第三节 • 线性代数

司继春

上海对外经贸大学统计与信息学院

1 向量与矩阵

对于两个集合 A, B，我们记 $A \times B = \{(a,b), \forall a \in A, b \in B\}$，即 $\times$ 运算定义了一个二元组的集合，我们称 $\times$ 为**笛卡尔乘积**（**Cartesian product**）。比如，如果我们选取 $A = \{\heartsuit, \spadesuit, \clubsuit, \diamondsuit\}, B = \{2, ..., 10, \mathrm{J}, \mathrm{Q}, \mathrm{K}, \mathrm{A}\}$，那么我们就得到了一副扑克牌共 52 张牌的集合。而如果选取 $A = \mathbb{R}, B = \mathbb{R}$，那么 $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ 为二维平面。

更一般的，我们可以记

$$\begin{aligned}\Omega_1 \times \Omega_2 \times \cdots \times \Omega_d &= \times_{i=1}^n \Omega_i \\ &= \{(\omega_1, \omega_2, ..., \omega_n), \omega_i \in \Omega_i, i = 1, ..., n\}\end{aligned}$$

特别的，$\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ 为 n 维的**欧几里得空间**（**Euclidean space**），其中 $\omega = (\omega_1, ..., \omega_n) \in \mathbb{R}^n$ 为**向量**（**vectors**）。向量可以按照行或者列排列，分别称为行向量（column vector）或者列向量（column vector），不失一般性，我们下面提到的向量都默认为列向量。行向量和列向量之间可以通过转置（transpose）得到，即对于一个列向量 ω，其转置记为 ω'，为行向量。特别的，我们记 0 为所有元素都等于 0 的向量，即零向量，而记 $\iota = [1, 1, ..., 1]'$ 为所有元素都为 1 的列向量：

$$0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \iota = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

令 $\alpha \in \mathbb{R}$ 为标量，$\omega, \eta \in \mathbb{R}^n$ 为 n 维向量，我们定义向量加法运算的数乘运算：

$$\omega + \eta = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} = \begin{bmatrix} \omega_1 + \eta_1 \\ \omega_2 + \eta_2 \\ \vdots \\ \omega_n + \eta_n \end{bmatrix}, \alpha\omega = \begin{bmatrix} \alpha\omega_1 \\ \alpha\omega_2 \\ \vdots \\ \alpha\omega_n \end{bmatrix}$$

即对向量的每个元素相加，或者都乘以这个标量。

如果 $x \in \mathbb{R}^n, y \in \mathbb{R}^n$ 我们可以定义**内积**（**inner product**）或者**点乘积**（**dot product**）为：

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i y_i$$

如果 $\langle x, y \rangle = 0$，我们称两个向量**正交**（**orthogonal**）。在几何上，正交意味着两个向量是垂直的。

有了内积之后，可以使用内积定义长度的概念，即所谓的（欧几里得）**范数**（**norm**）：

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

以及两个向量间的**距离**（**metric**）：

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

对于一个 n 维的欧几里得空间 $\mathbb{R}^n$，我们可以在这个空间上定义 Borel 域：

$$\mathscr{B}^n = \times_{i=1}^{n} \mathscr{B}_i = \sigma\left(\left\{\times_{i=1}^{n} A_i, A_i \in \mathscr{B}_i\right\}\right)$$

如果我们把 k 个 n 维向量按列摆放在一起，我们得到了一个 $n \times k$ 维的**矩阵**（**matrix**）$A_{n \times k} = [a_1, ..., a_k]$，其中 a_i 为 n 维向量。我们也可以把向量看成是一个 $n \times 1$ 维的矩阵。特别的，如果 $k = n$，即一个行数与列数相等的矩阵，我们称之为**方阵**（**square matrix**）。

对于矩阵，我们同样可以类似定义加法和数乘运算：

$$\begin{aligned} A + B &= \left[\begin{array}{ccc} a_1 & \cdots & a_k \end{array}\right] + \left[\begin{array}{ccc} b_1 & \cdots & b_k \end{array}\right] = \left[\begin{array}{ccc} a_1 + b_1 & , \cdots , & a_k + b_k \end{array}\right] \\ \alpha A &= \left[\begin{array}{ccc} \alpha a_1 & \cdots & \alpha a_k \end{array}\right] \end{aligned}$$

