i) \geq \prod_{i=1}^n a_i + \prod_{i=1}^n b_i$$

而左侧展开即包含右侧, 从而得证。

6. 若 n 阶方阵 A 半正定, 设其前 r 行 r 列构成的矩阵为 B , 后 $n-r$ 行 $n-r$ 列构成的矩阵为 C , 证明 $\det A \leq \det B \det C$ 。

- 先证明 A 正定时正确。此时由于 B, C 为主子矩阵可知 B, C 正定, 设

$$A = \begin{pmatrix} B & D \\ D^T & C \end{pmatrix}$$

利用 Schur 公式类似上一小节证明可知

$$\begin{pmatrix} I & \\ -D^T B^{-1} & I \end{pmatrix} \begin{pmatrix} B & D \\ D^T & C \end{pmatrix} \begin{pmatrix} I & -B^{-1}D \\ & I \end{pmatrix} = \begin{pmatrix} B & O \\ O & C - D^T B^{-1}D \end{pmatrix}$$

利用相合不改变正定性，考虑右侧主子矩阵可知 $C - D^T B^{-1} D$ 正定，且两端计算行列式有

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

由 B 正定，考虑特征值发现 B^{-1} 正定，设其为 $P^T P$ ，则 $D^T B^{-1} D = (PD)^T (PD)$ ，从而其半正定。利用上一题可知

$$\det C \geq \det(C - D^T B^{-1} D) + \det(D^T B^{-1} D) \geq \det(C - D^T B^{-1} D)$$

从而再由 $\det B \geq 0$ 即得

$$\det A \leq \det B \det C$$

- 对半正定的 A ，利用特征值可发现对任何 $\lambda > 0$ 有 $\lambda I + A$ 正定，由此由正定情况有

$$\det(\lambda I + A) \leq \det(\lambda I + B) \det(\lambda I + C)$$

对任何 $\lambda > 0$ 成立，由连续性取 $\lambda \rightarrow 0^+$ 得证。

* 这是半正定阵的摄动法的常用形式。

7. 若 n 阶方阵 A, B 正定，设 n 阶方阵 C 满足 $c_{ij} = a_{ij}b_{ij}$ ，证明 C 正定。

设 $A = P^T P$, $B = Q^T Q$, 且 P, Q 可逆，可发现

$$c_{ij} = \sum_{k,l=1}^n p_{ki}p_{kj}q_{li}q_{lj}$$

为了将其配凑为矩阵乘法的形式，我们先固定 k ，将它看成

$$\sum_{k=1}^n \left(\sum_{l=1}^n p_{ki}p_{kj}q_{li}q_{lj} \right) = \sum_{k=1}^n \left(\sum_{l=1}^n (p_{ki}q_{li})(p_{kj}q_{lj}) \right)$$

由此，记 $R_{(k)}$ 为第 i 行第 j 列为 $p_{kj}q_{ij}$ 的方阵，上述的乘法事实上可以写为

$$C = \sum_{k=1}^n R_{(k)}^T R_{(k)}$$

进一步地，由形式可以发现

$$R_{(k)} = Q \operatorname{diag}(p_{k1}, \dots, p_{kn})$$

设

$$R = \begin{pmatrix} R_{(1)} \\ \vdots \\ R_{(n)} \end{pmatrix} = \begin{pmatrix} Q & & \\ & \ddots & \\ & & Q \end{pmatrix} \begin{pmatrix} \operatorname{diag}(p_{11}, \dots, p_{1n}) \\ \vdots \\ \operatorname{diag}(p_{n1}, \dots, p_{nn}) \end{pmatrix}$$

可发现左侧可逆，右侧可行变换出 P ，从而 $\operatorname{rank} R = n$ ，其列满秩，而又由于 $C = R^T R$ ，即可从之前的性质得到正定。

* 这个结论的形式非常奇怪，但确实在进行一些估计时会有用。

§11.3 最小二乘

* 这部分在书上只作为例题进行了简单介绍，但实际上结论十分重要。

11.3.1 二次型的最值

在经过了正定相关的讨论后，我们可以对二次型的最值问题给出一个结论。设

$$Q(x) = x^T Ax$$

其中 A 对称，则 $Q(x)$ 的最小值存在当且仅当 A 半正定，此时最小值为 0； $Q(x)$ 的最大值存在当且仅当 A 半负定，此时最大值为 0。

这个结论的证明是自然的：由合同标准形，作非退化线性替换 $y = Px$ ，使得

$$x^T Ax = y^T \text{diag}(I_a, -I_b, O)y = \sum_{i=1}^a y_i^2 - \sum_{i=a+1}^{a+b} y_i^2$$

由于 $x = P^{-1}y$ ，任取 y_i 都可以构造出对应的 x 。从而只要 a 非零，就可以取 $y_1 \rightarrow \infty$ ，其余 y_i 为 0 使其趋于正无穷；反之，若 $a = 0$ ，利用半负定阵的等价定义可知 $x^T Ax \leq 0$ ，且 $x = 0$ 时取到。对 b 可以同理讨论得到结论。

比起这个容易得到的结论，我们往往更关心另一个式子的最值，也即

$$R(x) = \frac{x^T Ax}{x^T x}, \quad x \neq 0$$

这称为 **Rayleigh 商**。仍然假设 A 对称，由于 x 乘非零倍数不影响此式，我们有时也设其定义在单位球面 $\{x \mid \|x\| = 1\}$ 上。

由于这里需要控制分母 $x^T x$ 不变，我们可以想到通过正交换元构造。设 A 的正交相似对角化为 $P^T DP$ ，其中 P 为正交阵、 D 为对角阵，则

$$\frac{x^T Ax}{x^T x} = \frac{(Px)^T D(Px)}{x^T x}$$

进一步地，由于 P 正交， $x^T x = (Px)^T Px$ ，于是记 $y = Px$ ， D 的对角元（即 A 的特征值）为 $\lambda_1, \dots, \lambda_n$ ，可知原式化为

$$\frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2}$$

与刚才类似讨论可知 y 可取任何非零向量，而设 M 为特征值中的最大值， m 为特征值中的最小值，有

$$m \sum_{i=1}^n y_i^2 \leq \sum_{i=1}^n \lambda_i y_i^2 \leq M \sum_{i=1}^n y_i^2$$

且当 y 只在最大/最小的分量为 1，其余为 0 时可取等，由此即得 $R(x)$ 的最大/最小值为 A 的最大/最小特征值。

11.3.2 二次函数的极值与最值

在解决了二次型的最值问题后，我们来考虑更一般的 n 元二次函数，即

$$Q(x) = x^T Ax + b^T x + c$$

这里 A 为 n 阶实对称阵， $b \in \mathbb{R}^n$ ， c 为实数。

- $Q(x)$ 的驻点（即对各 x_i 导数均为 0 的点）为方程 $2Ax + b = 0$ 的解。

将其展开为求和，并看成关于每个 x_i 的函数求导即得。由于这不是本门课的重点，此处只给出一个简略的证明：

$$\begin{aligned}
 0 &= \frac{\partial}{\partial x_k} Q(x) \\
 &= \frac{\partial}{\partial x_k} \left(\sum_{i,j} a_{ij} x_i x_j + \sum_i b_i x_i + c \right) \\
 &= \frac{\partial}{\partial x_k} \left(a_{kk} x_k^2 + \sum_{i \neq k} 2a_{ik} x_i x_k + b_k x_k \right) \\
 &= 2a_{kk} x_k + \sum_{i \neq k} 2a_{ik} x_k + b_k \\
 &= 2 \sum_i a_{ik} x_k + b_k = (2Ax + b)_k
 \end{aligned}$$

- 当 A 不定时， $Q(x)$ 不存在最值点。

类似之前，我们设 $A = P^T \text{diag}(I_a, -I_b, O)P$ ，其中 P 为可逆阵，并记非退化线性替换 $y = Px$ ，则上式可以化为

$$Q(x) = \sum_{i=1}^a y_i^2 - \sum_{i=a+1}^{a+b} y_i^2 + b^T P^{-1} y + c$$

再记 $u = P^{-T} b$ ，其各元素为 u_i ，则可进一步写为

$$Q(x) = \sum_{i=1}^a y_i^2 - \sum_{i=a+1}^{a+b} y_i^2 + \sum_{i=1}^n u_i y_i + c$$

若 a, b 均非零，在 $y_1 \rightarrow \infty$ ，其余为 0 时 $Q(x) \rightarrow +\infty$ ，在 $y_{a+1} \rightarrow \infty$ ，其余为 0 时 $Q(x) \rightarrow -\infty$ ，因此无最值。

- 当 A 半正定时，若其驻点不存在，则其无最小值。

沿用上一部分证明的记号。由 A 半正定，可进一步写为

$$Q(x) = \sum_{i=1}^a y_i^2 + \sum_{i=1}^n u_i y_i + c$$

若存在 $k > a$ 且 $u_k \neq 0$ ，则 $Q(x)$ 可以看成 y_k 的线性函数，可以趋于正负无穷，从而不存在最小值。下面只需说明驻点存在等价于 $k > a$ 时 u_k 均为 0。

驻点存在即 $2Ax + b = 0$ 有解，考虑增广矩阵即

$$\text{rank } 2A = \text{rank}(2A, -b)$$

利用第三类初等变换得到这等价于

$$\text{rank } A = \text{rank}(A, b)$$

展开并利用 $b = P^T u$ 可知

$$a = \text{rank}(P^T \text{diag}(I_a, O)P, P^T u)$$

由 P^T 可逆，左乘其逆不影响秩，得到其等价于

$$a = \text{rank}(\text{diag}(I_a, O)P, u)$$

由于 P 可逆，考虑分块矩阵第三类初等列变换，第一列右乘 P^{-1} ，得到等价于

$$a = \text{rank}(\text{diag}(I_a, O), u)$$

而直接作列变换可知这即等价于 $k > a$ 时 $u_k = 0$ ，得证。

- 当 A 半正定时，若其驻点存在，则所有驻点都是最小值点。

沿用上一部分证明的记号。若已知 $k > a$ 时 $u_k = 0$ ，利用二次函数知识可知最小值等价于

$$y_i = -\frac{u_i}{2}, \quad i = 1, \dots, a$$

另一方面，驻点方程写为

$$2P^T \text{diag}(I_a, O)Px + P^T u = 0$$

消去并利用 $y = Px$ 得到

$$\text{diag}(I_a, O)y = -\frac{u}{2}$$

由于已知驻点存在， $k > a$ 时 $u_k = 0$ ，上述方程即与最小值条件完全相同，得证。

* 进一步利用二次函数知识可得此时最小值为

$$-\sum_{i=1}^a \frac{u_i^2}{4} + c = -\frac{1}{4}u^T u + c = -\frac{1}{4}b^T P^{-1}P^{-T}b + c$$

