

# 高等代数选讲

作者: ZZH

本讲义使用于 HEO 高等代数II 课程中,  
内容由  $\text{\LaTeX}$  编译, 图片使用 GeoGebra 绘制.  
参考多部书目, 仅用作学习讨论和笔记需求. 如有错误, 欢迎指正.  
使用时间 2025/3/17  
最后编译时间 March 17, 2025

每逢拾笔, 愿母亲安康.

本笔记摘选自巴猪数学讲义: 第二卷高等代数.

谨以彼书赠与女友与巴猪的陪伴.

祝诸位在数学上逢见挚爱.

摘自 Grassmann 扩张论. 您的理论如今已是大学入学必学的课程.

我始终坚信我在此科学上所付出的劳动不会白费, 它耗尽了我生命中最重要阶段, 让我付出了超常的努力. 我当然知道我给出的这门科学的形式还不完善, 它一定是不完善的. 但是, 我知道而且有义务在此声明 (可能有人会认为我很狂妄), 即使这一成果再过十七年或更长时间还不被使用, 也没有真正融入到科学的发展之中, 它冲出遗忘的尘埃现身的时候也一定会到来, 现在沉睡着的思想结出硕果的那一天一定会到来. 我知道, 如果我今天还不能 (如我至今徒劳地期望那样) 把学者们吸引到我的周围, 用这些思想帮助他们成果累累, 促使其进步, 丰富其学识, 那么这种思想在将来一定会重生, 或许以另一种新形式, 与时代发展水乳交融. 因为真理是永恒不灭的.

摘自 Grothendieck 丰收与播种. 愿数学的远端没有硝烟.

我可以用同样的坚果意象来说明第二种方法:

第一种类比: 我首先想到的是将坚果浸泡在某种软化液体中——为何不直接用水呢? 你偶尔摩擦坚果以促进液体渗透, 其余时间则静待其变. 经过数周甚至数月, 外壳逐渐变得柔软——当时机成熟, 仅需用手轻轻一压, 它便会如完美成熟的牛油果般自然裂开!

几周前, 我的脑海中闪现了另一幅图像.

那片等待被理解的未知, 仿佛一片坚硬的土地或泥灰岩, 抗拒着侵入... 而海水无声无息地悄然上涨, 看似毫无动静, 潮水遥远得几乎听不见声响... 但它终将温柔包裹住那顽固的物体.

代数就是这片 Rising Sea.

# Lecture 3 二次型与合同变换

## 3.1 引入

### 3.1.1 基本定义

在谈起高等代数的时候, 我们一般都喜欢从方程来出发. 在解方程的时候我们有矩阵表示形式  $\mathbf{Ax} = \mathbf{b}$ , 其中  $\mathbf{A} \in \mathcal{M}_{m,n}(F)$ ,  $\mathbf{x} \in F^n$ ,  $\mathbf{b} \in F^m$ . 这个过程我们可以用映射来看, 记映射

$$\begin{aligned}\varphi_{\mathbf{A}} : F^n &\rightarrow F^m \\ \mathbf{x} &\mapsto \mathbf{Ax}\end{aligned}\tag{3.1}$$

我们用这样的写法来表示一个映射, 映射从定义域射入一个集合的一个子集, 这个子集就是值域; 一个原象对应唯一的一个像. 于是, 本质上一个矩阵乘以一个向量, 是将它映入了另一个向量空间之中. 我们记值域中的一点  $\mathbf{b} \in F^m$  所对应的原象的集合为  $\varphi_{\mathbf{A}}^{-1}(\mathbf{b})$ , 从解方程的理论来看, 当然, 原象不一定只有一个点, 而找出这整个原象集的过程就是求解方程的过程.

特别地, 由于显式方程都可以被函数的零点所描述, 因此每个显式方程都可以写成原象集的语言. 从而方程和映射总是存在一个对偶的关系, 从这点而言, 分析同代数只不过是视角不同但殊途同归的数学分支罢了.

注意到 (3.1) 所规定的映射满足如下的性质, 都能够由矩阵的加法乘法数乘所立即得到:

$$\begin{aligned}\varphi_{\mathbf{A}}(\mathbf{x} + \mathbf{y}) &= \varphi_{\mathbf{A}}(\mathbf{x}) + \varphi_{\mathbf{A}}(\mathbf{y}) \\ \varphi_{\mathbf{A}}(k\mathbf{x}) &= k\varphi_{\mathbf{A}}(\mathbf{x})\end{aligned}$$

其中  $k \in F$  注意到由于向量空间运算的性质, 式子的右端也都是属于向量空间的. 我们称, 满足以上性质的映射  $\varphi_{\mathbf{A}}$  为**线性映射**.

为了满足运算封闭性的要求, 我们需要让线性映射是构建在两个**线性空间**之间的, 既然我们能够从矩阵乘向量推广出线性映射, 那么从向量空间过渡到线性空间也是十分自然的事. 具体的理论我们在后续的线性空间章节再叙.

事实上, 线性映射和矩阵是一一对应的, 矩阵乘法对应着线性映射的复合, 矩阵的转置则对应着线性空间的对偶空间思想, 而矩阵自身就是一个线性空间, 对应着两个线性空间之间的线性映射全体也是一个线性空间.