其中 a_i, b_i 为 $n \times 1$ 维的列向量。自然的，我们也可以定义矩阵的转置：

$$A' = \left[\begin{array}{c} a_1' \\ \vdots \\ a_k' \end{array}\right]$$

如果 A 为 $n \times k$ 的矩阵，那么其转置 A' 为 $k \times n$ 的矩阵。矩阵的数乘、加法、转置等满足一些良好的性质，比如：

- $(A + B) + C = A + (B + C)$
- $(\alpha + \beta) A = \alpha A + \beta A$

- $\alpha(A+B)=\alpha A+\alpha B$
- $\alpha(\beta A)=(\alpha\beta)A$
- $(A')'=A$
- $(A+B)'=A'+B'$

矩阵与向量之间的乘法运算可以定义为矩阵列向量的**线性组合**（**linear conbination**），即对于矩阵 $A_{n\times k}=[a_1,...,a_k]$ 以及 k 维向量 $x=[x_1,...,x_k]'$，可以定义矩阵与向量的乘法为：

$$Ax=x_1a_1+\cdots+x_ka_k=\sum_{j=1}^{k}x_ja_j$$

即 a_j 的一个线性组合。注意以上乘法的定义要求矩阵 A 的列数必须与向量 x 的维数一致。

如果存在一个非零向量 x，使得 $Ax=0$，那么我们称向量 $a_1,...,a_k$ 是**线性相依**的（**linearly dependent**）。线性相依意味着其中的某个向量一定可以被其他向量线性表示出。比如，不放假设 $Ax^*=0$，且 $x_1^*\neq 0$，那么必然有：

$$a_1=-\frac{x_2^*}{x_1^*}a_2-\cdots-\frac{x_k^*}{x_1^*}a_k$$

即 a_1 可以通过其他几个向量线性表示出。而如果只有当 $x=0$ 时，才有 $Ax=0$，那么我们称向量 $a_1,...,a_k$ 是**线性无关**或者**线性独立**的（**linearly independent**）。比如矩阵：

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

其任何一个列向量都不可以被其他两个列向量线性表示出。

使用矩阵与向量的乘法我们可以定义两个矩阵之间的乘法，对于矩阵 $A_{n\times k}=[a_1,...,a_k]$ 以及 $B_{k\times m}=[b_1,...,b_m]$，其乘法定义为：

$$AB=A[b_1,...,b_m]=[Ab_1,...,Ab_m]_{n\times m}$$

即一个 $n\times k$ 的矩阵与一个 $k\times m$ 的矩阵相乘，得到了一个 $n\times m$ 的矩阵。对于矩阵相乘而言，即使乘积都存在，交换律也并不成立，即 $AB\neq BA$，而其转置满足：$(AB)'=B'A'$。

对于一个矩阵 $A_{n\times k}=[a_1,...,a_k]$，其列向量的所有子集即可以是线性相关的，也可以是线性无关的。在矩阵列向量的所有子集中线性无关向量组所包含

向量个数的最大值，我们称之为**秩**（**rank**）。比如对于矩阵

$$A = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 2 & 3 & 5 \end{bmatrix}$$

其所有四个向量是线性相依的，任意三个列向量也都是线性相依的，而前两个列向量是线性无关的，因而其矩阵的秩为 2，我们记为 $\mathrm{rank}\,(A) = 2$。矩阵的秩有如下的性质：

1. $\mathrm{rank}\,(A_{n\times k}) \leq \min\,(n, k)$
2. $\mathrm{rank}\,(A) = \mathrm{rank}\,(A') = \mathrm{rank}\,(A'A) = \mathrm{rank}\,(AA')$
3. 若 A 为 $n \times k$ 的矩阵，而 B 为 $k \times k$ 的矩阵，且 $\mathrm{rank}\,(k) = k$，那么 $\mathrm{rank}\,(AB) = \mathrm{rank}\,(A)$
4. $\mathrm{rank}\,(A + B) \leq \mathrm{rank}\,(A) + \mathrm{rank}\,(B)$

对于方阵，如果其秩等于行（列）数，那么我们称其为**满秩**（**full rank**）矩阵。如果方阵 A 满足 $A = A'$，那么我们称其为**对称矩阵**（**symmetric matrix**）。如果一个方阵其非对角线元素都为 0，那么我们称其为对角矩阵。我们接下来会使用 $\mathrm{diag}\,()$ 符号将向量转化为对角阵，比如：

$$\mathrm{diag}\,([1, 2, 3]) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