- 当 A 半负定时，若其驻点不存在，则其无最大值。

考虑 $-Q(x) = x^T(-A)x + (-b)^T x - c$ ，此函数的驻点条件为 $-2Ax - b = 0$ ，即 $2Ax + b = 0$ ，且 $-A$ 半正定，从而由之前的部分得证。

- 当 A 半负定时，若其驻点存在，则所有驻点都是最大值点。

同样考虑 $-Q(x)$ 得证。

利用分析学的方法，可以通过 Hesse 阵证明 A 不定时 $Q(x)$ 也不存在极值点（类似教材 6.3 节定理 7）。对于其他情况，由于极值点一定是驻点（同样利用分析学方法），最值点一定是极值点，由于最值点等价于驻点，其自然也等价于极值点。由此，我们可以完全确定一个一般二次函数的所有最值点与极值点。

11.3.3 最小二乘解

在本讲义 1.4.1 中，我们其实已经介绍了最小二乘问题：给定 $A_{m \times n}$ 、 $b_{m \times 1}$ ，求 x 使得 $\|Ax - b\|$ 最小。由于 $\|Ax - b\|$ 最小等价于其平方最小，即对应求二次函数

$$Q(x) = (Ax - b)^T(Ax - b)$$

的最小值点。将其展开得到（利用一阶方阵必然对称， $b^T Ax = x^T A^T b$ ）

$$Q(x) = x^T A^T Ax - 2b^T Ax + b^T b$$

根据半正定阵的等价定义， $A^T A$ 是半正定的，从而 $Q(x)$ 的所有驻点都是最小值点，由之前讨论可知其驻点方程为

$$A^T Ax = A^T b$$

事实上，最小二乘问题比起一般的二次函数问题具有特殊性：其驻点必然存在，可直接利用秩相关的一些结论（主要来自教材 4.3 节）得到

$$\text{rank}(A^T A, A^T b) = \text{rank } A^T (A, b) \leq \text{rank } A = \text{rank } A^T A$$

而又由 $A^T A$ 是 $(A^T A, A^T b)$ 的子矩阵可知秩相等，从而此非齐次线性方程组有解。

* 最小二乘问题有很多算法，但人工计算一般直接求解齐次线性方程组是最方便的。

我们以上述结论的某种逆命题结束这次习题课：若 $Ax = b$ 有解，且 A 半正定，则存在矩阵 C, c 使得

$$A = C^T C, \quad b = C^T c$$

证明：设 $A = P^T \text{diag}(I_r, O)P$ ，则由 P 可逆，记 $y = Px$ ， $Ax = b$ 有解等价于

$$\text{diag}(I_r, O)y = P^{-T}b$$

有解，而这意味着 $P^{-T}b$ 的后 $n - r$ 个分量为 0，从而计算得 $P^{-T}b = \text{diag}(I_r, O)P^{-T}b$ 。取

$$C = \text{diag}(I, O)P, \quad c = P^{-T}b$$

可验证结论成立。

这意味着，二次函数若最小值存在，其最小值点一定可以看成某个最小二乘问题的解。

十二 期末总结

§12.1 习题解答

1. 习题 6.3-8

充分性：见本讲义 11.2.2。

必要性：由列满秩定义须有 $m \geq n$ 。由正定等价定义可知存在可逆矩阵 P_0 使得 $A = P_0^T P_0$ ，取

$$P = \begin{pmatrix} P_0 \\ O_{(m-n) \times n} \end{pmatrix}$$

由其存在 n 阶可逆子式可知列满秩，直接计算可知 $P^T P = P_0^T P_0 + O^T O = P_0^T P_0 = A$ ，得证。

2. 习题 6.3-10

见本讲义 11.2.2。

3. 计算奇异值分解

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix}$$

算法见本讲义 10.4.2，注释第二条中给出了常用算法，最终结果为

$$A = U \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & 0 \end{pmatrix} V^T, \quad U = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{2} \end{pmatrix}$$

* 奇异值唯一， U, V 未必唯一。

4. 设 A 是 n 级实对称矩阵，证明 A 的特征值都在 $[a, b]$ 中等价于 $A - tI$ 当 $t < a$ 时正定， $t > b$ 时负定。

利用特征值定义计算特征多项式可知 $A - tI$ 的特征值为 A 的每个特征值减 t ，由此由实对称阵正定等价于特征值均正、负定等价于特征值均负可得结论。

5. 补充题六-1

存在性: 由 A 可逆知 $A^T A$ 正定, 从而特征值均正, 于是 A 的奇异值均正, 奇异值分解为 $A = U\Sigma V^T$, 这里 Σ 为对角元均正的对角阵, U, V 正交。取 $T = UV^T$ 、 $S_1 = V\Sigma V^T$ 、 $S_2 = U\Sigma U^T$, 利用相合不改变正定性即得 S_1, S_2 正定, 而由正交阵乘积正交可知 T 正交, 从而符合要求。

唯一性: 若 $A = TS = T_0S_0$ 是两种极分解, 有 $TS = T_0S_0$, 进一步写为

$$T_0^T T = S_0 S_0^{-1}$$

设上式左右两侧的矩阵为 Q 。利用本讲义 11.2.3 第三题的证明过程, 由于 S_0, S_0^{-1} 正定, Q 相似于某正定阵, 因此可对角化且特征值均为正。另一方面, $Q = T_0^T T$ 为正交阵, 由本讲义 10.3.1 可知特征值模长均为 1, 因此只能所有特征值为 1。由于其可对角化, 特征值代数重数等于几何重数, 1 的几何重数也为 n , 即 $\text{rank}(Q - A) = 0$, 于是 $Q = I$, 也即 $T_0 = T$ 、 $S_0 = S$ 。

6. 习题 4.6-8

只对行证明, 对列完全类似。设 Q 为正交矩阵, 考虑第 i_1, \dots, i_k 行, 由 $QQ^T = I$ 可知

$$(QQ^T) \begin{pmatrix} i_1 & \dots & i_k \\ i_1 & \dots & i_k \end{pmatrix} = \det I_k = 1$$

于是由教材 4.3 节命题 1 可得

$$1 = (QQ^T) \begin{pmatrix} i_1 & \dots & i_k \\ i_1 & \dots & i_k \end{pmatrix} = \sum_{j_1, \dots, j_k} Q \begin{pmatrix} i_1 & \dots & i_k \\ j_1 & \dots & j_k \end{pmatrix} Q^T \begin{pmatrix} j_1 & \dots & j_k \\ i_1 & \dots & i_k \end{pmatrix}$$

而利用 Q^T 与 Q 的行列关系, 右侧即为 Q 第 i_1, \dots, i_k 行所有子式平方和, 得证。

7. 设 u, v 是 n 维列向量, $A = uv^T$ 。证明 u 非零时是 A 的特征向量, 并求 A 所有特征值与对应的特征向量。

直接计算可知

$$Au = uv^T u = (v^T u)u$$

由此 u 非零时 u 是特征值 $v^T u$ 的特征向量。

由于 $\text{rank } A \leq \text{rank } u \leq 1$, 0 的几何重数至少为 $n - 1$, 下面进行分类讨论:

(a) 当 u, v 有一个是零向量时, $A = O$, 其特征值为 n 个 0, 全空间一组基为特征向量。

(b) 当 u, v 均非零向量, 且 $v^T u = 0$ 时, 计算可发现此时 $A \neq O$, 于是 0 的几何重数为 $n - 1$ 。另一方面, 利用本讲义 8.1.3 第一题计算可知

$$\det(\lambda I - uv^T) = \lambda^{n-1} \det(\lambda I - v^T u) = \lambda^{n-1}(\lambda - v^T u) = \lambda^n$$

由此 0 的代数重数为 n , 不存在其他特征值。另一方面, $Ax = 0$ 等价于 $(v^T x)u = 0$, 由 u, v 非零这又等价于 $v^T x = 0$, 于是取 $\{x \mid v^T x = 0\}$ 的一组基即构成 0 的特征向量 (利用解空间维数定理可发现此空间维数的确为 $n - 1$)。

(c) 当 $v^T u \neq 0$ 时, 与上一种情况完全相同可知特征多项式为 $\lambda^{n-1}(\lambda - v^T u)$, 特征值 0 对应的特征子空间是 $\{x \mid v^T x = 0\}$, 特征值 $v^T u$ 的代数重数为 1, 对应特征向量为 u 。

* 可以发现 (a)、(c) 两种情况均可对角化, 而 (b) 不可对角化, 这就完全解决了秩 1 矩阵的相似对角化问题。

8. 已知 $x_1^2 + x_2^2 + x_3^2 = 1$, 求 $x_1^2 + x_2^2 + 2x_3^2 - 2x_1x_2 + 4x_1x_3 + 4x_2x_3$ 的最大/最小值与取到最大/最小值的 x_i 。

算法见本讲义 11.3.1 或教材 6.1.2 例 11。以教材的记号为例, 由证明过程可发现当向量 $b = e_1$, 即 $\alpha = Te_1$ 时取到最大值 λ_1 , $b = e_n$, 即 $\alpha = Te_n$ 时取到最小值 λ_n 。由此, 写出二次型对应的矩阵

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

计算对应的正交相似对角化

$$T^T AT = \text{diag}(4, 2, -2), \quad T = \begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix}$$

