MATH10003.01 **高等线性代数** Fall 2025

Lecture 6 线性映射、二次型、正定矩阵 — 2025.10.14

教授：邵美悦

Scribe: 路人局

1 大纲

1. 线性映射的应用案例
2. 二次型与正定矩阵

2 线性映射的应用案例

例 2.1. 求解常微分方程

$$y'' - 2y' + y = x^4$$

的一组特解。

解. 设求导算子 D 作用在 $\mathbb{R}$ 上的不超过四次的多项式空间 $P(\mathbb{R}^4)$ 上，即 $D : P(\mathbb{R}^4) \to P(\mathbb{R}^4), y \mapsto y'$. 则原方程可写为

$$(D^2 - 2D + I)y = x^4,$$

亦即

$$(I - D)^2 y = x^4.$$

如果 $I - D$ 可逆，则

$$y = (I - D)^{-2} x^4.$$

下面我们来验证 $I - D$ 的可逆性。注意到

$$(I - D)(I + D + D^2 + D^3 + D^4) = I - D^5 = I,$$

因此 $I - D$ 可逆，且

$$(I - D)^{-1} = I + D + D^2 + D^3 + D^4.$$

这是因为 D^5 作用在 V 上恒为零映射。

接下来我们分两步来计算 $(I - D)^{-2} x^4$.

第一步，我们先计算 $z=(I-D)^{-1}x^4$.

$$z=(I+D+D^2+D^3+D^4)x^4=x^4+4x^3+12x^2+24x+24.$$

第二步，我们计算 $y=(I-D)^{-1}z$.

$$y=(I+D+D^2+D^3+D^4)z=x^4+8x^3+36x^2+96x+120.$$

这就是原方程的一组特解。

上面我们看到了线性算子在分析学中的一个巧妙应用。接下来，让我们转向一个完全不同的领域——组合数学，看一看线性代数是如何帮助我们理解图的结构的。

我们知道，给定一张图，我们可以利用邻接矩阵来表示这张图的结构。图分为两类：有向图和无向图。它们的邻接矩阵表示是不同的：

例 2.2. 设有无向图 $G=(V,E)$，其中 $V=\{1,2,3\}$，$E=\{(1,2),(2,3)\}$。则其邻接矩阵为

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

设有向图 $G'=(V',E')$，其中 $V'=\{1,2,3\}$，$E'=\{(1,2),(2,3)\}$。则其邻接矩阵为

$$A'=\begin{bmatrix}0&0&0\\1&0&0\\0&1&0\end{bmatrix}.$$

初看起来，邻接矩阵似乎只是将图的结构用一种新的符号来表示，但它真正的威力在于——它可以进行矩阵运算，从而揭示图的深层次性质。邻接矩阵的幂次具有明确的组合意义。在上面这个问题中，如果我们把邻接矩阵 A 看作是一个线性映射 $T:\mathbb{R}^3\to\mathbb{R}^3$，那么，它就把标准正交基 $\{e_1,e_2,e_3\}$ 映射为

$$T(e_1)=a_1,\quad T(e_2)=a_2,\quad T(e_3)=a_3,$$

其中 a_i 是邻接矩阵 A 的第 i 列。因此，T 的作用就是把每个顶点映射到它能够直接到达的顶点的集合上。相应地，矩阵 A 的第 (i,j) 元素的意义就是：从顶点 j 出发，能否直接到达顶点 i.

那么，线性映射 T^2 的作用是什么呢？现在我们就来研究一下 T^2 究竟能把标准正交基映射到哪里。例如，对于 e_1，我们有

$$T^2(e_1)=A^2e_1=A(Ae_1)=Aa_1.$$

在上式中，a_1 是从顶点 1 出发能够直接到达的顶点的集合，而 Aa_1 自然就是从这些顶点出发，能够直接到达的顶点的集合—— 这也就是从顶点 1 出发，经过两步能够到达的顶点的集合。同理，

A^2e_i 就是从顶点 i 出发，经过两步能够到达的顶点的集合。因此，线性映射 T^2 的作用就是把每个顶点映射到它经过两步能够到达的顶点的集合上。相应地，矩阵 A^2 的第 (i,j) 元素的意义就是：从顶点 j 出发，经过两步到达顶点 i 的路径数。类似地，线性映射 T^k 的作用就是把每个顶点映射到它经过 k 步能够到达的顶点的集合上，而矩阵 A^k 的意义也就不言而喻了——它的第 (i,j) 元素表示从顶点 j 出发，经过 k 步到达顶点 i 的路径数。

这种精确的组合意义是邻接矩阵在图论中非常有用的原因之一。它不仅告诉我们“能不能到”，还告诉我们“有多少种走法”。

但是，在上面的讨论中，我们绕不开一个问题：怎么计算矩阵的幂次呢？对于一般的矩阵，这个问题并不容易，我们需要求解特征值、特征向量，并对特征向量进行正交化处理，才能最终得到矩阵的幂次表达式。但对一些具有特殊结构的矩阵，我们就可以利用它们的特殊性质来简化计算。

例 2.3. 设无向图 $G=(V,E)$，其中 $V=\{1,2,3,4\}, E=\{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$. 则其邻接矩阵为

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

试计算 A^n.