特别的，我们令 $I = \mathrm{diag}\,(\iota)$ 为对角线元素都为 1 的对角阵，称为**单位阵**（**identity matrix**）。

此外，如果一个方阵 A 满足 $AA' = A'A = I$，那么我们称 A 为**正交矩阵**（**orthogonal matrix**）。实际上，如果将矩阵 A 看成是列向量的组合，$A = [a_1, ..., a_n]$，那么：

$$A'A = \begin{bmatrix} a_1' \\ \vdots \\ a_n' \end{bmatrix} [a_1, ..., a_n] = \begin{bmatrix} a_1'a_1 & \cdots & a_1'a_n \\ \vdots & \ddots & \vdots \\ a_n'a_1 & \cdots & a_n'a_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

意味着 A 的列向量其范数都等于 1，且两两正交。显然，单位阵满足上述条件，为正交矩阵。

对于方阵 $A_{n\times n}$，我们可以定义其行列式（determinant）：

$$|A| = \sum (-1)^{\phi(j_1,...,j_n)} \prod_{i=1}^{n} a_{ij_i}$$

其中 $(j_1, ..., j_n)$ 为 $(1, ..., n)$ 的排列（permutations），而 $\phi(j_1, ..., j_n)$ 为 $(1, ..., n)$ 变换到 $(j_1, ..., j_n)$ 所需要交换的次数。实际上，矩阵的行列式与面积、体积紧密相关，比如，对于一个 2×2 的矩阵：

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

其行列式的绝对值等于以 $(0,0), (1,2), (3,1)$ 为顶点的矩形的面积（或者三角形面积的两倍）。如果方阵为不满秩，即矩阵的某一列可以被其他列向量表示出，那么其「体积」等于 0，因而其行列式等于 0。对于此类矩阵，我们称之为**奇异阵**（**singular matrix**）。行列式有以下性质：

- $|AB| = |A| |B|$
- $|A'| = |A|$
- $|\alpha A| = \alpha^n |A|$

此外，对于方阵，我们还可以定义**迹**（**trace**），即对角线元素的和：

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

矩阵的迹有如下结论：

- $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$
- $\mathrm{tr}(\lambda A) = \lambda \mathrm{tr}(A)$
- $\mathrm{tr}(A') = \mathrm{tr}(A)$
- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

此外，显然对于一个标量（1×1），其迹的值就等于他本身。我们可以使用迹的概念定义一个矩阵的范数：

$$\|A\| = \sqrt{\mathrm{tr}(A'A)} = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2}$$

如果 $\|A - B\| = 0$ 那么 $A = B$。

最后，对于方阵，如果其行列式不等于 0，即非奇异矩阵，那么可以定义其逆矩阵（inverse matrix）。对于一个 $n \times n$ 的非奇异矩阵，如果矩阵 B 满足：

$$AB = BA = I$$

那么我们称矩阵 B 为矩阵 A 的逆矩阵，记为 $B = A^{-1}$。对于逆矩阵，有：

- $\left(A^{-1}\right)' = \left(A'\right)^{-1}$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $\left|A^{-1}\right| = |A|^{-1}$
- 若 A 为 $n \times n$ 的非奇异矩阵，且：

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

其中 A_{11} 为 $r \times r$ 的非奇异矩阵，A_{22} 为 $k \times k$ 的方阵，A_{12} 为 $r \times k$ 的矩阵，令 $D = A_{22} - A_{21}A_{11}^{-1}A_{12}$，且 D 也是非奇异矩阵，那么：

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}D^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}D^{-1} \\ -D^{-1}A_{21}A_{11}^{-1} & D^{-1} \end{bmatrix}$$

或者令 $E = A_{11} - A_{12}A_{22}^{-1}A_{21}$，且 E 也是非奇异矩阵，那么：

$$A^{-1} = \begin{bmatrix} E^{-1} & -E^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}E^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}E^{-1}A_{12}A_{22}^{-1} \end{bmatrix}$$