从而 $x_1 = x_2 = \frac{1}{\sqrt{6}}$ 、 $x_3 = \frac{2}{\sqrt{6}}$ 时最大为 4, $x_1 = x_2 = -\frac{1}{\sqrt{3}}$ 、 $x_3 = \frac{1}{\sqrt{3}}$ 时最小为 -2。

* 同样利用证明过程可发现, 当 x_i 全变为相反数时仍取到对应的最大/最小值, 其他情况不为最大/最小。

9. 证明 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 可对角化当且仅当 $(a-d)^2 + 4bc \neq 0$ 或 $a = d$ 、 $b = c = 0$ 。

先计算其特征多项式为

$$\det(\lambda I - A) = (\lambda - a)(\lambda - d) - bc = \lambda^2 - (a+d)\lambda + ad - bc$$

若两特征值不同, 考虑判别式即 $(a-d)^2 + 4bc \neq 0$, 其必然可对角化。否则, 由于某特征值 λ 代数重数为 2, 其可对角化当且仅当几何重数也为 2, 也即 $\text{rank}(\lambda I - A) = 0$, 于是 $A = \lambda I$, 此时对应 $a = d$ 、 $b = c = 0$ 。

反之, 若 A 可对角化, 必然落入上述两种情况的一种, 由此仍然成立。

§12.2 拾遗

12.2.1 不变子空间

之前我们提到过, 分块与限制映射的本质是统一的, 现在我们来具体叙述这种统一性。总体来说, 可以归为如下定理:

设线性映射 $f: U \rightarrow V$ 满足 $f(U_0) \subset V_0$, 其中 U_0 为 U 某子空间, V_0 为 V 某子空间, 则设 U_0 一组基为 u_1, \dots, u_r , 扩充为 U 一组基 $u_1, \dots, u_r, u_{r+1}, \dots, u_n$; V_0 一组基为 v_1, \dots, v_s , 扩充为 V 一组基 $v_1, \dots, v_s, v_{s+1}, \dots, v_m$ 。设 $f|_{U_0 \rightarrow V_0}$ 在上述基下的矩阵表示为 $B_{s \times r}$, f 在上述基下的矩阵表示为 $A_{m \times n}$, 则有

$$A = \begin{pmatrix} B & O \\ X & Y \end{pmatrix}$$

这个定理的证明是可以通过矩阵表示的定义直接进行的。 $f|_{U_0 \rightarrow V_0}$ 的矩阵表示为 B 的含义是

$$\forall i = 1, \dots, r, \quad f(u_i) = \sum_{j=1}^s b_{ij} v_i$$

而 f 的矩阵表示为 A 的含义是

$$\forall i = 1, \dots, n, \quad f(u_i) = \sum_{j=1}^m b_{ij} v_i$$

对 i 为 1 到 r (也即 A 的前 r 行), 对比系数即可发现结论。

因此, 利用线性映射的基变换等同于相抵, 限制映射可以看作相抵为分块上三角阵。

同样, 若 $U = V$, 上述定理对线性变换仍然成立: 设线性变换 $f: U \rightarrow U$ 满足 $f(U_0) \subset U_0$, 其中 U_0 为 U 某子空间, 则设 U_0 一组基为 u_1, \dots, u_r , 扩充为 U 一组基 $u_1, \dots, u_r, u_{r+1}, \dots, u_n$ 。设 $f|_{U_0 \rightarrow U_0}$ 在上述基下的矩阵表示为 $B_{r \times r}$, f 在上述基下的矩阵表示为 $A_{n \times n}$, 则有

$$A = \begin{pmatrix} B & O \\ X & Y \end{pmatrix}$$

利用线性变换的基变换等同于相似, 线性变换的限制映射可以看作相似为分块上三角阵。

* 满足上述条件的 U_0 称为 f 的不变子空间, 对于后续的相似标准形分析非常重要。这里我们其实给出了不变子空间在矩阵视角的含义。

最后, 考虑去除限制映射条件的一般分块: 考虑线性映射 $f: U \rightarrow V$ 与 U 某子空间 U_0 , V 某子空间 V_0 。设 U_0 一组基为 u_1, \dots, u_r , 扩充为 U 一组基 $u_1, \dots, u_r, u_{r+1}, \dots, u_n$; V_0 一组基为 v_1, \dots, v_s , 扩充为 V 一组基 $v_1, \dots, v_s, v_{s+1}, \dots, v_m$, 设 f 在上述基下的矩阵表示为 $A_{m \times n}$, 分块为

$$A = \begin{pmatrix} B_{s \times r} & C \\ X & Y \end{pmatrix}$$

可以发现, 这时的 B 意味着 $f(U_0)$ 在 V_0 中的分量。例如, B 的第一列表示 $f(u_1)$ 在基 v_1, \dots, v_n 下的前 s 个坐标。

但是, 这样的定义价值并不大, 因为这前 s 个坐标并不只与 v_1, \dots, v_s 相关。例如。考虑 $f(x, y) = (x, y)$ 、 $U = V = \mathbb{R}^2$, $U_0 = \{(a, a)\}$ 、 $V_0 = \{(a, 0)\}$, $u_1 = (1, 1)$ 、 $v_1 = (1, 0)$ 。

当 $v_2 = (0, 1)$ 时, 有

$$f(u_1) = v_1 + v_2$$

但当 $v_2 = (1, 1)$ 时, 有

$$f(u_1) = v_2$$

两种情况对应的 $B_{1 \times 1}$ 分别为 1 与 0。这就意味着, 一般情况下, 只通过 U_0, V_0 的基无法确定矩阵表示的分块 B , 必须给出 V_0 扩充后的基。

利用本讲义 4.2.2 对补空间的定义 (交为 $\{0\}$ 、和为全空间), 上述问题的出现本质是由于 V_0 的补空间不唯一, 例如上方的两个 v_2 生成的空间均为 V_0 的补空间。因此, 我们需要定义更强的结构, 这就是下面要说的正交补。

12.2.2 正交补空间

正交补空间的定义如下: 设 U 为 \mathbb{R}^n 某子空间, 其正交补空间为

$$U^\perp = \{x \mid x^T u = 0, \forall u \in U\}$$

根据定义, 正交补空间自然是唯一的。我们下面说明其性质 (下方运用一个简单的结论: 若 v 与一些向量正交, 则 v 与它们的线性组合也正交, 可见教材习题 4.6-2, 本讲义 9.1 有解答):

1. U^\perp 是 U 的补空间。

利用 $\dim(U + V) = \dim U + \dim V - \dim U \cap V$, U^\perp 是 U 的补空间等价于 $U^\perp \cap U = \{0\}$ 且 $\dim U^T + \dim U = n$ 。

若 $x \in U^T \cap U$, 由定义有 $x^T x = 0$, 从而 $x = 0$ 。另一方面, $x \in U^\perp$ 等价于 x 与 U 的一组基正交, 将这组基作为行向量拼成矩阵 A , 即等价于 $Ax = 0$, 于是 U^\perp 是 $Ax = 0$ 的解空间。由行秩定义可发现 $\text{rank } A = \dim U$, 而根据解空间维数定理即得 $\dim U^T + \dim U = n$, 得证。

2. 若 V 是 U 的补空间且对任何 $v \in V$ 、 $u \in U$ 有 $v^T u = 0$, 则 $V = U^\perp$ 。

由对任何 $v \in V$ 、 $u \in U$ 有 $v^T u = 0$, 根据正交补空间定义可知 $V \subset U^\perp$, 但根据补空间的维数性质 $\dim V = \dim U^\perp$, 从而只能相同。

3. $(U^\perp)^\perp = U$, 由此可以说两个空间互为正交补。

由定义对任何 $v \in U$ 、 $u \in U^\perp$ 有 $v^T u = u^T v = 0$, 而由补空间定义, 从 U^\perp 是 U 的补空间可知 U 是 U^\perp 的补空间, 由上一个性质得证。

4. 对任何矩阵 A , $Ax = 0$ 的解空间 $N(A)$ 和 A 的行空间 $R(A)$ 互为正交补。

设 $A \in \mathbb{R}^{m \times n}$, 假设其行向量为 $\alpha_1^T, \dots, \alpha_m^T$, 则计算有

$$N(A) = \{x \mid \alpha_i^T x = 0, \forall i = 1, \dots, m\}$$

另一方面, 与 $\alpha_1, \dots, \alpha_m$ 正交等价于与它们的任何线性组合正交, 而它们生成的空间是 $R(A)$, 于是

$$N(A) = \{x \mid \alpha^T x = 0, \forall \alpha \in R(A)\}$$

这就是正交补空间的定义。

* 回看本讲义 6.1 第五题。

除了定义以外, 我们还有另一个常用的计算正交补空间的方法: 任意取出 U 的一组基 $\alpha_1, \dots, \alpha_r$, 将其扩充为 \mathbb{R}^n 的一组基 $\alpha_1, \dots, \alpha_n$, 并进行 Schmidt 正交化得到 \mathbb{R}^n 的标准正交基 β_1, \dots, β_n , 则 β_1, \dots, β_r 构成 U 的一组标准正交基, $\beta_{r+1}, \dots, \beta_n$ 构成 U^\perp 的一组标准正交基。

证明: 首先, 根据 Schmidt 正交化的性质可知 β_1, \dots, β_r 与 $\alpha_1, \dots, \alpha_r$ 等价, 且标准正交, 于是它们构成 U 的标准正交基。利用标准正交基的性质, $\beta_{r+1}, \dots, \beta_n$ 都与 β_1, \dots, β_r 正交, 于是与 U 正交, 又由它们线性无关, 生成空间维数为 $n - r$, 因此利用上方性质 2 可知它们生成的空间就是 U^\perp 。由 Schmidt 正交化的性质, $\beta_{r+1}, \dots, \beta_n$ 也标准正交, 从而得证。