我们再来看另一种乘法:

$$\begin{aligned} \langle \cdot, \cdot \rangle_A: F^m \times F^n &\rightarrow F \\ (\alpha, \beta) &\mapsto \langle \alpha, \beta \rangle_A = \alpha^T A \beta \end{aligned} \quad (3.2)$$

根据我们刚才所定义的线性映射的概念, 不难发现  $\langle \cdot, \cdot \rangle_A$  如果在固定一个分量成为  $\langle \alpha_0, \cdot \rangle_A$  或者  $\langle \cdot, \beta_0 \rangle_A$  时, 就会成为关于另外一个变量的线性映射. 我们称这种固定某一变量后就会成为线性映射的二元映射为**双线性映射**, 特别地, 就像此时我们直接将域上的向量空间射向了域  $F$ , 我们此时称  $\langle \cdot, \cdot \rangle_A$  为**双线性型**.

当双线性型的两个变量都取自同一向量空间  $F^n$  时, 此时矩阵  $A \in \mathcal{M}_n(F)$ , 我们可以来讨论它的对称性. 当双线性型满足对任意的  $\alpha, \beta \in F^n$  均满足

$$\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$$

时, 我们称其为**对称的**.

我们随即就能给出一个对称的双线性型的例子, 它叫做**内积**, 是我们将后面的**内积空间**中着重研究的. 对于向量空间而言, 内积就是普通的标准内积, 即

$$\langle \alpha, \beta \rangle = \alpha^T \beta$$

实际就是将我们约定中的矩阵  $A$  取为单位矩阵  $I_n$  罢了.

而对于我们已给定的双线性型 (3.2), 由于一个数的转置自然是不变的, 有  $\alpha^T A \beta = \beta^T A^T \alpha$  并且注意到  $e_i^T A e_j = a_{ij}$  从而 (3.2) 构成双线性型当且仅当  $A^T = A$  即  $A$  为**对称矩阵**.

对于一个对称的双线性型, 我们取对角线映射复合双线性型:

$$\begin{aligned} F^n &\rightarrow F^n \times F^n \rightarrow F \\ \alpha &\mapsto (\alpha, \alpha) \mapsto \langle \alpha, \alpha \rangle_A \end{aligned} \quad (3.3)$$

我们称这样的映射为**二次型**. 一个  $n$  元二次型即为

$$f = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{1 \leq i < j \leq n} a_{ij} x_i x_j$$

由于  $i, j$  的置换对整个二次型的值完全不影响, 因此即使我们可以选取不对称的矩阵来表示二次型, 我们还是使用对称矩阵来表示二次型.

**Example 3.1.1.** 一个二维情形的二次型为  $f = a_{11}x_1^2 + a_{22}x_2^2 + bx_1x_2$  则它的二次型可以用双线性型来表示, 但是只有对于对称的双线性型, 这种表示是唯一的. 例如

$$f = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & b \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

再例如

$$f = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & c \\ b-c & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

只要  $c$  任意取值, 上式都成立. 于是我们更喜欢取  $a_{12} = a_{21} = b/2$ , 从这一步看已经表明唯一性, 得到

$$f = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

由上例很容易可以看出下面的结论.

**Theorem 3.1.1.** 一个矩阵  $A \in \mathcal{M}_n(F)$  可与一个二次型  $f = \mathbf{x}^T A \mathbf{x}$  对应; 而一个对称矩阵  $A$  与一个二次型  $f = \mathbf{x}^T A \mathbf{x}$  1-1 对应; 从而一个矩阵  $A$  可以对应一个对称矩阵  $(A^T + A)/2$ .

由此, 我们对一个二次型的研究就可以只关注它所对应的对称矩阵.

读者很容易联想到, 我们只需要约定矩阵的形状, 例如约定二次型对应的矩阵为一个上三角阵, 那么也自然能够得到一个 1-1 对应的关系. 事实上的确如此, 选用对称矩阵主要在于二次型本身具备的对称意义, 以及用对称矩阵能够非常自然地解决正定性的问题, 更进一步, 对称矩阵可以进行谱分解, 从而转化成相当好的标准型. 这些都是我们本章节的目标.

### 3.1.2 研究目的

不同于先前所学的矩阵和多项式的理论, 实际二次型的研究是比较专门的一块领域, 对它的推广的**多重线性函数**则是代数学的重点, 不过也是交由研究生阶段的. 对二次型本体的研究对代数并无太大的作用, 唯一的意义就是对二次方程的分类, 而受限于它只能处理二次的情形, 除却在几何上的意义, 并无多大价值. 但是, 就像工科生需要学习高等数学和线性代数一样, 他们所学的数学基础是为了工科需要的其他课程作为根本的底子, 诸如通信所需要的 Fourier 分析, 金融所需要的概率论和 PDE, 化学所需要的群论等等, 更不用提几乎处处都在的微积分和矩阵运算. 二次型研究的意义主要在于数学的其他分支上.