对于任意两个矩阵 $A_{m\times n}$ 和 $B_{p\times q}$，我们还可以定义其 Kronecker 乘积：

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

即得到了一个 $mp \times nq$ 的矩阵。此外，我们还可以定义 $\text{vec}\,()$ 操作符。对于一个 $n \times k$ 矩阵 $A = [a_1, ..., a_k]$，那么

$$\text{vec}\,(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}$$

即得到了一个 $nk \times 1$ 的向量。Kronecker 乘积和 $\text{vec}\,()$ 操作符有如下性质：

- $A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C)$
- $(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D$
- $(A \otimes B) \otimes (C \otimes D) = AC \otimes BD$ 如果 AC、BD 存在

- $\alpha \otimes A = \alpha A = A \otimes \alpha$
- 对于两个向量 a, b，$a' \otimes b = ba' = b \otimes a'$
- $(A \otimes B)' = A' \otimes B'$
- $\mathrm{tr}\,(A \otimes B) = \mathrm{tr}\,(A)\,\mathrm{tr}\,(B)$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- $\mathrm{rank}\,(A \otimes B) = \mathrm{rank}\,(A)\,\mathrm{rank}\,(B)$
- 对于向量 a，$\mathrm{vec}\,(a') = \mathrm{vec}\,(a) = a$
- 对于两个向量 a, b，$\mathrm{vec}\,(ab') = b \otimes a$
- $\mathrm{vec}\,(A)'\,\mathrm{vec}\,(B) = \mathrm{tr}\,(A'B)$
- $\mathrm{vec}\,(ABC) = (C' \otimes A)\,\mathrm{vec}\,(B)$
- $\mathrm{tr}\,(ABCD) = (\mathrm{vec}\,(D'))'\,(C' \otimes A)\,\mathrm{vec}\,(B) = (\mathrm{vec}\,(D'))'\,(A \otimes C')\,\mathrm{vec}\,(B')$

2 线性空间与线性变换

如果我们将矩阵 A 左乘一个 n 维向量 x，那么 $y = Ax$ 为一个 k 维向量。现在我们可以把矩阵左乘向量视为一个函数，即 $y = A(x) = Ax$，易知 $A(x_1 + x_2) = Ax_1 + Ax_2$，以及 $A(\alpha x) = \alpha Ax$，我们一般把符合如上两个性质的函数成为**线性映射**（**linear mapping**）。特别的，当 $k = n$，即 A 为 $n \times n$ 维方阵时，线性映射 A 将 $\mathbb{R}^n$ 上的一个向量 x 映射到 $\mathbb{R}^n$ 上的另外一个向量 y，此时我们称 A 为**线性变换**（**linear transformation**）。比如变换：

$$A = \begin{bmatrix} cos\,(\theta) & -sin\,(\theta) \\ sin\,(\theta) & cos\,(\theta) \end{bmatrix}$$

就将一个二维空间 $\mathbb{R}^2$ 上的向量逆时针旋转 θ 度。取 $\theta = \frac{\pi}{2}$，那么：

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

取 $x = [1, 0]'$，那么 $y = Ax = [0, 1]'$，逆时针旋转了 90 度。

使用分块矩阵，如果令 $x = [x_1, x_2, ..., x_n]' \in \mathbb{R}^n$，$A_{k \times n} = [a_1, ..., a_n]$，其中 a_i 为 k 维列向量，那么：

$$y = Ax = [a_1, ..., a_n] \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i a_i$$

也就是说线性映射的结果 y 实际上是矩阵 A 的列向量 a_i 的一个线性组合。因而矩阵 A 的秩 $\operatorname{rank}(A)$，即矩阵 A 的列向量的极大线性无关组，也就是对于所有 $x \in \mathbb{R}^n$，所有 $y = Ax$ 的极大线性无关组，或者所有向量 $\{y = Ax, \forall x \in \mathbb{R}^n\}$ 这个线性空间的维数。

3 线性型与二次型

对于任何一个向量 $x \in \mathbb{R}^n$，**线性型**（**linear form**）即一个关于 x 的线性函数 $f(x) : \mathbb{R}^n \to \mathbb{R}$：

$$f(x) = a_1 x_1 + \cdots + a_n x_n = a' x$$

其中：

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

而**二次型**（**quadratic form**）则是一个关于 x 的二次函数 $g(x) : \mathbb{R}^n \to \mathbb{R}$：

$$\begin{aligned} g(x) &= a_{11} x_1^2 + a_{12} x_1 x_2 + \cdots + a_{1n} x_1 x_n + a_{21} x_2 x_1 + \cdots + a_{nn} x_n^2 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \\ &= x' A x \end{aligned}$$