解. 我们的思路仍然是通过谱分解来计算 A^n. 不过，我们无需硬算特征值和特征向量。

注意到 $A=ee^{\mathrm{T}}-I$，其中 $e=[1,1,1,1]^{\mathrm{T}}$.

对于矩阵 $M=ee^{\mathrm{T}}$，我们可以注意到

$$M^2=(ee^{\mathrm{T}})(ee^{\mathrm{T}})=e(e^{\mathrm{T}}e)e^{\mathrm{T}}=4ee^{\mathrm{T}}=4M.$$

因此它的特征值为 $0,4$，并且由于 M 是秩一矩阵，我们可以得到：4 的代数重数是 1，而 0 的代数重数是 3. 因此，矩阵 A 的特征值为 $3,-1$，其中 3 的代数重数是 1，而 -1 的代数重数是 3.

现在我们来计算每个特征值对应的特征向量。由于 A 的每一行的和都是 3，因此我们有

$$Ae=e\cdot 3,$$

即 3 的一个特征向量为 e. 当然，为了之后的计算方便，我们把它归一化为

$$q=\frac{1}{2}e=\begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{bmatrix}.$$

特征值 -1 对应的特征向量并不好求，但我们也没必要求。我们只需要找到 q 的正交补空间中的三个线性无关的向量即可。或者，我们也没必要把这三个向量都找出来，我们只需要把它们构成的空间表示出来就行了。这样，问题就变得非常简单了。

我们知道，谱分解可以写成分量形式：

$$A = \sum_{i=1}^{4} \lambda_i q_i q_i^{\mathrm{T}},$$

其中 λ_i 是 A 的特征值，q_i 是对应的单位特征向量。因为 $q_i q_i^{\mathrm{T}}$ 是投影算子，我们也可以将上式理解为三个投影算子的线性组合。现在我们已经知道 $\lambda_1 = 3$，$q_1 = q$. 接下来我们就要来处理剩下三项。

首先，它们的特征值都是 -1，因此 A 又可以写成

$$A = 3qq^{\mathrm{T}} - \sum_{i=2}^{4} q_i q_i^{\mathrm{T}}.$$

后面的求和项是三个投影算子的和——这个和算子的作用就相当于把一个向量投影到 q_2, q_3, q_4 张成的子空间上，而 q_2, q_3, q_4 张成的子空间正好是 q 的正交补空间。因此，这个和算子的作用就是把一个向量投影到 q 的正交补空间上。 $\sum_{i=2}^{4} q_i q_i^{\mathrm{T}}$ 就正是这个投影算子的表示矩阵。我们的目标就是把这个矩阵求出来。

其次，我们注意到

$$\sum_{i=2}^{4} q_i q_i^{\mathrm{T}} = I - qq^{\mathrm{T}}.$$

这是因为 $P_1 = qq^{\mathrm{T}}$ 是到特征向量 q 所在的一维子空间上的正交投影矩阵，而 $P_2 = I - qq^{\mathrm{T}}$ 是到特征向量 q 的正交补空间上的正交投影矩阵，这正是我们需要的矩阵！因此 $\sum_{i=2}^{4} q_i q_i^{\mathrm{T}} = I - qq^{\mathrm{T}}$.
综上所述，我们有

$$A = 3qq^{\mathrm{T}} - (I - qq^{\mathrm{T}}).$$

现在我们可以计算 A^n 了。投影矩阵具有两个关键的性质：

1. 幂等性：$P^2 = P$.
2. 正交性：如果 P_1, P_2 是两个正交的投影矩阵，那么 $P_1 P_2 = P_2 P_1 = 0$.

由这两条性质，我们可以得到：

$$A^n = 3^n (qq^{\mathrm{T}}) + (-1)^n (I - qq^{\mathrm{T}}).$$

谱分解定理 在上面这道例题中，我们说 A 是可以进行谱分解的。但我们并没有严格地证明。现在，我们来补充一下这个细节。

定理 2.4 (Hermite 矩阵的谱分解定理)**.** *设 $A \in \mathbb{C}^{n\times n}$ 是一个 Hermite 矩阵，即 $A = A^\star$. 则 A 可以进行谱分解，即存在酉矩阵 Q 和对角矩阵 Λ，使得*

$$A = Q\Lambda Q^\star.$$

证明. 我们之前证明过任意复方阵 $A \in \mathbb{C}^{n\times n}$ 都酉相似于上三角阵。在这里，我们采取同样的证明方法（归纳法），来证明 Hermite 矩阵酉相似于对角阵。

当矩阵的阶数 $n = 1$ 时，结论显然成立。假设当矩阵的阶数小于 n 时，结论成立。现在考虑阶数为 n 的 Hermite 矩阵 A.