1. **多元函数的 Taylor 公式:** 读者或许此时还未涉足多元函数, 但是读者可以尝试想象到, 多元函数的 Taylor 公式有如下的样貌:

$$\begin{aligned} f(x_0 + \Delta x, y_0 + \Delta y) &= f(x_0, y_0) + \left( \Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y} \right) (x_0, y_0) \\ &+ \frac{1}{2!} \left( \Delta x^2 \frac{\partial^2 f}{\partial x^2} + 2\Delta x \Delta y \frac{\partial^2 f}{\partial x \partial y} + \Delta y^2 \frac{\partial^2 f}{\partial y^2} \right) (x_0, y_0) + o(\sqrt{\Delta x^2 + \Delta y^2}) \\ &= f(x_0, y_0) + \left( \Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y} \right) (x_0, y_0) + \frac{1}{2!} \left( \Delta x \frac{\partial^2}{\partial x} + \Delta y \frac{\partial}{\partial y} \right)^2 f(x_0, y_0) \\ &+ o(\sqrt{\Delta x^2 + \Delta y^2}) \end{aligned}$$

注意到其中的二次项本质上是如此的二次型:

$$\begin{pmatrix} \Delta x & \Delta y \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}|_{(x_0, y_0)} & \frac{\partial^2 f}{\partial x \partial y}|_{(x_0, y_0)} \\ \frac{\partial^2 f}{\partial y \partial x}|_{(x_0, y_0)} & \frac{\partial^2 f}{\partial y^2}|_{(x_0, y_0)} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

而在数学分析中, 往往由于函数具有连续的偏导数, 中间的矩阵总是对称的, 我们称其为 **Hesse 矩阵**.

正如我们在一元函数中有极值点判断的原理: 首先由 Fermat 引理我们知道函数的极值点一定是导函数的零点, 从而找到一系列的**稳定点**; 但是稳定点不一定全是极值点, 例如  $f(x) = x^3$  中的零点, 为此我们需要去观察其二阶导数, 对应的就是函数的**凹凸性**, 我们需要保证二阶导数在该点上非 0, 进而由二阶导数的正负性来判定此点是极小值点还是极大值点, 若为正, 则为极小值点.

对应的, 我们在多元函数中也是重复这种操作. 首先我们找到所有偏导数都为 0 的稳定点; 其次我们验证其 Hesse 矩阵的**正定性**, 正定性是保证这个二次型是否对任意的  $(\Delta x, \Delta y)$  (由导数的性质来看, 这是必然的约束), 都能够保证二次型的正号. 只有当 Hesse 矩阵为正定矩阵的时候, 该点才为极小值点.

多元函数极值的研究是数学分析多元微积分中极为重要的一块内容.

2. **随机向量的协方差矩阵** 在概率论与数理统计中, 我们实际经常接触的也是具有这种高维的**随机向量**, 我们高中就已经学过了, 一组数据之间是具有**方差**这种性质的, 描述的是样本点距离样本均值偏离的程度. 高维的随机向量也具备这种方差, 称之为**协方差矩阵**

$$\Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n) \end{pmatrix}$$

其中  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  为一个  $n$  维随机向量, 其中的每个  $X_i$  为一个随机变量, 协方差  $\text{Cov}(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])]$ , 当  $X_i = X_j$  时,  $\text{Var}(X_i) = \text{Cov}(X_i, X_i)$ , 很容易看出这和高中所定义的方差是相容的. 而容易看出协方差矩阵的矩阵刻画如下:

$$\Sigma = \text{Cov}(\mathbf{X}, \mathbf{X}) = E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^T]$$

所谓协方差矩阵, 也不过是随机向量的方差罢了. 而方差自然是非负的, 当且仅当所有样本点都集中在均值上; 将这性质推广到协方差矩阵上, 就是说  $\Sigma$  是非负定的, 借由我们接下来所学的二次型和一点点的概率论定义, 这个证明是很自然的.

此外, 大学本科的概率论和高中所学的概率论的最大不同是: 随机变量不仅仅是一个个离散的样本点的形态 (对应于二项分布), 而更经常显现出连续的态势 (对应于正态分布), 由此, 微积分恰然地嵌入进现代概率论之中, 并远不止于此.

概率论可以说是数学基础分支中最年轻的一支, 这说的是概率论的公理体系直到不到一百年前, 才由 Kolmogorov 建立的. (一百年后, 一些做神经网络的人将他们的模型命名为 KAN: Kolmogorov - Arnold Network, 以纪念两人完成的一则表示

定理) 在此之前的概率论是由赌博发展出来的, 从离散的一些分布渐渐注意到一些非常有效的连续分布, 将数学分析的工具嵌入进去, 数学家们在 19 世纪证明了大数定律和中心极限定理. Kolmogorov 所做的更进一步, 他将分析学的工具直接跨了一步, 用实分析和测度论来构造概率论的公理体系, 用测度来表示概率, 让概率论一下子拥有了稳固的根基, 进而不断地迸发出蓬勃的生命力, 此后, 随机过程和高等数理统计也发展得愈发成熟. 如今, SDE 随机微分方程依然是数学的一大热门分支, 因为人工智能的发展, 概率论的前景依然十分辉煌.

3. 二阶线性偏微分方程分类 设  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  表示空间坐标, 一般问题中  $n \leq 3$ , 除此之外还有一个时间变量  $t$ . 无需多说, 偏微分方程都有很强的物理和工程背景, 因此变量的选择都有其意义. 对于此  $\mathbb{R}^{n+1}$  中 (我们视  $x_{n+1} = t$ ) 为其中一个元素, 二阶线性偏微分方程的一般形式为:

$$\sum_{i,j=1}^{n+1} a_{ij}(\mathbf{x}, t) u_{x_i x_j} + \sum_{i=1}^{n+1} b_i(\mathbf{x}, t) + c(\mathbf{x}, t) u(\mathbf{x}, t) = f(\mathbf{x}, t)$$

我们将二次项展开成矩阵形式就是:

$$\begin{pmatrix} \frac{\partial}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} & \frac{\partial}{\partial t} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} & a_{1,n+1} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & a_{n,n+1} \\ a_{n+1,1} & \cdots & a_{n+1,n} & a_{n+1,n+1} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \cdots \\ \frac{\partial}{\partial x_n} \\ \frac{\partial}{\partial t} \end{pmatrix} u(x_1, \dots, x_n, t)$$

分量形式是已经显见其二次型本质, 因此此处的矩阵  $\mathbf{A} = (a_{ij})$  也是一个对称矩阵.