其中：

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

不失一般性，我们一般将以上二次型写成一个对称矩阵的形式，即假定 A 为一个实对称矩阵。

对于一个方阵 $A_{n \times n}$，如果一个向量 $v \in \mathbb{R}^n$ 满足：

$$Av = \lambda v$$

那么我们称 v 为 A 的一个**特征向量**（**eigenvector**），而 λ 则称为一个**特征值**（**eigenvalue**）。如果我们将矩阵 A 看做是一个线性变换，那么特征向量即经过 A 的变换以后，保持方向不变的向量。实际上，如果 v 是 A 的一个特征向量，那么将其乘以一个常数 c，cv 也满足：$Acv = cAv = c\lambda v = \lambda cv$，因而 cv 也是矩阵 A 的一个特征向量。

如果矩阵 $A_{n\times n}$ 有 n 个线性独立的特征向量：$\left\{v^1,...,v^n\right\}$，其对应的特征值为：$\{\lambda_1,...,\lambda_n\}$，记矩阵：

$$\Gamma=\left[\begin{array}{ccc} v^1 & \cdots & v^n \end{array}\right]^{-1}$$

那么矩阵 A 可以被分解为：

$$A=\Gamma^{-1}\mathrm{diag}\left([\lambda_1,...,\lambda_n]\right)\Gamma\triangleq\Gamma^{-1}\Lambda\Gamma$$

以上分解我们称之为**特征分解**（**eigendecomposition**）。根据行列式的性质，我们有：

$$|A|=\left|\Gamma^{-1}\Lambda\Gamma\right|=\left|\Gamma^{-1}\right||\Lambda||\Gamma|=|\Lambda|=\prod_{i=1}^{n}\lambda_i$$

而根据迹的性质，有：

$$\mathrm{tr}\left(A\right)=\mathrm{tr}\left(\Gamma^{-1}\Lambda\Gamma\right)=\mathrm{tr}\left(\Lambda\Gamma\Gamma^{-1}\right)=\mathrm{tr}\left(\Lambda\right)=\sum_{i=1}^{n}\lambda_i$$

注意并非任何一个方阵 A 都可以进行特征分解，不过有一类特殊的矩阵，即实对称矩阵，必然可以被特征分解，且矩阵 Γ 可以变换为一个正交矩阵。由于正交矩阵的逆矩阵等于其对称矩阵，因而实对称矩阵 A 总是可以被分解为如下形式：

$$A=\Gamma'\Lambda\Gamma$$

其中 Γ 为正交矩阵。

由于正交矩阵 $\Gamma'\Gamma=I$，因而矩阵 Γ 的列向量（**特征向量**，**eigenvector**）是两两正交的，且每个列向量的范数为 1。比如矩阵：

$$\Gamma=\left[\begin{array}{cc} cos\left(\theta\right) & -sin\left(\theta\right) \\ sin\left(\theta\right) & cos\left(\theta\right) \end{array}\right]$$

为正交矩阵，其每个列向量都是正交的且范数为 1。这类矩阵对应着等距变换（isometry），即两个点经过正交矩阵 Γ 的变换之后，$d\left(\Gamma x,\Gamma y\right)=\sqrt{x'\Gamma'\Gamma y}=\sqrt{x'y}=d\left(x,y\right)$。正交矩阵对应着旋转、翻转等变换；而相应的，对角矩阵 Λ 则对应着在不同的方向上的拉伸变换。因而实对称矩阵的特征值分解将一个线性变换 A 分解为了一个等距变换 Γ 和一个放缩变换 Λ。

如果对于任何一个向量 $x\in\mathbb{R}^n$，二次型 $x'Ax>0$，我们称矩阵 A 为**正定矩阵**（**Positive-definite matrix**）；如果满足 $x'Ax\geq 0$，则成为**半正定矩阵**（**Positive semi-definite matrix**）；负定矩阵和半负定矩阵可以类似定义。

对于一个实对称矩阵 A，其二次型：

$$x'Ax=x'\Gamma'\Lambda\Gamma x=\left(\Gamma x\right)'\Lambda\left(\Gamma x\right)$$