首先，如果我们设 A 的一个特征值是 λ，对应的单位特征向量是 q_1，那么我们可以构造一个酉矩阵 Q_1，

$$Q_1 = [q_1, \star],$$

使得

$$AQ_1 = Q_1 \begin{bmatrix} \lambda & \star \\ 0 & B \end{bmatrix}.$$

这里 B 是一个 $(n-1)\times(n-1)$ 的矩阵。接下来，我们来证明 B 也是一个 Hermite 矩阵。我们先将上式进行变形：

$$Q_1^\star AQ_1 = \begin{bmatrix} \lambda & \star \\ 0 & B \end{bmatrix}.$$

由于 A 是 Hermite 矩阵，我们有：

1. λ 是实数；
2. $\star$ 部分是一个零向量；
3. B 是一个 Hermite 矩阵。

根据归纳假设，B 可以进行谱分解，即存在酉矩阵 Q_2' 和对角矩阵 Λ_2，使得

$$B = Q_2'\Lambda_2 Q_2'^\star.$$

现在我们构造酉矩阵

$$Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q_2' \end{bmatrix}.$$

则有

$$Q_2^\star Q_1^\star AQ_1Q_2 = \begin{bmatrix} \lambda & \star \\ 0 & \Lambda_2 \end{bmatrix}.$$

如果我们记 $Q = Q_1 Q_2$，则 Q 也是酉矩阵，并且

$$Q^\star A Q = \begin{bmatrix} \lambda & \star \\ 0 & \Lambda_2 \end{bmatrix} := \Lambda,$$

亦即

$$A = Q \Lambda Q^\star.$$

这就完成了证明。

从上面的证明中我们还可以发现，Λ 是 A 的特征值构成的对角矩阵，它是一个实矩阵；而 Q 的列向量是特征值对应的单位特征向量。 □

有了 Hermite 阵的谱分解定理，我们自然也会好奇，实对称矩阵是否也可以进行谱分解。答案是肯定的。实对称矩阵的谱分解定理如下：

定理 2.5 (实对称矩阵的谱分解定理)**.** *设 $A \in \mathbb{R}^{n\times n}$ 是一个实对称矩阵，即 $A = A^{\mathrm{T}}$. 则 A 可以进行谱分解，即存在实正交矩阵 Q 和实对角矩阵 Λ，使得*

$$A = Q \Lambda Q^{\mathrm{T}}.$$

证明. 实对称矩阵是一种特殊的 Hermite 矩阵，因此它的特征值都是实数。由特征方程

$$(\lambda I - A)x = 0,$$

系数矩阵 $\lambda I - A$ 是一个实矩阵，因此对特征值 λ_1，我们必然可以找到一个实数解 x_1. 因此我们可取 Q_1 为

$$Q_1 = [x_1, \star] \in \mathbb{R}^{n\times n},$$

使得

$$AQ_1 = Q_1 \begin{bmatrix} \lambda_1 & \star \\ 0 & B \end{bmatrix}.$$

同样地，由 A 的实对称性，我们有：

1. $\star$ 部分是一个零向量；
2. B 是一个实对称矩阵。

由归纳假设，B 可以进行谱分解，即存在实正交矩阵 Q_2' 和实对角矩阵 Λ_2，使得

$$B = Q_2' \Lambda_2 Q_2'^{\mathrm{T}}.$$

现在我们构造实正交矩阵

$$Q = Q_1 \begin{bmatrix} 1 & 0 \\ 0 & Q_2' \end{bmatrix},$$

则 Q 也是实正交矩阵，并且

$$Q^{\mathrm{T}} A Q = \begin{bmatrix} \lambda_1 & \star \\ 0 & \Lambda_2 \end{bmatrix} := \Lambda,$$

亦即

$$A = Q \Lambda Q^{\mathrm{T}}.$$

这就完成了证明。 □

谱分解定理有几种等价的形式：

推论 2.6. *1. $Q^\star A Q = \Lambda$;*

2. $AQ = Q\Lambda$;

3. $Aq_i = \lambda_i q_i, \forall i = 1, 2, \ldots, n$.

4. $A = \sum_{i=1}^n \lambda_i q_i q_i^\star$.