在学习偏微分方程 PDE 之初, 我们会先学习几个特定形式的偏微分方程以入门, 对这三种偏微分方程的研究也形成了一个专门的课程, 叫做**数学物理方法**. 在接触偏微分方程之前的必备前置课程是常微分方程 ODE, 其中有一半篇幅是对于各种各样的常微分方程的求解, 后半段则是对方程的性质的研究. 在 PDE 上则并不如此, 因为我们所能够求解的偏微分方程, 几乎可以说就是在数学物理方程这门课中所能学习到的了. 他们分别是:

**位势方程:** 一个著名例子是极小曲面, 一个肥皂泡在环境之中稳定下来时的形状时怎样的, 这类问题考虑的是稳定状态, 因此时间不考虑其中. 方程形如

$$-\Delta u = f(\mathbf{x})$$

其中

$$\Delta = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$$

叫做 Laplace 算子. 对应的一阶项  $b_i(\mathbf{x}) \equiv 0, i = 1, \dots, n$ , 零阶项  $c(\mathbf{x}) \equiv 0$ ,

$$\mathbf{A}_1 = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix}$$

注意这里矩阵  $\mathbf{A}_1$  是  $n$  阶的.

**热传导方程:** 如其名, 最主要的例子是热在物体中的传播, 方程形如

$$u_t - a^2 \Delta u = f(\mathbf{x}, t), \quad a > 0$$

$a$  为常数. 对应的一阶项  $b_i(\mathbf{x}) \equiv 0, i = 1, \dots, n, b_{n+1} \equiv 1$  零阶项  $c(\mathbf{x}) \equiv 0$ ,

$$\mathbf{A}_2 = \begin{pmatrix} -a^2 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & -a^2 & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

注意这里矩阵  $\mathbf{A}_2$  是  $n+1$  阶的.

**波动方程:** 如其名, 最主要的例子是弦振动, 方程形如

$$u_{tt} - a^2 \Delta u = f(\mathbf{x}, t), \quad a > 0$$

$a$  为常数. 对应的一阶项  $b_i(\mathbf{x}) \equiv 0, i = 1, \dots, n+1$ , 零阶项  $c(\mathbf{x}) \equiv 0$ ,

$$\mathbf{A}_3 = \begin{pmatrix} -a^2 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & -a^2 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

注意这里矩阵  $\mathbf{A}_3$  是  $n+1$  阶的.

由于此处常系数矩阵都是对角阵, 我们很容易看出它们的差异. 其中,  $\mathbf{A}_1$  的  $n$  个特征值都为负数, 我们称其 **负定**, 称对应的二阶线性偏微分方程 (以下简称为方程) 叫做 **椭圆型**的;  $\mathbf{A}_2$  的  $n$  个特征值为负数, 有一个特征值为 0, 称对应的方程叫做 **抛物型**的;  $\mathbf{A}_3$  的  $n$  个特征值为负数, 有一个特征值为正数, 称对应的方程叫做 **双曲型**的. 如此命名的意义就在于我们后面要讲的几何意义.

运用我们接下来所要学习的**合同变换**, 很容易证明: 若二阶线性偏微分方程的系数矩阵  $\mathbf{A}$  是常数矩阵, 且它只有  $n$  个特征值为负数, (从而它是上述的三种方程中的一种), 则存在一个非退化的自变量变换把方程的二阶项化成我们刚才所述的三种标准型.

换言之, 我们用二次型的理论让 PDE 中的部分方程进行分类, 遇到相应的方程, 只要按照分类标准对应到某一形式, 我们就可以将其简化为一个标准型, 而对于这



类标准型的研究已经是固定套路了.

4. **二次曲面的分类:** 如 PDE 中的理论, 对一个数学对象的分类是非常重要的工作, 这恰恰也是几何所做的工作.

### 3.1.3 章节安排

由于上面的一些对应用的讨论, 接下来我们的主线非常明晰: 我们要找到一个对二次型分类的方法, 它的思想和几何中以及 PDE 中是契合的.

二次项中, 交叉项总是麻烦的, 为了寻找这个分类的方法, 我们首先要做的就是将所有交叉项消除, 只保留平方项. 我们首先注意到的一个方法就是二次多项式的配方法, 出于对配方法的简化考虑 (就像我们对 Gauss 消元法做的), 我们提出矩阵的合同变换, 并直接应用合同变换证明配方法的可实施性.

分类的本质是寻找到一种简化的变换, 让某一希望研究的量在变换下保持不变, 我们所希望的这种不变量就是惯性指数. 我们将证明惯性指数是在合同变换下的不变量. 从而我们可以以惯性指数来对二次型分类, 由于合同关系是一个等价关系, 划分出来的类是互不相交的等价类.

经过分类之后, 可以得到一类比较特殊的二次型, 旋即为了研究它良好的性质, 我们开始对正定型的研究.

最后, 需要指出的是配方法对应的是一种会压缩的投影, 如果我们希望绘制的图形不会因为变换而变形, 而仅仅是在坐标轴上转动了几下, 换了个地方, 那我们就需要正交变换的工具. 这部分内容并非二次型的内容, 核心的部分交由内积空间上的自伴随算子.