由补空间的性质, 任何向量 $x \in \mathbb{R}^n$ 可以唯一分解为 $y + z$, 其中 $y \in U$ 、 $z \in U^\perp$, 将 y 称为 x 在 U 上的投影。在本部分的最后, 我们证明 y 是 U 里满足 $\|x - y\|$ 最小的元素。

对任何 $w \in U$, 有

$$(x - w)^T (x - w) = (z + (y - w))^T (z + (y - w))$$

由于 $y - w \in U$, 其与 z 正交, 因此直接计算可知

$$(x - w)^T (x - w) = z^T z + (y - w)^T (y - w) \geq z^T z = (x - y)^T (x - y)$$

从而得证。

* 注意到, 只要给定 U 的子空间 U_0 , 任何向量 x 在其上的投影都是确定的, 无需再给出其他的基。由此, 对标准正交基下的矩阵表示, 其任何分块都可以被子空间确定。不过, 这部分内容已经不在本学期的范围中了, 之后有机会时我们将继续讨论。

12.2.3 QR 分解

在本讲义 10.2.2 中, 我们已经给出了 QR 分解的定义: 任何一个列满秩矩阵都可以分解成 QR , 其中 Q 的列均为单位向量且相互正交, 而 R 为对角元均正的 (可逆) 上三角阵。

我们先证明它的唯一性。若列满秩矩阵 $A_{m \times n}$ 有两种 QR 分解

$$A = Q_1 R_1 = Q_2 R_2$$

先将上式改写为

$$Q_1 = Q_2 R_2 R_1^{-1}$$

两侧同时左乘转置得到

$$Q_1^T Q_1 = R_1^{-T} R_2^T Q_2^T Q_2 R_2 R_1^{-1}$$

设

$$Q_1 = (\beta_1, \dots, \beta_n)$$

利用分块计算可发现

$$(Q_1^T Q_1)_{ij} = \beta_i^T \beta_j$$

由此, 由于 β_i 均为单位向量且相互正交, $Q_2^T Q_2$ 为单位阵 I , 同理 $Q_1^T Q_1 = I$, 因此

$$I = R_1^{-T} R_2^T R_2 R_1^{-1}$$

设 $R = R_2 R_1^{-1}$, 则利用三角阵的乘法与逆性质可知 R 为对角元均正的上三角阵, 由 $(XY)^T = Y^T X^T$, 上式可改写为

$$R^T R = I$$

由此, R 为正交的上三角阵。从 $R = R^{-T}$, 左侧为上三角阵、右侧为下三角阵可知 R 为对角阵, 又由正交可知每个对角元平方均为 1, 再根据对角元均正即得 $R = I$, 从而 $R_1 = R_2$, 进一步有

$$Q_1 = A R_1^{-1} = A R_2^{-1} = Q_2$$

* 注意对列均为单位向量且相互正交的矩阵 Q , 有 $Q^T Q = I$, 但 $Q Q^T \neq I$ 。

最后, 我们给出其对最小二乘问题的应用。假设 A 为一个列满秩方阵, 根据本讲义 11.3.3 的讨论, x 使得 $\|Ax - b\|$ 最小当且仅当

$$A^T A x = A^T b$$

作分解 $A = QR$ 后, 利用 $Q^T Q = I$, 上述方程即改写为

$$R^T R x = R^T Q^T b$$

即

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

如之前所说, 此算法的计算过程过于复杂, 一般并不适合人类使用, 但是对计算机有显著的效果: 计算机可以以较低复杂度得到某个近似分解 $A_{m \times n} \approx \tilde{Q} \tilde{R}$, 当已知 b 时, 计算 $\tilde{Q}^T b$ 需要 mn 量级次计算, 而上三角矩阵 \tilde{R} 的逆需要 n^2 量级次计算, 再由列满秩可知 $m \geq n$, 于是最终运算次数为 mn 量级, 低于正常求解。

§12.3 知识与技巧整理

- * 仍然以节为单位，梳理重要的定理与技巧，并只介绍**基本练习题**。与期中前会有部分重合内容，但关注点并不相同，复习时务必优先将下方列举的内容梳理清楚，对应有问题再查询期中前内容。
- * 在开始之前，请先对教材 5.1 节的内容有基本了解，之后的整理都是在这种“等价类”思想下进行的。

12.3.1 相抵

- * 从 4.3 到 4.5 的三节是之后一切分析的基础，**务必充分熟悉**。

3.5 矩阵的秩

定义：行秩、列秩、秩、行/列满秩、满秩

定理：行列秩相等、秩的非零子式定义

练习：3.5.2 例 8、3.5.2 例 9、习题 3.5-14 (均为结论重要，基本的估算方式)

* 在学习相抵标准形后建议尝试用标准形**暴力破解**本节的例题与练习题，这样的分块操作是之后很多证明的基础。

4.3 矩阵乘积的秩与行列式

定理： $\text{rank}(AB) \leq \min(\text{rank } A, \text{rank } B)$ ，Binet-Cauchy 公式、Cauchy 不等式 (4.3.2 例 10)

* 由于期中后出现了大量与子式相关的估算，Binet-Cauchy 公式与其下方的命题 1 都值得记忆。此外，Cauchy 不等式同样对期中后的估算**非常重要**，须记得结论。

练习：4.3.2 例 3 (方程组视角是之后二次型分析的基础)

4.4 可逆矩阵

定义：可逆矩阵、逆矩阵、伴随矩阵

算法：通过伴随方阵计算逆、初等变换法计算逆

定理：本节的所有命题与性质均为逆的基本性质

* 好用的结论： A 的多项式的逆一定是 A 的多项式，由此可以简化待定系数，如 4.4.2 例 7。

练习：4.4.2 例 2 (配凑思路整体求逆)、4.4.2 例 4 (关于逆的封闭性结论)、4.4.2 例 5 (同样为封闭性结论)、习题 4.2-15 (熟悉整体的计算)

* 4.4.2 例 12 的结论与过程对于下学期很重要，关乎根子空间与循环子空间。

4.5 分块矩阵

定义：分块矩阵、分块初等变换、分块对角阵、分块三角阵

定理：分块矩阵乘法 (看着能乘就能乘，详见本讲义 8.1.3)、 $\text{diag}(I_n, I_m - AB)$ 到 $\text{diag}(I_m, I_n - AB)$ 的初等变换 (会对之后 $I_m - AB$ 与 $I_n - BA$ 相关的分析有很大作用，用例见复习题)、Schur 公式 (4.5.2 例 16，最重要的分块行列变换应用)、Sylvester 秩不等式/Frobenius 秩不等式 (4.5.2 例 2/习题 4.5-3，稍复杂的秩估计常用)

练习：4.5.2 例 3 (分块初等变换计算秩)、习题 4.5-18 (分块初等变换计算行列式)、4.5.2 例 17 (摄动法处理不可逆情况，伴随矩阵相关问题也常用)、4.5.2 例 21 (行列变换的相似操作)、习题 4.5-13 (常用结论、分块对角的计算)

5.2 相抵标准形

定义：相抵标准形

算法: 矩阵相抵标准形与相应过渡矩阵 P, Q 的计算 (定理 1 证明过程)

练习: 5.2.2 例 3 (满秩分解, 结论重要, 尤其秩为 1 时)、习题 5.2-6 (不用标准形可能存在很简单的做法)

* 更多习题可见本讲义 4.2.2, 有大量相抵标准形与逆矩阵结合的操作, 一些结论也会有应用, 如将 $\text{rank } A^2 = \text{rank } A$ 时 A 的相似情况用于习题 5.2-7。

12.3.2 相似

* 这里的 5.5 节中融合了 5.7 节里作为例题的相似对角化, 此节作为相似的核心内容, 存在大量技巧。

4.7 线性映射

定义: 线性映射、像空间、核空间

练习: 4.7.2 例 5 (矩阵与线性映射关系)、习题 4.7-6 (乘可逆矩阵换元的基础)

* 即使前半学期并未学习空间技巧这条路, 学习相似前务必对矩阵看作线性映射有基本的了解。

5.4 相似的概念

定义: 相似、迹、可对角化、置换矩阵

定理: $\text{tr}(AB) = \text{tr}(BA)$ 、相似不变量与基本性质 (性质 1 到 5、5.4.2 例 1 到 4、习题 5.4-1 到 7)、可对角化的基本等价性质、置换相似 (5.4.2 例 7 到 10)

* 对用定义推导相似基本性质需要充分熟悉, 这是因为我们目前并未学习相似的标准形, 能依赖的只有定义。

* 不要忽视本节对可对角化的基本讨论, 此外, 之前习题 4.5-13 的结论结合相似对角化后会有大量作用。

练习: 习题 5.4-11 (有 $AB - BA$ 这类结构时可尝试用 tr 分析)、5.4.2 例 13 (结论重要, 注意数域影响)、习题 5.4-12 (相似初等变换的构造)

5.5 特征值与特征向量

定义: 特征值、特征向量、特征多项式、特征子空间、几何重数、代数重数

算法: 复矩阵的相似三角化 (5.7.2 例 6)

* 相似三角化这个结论建议在此处直接学习, 它可以得到一个极为重要的性质: A 特征值为 $\lambda_1, \dots, \lambda_n$, f 为多项式, 则 $f(A)$ 特征值为 $f(\lambda_1), \dots, f(\lambda_n)$, 以此可以解决很多问题。下方列举的练习不包含用此性质容易解决的。

定理: 特征系统相关的相似不变量、特征多项式各项系数 (最常用为 $\det A$ 等于特征值乘积、 $\text{tr } A$ 等于特征值求和)、几何重数不超过代数重数 (注意证明中结合空间技巧)

练习: 5.5.2 例 2 (实矩阵复特征值成对)、5.5.2 例 3 (注意数域影响特征值)、5.5.2 例 4 (特征方程出发的操作, 各种特殊矩阵特征值性质往往用其进行证明, 建议多练几个)、5.5.2 例 6 (结论重要, 逆的特征值)、5.5.2 例 9 (AB 与 BA 的重要关系)、习题 5.5-12 (利用解空间的定义计算)、习题 5.5-19 (空间的基本技巧)