在以上等价形式中，或许大家最不熟悉的是最后一种。它的意义是：A 可以看作是 n 个投影算子的线性组合。这里的投影算子是指 $P_i = q_i q_i^\star$，它的作用是把一个向量投影到 q_i 所在的一维子空间上。在例题 2.3 中，我们正是利用了这种投影算子的线性组合的形式，来简化计算的过程。在那里，我们还陈述了投影算子的两个重要性质：幂等性和正交性。这两个性质使得投影算子的线性组合的幂次计算变得非常简单。

一般地，如果 A 是 n 阶 Hermite 矩阵，那么它的幂可以写成

$$A^k = \sum_{i=1}^n \lambda_i^k q_i q_i^\star, k = 1, 2, \ldots$$

3 二次型与正定矩阵

二次型 我们把 $f(x_1, \ldots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} x_i x_j$ 称为二次型。它有着几种等价的矩阵表示形式。我们将一一介绍。最直接的表示形式就是

$$f(x) = x^{\mathrm{T}} A x,$$

其中 $x = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$，$A = (a_{i,j})_{n \times n}$. 当然，这里 A 不一定是对称矩阵。这是因为 $a_{i,j}$ 可能不等于 $a_{j,i}$，我们完全可以分配不一样的系数。

由这种表示方法，我们还可以继续延伸，得到二次多项式的矩阵表示：设有 $g(x)=x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c$，如果我们想要把它写成矩阵形式，我们可以引入一个新的变量 $x_{n+1}=1$，并构造新的向量

$$y=\begin{bmatrix}x\\1\end{bmatrix}\in\mathbb{R}^{n+1},$$

以及新的矩阵

$$B=\begin{bmatrix}A & b\\0 & c\end{bmatrix},$$

则有

$$g(x)=y^{\mathrm{T}}By.$$

这种表示方法的缺陷在于：一方面，表示方法不唯一，这是因为尽管 x_ix_j 和 x_jx_i 是一个东西，但赋给它们的系数可能不同，只要满足系数和 $a_{i,j}+a_{j,i}$ 固定，我们就有无数种赋值方法，对应有无数种矩阵 B；另一方面，也是因为 $x_ix_j=x_jx_i$，因此我们完全不需要一整块矩阵来表示 $f(x)$ 和 $g(x)$，我们只需要矩阵的上三角或下三角部分，就足够表示二次型了。

第二种方案就是用上三角矩阵（或下三角矩阵）来表示二次型。但是，这种表示方法也有一个缺陷：它不够对称。那么，不对称究竟会带来什么问题呢？假如我们要做坐标变换 $x=Cy$，则有

$$x^{\mathrm{T}}Ax=y^{\mathrm{T}}(C^{\mathrm{T}}AC)y.$$

我们知道 A 是一个上三角矩阵，但变换后的矩阵 $C^{\mathrm{T}}AC$ 可能不再是上三角矩阵；如果我们采取上三角的表示方法，我们就需要把 $C^{\mathrm{T}}AC$ 再变回上三角矩阵，这样就很麻烦了。因此，我们需要一种更好的表示方法。

第三种方案就是用对称矩阵来表示二次型。这是因为，虽然 $a_{i,j}$ 可能不等于 $a_{j,i}$，但它们的和 $a_{i,j}+a_{j,i}$ 是唯一确定的。因此，我们可以把 $a_{i,j}$ 和 $a_{j,i}$ 都设为它们的平均值 $\frac{a_{i,j}+a_{j,i}}{2}$. 这样，我们就得到了一个对称矩阵 $B=(b_{i,j})$，其中

$$b_{i,j}=\begin{cases}a_{i,j}, & i=j;\\ \frac{a_{i,j}+a_{j,i}}{2}, & i\neq j.\end{cases}$$

则有

$$f(x)=x^{\mathrm{T}}Bx.$$

这种表示方法的好处在于：如果我们对 x 进行坐标变换，得到的矩阵仍然是对称矩阵。因此，我们以后讨论二次型时，默认它们都是用对称矩阵来表示的。

例 3.1. *$f(x,\ y,\ z)=2xy+2yz+2xz$ 可以表示为*

$$f(x,y,z)=\begin{bmatrix}x & y & z\end{bmatrix}\begin{bmatrix}0 & 1 & 1\\1 & 0 & 1\\1 & 1 & 0\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}.$$

合同变换 我们刚刚通过坐标变换得到了一个新的矩阵 $C^{\mathrm{T}}AC$. 这样的变换就叫做合同变换。

定义 3.2 (合同变换)**.** *设 A 和 B 是两个同型的矩阵，如果存在可逆矩阵 C，使得 $B = C^{\mathrm{T}}AC$，则称 A 和 B 是合同的。*

命题 3.3. *合同关系是一个等价关系。*

请注意：合同变换和相似变换是两回事。相似变换是 $B = C^{-1}AC$，而合同变换是 $B = C^{\mathrm{T}}AC$. 对正交矩阵来说，二者是等价的；但对一般矩阵来说，二者是完全不同的。

我们引入合同变换和坐标变换的目的，是为了化简二次型。具体来说，我们想要将二次型 $x^{\mathrm{T}}Ax$ 化简成若干个平方项的和，消去所有的交叉项。也就是说，我们想要把矩阵 A 通过合同变换化简成一个对角矩阵。而我们知道，对称矩阵是可以进行谱分解的，因此我们可以利用谱分解来实现这个目标。但是，谱分解毕竟还是比较麻烦的，我们需要求特征值、求特征向量、正交化处理等等。那么，有没有一种更简单的方法呢？

我们可以注意到，如果我们取 C 为初等矩阵，那么合同变换 $C^{\mathrm{T}}AC$ 就是对矩阵 A 进行初等变换：先做行变换，再做列变换。因此，我们可以考虑通过初等变换，换言之，Gauss 消去法，来化简矩阵 A.