因此, 主轴如下

化简: 配方法/合同变换  $\rightarrow$  二次型的分类: 惯性定理  $\rightarrow$  特殊类: 正定型

## 3.2 配方法与矩阵的合同

我们的问题是: 如何将二次型化为只含平方项的标准型? 并且我们希望, 这种变换是可逆的.

即: 对于一个二次型  $f = \mathbf{x}^T \mathbf{A} \mathbf{x}$ , 我们希望做一个非退化的变换  $\mathbf{x} = \mathbf{C} \mathbf{y}$ , 非退化的含义是  $\mathbf{C}$  是一个可逆矩阵. 从而将二次型化成

$$f = \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{y}^T (\mathbf{C}^T \mathbf{A} \mathbf{C}) \mathbf{y}$$

使得其中的矩阵  $\mathbf{C}^T \mathbf{A} \mathbf{C}$  是一个对角阵.

### 3.2.1 配方法

配方法的核心是对于多项式的平方展开的应用, 即:

$$(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_m)^2 = \sum_{i=1}^m \mathbf{x}_i^2 + 2 \sum_{1 \leq i < j \leq m} \mathbf{x}_i \mathbf{x}_j$$

我们用例子来体悟.

**Example 3.2.1.** 将下列二次型化简为对角形.

$$f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 + 2x_1x_3 + 6x_2x_3$$

*Solution.* 首先, 我们按照变量  $x_1$  进行配方:

$$\begin{aligned} f(x_1, x_2, x_3) &= x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 + 2x_1x_3 + 6x_2x_3 \\ &= (x_1^2 + 2x_1(x_2 + x_3)) + 2x_2^2 + 5x_3^2 + 6x_2x_3 \\ &= (x_1 + (x_2 + x_3))^2 - (x_2 + x_3)^2 + 2x_2^2 + 5x_3^2 + 6x_2x_3 \\ &= (x_1 + x_2 + x_3)^2 - (x_2^2 + 2x_2x_3 + x_3^2) + 2x_2^2 + 5x_3^2 + 6x_2x_3 \\ &= (x_1 + x_2 + x_3)^2 + x_2^2 + 4x_2x_3 + 4x_3^2 \end{aligned}$$

然后, 对含有  $x_2$  的项进行配方:

$$\begin{aligned} f(x_1, x_2, x_3) &= (x_1 + x_2 + x_3)^2 + (x_2^2 + 4x_2x_3 + 4x_3^2) \\ &= (x_1 + x_2 + x_3)^2 + (x_2 + 2x_3)^2 \end{aligned}$$

最后我们进行变量替换,

$$\begin{cases} y_1 = x_1 + x_2 + x_3 \\ y_2 = x_2 + 2x_3 \\ y_3 = x_3 \end{cases}$$

则原二次型  $f(x_1, x_2, x_3)$  化为标准形  $f = y_1^2 + y_2^2$ .

对应的, 我们做了非退化线性变换

$$\mathbf{y} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x}$$

这里的变换矩阵  $\mathbf{C}$  是取逆求出来的. □

如果已知的二次型中没有平方项, 我们可以采用下面例子中的方法. 运用的原理也不过是平方差公式罢了.

**Example 3.2.2.** 将下列二次型化简为对角形.

$$f = 2x_1x_2 + 2x_1x_3 - 6x_2x_3$$

*Solution.* 因为二次型中没有平方项, 我们先通过变量替换构造出平方项. 令

$$\begin{cases} x_1 = y_1 + y_2 \\ x_2 = y_1 - y_2 \\ x_3 = y_3 \end{cases}$$

将其代入原二次型  $f(x_1, x_2, x_3)$  中:

$$\begin{aligned} f(x_1, x_2, x_3) &= 2(y_1 + y_2)(y_1 - y_2) + 2(y_1 + y_2)y_3 - 6(y_1 - y_2)y_3 \\ &= 2(y_1^2 - y_2^2) + 2y_1y_3 + 2y_2y_3 - 6y_1y_3 + 6y_2y_3 \\ &= 2y_1^2 - 2y_2^2 - 4y_1y_3 + 8y_2y_3 \end{aligned}$$

接下来, 按照  $y_1$  进行配方:

$$\begin{aligned} f &= 2y_1^2 - 2y_2^2 - 4y_1y_3 + 8y_2y_3 \\ &= 2(y_1^2 - 2y_1y_3) - 2y_2^2 + 8y_2y_3 \\ &= 2(y_1 - y_3)^2 - 2y_3^2 - 2y_2^2 + 8y_2y_3 \end{aligned}$$

再对含有  $y_2$  的项进行配方:

$$\begin{aligned} f &= 2(y_1 - y_3)^2 - 2y_3^2 - 2(y_2^2 - 4y_2y_3) \\ &= 2(y_1 - y_3)^2 - 2y_3^2 - 2(y_2 - 2y_3)^2 + 8y_3^2 \\ &= 2(y_1 - y_3)^2 - 2(y_2 - 2y_3)^2 + 6y_3^2 \end{aligned}$$

令

$$\begin{cases} z_1 = y_1 - y_3 \\ z_2 = y_2 - 2y_3 \\ z_3 = y_3 \end{cases}$$

则原二次型  $f(x_1, x_2, x_3)$  化为标准形

$$f = 2z_1^2 - 2z_2^2 + 6z_3^2.$$

第一次变量替换对应的矩阵为:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

记变换矩阵为  $C_1$  第二次变量替换对应的矩阵为:

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

记变换矩阵为  $C_2$ . 于是变量替换为  $z = C_2 y = C_2 C_1 x$  求解其逆矩阵, 得到

$$x = Cz, \quad C = C_1^{-1} C_2^{-1}$$

其中

$$C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 3 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

□

于是我们可以这样叙述配方法的步骤:

**Step 1:** 检查二次型中是否有平方项, 如果没有平方项, 则对某一个混合项构造平方差公式代换, 从而构造出平方项.

**Step 2:** 循着某一个平方项  $x_i$ , 将所有包含  $x_i$  的项放在一块, 形如

$$[a_i x_i^2 + (b_{i1}x_1 \cdots + b_{i,i-1}x_{i-1} + b_{i,i+1}x_{i+1} + \cdots + b_{in}x_n)x_i]$$

将其用平方式合并:

$$a_i(x_i + c/2)^2 - a_i c^2/4$$

其中  $c = b_{i1}x_1 \cdots + b_{i,i-1}x_{i-1} + b_{i,i+1}x_{i+1} + \cdots + b_{in}x_n$  对每个平方项都做这样的操作, 并作非退化的变量替换  $y_i = x_i + c/2$ .

两个步骤是一个循环的过程, 后面我们也确实能够证明, 配方法一定是能够将二次型化成对角型.

可问题也出现在步骤之中, 非退化的变量替换并非那么显然就能看出可行性, 如下例所示:

### Example 3.2.3.

$$\begin{aligned} f &= 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 + 2x_1x_3 + 2x_2x_3 \\ &= (x_1 - x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2 \end{aligned}$$

若令  $y_1 = x_1 - x_2$ ,  $y_2 = x_1 + x_3$ ,  $y_3 = x_2 + x_3$ , 则得到  $f = y_1^2 + y_2^2 + y_3^2$ , 一切看起来简单而自然. 可是一旦我们查看对应的矩阵

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

不可逆, 因此如此轻松的结论是错误的.

所以我们需要去找一个更加标准的手段. 既然问题出现在找到的矩阵不可逆上, 并且我们每一步的变换都是相乘一个可逆矩阵进行的, 而我们最终的目标也是将原先的二次型相伴的矩阵化为对角阵, 那么我们何不直接用矩阵来操作呢?

## 3.2.2 矩阵的合同

我们先再看一眼二次型化简所做的事:

$$f = \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{y}^T (\mathbf{C}^T \mathbf{A} \mathbf{C}) \mathbf{y}$$

我们将变量改变了, 但是因为二次型里面的未知量都为未定元, 因此其实质是没有改变的. 二次型和伴随矩阵之间又有着 1-1 对应的关系, 我们关心的是  $A$  和  $C^T AC$  之间的关系, 也就是它们是拥有怎样共同的实质.

**Definition 3.2.1** (合同 congruence). 设  $A, B \in \mathcal{M}_n(F)$ , 若存在  $n$  阶可逆矩阵  $C$  使得

$$B = C^T AC$$

则称  $A$  和  $B$  是合同的.

由定义立刻可以得到:

**Proposition 3.2.1.** 合同关系是一个等价关系.

注意这里并未约束矩阵是对称矩阵.

回忆一下, 其实相似关系也是一个等价关系. 我们在构造矩阵相似直至找到对角化的时候, 也是在寻找一个可逆矩阵  $P$ , 使得  $P^{-1}AP = I_n$ . 彼时遇到的最麻烦的问题就是矩阵不一定可对角化.

万幸的是, 我们需要研究的对称矩阵正是这种 "可对角化" 的.

我们现在先来做化简的矩阵手段, 其实质也不超出 Gauss 消元法.

**Lemma 3.2.1.** 对称矩阵  $A$  的下列变换都是合同变换, 即变换后的矩阵与其合同:

- (1) 取  $C = E_{ij}$ , 对应的变换为对换  $A$  的第  $i$  行和第  $j$  行, 再对换第  $i$  列和第  $j$  列;
- (2) 取  $C = E_{k(i)}$ , 对应的变换为将非零常数  $k$  乘以  $A$  的第  $i$  行, 再将  $k$  乘以  $A$  的第  $i$  列;
- (3) 取  $C = E_{k(i),j}$ , 对应的变换为将常数  $k$  乘以第  $i$  行加到第  $j$  行上, 再将  $k$  乘以第  $i$  列加到第  $j$  列上.

*Proof.* 代入进去即可验证. □

**Lemma 3.2.2.** 设  $A \in \mathcal{M}_n(F)$  为非 0 对称矩阵, 则必存在可逆矩阵  $C$ , 使得  $C^T AC$  的第  $(1,1)$  个元素不等于 0.

*Proof.* 倘若对角线上有某一个元素  $a_{ii} \neq 0$ , 则我们用第一种合同变换, 可以将其换到  $(1,1)$  的位置上来.

倘若对角线上所有元素都为 0, 则我们必然能够找到一个元素  $a_{ij} \neq 0$ . 我们取第三种合同变换, 将第  $i$  行加到第 1 行, 将第  $i$  列加到第 1 列上, 于是得到  $a'_{11} = 2a_{ij} \neq 0$ .