* 本节还有一些正交阵相关的特征值结论, 注意谈论正交阵的特征值是在实数范围内, 可以等 4.6 节复习完后回看。

5.6 可对角化矩阵

算法: 可对角化阵的对角化 (求解特征值与特征向量)

定理: 不同特征值对应的特征向量线性无关 (特征子空间为直和)、可对角化的等价判定 (与常用的充分条件: n 个特征值互不相同)

练习: 5.6-2 例 1 (通过秩确定几何重数以说明可对角化)、5.6-2 例 2 (相似结果结合相似不变量)、5.6-2 例 8 (多项式结合相似进行操作)、5.6-2 例 16 (特征多项式相关往往需要直接通过特征值处理)、习题 5.6-14 (熟悉 $tI + X$ 类型的矩阵分解, 这里 X 为低秩矩阵, 可进一步满秩分解)

12.3.3 正交阵与对称阵

4.6 内积与正交阵

定义: 正交矩阵、内积、正交、标准正交基、正交补空间 (4.6.2 例 17)、正交投影 (4.6.2 例 18)

* 教材对于正交补与正交投影的讨论其实不完全严谨 (缺少一些存在唯一性), 可参考本讲义 12.2.2。

算法: Schmidt 正交化 (也对应 QR 分解算法, 见 4.6.2 例 5)、最小二乘解法 (优先 4.6.2 例 19)

定理: 正交阵对乘法与逆的封闭性、正交阵等价定义 (行/列为标准正交基)

练习: 4.6.2 例 13 (结论重要, 映射角度的正交阵)、习题 4.6-3 (基本性质)、习题 4.6-7 (常用的化简方式)、习题 4.6-11 到 13 (一些基本的正交相关等式与不等式)

5.7 实对称阵的正交相似

定义: 正交相似

算法: 实对称阵的正交相似对角化 (求解特征值与特征向量并正交化)

定理: 实对称阵特征值为实数 (特征方程与转置结合操作)、实对称阵不同特征值对应的实特征向量正交

* 注意由于对称阵正交相似仍对称, 能正交相似对角化的矩阵一定为实对称阵。

练习: 5.7.2 例 3 (结论重要、特征方程操作)、5.7.2 例 7 (斜对称阵基本性质)、习题 5.7-3 (一般数域相似三角化)、习题 5.7-7 到 8 (注意推广到复数域后如何将转置替换为共轭转置类似处理)

* 建议时间充裕时将教材第五章习题都过一遍以加深对相似操作的印象。

补充 正交相抵

定义: 奇异值、正交相抵

算法: 奇异值分解的计算

* 本段不在教材中, 但按去年确实在考纲内, 可参考本讲义 10.4.2 (或自行搜索其他资料), 注释中有常用计算方法。

6.1 二次型

定义: 二次型、二次型的矩阵、等价、合同、非退化线性替换、正交替换、二次型的标准形、合同标准形

算法: 成对初等行列变换计算对称阵合同对角化

定理: 成对版本 Schur 公式 (6.1.2 例 9)、斜对称阵的合同标准形 (6.1.2 例 10)

练习: 6.1.2 例 11 (正交替换的用法)、6.1.2 例 14 (同时对角化的相关结论)、习题 6.1-5 (标准形后拆分)

* 非常建议学习第六章时以矩阵相关的理论作为基础, 再应用到二次型上, 具体视角可参考本讲义。

6.2 对称阵的合同

定义：规范形、惯性指数、符号差

算法：实(复)对称阵的规范形计算

* 一般不太讨论复二次型，因为复数不能提供不等式信息。

定理：惯性定理(结论与证明思想都挺重要)

练习：6.2.2 例 3(合同类计数)、6.2.2 例 5(一些不太容易用矩阵说明的情况)、习题 6.2-4(合同相关的操作)、习题 6.2-5(通过二次型性质取定符合要求的向量)

6.3 正定与半正定

定义：正定、半正定、负定、半负定

定理：正定与半正定的若干等价定义(教材中相对分散，可参考本讲义 11.2.2，事实上最常用的三种为正交相似对角化、合同标准形与二次型版本定义)、同时相合对角化(6.3.2 例 9)

练习：6.3.2 例 14(利用相合进行不等式分析)、6.3.2 例 16(分块初等变换与不等式)

* 正定与半正定相关的题目可能非常灵活，需要灵活切换各种等价定义与性质，很多等价性的证明本身也是练习，更多练习可见本讲义 11.2.3 与补充题六。

十三 补充：期末复习

§13.1 复习题

* 与期中相同，这里只讲解证明题，但请务必对常考的计算进行一些练习。此外，本讲义第九章的习题也值得一看。

1. 证明 $A_{m \times n} X_{n \times p} B_{p \times q} = O$ 只有零解(即此式成立等价于 $X = O$) 当且仅当 A 列满秩、 B 行满秩。

2. (a) 设

$$M = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}$$

其中 A, B 为 n 阶方阵，且

$$M^{-1} = \begin{pmatrix} C & D \\ E & F \end{pmatrix}$$

其中 C, D, E, F 为 n 阶方阵，求证 $C = F, D = -E$ 。

(b) 计算(用不含分块矩阵的行列式表示)

$$\det \begin{pmatrix} A & -B \\ B & A \end{pmatrix}$$

并证明当 A, B 为实方阵时此行列式非负。

(c) 设

$$M = \begin{pmatrix} A & B & O \\ C & D & C \\ O & -B & A \end{pmatrix}$$

其中 A 为 n 阶可逆方阵， D 为 m 阶方阵。证明 $\text{rank } M \geq 2n$ ，并给出 M 可逆的充要条件(用 A, B, C, D 的秩表示)，在可逆时计算其逆。

3. (a) 设 $A_{m \times n}, B_{n \times m}$, 求证 AB 与 BA 非零特征值相同, 且对应的代数重数、几何重数均相同。
 (b) 证明或否定: 对同阶方阵 A, B , AB 与 BA 相似。
 (c) 证明或否定: 对 $A_{m \times n}, B_{n \times m}$, 若 AB 与 BA 均无零特征值, 则它们相似。
4. (a) 若实方阵 A 满足 $A^T = -A$ 且可对角化, 证明 $A = O$ 。
 (b) 若实方阵 A 满足 $A^T = -A$ 且特征值全为 0, 证明 $A = O$ 。
 (c) 若复方阵 A_1, A_2, \dots, A_n 可对角化, 且 $A_i A_j = A_j A_i$ 对任何 $1 \leq i, j \leq n$ 成立, 证明存在 P 使得对任何 i , $P^{-1} A_i P$ 为对角阵。
5. (a) 设 A 为三阶实对称阵, 满足 $\det(\lambda I - A) = (\lambda - 1)^2(\lambda - 2)$, 且

$$A \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

求所有符合条件的 A 。

- (b) 证明 n 阶实对称矩阵 A 满足 $\text{rank } A = r$ 当且仅当其存在 r 阶非零主子式且不存在更高阶非零主子式。
 (c) 证明或否定: n 阶实对称矩阵 A 满足 $\text{rank } A = r < n$ 当且仅当其存在 r 阶非零主子式且不存在 $r+1$ 阶非零主子式。
6. 设 $x = (x_1, \dots, x_n)$, 且 A 正定, 证明下方定义的 $f(x)$ 为二次型, 且负定:

$$f(x) = \det \begin{pmatrix} A & x^T \\ x & 0 \end{pmatrix}$$

7. (a) 设 e_1, \dots, e_n 是 \mathbb{R}^n 的一组标准正交基, β_1, \dots, β_n 满足

$$\forall 1 \leq i \leq n, \quad \|e_i - \beta_i\| < \frac{1}{\sqrt{n}}$$

求证 β_1, \dots, β_n 构成 \mathbb{R}^n 的一组基。

- (b) 设 n 阶实方阵 A 分块为 (B, C) , 其中 B 为 $n \times m$ 矩阵, $m < n$, 求证

$$(\det A)^2 \leq \det(B^T B) \det(C^T C)$$

- (c) 若实方阵 A 不对称, 且 $A + A^T$ 正定, 求证

$$\det(A + A^T) < \det(2A)$$

8. 设 A 是实对称方阵, 其特征值为 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n$, 求证

$$\lambda_k = \max_{\dim V=k} \min_{0 \neq x \in V} \frac{x^T A x}{x^T x}$$

这里第一个 \max 代表 V 取遍 \mathbb{R}^n 的 k 维子空间。

§13.2 解答

1. 设 $\text{rank } A = r$ 、 $\text{rank } B = s$ ，并设可逆阵 P_A, P_B, Q_A, Q_B 使得

$$A = P_A \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Q_A, \quad B = P_B \begin{pmatrix} I_s & O \\ O & O \end{pmatrix} Q_B$$

则消去可知 $AXB = O$ 等价于

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Q_A X P_B \begin{pmatrix} I_s & O \\ O & O \end{pmatrix} = O$$

设 $Y = Q_A X P_B$ ，则 $X = Q_A^{-1} Y P_B^{-1}$ ，由此 X, Y 或同时为 O ，或同时非 O ，上述方程只有零解等价于

$$\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Y \begin{pmatrix} I_s & O \\ O & O \end{pmatrix} = O$$

只有零解。

直接计算可发现上方方程左侧的左上角 $r \times s$ 阶子矩阵与 Y 相同，其他元素为 0，由此只要 $r \neq n$ 或 $m \neq p$ ， Y 就有元素可任取，原方程有非零解；反之， $r = n, m = p$ 时，要求即化为 $Y = O$ ，只有零解。由此，原方程只有零解等价于 $r = n, m = p$ ，这就是结论。

2. (a) 我们介绍四种处理思路，从前往后越来越整体化：

i. 由 $MM^{-1} = I$ 可以得到

$$\begin{cases} AC - BE = I \\ AD - BF = O \\ BC + AE = O \\ BD + AF = I \end{cases}$$