例 3.4. *$f(x, y, z) = x^2 + 2y^2 + 8z^2 + 2xy + 2xz + 4yz$.*

解. *该二次型对应的矩阵为*

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 8 \end{bmatrix}.$$

现在，我们利用 Gauss 消去法来化简 A.

$$
\begin{aligned}
A &= \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & 1 \\ & 1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 6 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & 1 \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & \sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & \sqrt{6} \end{bmatrix} \\
&= C^{\mathrm{T}} I C.
\end{aligned}
$$

因此，令

$$
\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & \sqrt{6} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},
$$

如果我们将 $[x, y, z]^{\mathrm{T}}$ 简记为 θ，则有

$$
\begin{aligned}
f(x, y, z) &= \theta^{\mathrm{T}} A \theta \\
&= \theta^{\mathrm{T}} C^{\mathrm{T}} I C \theta \\
&= (C\theta)^{\mathrm{T}} I (C\theta) \\
&= x'^2 + y'^2 + z'^2.
\end{aligned}
$$

现在，我们从另一个视角来看这个问题。

解. 我们采用 *Lagrange* 配方法来化简二次型。方法是：我们先选一个主元，然后找到所有包含主元的一次项，把它们配方成一个平方项。然后，我们再处理剩下的部分，直到所有的交叉项都被消去为止。

$$
\begin{aligned}
f(x, y, z) &= x^2 + 2(y+z)x + (y+z)^2 + y^2 + 7z^2 + 2yz \\
&= (x+y+z)^2 + y^2 + 2yz + z^2 + 6z^2 \\
&= (x+y+z)^2 + (y+z)^2 + 6z^2.
\end{aligned}
$$

现在我们来对比一下这两种方法。我们在利用 Gauss 消去法消去第一列和第一行元素时，我们实际上是在处理所有包含 x 的交叉项。消去完成后，我们就把所有包含 x 的交叉项都消灭了。然

后，我们再利用 Gauss 消去法消去其他的列和行，直到所有的交叉项都被消去为止。这在本质上和 Lagrange 配方法是一样的。

我们之前在使用 Gauss 消去法时，经常会遇到要选主元的情况。在做合同变换时，我们也会遇到类似的问题。

例 3.5. $f(x,y,z)=2xy+2yz+2zx.$

解. 该二次型对应的矩阵为

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

这里我们需要选取主元。但是，对角线上全是零，我们该怎么办呢？

我们可以先做变换

$$\begin{cases}u=x+y\\v=x-y\end{cases}.$$

然后，我们就有

$$f(x,y,z)=\frac{1}{2}(u^2-v^2)+2uz.$$

接下来我们就可以用 *Gauss* 消去法来化简了。以下步骤我们省略。

上面这个例子中，我们巧妙地通过换元，解决了主元为零的问题。但是，这种方法似乎是需要我们拍拍脑袋的。那么，有没有更加机械化的方法呢？

Hermite 型的合同变换

定义 3.6 (Hermite 型)**.** 设 $A\in\mathbb{C}^{n\times n}$ 是 *Hermite* 阵，则称

$$f(x)=x^\star Ax,x\in\mathbb{C}^n$$

为 *Hermite* 型。我们可以注意到，

$$[f(x)]^\star=x^\star Ax=f(x),$$

因此 *Hermite* 型是一个复变量实值函数。

定理 3.7. 设 $A\in\mathbb{C}^{n\times n}$ 是 *Hermite* 阵，则存在可逆矩阵 C，使得

$$C^\star AC=\mathrm{diag}\{I,-I,O\},$$

我们称得到的矩阵为合同标准型。

证明. 我们先证明 A 经过合同变换后，可以化简为 $\operatorname{diag}\{\pm 1, 0, \ldots\}$. 如果这个结论成立，我们再乘上适当的排列矩阵，就可以得到合同标准型。我们采用归纳法来证明这个结论。

当 $n=1$ 时，若 $A=[0]$，结论显然成立。当 $A \neq [0]$ 时，设 $A=[a], a \neq 0$. 我们可以取 $C=[\sqrt{|a|}]$，则有

$$C^{\star} A C=[\pm 1].$$

因此当矩阵阶数 $n=1$ 时，结论成立。

假设当矩阵阶数小于 n 时，结论成立。现在考虑阶数为 n 的 Hermite 矩阵 A. 我们可以尝试在矩阵的对角线上寻找一个非零元 $a_{i,i}$.