根据引理 3.2.1, 以上操作得到的矩阵与原矩阵合同. □

**Theorem 3.2.1.** 设  $A \in M_n(F)$  为对称矩阵, 则必存在可逆矩阵  $C$ , 使得  $C^T A C$  为对角阵.

*Proof.* 通常的证明方法都是归纳假设法, 由于矩阵的阶数是一个确定的值, 我们直接提供一个算法, 运用它可以在有限步骤内直接计算出合同的对角矩阵, 从而证明定理. 当然, 本质上这和归纳假设法并无区别.

**Step 1:** 如果此时  $a_{11} = 0$ , 则我们可用上述引理中的方法将其化为一个非 0 的数, 得到矩阵  $A^{(1)} = (a_{ij}^{(1)})$ ; 接下来对每一行  $j = 2, 3, \dots, n$ , 若  $a_{j1}^{(1)} = 0$  则不做操作, 如果第  $a_{j1}^{(1)} \neq 0$ , 则我们取第 1 行乘以  $-a_{j1}^{(1)}/a_{11}^{(1)}$  加到第  $j$  行上, 将其化为 0. 最后, 我们能够得到一个第一行第一列除了  $a_{11}^{(1)}$  之外都为 0 的矩阵

$$B^{(1)} = \begin{pmatrix} a_{11}^{(1)} & 0 \\ 0 & C^{(1)} \end{pmatrix}$$

其中  $A_2$  为一个  $n-1$  阶矩阵.

**Step  $i = 2, \dots, n-1$ :** 如果此时  $C_{i-1}$  第  $(i, i)$  个元素  $c_{i,i}^{(i-1)}$  等于 0, 则我们可用上述引理中的方法将其化为一个非 0 的数, 得到矩阵  $A^{(i)} = (a_{ij}^{(i)})$ ; 接下来对每一行  $j = i+1, i+2, \dots, n$ , 若  $a_{ji}^{(i)} = 0$  则不做操作, 如果第  $a_{ji}^{(i)} \neq 0$ , 则我们取第  $i$  行乘以  $-a_{ji}^{(i)}/a_{ii}^{(i)}$  加到第  $j$  行上, 将其化为 0. 最后, 我们能够得到一个第  $i$  行第  $i$  列除了  $a_{ii}^{(i)}$  之外都为 0 的矩阵

$$B^{(i)} = \begin{pmatrix} \text{diag}(a_{11}^{(1)}, \dots, a_{ii}^{(i)}) & O \\ O & C^{(i)} \end{pmatrix}$$

**Step  $n$ :** 已经得到了所需的矩阵

$$B = B^{(n-1)} = \begin{pmatrix} a_{11}^{(1)} & 0 & \cdots & 0 \\ 0 & a_{22}^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}^{(n-1)} \end{pmatrix}$$

□

在实际计算中, 我们是不断地对矩阵进行合同变换的, 我们可以在矩阵的下方写一个单位矩阵, 对矩阵做合同变换的时候对下方的矩阵仅仅做列变换, 这样我们就将列变换全部写入下方矩阵, 即

$$\begin{pmatrix} A \\ I_n \end{pmatrix} \rightarrow \begin{pmatrix} C^T A C \\ C \end{pmatrix}$$

这样我们还直接省去了计算一些逆矩阵乘积的步骤.

### 3.3 二次型的分类与惯性定理

#### 3.3.1 不变量的研究

在研究相似变换的时候, 我们发现了特征值不会随着相似变换而改变. 进一步, 为了发现真正在相似变换中起到决定作用的结构, 并从中导出我们的相似标准型, 我们找到了初等因子组和不变因子组, 他们是起着真正分类作用的不变量.

既然我们要分类二次型, 那么寻找合同变换下的不变量是必经之路.

**Proposition 3.3.1.** 秩是合同变换下的不变量. 即: 如果  $A$  和  $B$  合同, 则  $\text{rank}(A) = \text{rank}(B)$ .

*Proof.* 只需注意到合同变换是相抵变换的一种即可.  $\square$

下面的引理不能直接推广到一般的域上, 因为至少要使对正数  $\sqrt{\cdot}$  运算封闭. 而  $\mathbb{Q}$  显然不满足这一性质.

**Lemma 3.3.1.** 设  $A \in \mathcal{M}_n(\mathbb{R})$  为对称矩阵, 则必存在可逆矩阵  $C$ , 使得  $C^T A C$  为一个元素仅为  $1, -1, 0$  的对角矩阵.

它表示为:

$$\begin{pmatrix} I_p & O & O \\ O & -I_q & O \\ O & -O & O_{n-r} \end{pmatrix}$$

*Proof.* 由定理 3.2.1, 可将  $A$  与一对角矩阵  $\text{diag}(b_1, b_2, \dots, b_n)$  合同. 我们再利用第一种合同变换, 按照正负性排序, 得到

$$\text{diag}(d_1, \dots, d_p, -d_{p+1}, \dots, -d_{p+q}, 0, \dots, 0)$$

其中  $d_i > 0, i = 1, \dots, p+1$ . 我们做第二种合同变换, 让第  $i$  行第  $i$  列均乘以  $1/\sqrt{d_i}$ , 于是得到了所需要的矩阵.  $\square$

下面的命题是这个引理的一个应用.