直接对比可发现 $(C, -E)$ 与 (F, D) 均为方程组

$$\begin{cases} AX + BY = I \\ AY = BX \end{cases}$$

的解 (X, Y) ，由此即要证明此矩阵方程组的解至多唯一。

设 X 的每一列为 x_1, \dots, x_n ， Y 的每一列为 y_1, \dots, y_n ，可发现上述方程组能拆分为

$$\forall i = 1, \dots, n \quad Ax_i + By_i = e_i, \quad Ay_i = Bx_i$$

而这又可以直接写为

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} e_i \\ 0 \end{pmatrix}$$

由系数矩阵可逆知解唯一，得证。

ii. 设 C, D, E, F 的每列为 x_i, y_i, z_i, w_i ，由 $MM^{-1} = I$ 可以得到

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} x_i \\ z_i \end{pmatrix} = \begin{pmatrix} e_i \\ 0 \end{pmatrix}, \quad \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} y_i \\ w_i \end{pmatrix} = \begin{pmatrix} 0 \\ e_i \end{pmatrix}$$

将第二个方程组的上下两行交换（即左乘对应分块行变换阵）可得

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} x_i \\ z_i \end{pmatrix} = \begin{pmatrix} e_i \\ 0 \end{pmatrix} = \begin{pmatrix} -B & A \\ A & B \end{pmatrix} \begin{pmatrix} y_i \\ w_i \end{pmatrix}$$

利用条件 $M^{-1}M = I$ 可得

$$\begin{cases} CA - DB = I \\ CB + DA = O \\ EA - FB = O \\ EB + FA = I \end{cases}$$

尝试可发现此时直接同时左乘 M^{-1} 无法消去化简，需要利用第二行为 0 进一步得到

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} x_i \\ z_i \end{pmatrix} = \begin{pmatrix} e_i \\ 0 \end{pmatrix} = \begin{pmatrix} -B & A \\ -A & -B \end{pmatrix} \begin{pmatrix} y_i \\ w_i \end{pmatrix}$$

这时再同左乘 M^{-1} ，可计算得

$$\begin{pmatrix} x_i \\ z_i \end{pmatrix} = \begin{pmatrix} C & D \\ E & F \end{pmatrix} \begin{pmatrix} -B & A \\ -A & -B \end{pmatrix} \begin{pmatrix} y_i \\ w_i \end{pmatrix} = \begin{pmatrix} O & I \\ -I & O \end{pmatrix} \begin{pmatrix} y_i \\ w_i \end{pmatrix} = \begin{pmatrix} w_i \\ -y_i \end{pmatrix}$$

这就是所求结论。

iii. 由 $MM^{-1} = I$ 可以得到

$$\begin{cases} AC - BE = I \\ AD - BF = O \\ BC + AE = O \\ BD + AF = I \end{cases}$$

进一步计算可发现

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix} \begin{pmatrix} F & -E \\ -D & C \end{pmatrix} = \begin{pmatrix} I & O \\ O & I \end{pmatrix}$$

于是利用逆的唯一性可知

$$\begin{pmatrix} C & D \\ E & F \end{pmatrix} = \begin{pmatrix} F & -E \\ -D & C \end{pmatrix}$$

这就得到了结论。

iv. 先用分块初等变换刻画 M 的对称性：其在两行两列相互交换后，将第一行、第一列乘 -1 仍可变回原矩阵。由此，将此描述用行列变换写出得到

$$\begin{pmatrix} -I & O \\ O & I \end{pmatrix} \begin{pmatrix} O & I \\ I & O \end{pmatrix} M \begin{pmatrix} O & I \\ I & O \end{pmatrix} \begin{pmatrix} -I & O \\ O & I \end{pmatrix} = M$$

两边同取逆，利用分块初等变换的逆可得

$$\begin{pmatrix} -I & O \\ O & I \end{pmatrix} \begin{pmatrix} O & I \\ I & O \end{pmatrix} M^{-1} \begin{pmatrix} O & I \\ I & O \end{pmatrix} \begin{pmatrix} -I & O \\ O & I \end{pmatrix} = M^{-1}$$

也即 M^{-1} 的对称性与 A 相同，再代入分块形式得结论。

(b) * 思路：回想练习题中行列变换可得到 (两个等号分别为第一行加上第二行、第一列加上第二列)

$$\det \begin{pmatrix} A & B \\ B & A \end{pmatrix} = \det \begin{pmatrix} A+B & A+B \\ B & A \end{pmatrix} = \det \begin{pmatrix} A+B & O \\ B & A-B \end{pmatrix} = \det(A+B) \det(A-B)$$

为配出这种形式，我们事实上需要第一行加第二行的 t 倍，使得此时第二列是第一列的 s 倍，从而可以消去。由此，第一行加第二行 t 倍、第二列减第一列 s 倍得到

$$\det \begin{pmatrix} A & -B \\ B & A \end{pmatrix} = \det \begin{pmatrix} A+tB & -B+tA \\ B & A \end{pmatrix} = \det \begin{pmatrix} A+tB & -B+tA-s(A+tB) \\ B & A-sB \end{pmatrix}$$

若希望右上角为 O , 对比系数可发现须 $s = t$ 、 $st = -1$, 也即可取 $s = t = i$, 最终得到原行列式为

$$\det \begin{pmatrix} A + iB & O \\ B & A - iB \end{pmatrix} = \det(A + iB) \det(A - iB)$$

由于共轭的加法/乘法为加法/乘法共轭, 可知 A, B 为实矩阵时 $\det(A + iB) = \overline{\det(A - iB)}$, 从而右侧为 $|\det(A + iB)|^2$ 非负。

(c) 利用 Schur 公式操作可得 (可将下面每个分块矩阵看作某行/列减某行/列的倍数)

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

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

由于每步均乘可逆矩阵, 不改变秩, 可得 (由 A 可逆)

$$\text{rank } M = 2 \text{rank } A + \text{rank } D = 2n + \text{rank } D$$

于是 $\text{rank } M \geq 2n$, 且其可逆等价于 $\text{rank } M = 2n + m$, 即 $\text{rank } D = m$, D 可逆。利用 $(XY)^{-1} = Y^{-1}X^{-1}$ 可知在 M 可逆时其逆为

$$\begin{pmatrix} I & -A^{-1}B & \\ & I & \\ & & I \end{pmatrix} \begin{pmatrix} I & & \\ & I & \\ & & A^{-1}B & I \end{pmatrix} \begin{pmatrix} A^{-1} & & \\ & D^{-1} & \\ & & A^{-1} \end{pmatrix} \begin{pmatrix} I & & \\ & I & -CA^{-1} \\ & & I \end{pmatrix} \begin{pmatrix} I & & \\ -CA^{-1} & I & \\ & I & I \end{pmatrix}$$

3. (a) 不妨设 $m \geq n$ 。

由教材 4.5 节 $\det(I - AB) = \det(I - BA)$, 从而 $\lambda \neq 0$ 时

$$\det(\lambda I - AB) = \lambda^m \det(I - (\lambda^{-1}A)B) = \lambda^m \det(I - B(\lambda^{-1}A)) = \lambda^{m-n} \det(\lambda I - BA)$$

利用连续性, $\lambda = 0$ 时两侧也相等, 由此, AB 的特征多项式为 BA 的特征多项式乘 λ^{m-n} , 这就已经说明了非零特征值的代数重数相同。另一方面, 在 4.5 节的操作中由乘可逆阵不改变秩可以得到

$$\text{rank}(I - AB) = \text{rank}(I - BA) + m - n$$

从而在 $\lambda \neq 0$ 时

$$\text{rank}(\lambda I - AB) = \text{rank}(I - (\lambda^{-1}A)B) = \text{rank}(I - B(\lambda^{-1}A)) + m - n = \text{rank}(\lambda I - BA) + m - n$$

利用解空间维数定理即说明了几何重数相同。

* 一般矩阵相似的相关问题由于尚未学习标准形结论, 往往需要直接从定义推导。这也是后半学期唯一一个没有学习标准形的情况。

(b) 结论是错误的, 如

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

可验证 $AB = O$ 但 $BA \neq O$, 不相似。

(c) 结论成立。由于不存在零特征值等价于可逆, 条件即 AB, BA 均可逆。利用 $\text{rank}(AB) \leq \text{rank } A$ 、 $\text{rank}(BA) \leq \text{rank } A$ 可知只能 A, B 均为方阵, 由此 AB 可逆能推出 A 可逆, 有

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

4. (a) 提供两种不同的思路:

i. 由 A 可对角化, 设 $A = P^{-1}DP$, D 为对角阵, 则条件化为 (对角阵转置为自身)

$$P^TDP^{-T} = -P^{-1}DP$$

也即

$$D = -P^{-T}P^{-1}DPP^T$$

设 $B = PP^T$, 上式可以化为 $BD + DB = O$, 且由 P 可逆知 B 正定。由于 B 的对角元均为正, 考察 $BD + DB$ 的对角元可发现必须 D 对角元全为 0, 从而 $D = O$, 即得 $A = O$ 。

* 利用正交相似标准形事实上可以证明, 若 B 正定, 则 $BX + XB = O$ 当且仅当 $X = O$ 。

ii. 由 A 为实斜对称阵, 先证明其特征值均为 ci , 其中 c 为某实数。

考虑特征方程 $Ax = \lambda x$, 两侧同作共轭转置得到 (实方阵的共轭为自身)

$$-x^H A = \bar{\lambda} x^H$$

第一式同左乘 x^H 、第二式同右乘 x 有

$$x^H A x = \lambda x^H x = -\bar{\lambda} x^H x$$

由于 x 为非零复向量, $x^H x > 0$, 从而 $\lambda = -\bar{\lambda}$, 设 $\lambda = a + bi$, a, b 为实数即可发现 $a = 0$, 得证。