情况一：如果找到了，我们就可以通过初等变换把它换到 $(1,1)$ 位置。现在我们设 $a_{1,1} \neq 0$. 我们可以通过合同变换将 $(1,1)$ 位置变换为 ± 1.

$$\begin{aligned} A &= \begin{bmatrix} \times & \times & \cdots & \times \\ \times & \times & \cdots & \times \\ \vdots & \vdots & & \vdots \\ \times & \times & \cdots & \times \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{|\alpha|} & & & \\ \times & 1 & & \\ \vdots & & \ddots & \\ \times & & & 1 \end{bmatrix} \begin{bmatrix} \pm 1 & & & \\ & \times & \cdots & \times \\ & \vdots & & \vdots \\ & \times & \cdots & \times \end{bmatrix} \begin{bmatrix} \sqrt{|\alpha|} & \times & \cdots & \times \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}. \end{aligned}$$

现在，我们得到了一个 $(n-1) \times (n-1)$ 的矩阵 B. 我们可以验证：B 同样是一个 Hermite 矩阵。由归纳假设，存在可逆矩阵 D，使得

$$D^{\star} B D=\operatorname{diag}\{\pm 1, 0, \ldots\}.$$

现在我们取

$$C=\begin{bmatrix} \sqrt{|\alpha|} & \times & \cdots & \times \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D \end{bmatrix},$$

则 C 是可逆矩阵，并且

$$C^{\star} A C=\operatorname{diag}\{\pm 1, 0, \ldots\}.$$

情况二：如果没有找到，这就说明 A 的对角线全是零。现在，我们可以在第一列寻找一个非零元 $a_{i,1}$，利用初等变换，把它换到 $(2,1)$ 位置。如果找不到，那就说明 A 的第一列全是零，因此 A 的第一行也全是零。 A 就不可能是 n 阶矩阵，这与 A 是 n 阶矩阵的假设矛盾。因此，我们可以找到一个非零元 $a_{i,1}$.

现在，我们设 $a_{2,1} = a_{1,2} \neq 0$. 现在，我们有一个可逆的块：

$$A_1 = \begin{bmatrix} 0 & a_{1,2} \\ a_{2,1} & 0 \end{bmatrix}.$$

我们利用这个块，对矩阵做块 Gauss 消去。

$$\begin{aligned} A &= \begin{bmatrix} \boxed{A_1} & \star \\ \star & \star \end{bmatrix} \\ &= \begin{bmatrix} I_2 & \\ M & I_{n-2} \end{bmatrix} \begin{bmatrix} A_1 & O \\ O & \boxed{\star} \end{bmatrix} \begin{bmatrix} I_2 & M^\star \\ & I_{n-2} \end{bmatrix}. \end{aligned}$$

这里的 $\star$ 部分我们也可以证明是一个 Hermite 矩阵。由归纳假设，存在可逆矩阵 D，使得

$$D^\star \boxed{\star} D = \operatorname{diag}\{\pm 1, 0, \ldots\}.$$

而对二阶矩阵 A_1，我们别无他法，只能利用谱分解暴力求解。好在二阶矩阵的谱分解并不复杂。假设我们得到的结果是

$$U^\star A_1 U = \operatorname{diag}\{1, -1\}.$$

现在我们取

$$C = \begin{bmatrix} I_2 & \\ M & I_{n-2} \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & D \end{bmatrix},$$

则 C 是可逆矩阵，并且

$$C^\star A C = \operatorname{diag}\{\pm 1, 0, \ldots\}.$$

这就完成了证明。 □

在定理 3.7 中，我们不仅证明了合同标准型的存在性，还给出了一个构造合同变换矩阵 C 的方法。我们可以将这个方法总结为以下步骤：

1. 寻找主元：
   - 如果对角线上有非零元，选取一个非零元作为主元；
   - 否则，在第一列寻找一个非零元，选取它作为主元。
2. 利用主元，对矩阵做初等变换，消去主元所在行和列的其他元素。
3. 对剩下的部分，重复步骤 1 和 2，直到矩阵化简为对角矩阵。
4. 将对角矩阵化简为合同标准型 $\operatorname{diag}\{I, -I, O\}$.