**Proposition 3.3.2.** 秩等于  $r$  的对称矩阵等于  $r$  个秩等于 1 的对称矩阵之和.

*Proof.* 设  $n$  阶矩阵  $A$  的秩为  $r$ . 由上述引理, 存在存在可逆矩阵  $C$ , 使得

$$\begin{aligned} C^T A C &= \text{diag}(d_1, \dots, d_r, 0, \dots, 0) \\ &= D_1 + D_2 + \dots + D_r \end{aligned}$$

其中  $D_i = d_i \mathbf{1}_{ij}$ , 因此我们可以得到:

$$A = (C^{-1})^T D_1 C^{-1} + \dots + (C^{-1})^T D_r C^{-1}$$

由命题 3.3.1, 秩是合同不变量, 因此  $(C^{-1})^T D_i C^{-1}$  的秩为1.  $\square$

引理 3.3.1 给出的矩阵已经足够简单, 我们将其作用在二次型上, 所得到的二次型已经完全符合我们的需求.

除此之外, 特征值也是我们关系的内容. 如果我们想要保证特征值在合同变换中不变, 我们最好的想法就是让合同变换成为一种相似变换, 对应的就是让  $C^T = C^{-1}$ . 我们称这样的矩阵叫做**正交矩阵**, 对应的变换叫做**正交变换**.

正交变换是比合同变换更好的一种变换, 几何直观上正交变换并不破坏几何的图形, 而只是挪一挪位置罢了. 为了我们对不变量的研究, 我们先叙述一个关于正交变换的重要定理.

**Lemma 3.3.2.** 设  $A \in \mathcal{M}_n(\mathbb{R})$  为实对称矩阵, 则  $A$  的特征值都为实数.

*Proof.* 设  $\lambda_0$  为  $A$  的任一特征值,  $\alpha$  为其对应的非零特征向量. 则有

$$A\alpha = \lambda_0\alpha.$$

一方面, 我们取其共轭 (实矩阵  $A$  不变), 并左乘  $\alpha^T$ , 得到

$$\alpha^T \bar{A} \bar{\alpha} = \bar{\lambda}_0 \alpha^T \bar{\alpha}$$

另一方面, 我们取其转置 (对称矩阵  $A$  不变), 并右乘  $\bar{\alpha}$ , 得到

$$\alpha^T \bar{A} \bar{\alpha} = \lambda_0 \alpha^T \bar{\alpha}$$

比较即可得到  $\bar{\lambda}_0 = \lambda_0$ , 从而特征值为实的.  $\square$

**Theorem 3.3.1.** 设  $A \in \mathcal{M}_n(\mathbb{R})$  为实对称矩阵, 则存在实正交矩阵  $Q$  使得

$$Q^T A Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

其中  $\lambda_i$  为  $A$  的特征值,  $i = 1, 2, \dots, n$ .

定理的证明交由后续章节.

有了此定理, 再由合同变换, 很容易得到以下定理:

**Theorem 3.3.2.** 设  $A \in \mathcal{M}_n(\mathbb{R})$  为实对称矩阵, 则记矩阵的秩为  $r$ , 正特征值为  $p$ , 负特征值  $q$ , 零特征值的个数为  $n - r$ , 他们均为合同变换下的不变量.

*Proof.* 首先, 由引理 3.3.2 可知实对称矩阵的特征值均为实数, 因此才可以判断正负性. 其次, 由定理 3.3.1 可知  $A$  于一个由其特征值组成的对角矩阵, 再经由合同变换, 可将其化为引理 3.3.1 的形式.  $\square$



### 3.3.2 实二次型与惯性定理

于上一小节中, 我们已将不变量的理论推导得十分充分, 接下来的不过是定义与语言, 或者说, 是我们享受推导的成果的时刻.

**Theorem 3.3.3. [惯性定理]** 设  $f(x_1, x_2, \dots, x_n)$  是一个  $n$  元实二次型, 且  $f$  可化为两个标准型:

$$c_1 y_1^2 + \dots + c_p y_p^2 - c_{p+1} y_{p+1}^2 - \dots - c_r y_r^2,$$

$$d_1 z_1^2 + \dots + d_k z_k^2 - d_{k+1} z_{k+1}^2 - \dots - d_r y_z^2,$$

其中  $c_i > 0, d_i > 0$ , 则必有  $p = k$ .

*Proof.* 首先引理 3.3.1 给出了实对称矩阵合同标准型. 而二次型与对称矩阵是 1-1 对应的, 实际上给出的也是二次型的标准型.

因为合同关系是等价关系, 由等价类的划分定理, 等价类之间是无交集的, 因此一个二次型只有一种合同标准型.  $\square$

由于定理 3.3.2 证明了正负特征值的个数也为合同不变量, 而它又牢牢地与合同标准型绑定在一起, 于是下面的定义就非常简单了.

**Definition 3.3.1.** 设  $f(x_1, x_2, \dots, x_n)$  为一个实二次型, 若其对应的实对称矩阵的秩为  $r$ , 则称  $r$  是  $f$  的秩; 正特征值个数为  $p$ , 称为  $f$  的正惯性指数; 负特征值个数  $q = r - p$ , 称为负惯性指数;  $s = p - q$  称为  $f$  的符号差.

这种定义与标准型的定义方式是等价的. 它也意味着我们不需要去计算合同变换, 而只需要估计特征值的正负性即可, 此时特征值计算的各种理论 (如 Gerschgorin 圆盘定理) 嵌入其中, 大大地降低了复杂度.

### 3.3.3 复二次型

复二次型因为对任意的数都可以开方, 它的标准型只有  $z_1^2 + z_2^2 + \dots + z_r^2$ . 复对称矩阵的合同关系只有一个不变量, 秩.