由于实方阵 A 可对角化 (注意谈论实方阵的可对角化是在实数域中), 其特征值必然全为实数, 由此只有 ci 均为 0, 也即其特征值全为 0, 从而考虑对角化知

$$A = P^{-1}OP = O$$

(b) 由 A^2 的特征值为 A 特征值的平方 (利用相似三角化可证明) 得到 A^2 特征值全为 0, 因此利用 tr 为特征值之和可得 $\text{tr}(A^2) = 0$, 替换其中一个 A 为 $-A^T$ 可得

$$\text{tr}(A^T A) = -0 = 0$$

而直接计算可发现 $\text{tr}(A^T A)$ 为 A 所有元素的平方和, 因此只能 $A = O$ 。

(c) 先证明两个方阵 A, B 的情况。由 A 可对角化设

$$R^{-1}AR = \text{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \dots, \lambda_k I_{n_k})$$

这里 $\lambda_1, \dots, \lambda_k$ 互不相同。记 $B_0 = R^{-1}BR$, 由 $AB = BA$ 直接计算可得 $R^{-1}ARR^{-1}BR = R^{-1}BRR^{-1}AR$, 即

$$\text{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \dots, \lambda_k I_{n_k})B_0 = B_0 \text{diag}(\lambda_1 I_{n_1}, \lambda_2 I_{n_2}, \dots, \lambda_k I_{n_k})$$

设 B_0 第 i 行第 j 列为 b_{ij} , 计算上述矩阵乘法可发现上式成立当且仅当

$$B_0 = \text{diag}(B_1, B_2, \dots, B_k)$$

其中 B_i 为 n_i 阶方阵。

由 B 可对角化、 B_0 与 B 相似可知 B_0 可对角化，即其任何特征值的代数重数等于几何重数。考虑特征值 λ ，由定义可发现其在 B_0 中的代数重数等于其在 B_i 中的代数重数之和，在 B_0 中的几何重数等于其在 B_i 中的几何重数之和（若不为特征值则代数/几何重数为 0）。由于在每个 B_i 中， λ 的几何重数均不超过代数重数，若有某个 B_i 使得小于号成立， λ 在 B_0 的几何重数应小于代数重数，矛盾。由此，每个 B_i 均可对角化，设 $Q_i^{-1}B_iQ^i$ 为对角阵 D_i ，并记 $Q = \text{diag}(Q_1, Q_2, \dots, Q_k)$ ，计算有

$$Q^{-1}B_0Q = \text{diag}(D_1, D_2, \dots, D_k)$$

$$Q^{-1} \text{diag}(\lambda_1 I_{n_1}, \dots, \lambda_k I_{n_k}) Q = \text{diag}(\lambda_1 Q_1^{-1} Q_1, \dots, \lambda_k Q_k^{-1} Q_k) = \text{diag}(\lambda_1 I_{n_1}, \dots, \lambda_k I_{n_k})$$

于是由 D_i 对角可知 $Q^{-1}R^{-1}BRQ$ 与 $Q^{-1}R^{-1}ARQ$ 均为对角阵，取 $P = RQ$ 即满足要求。

接着进行归纳。若结论对任何 $n-1$ 个方阵成立，对 n 个方阵时，取 R 使得 $R^{-1}A_iR$ 对 $i = 1, \dots, n-1$ 为对角阵 F_i ，并记 $A_0 = R^{-1}A_nR$ 。直接计算可发现 A_0 与任何 F_i 可交换，从而 $(A_0)_{st}$ 非零当且仅当任何 F_i 的第 s 个对角元等于第 t 个对角元（这里称为 s, t 等价）。由此，进一步将每个 s 所在等价类对应的主子矩阵进行对角化（类似上方将 B_i 对角化），并拼接 Q_i 成为新的 Q ，即可计算发现 $P = RQ$ 可将 A_1 到 A_n 同时对角化，得证。

* 归纳步的详细说明较复杂，这里只进行简单介绍，能感受到正确即可。

5. (a) 我们介绍两种方法，第一种方法较为直接操作性，第二种方法则更为本质：

i. 由于其特征值为两个 1、一个 2，利用实对称阵可正交相似对角化设其为

$$A = Q^T \text{diag}(2, 1, 1)Q$$

其中 Q 正交。将第二条件代入并两侧左乘 Q 得到

$$\text{diag}(2, 1, 1)Q \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = Q \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = 2Q \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

由此， $Q(1, 1, 1)^T$ 是 $\text{diag}(2, 1, 1)$ 属于特征值 2 的特征向量（由 Q 可逆，其非零），从而直接计算可得

$$Q \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

* 至此，若只需要求出一个解，利用本讲义 10.4.1 的 Householder 阵进行构造即可，如。下面观察全部解的求法。

考虑两侧的模长，由乘正交阵不改变模长 ($x^T x = x^T Q^T Q x$) 可得 $3 = a^2$ ，于是 $a = \pm\sqrt{3}$ 。若 $a = -\sqrt{3}$ ，记 $\tilde{Q} = -Q$ ，可发现 $\tilde{Q}^T \text{diag}(2, 1, 1) \tilde{Q} = Q^T \text{diag}(2, 1, 1)Q$ ，且

$$\tilde{Q} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \sqrt{3} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

由此，想要求解不重复的 A ，只需考虑 $a = \sqrt{3}$ 的情况即可。由于上式可以改写为

$$\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = Q^T \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

为配凑出此形式，我们重新改写对角化

$$A = Q^T(I + \text{diag}(1, 0, 0))Q = I + Q^T \text{diag}(1, 0, 0)Q$$

由于

$$\text{diag}(1, 0, 0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^T$$

即可得到

$$A = I + \left(Q^T \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right) \left(Q^T \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right)^T = I + \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}^T = \frac{1}{3} \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}$$

于是这是 A 的唯一解。

- ii. 由 A 为实对称矩阵，其不同特征值对应的特征向量正交，且由其可对角化所有特征子空间直和为全空间。由第一个条件 2 的代数重数为 1，而由第二个条件可知 $(1, 1, 1)^T$ 是 2 的一个特征向量，于是 2 的特征子空间即为 $(1, 1, 1)^T$ 生成的空间

$$U_2 = \langle (1, 1, 1)^T \rangle$$

由 1 的特征向量与 2 的特征向量正交，设 1 的特征子空间为 U_1 ，有 $U_1 \subset U_2^\perp$ ，而由可对角化 $\dim U_1 + \dim U_2 = 3$ ，于是只能 $U_1 = U_2^\perp$ 。利用本讲义 12.2.2 可构造出 U_1 、 U_2 的标准正交基

$$U_1 = \left\langle \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \end{pmatrix} \right\rangle, \quad U_2 = \left\langle \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} \right\rangle$$

将此三个向量作为列向量顺次拼成正交矩阵 T ，利用特征向量的定义有

$$AT = \text{diag}(1, 1, 2)T$$

从而直接计算可得

$$A = T^T \text{diag}(1, 1, 2)T = \frac{1}{3} \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}$$

下面来证明唯一性，我们说明更一般的结论：若两可对角化矩阵的特征值与对应的特征子空间均相同，则它们相同。这样，由于 U_1 、 U_2 都是确定的，即可说明 A 唯一。

设上述两个可对角化矩阵为 X 、 Y ，由可对角化， X 的特征向量中可选出全空间一组基，设它们为 $\alpha_1, \dots, \alpha_n$ ，对应特征值为 $\lambda_1, \dots, \lambda_n$ ，由定义可知

$$X\alpha_i = \lambda_i \alpha_i$$

于是将 $\alpha_1, \dots, \alpha_n$ 拼成可逆矩阵 P 后有

$$XP = \text{diag}(\lambda_1, \dots, \lambda_n)P$$

也即

$$X = P^{-1} \text{diag}(\lambda_1, \dots, \lambda_n)P$$

另一方面，由于 α_1 属于 X 关于 λ_1 的特征子空间，由 X, Y 特征值与对应特征子空间相同， α_1 也属于 Y 关于 λ_1 的特征子空间，即

$$Y\alpha_1 = \lambda_1\alpha_1$$

同理，此式对任何 i 成立，于是与上方完全类似得到

$$Y = P^{-1} \operatorname{diag}(\lambda_1, \dots, \lambda_n) P = X$$

(b) • 仅当

由于 $\operatorname{rank} A < r + 1$ ，其不存在 $r + 1$ 阶非零子式，自然也不存在 $r + 1$ 阶非零主子式。由此，只需证明 A 存在 r 阶非零主子式。

设 A 的列向量的一个极大线性无关组是第 i_1, \dots, i_r 列，则利用对称性其 i_1, \dots, i_r 行构成行向量的极大线性无关组，由此利用期中考试第 5 题（证明见本讲义 7.1.2）可得这些行、列相交处的 r 阶主子式非零，得证。

• 当

由于其存在 r 阶可逆主子式，秩至少为 r ，若 $\operatorname{rank} A = r_0 > r$ ，由仅当部分可知其存在 r_0 阶非零主子式，矛盾。

(c) 结论是错误的，如

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

它存在 2 阶可逆主子式，不存在 3 阶可逆主子式，且秩为 4。

6. 利用 Schur 公式可得

$$f(x) = \det \begin{pmatrix} I & \mathbf{0} \\ -xA^{-1} & 1 \end{pmatrix} \begin{pmatrix} A & x^T \\ x & 0 \end{pmatrix} \begin{pmatrix} I & -A^{-1}x^T \\ \mathbf{0} & 1 \end{pmatrix} = \det \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & -xA^{-1}x^T \end{pmatrix} = -(\det A)x A^{-1}x^T$$

从此形式可直接得到

$$f(x) = x(-(\det A)A^{-1})x^T$$

从而其为二次型。由正定设 $A = P^T P$ ， P 可逆，则利用行列式与逆的性质有

$$f(x) = -(\det P)^2 x(P^{-1}P^{-T})x^T = -(\det P)^2 (xP^{-1})(xP^{-1})^T$$

由 P^{-1} 可逆，当 $x \neq 0$ 时 $xP^{-1} \neq 0$ ，此时利用内积性质即得 $(xP^{-1})(xP^{-1})^T > 0$ ，再由 $\det P \neq 0$ 可知 $f(x) < 0$ ，从而得证负定。