正定矩阵　多元函数的泰勒展开公式如下：

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}(x-x_0)^{\mathrm{T}}H(x-x_0)+o(\|x-x_0\|^2),$$

其中 H 是 Hessian 矩阵。

如果 x_0 是 $f(x)$ 的一个驻点，即 $f'(x_0)=0$，那么 $f(x)$ 在 x_0 附近的变化就主要由二次项决定。如果这个二次型恒大于零，那么 $f(x)$ 在 x_0 附近就总是大于 $f(x_0)$，因此 x_0 是一个局部极小值点。反之，如果这个二次型恒小于零，那么 $f(x)$ 在 x_0 附近就总是小于 $f(x_0)$，因此 x_0 是一个局部极大值点。因此，研究二次型的恒正恒负性质，就变得非常重要。

定义 3.8 (正定矩阵、半正定矩阵、负定矩阵、半负定矩阵). *若矩阵 A 的合同标准型是 I，则称 A 是正定矩阵。若矩阵 A 的合同标准型是 $diag\{I,O\}$，则称 A 是半正定矩阵。若矩阵 A 的合同标准型是 $-I$，则称 A 是负定矩阵。若矩阵 A 的合同标准型是 $diag\{-I,O\}$，则称 A 是半负定矩阵。*

由于定义的对偶性，我们下面只研究正定矩阵和半正定矩阵。

命题 3.9 (正定矩阵、半正定矩阵的等价定义). *A 是正定矩阵的充要条件是*

$$x^\star Ax>0,\forall x\in\mathbb{C}^n,x\neq 0.$$

A 是半正定矩阵的充要条件是

$$x^\star Ax\geq 0,\forall x\in\mathbb{C}^n.$$

证明. 我们只证明正定矩阵的情况，半正定矩阵的情况类似。 **充分性**：如果 A 是正定矩阵，则存在可逆矩阵 C，使得

$$A=C^\star IC.$$

因此，对于任意非零向量 x，我们有

$$x^\star Ax=(Cx)^\star I(Cx)=\|Cx\|^2\geq 0.$$

现在我们来验证一下等号能不能成立。由于 C 是可逆矩阵，因此 $Cx\neq 0$，从而 $\|Cx\|^2>0$. 因此，充分性得证。

必要性：由定理 3.7，存在可逆矩阵 C，使得

$$C^\star AC=\mathrm{diag}\{I,-I,O\}.$$

现在，我们迫切地想要在 $x^\star Ax$ 中引入一个矩阵 C. 一个合适的引入方法是：看看是否存在坐标变换 $x=Cy$. 事实上，这个变换是肯定存在的，因为 C 是可逆矩阵，我们只需令 $y=C^{-1}x$ 即可。

现在，我们有

$$\begin{aligned} x^\star A x &= (Cy)^\star A (Cy) \\ &= y^\star (C^\star A C) y \\ &= y^\star \begin{bmatrix} I & & \\ & -I & \\ & & O \end{bmatrix} y. \end{aligned}$$

既然 $(C^\star AC)$ 分成了三块，我们不妨把 y 也分成三块：

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

则有

$$\begin{aligned} x^\star A x &= \begin{bmatrix} y_1^\star & y_2^\star & y_3^\star \end{bmatrix} \begin{bmatrix} I & & \\ & -I & \\ & & O \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= \|y_1|^2 - |y_2|^2 > 0. \end{aligned}$$

我们若要让最后的大于号恒成立，就必须让 y_2 的维数被压到零。这也就是说，合同标准型中不能有 $-I$ 和 O 这两块。因此，合同标准型只能是 I. 这就完成了证明。 □

命题 3.10. 正定矩阵所有对角线元素大于零。

正定矩阵所有主子阵都正定。反之，若一个矩阵所有主子阵都正定，则该矩阵正定。同样地，半正定矩阵所有主子阵都半正定。反之，若一个矩阵所有主子阵都半正定，则该矩阵半正定。

证明. 我们只证明正定的情况，半正定的情况类似。

$$a_{i,i} = e_i^\star A e_i > 0.$$

这里的 e_i 是标准正交基。

由于 $a_{1,1} > 0$，我们可以找到一个可逆阵 L_1，使得

$$L_1^\star A L_1 = \begin{bmatrix} 1 & 0 \\ 0 & A_1 \end{bmatrix}.$$

接下来，我们考虑

$$\begin{bmatrix} 0 & x_2^\star \end{bmatrix} \begin{bmatrix} 1 & \\ & A_1 \end{bmatrix} \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x_2^\star A_1 x_2 > 0, \forall x_2 \neq 0.$$

这说明 A_1 也是正定矩阵。依此类推，我们可以证明正定矩阵的所有的主子阵都正定。

反过来，如果一个方阵所有主子阵正定，由于它本身也是它的一个主子阵，因此它本身也正定。

□

由正定矩阵的定义及定理 3.7，我们可以得到一个重要的分解：

定理 3.11 (Cholesky 分解)**.** *设 $A \in \mathbb{C}^{n\times n}$ 是一个正定矩阵，则存在唯一的下三角矩阵 L，使得*

$$A = LL^\star.$$

证明. 我们先证明存在性。

由定理 3.7，存在下三角阵 L（构造过程保证了它是下三角的），使得

$$A = L^\star IL = L^\star L.$$

现在我们来证明唯一性。假设存在另一个下三角阵 M，使得

$$A = MM^\star.$$

则有

$$LL^\star = MM^\star.$$

因此

$$(L^{-1}M)(L^{-1}M)^\star = I.$$

这说明 $L^{-1}M$ 是一个酉矩阵。由于 $L^{-1}M$ 还是一个下三角矩阵，因此它只能是一个对角阵。设 $L^{-1}M = D$，则 D 是一个对角阵，并且

$$DD^\star = I.$$

这说明 D 的对角线元素的模长都为 1.