7. (a) 由于只知道 β_i 与 e_i 的关联，设

$$\beta_i = \sum_{j=1}^n b_{ij} e_j$$

则条件即

$$(\beta_i - e_i)^T (\beta_i - e_i) < \frac{1}{n}$$

利用 $i \neq j$ 时 $e_i^T e_j = 0$ 、 $e_i^T e_i = 1$ ，可直接计算得上式等价于

$$\sum_{j \neq i} b_{ij}^2 + (1 - b_{ii})^2 < \frac{1}{n}$$

而假设 b_{ij} 拼成矩阵 B , 由教材结论, 证明 β_1, \dots, β_n 为一组基只需 B 可逆。

* 注意到条件意味着 b_{ii} 接近 1, 其他 b_{ij} 接近 0, 且对每行独立, 可以联想到学习过的主对角占优阵的可逆性 (见教材 3.3.2 例 11)。

只需证明

$$|b_{ii}| > \sum_{j \neq i} |b_{ij}|$$

利用 Cauchy 不等式可知

$$\left(\sum_{j \neq i} |b_{ij}| \right)^2 \leq \sum_{j \neq i} 1^2 \sum_{j \neq i} b_{ij}^2 = (n-1) \sum_{j \neq i} b_{ij}^2$$

再由条件即可知

$$\left(\sum_{j \neq i} |b_{ij}| \right)^2 < (n-1) \left(\frac{1}{n} - (1-b_{ii})^2 \right)$$

由此只需证明当 $(1-b_{ii})^2 < 1/n$ 时

$$b_{ii}^2 \geq (n-1) \left(\frac{1}{n} - (1-b_{ii})^2 \right)$$

由于此为开头向上的二次函数 ≥ 0 问题, 直接求导可发现其最小值在 $b_{ii} = \frac{n-1}{n}$ 取到, 为 0, 从而得证。

* 下半学期必须对 Cauchy 不等式与一般形式 Binet-Cauchy 公式充分熟悉, 以处理不等式问题。

(b) 此题为教材习题, 我们给出前半学期与后半学期的不同做法:

i. 对右侧利用 Binet-Cauchy 公式展开可得 $\det(B^T B)$ 为

$$\sum_{1 \leq v_1 < \dots < v_m \leq n} B^T \begin{pmatrix} 1 & \dots & m \\ v_1 & \dots & v_m \end{pmatrix} B \begin{pmatrix} v_1 & \dots & v_m \\ 1 & \dots & m \end{pmatrix} = \sum_{1 \leq v_1 < \dots < v_m \leq n} B \begin{pmatrix} v_1 & \dots & v_m \\ 1 & \dots & m \end{pmatrix}^2$$

同理 $\det(C^T C)$ 为

$$\sum_{1 \leq u_1 < \dots < u_{n-m} \leq n} C \begin{pmatrix} u_1 & \dots & u_{n-m} \\ 1 & \dots & n-m \end{pmatrix}^2$$

利用 Cauchy 不等式, 将 v_1, \dots, v_m 与 u_1, \dots, u_{n-m} 可以拼成 $1, \dots, n$ 的项配对, 可得右侧大于等于

$$\left(\sum_{\substack{1 \leq v_1 < \dots < v_m \leq n \\ 1 \leq u_1 < \dots < u_{n-m} \leq n \\ v_i \neq u_j}} B \begin{pmatrix} v_1 & \dots & v_m \\ 1 & \dots & m \end{pmatrix} C \begin{pmatrix} u_1 & \dots & u_{n-m} \\ 1 & \dots & n-m \end{pmatrix} \right)^2$$

而根据 Laplace 展开定理, 这就是 $(\det A)^2$, 得证。

ii. 由于 $(\det A)^2 = \det A^T A$, 直接分块计算可知左侧为

$$\det \begin{pmatrix} B^T B & B^T C \\ C^T B & C^T C \end{pmatrix}$$

由于 A 半正定, 利用本讲义 11.2.3 第六题可得结论成立。

(c) 记 $X = A + A^T$ 、 $Y = A - A^T$, 则 X 实对称, Y 斜对称, 且根据条件, X 正定, $Y \neq O$ 。要证的结论化为

$$\det X \leq \det(X + Y)$$

设 $X = P^T P$, 其中 P 可逆, 则上方不等式即

$$\det(P^T P) \leq \det(P(I + P^{-1} Y P^{-T}) P^T)$$

利用 Binet-Cauchy 公式与 $\det P \neq 0$ 即得要证

$$\det(I + P^{-1} Y P^{-T}) \geq 1$$

设 $Y_0 = P^{-1} Y P^{-T}$, 由其与 Y 相合可知为非零斜对称阵。在第四题的第二种思路中, 我们已经证明了 Y_0 的特征值均为纯虚数或 0。另一方面, 由 Y_0 为实方阵, 设其特征多项式 $f(\lambda) = \det(\lambda I - Y_0)$, 可计算知

$$\overline{f(\lambda)} = f(\bar{\lambda})$$

于是利用特征多项式, Y_0 的特征值 $\lambda = ci$ 与 $\bar{\lambda} = -ci$ 的代数重数相同。

* 对实矩阵 A , 上述方法可说明其特征值 $a + bi$ 与 $a - bi$ ($a, b \in \mathbb{R}$) 的代数重数相同, 类似可证几何重数相同, 此性质称为虚特征值成对出现。

综合以上, 设 Y_0 的特征值为 $c_1i, -c_1i, c_2i, -c_2i, \dots, c_ki, -c_ki, 0, \dots, 0$, 可得 $I + Y_0$ 的特征值为每个特征值对应加 1, 从而根据行列式为特征值乘积可知

$$\det(I + Y_0) = \prod_{j=1}^k (1 + c_j i)(1 - c_j i) = \prod_{j=1}^k (1 + c_j^2)$$

只需证明非零斜对称方阵的特征值不可能全为 0, 即可得到上式严格大于 1, 而这已经在 4(b) 说明了。

8. 设 $A = Q^T D Q$, 其中 $D = \text{diag}(\lambda_1, \dots, \lambda_n)$, Q 为正交阵, 则计算得原式右侧化为

$$\max_{\dim V=k} \min_{0 \neq x \in V} \frac{(Qx)^T D (Qx)}{(Qx)^T (Qx)}$$

下面先证明, \mathbb{R}^n 的子空间之间的映射

$$\mathcal{Q}(V) = \{Qx \mid x \in V\}$$

是所有 k 维子空间之间的双射, 由此, 换元 $Qx = y$ 只是换了一个空间进行比较, 但对所有 k 维子空间取最大值结果不变:

- 设 V 的一组基为 $\alpha_1, \dots, \alpha_k$, 由定义可知 $Q\alpha_1, \dots, Q\alpha_k$ 可生成 $\mathcal{Q}(V)$, 进一步通过 Q 可逆验证线性无关即得它们构成 $\mathcal{Q}(V)$ 的一组基, 于是其的确为 k 维子空间。
- 给定任何子空间 V , 设其一组基为 $\alpha_1, \dots, \alpha_k$, 考虑 $Q^{-1}\alpha_1, \dots, Q^{-1}\alpha_k$ 生成的子空间即为其原像, 于是 \mathcal{Q} 为满射。
- 若两个空间满足 $\mathcal{Q}(V_1) = \mathcal{Q}(V_2)$, 也即

$$\{Qx \mid x \in V_1\} = \{Qx \mid x \in V_2\}$$

设此空间为 V , 对任何 $x \in V$, 有 $Q^{-1}x \in V_1 \cap V_2$, 而与第一步证明相同得所有 $Q^{-1}x$ 为 k 维子空间, 于是只能 $V_1 \cap V_2 = V_1 = V_2$, 于是 \mathcal{Q} 为单射。

这样, 换元后原式右侧就化为了

$$\max_{\dim V=k} \min_{0 \neq y \in V} \frac{y^T D y}{y^T y}$$

我们最后分两步证明:

- 先证明存在一个 k 维子空间 V_0 使得

$$\lambda_k = \min_{0 \neq y \in V_0} \frac{y^T Dy}{y^T y}$$

考虑 V_0 为后 $n - k$ 个分量均为 0 的 n 维向量构成的线性空间，则 V_0 中

$$\frac{y^T Dy}{y^T y} = \frac{\sum_{i=1}^k y_i^2 \lambda_i}{\sum_{i=1}^k y_i^2} \geq \frac{\sum_{i=1}^k y_i^2 \lambda_k}{\sum_{i=1}^k y_i^2}$$

且当 $y = e_k$ 时可取到。

- 再证明对任何 k 维子空间 V 有

$$\lambda_k \geq \min_{0 \neq y \in V} \frac{y^T Dy}{y^T y}$$

我们证明，任何 k 维子空间 V 中都有一个非零向量的前 $k - 1$ 个分量均为 0，设此向量为 z ，即有

$$\min_{0 \neq y \in V} \frac{y^T Dy}{y^T y} \leq \frac{z^T Dz}{z^T z} = \frac{\sum_{i=k}^n z_i^2 \lambda_i}{\sum_{i=k}^n z_i^2} \leq \frac{\sum_{i=k}^n z_i^2 \lambda_k}{\sum_{i=k}^n z_i^2} = \lambda_k$$

设 V 的一组基为 $\alpha_1, \dots, \alpha_k$ ， α_j 的第 i 个分量为 a_{ij} 。考虑关于 $\lambda_1, \dots, \lambda_k$ 的方程组

$$\forall i = 1, \dots, k-1, \quad \sum_{j=1}^k a_{ij} \lambda_j = 0$$

由于这是一个未知数个数比方程个数多的齐次线性方程组，一定存在非零解 $\lambda_1, \dots, \lambda_k$ ，取

$$z = \lambda_1 \alpha_1 + \dots + \lambda_k \alpha_k$$

由方程组可发现 z 的前 $k - 1$ 个分量为 0，再由 λ_i 不全为 0 与 α_i 为一组基可知 z 非零，从而得证。

* 最后一部分证明即为期中复习题第四题结论的应用。

* 有能用的标准形尽量化为标准形，其他情况只能根据经验进行处理，熟悉一些常用的小结论会很有帮助。
* 祝大家期末好运。