另一方面，我们的构造过程保证了 L 和 M 的对角线元素都是正数，因此 D 的对角线元素也是正数，从而只能是 1. 这说明 $D = I$，从而 $L = M$. □

注：我们还可以得到另一个分解 $A = LDL^\star$，其中 L 是单位下三角阵，D 是对角阵。这其实就是把 Cholesky 分解中 L 的对角元拆成一个单独的对角阵 D. 这个分解相比 Cholesky 分解的好处是，我们不需要计算平方根，因此数值稳定性更好。当然，现在许多计算机计算平方根也能算得非常准了。

定理 3.12 (正定矩阵的等价定义)**.** *A 正定等价于 A 的所有顺序主子式大于零。*

A 半正定等价于 A 的所有主子式大于等于零。

证明. 我们先证明正定的情况。

充分性：设 A 的所有顺序主子式大于零。这说明，我们可以通过不换主元的 Gauss 消去法，将 A 分解为 $A = LDL^\star$. 在消元过程中，D 的对角线元素与顺序主子式直接相关：如果我们取前 k 阶

主子式，由于 $\det(A_k)=\det(L_k)^2\det(D_k)=\det(D_k)$，因此 D_k 的行列式等于 A_k 的行列式。由 D 的第一个对角元 $d_1=\Delta_1$，我们可以归纳得到第 k 个对角元的表达式：

$$d_k=\Delta_k/\Delta_{k-1}>0.$$

现在，我们可以利用 $LDL^\star$ 分解来证明 A 正定。我们考虑

$$\begin{aligned}x^\star Ax&=x^\star LDL^\star x\\&=(L^\star x)^\star D(L^\star x)\\&:=y^\star Dy\\&=\sum_{i=1}^n d_i|y_i|^2\geq 0,\forall x\neq 0.\end{aligned}$$

现在，我们只需说明 $y\neq 0$，即可说明 $x^\star Ax>0$. 由于 L 是可逆矩阵，因此 $L^\star$ 也是可逆矩阵，从而 $y=L^\star x\neq 0$. 这就完成了充分性的证明。

必要性：设 A 是正定矩阵。我们可以同样由 $\det(A_k)=\det(L_k)^2\det(D_k)=\det(D_k)=d_1\cdots d_k$，得到 A 的顺序主子式都大于零。

现在我们来证明半正定矩阵的情况。

必要性：设 A 是半正定矩阵。我们知道半正定矩阵的任意主子阵都是半正定的。现在我们取 A 的一个主子阵 A_k，设它的特征值为 $\lambda_1,\ldots,\lambda_k$.

由半正定矩阵的特征值非负，我们有

$$\det(A_k)=\lambda_1\cdots\lambda_k\geq 0.$$

因此 A 的任意主子式都大于等于零。

充分性：考虑特征多项式

$$p(\lambda)=\det(\lambda I-A)=\lambda^n-s_{n-1}\lambda^{n-1}+\cdots+(-1)^{n-1}s_1\lambda+(-1)^n s_0.$$

其中 s_k 是 A 的所有 $n-k$ 阶主子式之和。由于 A 的所有主子式都大于等于零，因此 $s_k\geq 0,k=0,1,\ldots,n-1$.

由 Viète 定理，我们有

$$\begin{gathered}\lambda_1+\lambda_2+\cdots+\lambda_n=s_{n-1}\geq 0,\\ \lambda_1\lambda_2+\lambda_1\lambda_3+\cdots+\lambda_{n-1}\lambda_n=s_{n-2}\geq 0,\\ \vdots\\ \lambda_1\lambda_2\cdots\lambda_n=s_0\geq 0.\end{gathered}$$

因此 $\lambda_i \geq 0, \forall i \in \{1, \ldots, n\}$. 否则，一定会存在某个 $s_k < 0$. 下面我们来证明这个事实。

考虑

$$p(-\lambda) = (-1)^n (\lambda^n + s_{n-1}\lambda^{n-1} + \cdots + s_1\lambda + s_0).$$

显然 $p(-\lambda)$ 的所有系数都同号。因此 $p(-\lambda)$ 不可能有正根。而 $p(-\lambda)$ 的零点正是 $-\lambda_i$，因此 λ_i 不可能为负。这就完成了充分性的证明。 $\square$