注意由于 $y=\Gamma x$ 为一个 n 维向量，因而上式可以写为：

$$x'Ax=y'\Lambda y=\sum_{i=1}^{n}\lambda_i y_i^2>0(\geq 0)$$

因而如果实对称矩阵 A 的所有特征值都 >0（≥ 0），那么这个矩阵为正定矩阵（半正定矩阵）。

4　投影与幂等矩阵

在实对称矩阵中，有一类矩阵是我们接下来非常频繁使用的，即**幂等矩阵**（**Idempotent matrix**）。如果一个方阵 P 满足 $P^2=P$，那么我们称矩阵 P 为幂等矩阵。例如，矩阵：

$$P=\frac{1}{4}\cdot\begin{bmatrix}-4 & 6 & 2\\ -12 & 13 & 3\\ 20 & -15 & -1\end{bmatrix}$$

为幂等矩阵，可以验证 $P^2=P$。

特别的，当 P 为实对称矩阵时，我们称其为**投影矩阵**（**Projection matrix**）。由于所有实对称矩阵都可以被对角化，所以对于任意的投影矩阵，都可以写为：

$$P=\Gamma'\Lambda\Gamma$$

而由于 $P^2=\Gamma'\Lambda\underbrace{\Gamma\Gamma'}_{I}\Lambda\Gamma=\Gamma'\Lambda^2\Gamma=\Gamma'\Lambda\Gamma$，且 Γ 为可逆矩阵，所以 $\Lambda^2=\Lambda$。由于 Λ 为对角阵，所以 Λ 的对角元必为 0 或者 1。因而 $\operatorname{rank}(P)=\operatorname{rank}(\Lambda)=\operatorname{tr}(\Lambda)$。

投影矩阵顾名思义，与**投影**（**Projection**）的概念密不可分。如果把投影矩阵 P 视为线性变换，幂等矩阵的定义意味着一个向量经过 P 的变换以后，再次经过 P 的变换仍然保持不变，即 $P(Px)=P^2x=Px$。比如矩阵：

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

即把一个 $x-y-z$ 三维坐标系中的一个向量 $x=(x_1,x_2,x_3)'$ 映射到 $x-y$ 二维平面上的点 Px，而一个本身就在 $x-y$ 二维平面的点，如 Px，再次经过 P 的映射，还是在 $x-y$ 二维平面上，且就是其本身。可以验证，$P^2=P$。类似的，矩阵：

$$P=\begin{bmatrix}0.5 & 0.5 & 0\\ 0.5 & 0.5 & 0\\ 0 & 0 & 0\end{bmatrix}$$

则把一个三维向量 $x=(x_1,x_2,x_3)'$ 映射到 $y=x$ 这条直线上，同样有 $P^2=P$。

如果定义 $M=I-P$，那么 $M^2=(I-P)(I-P)=I-P-P+P^2=I-P=M$，即 $M=I-P$ 也为投影矩阵。注意 $(Mx)'Px=x'(I-P)Px=x'\left(P-P^2\right)x=0$，因而 Px 与 Mx 是正交的。也就是说，幂等矩阵把一个向量 x 分解成了正交的两个部分：Px 和 Mx，$x=Px+Mx$ 且 $\langle Mx,Px\rangle=0$。

例 1. 令 $\iota\in\mathbb{R}^n,\iota=(1,1,...,1)'$，那么矩阵 $P_0=\frac{1}{n}\iota\iota'$ 为投影矩阵，即 $P_0'=P_0$，且 $P_0^2=\frac{1}{n^2}\iota\underbrace{\iota'\iota}_{n}\iota'=\frac{1}{n}\iota\iota'=P_0$。对于一个向量 x，

$$P_0x=\frac{1}{n}\iota\iota'x=\frac{1}{n}\iota\cdot\sum_{i=1}^{n}x_i=\iota\cdot\bar{x}=\begin{pmatrix}\bar{x}\\ \vdots\\ \bar{x}\end{pmatrix}$$

即 P_0 将一个向量投影变换为其均值向量。易知 $\operatorname{rank}(P_0)=\operatorname{tr}(P_0)=\operatorname{tr}\left(\frac{1}{n}\iota\iota'\right)=\frac{1}{n}\operatorname{tr}(\iota'\iota)=1$。如果令 $M_0=I-P_0$，根据上述结论，易知 M_0 也是幂等矩阵，且 $\operatorname{rank}(M_0)=\operatorname{tr}(M_0)=\operatorname{tr}(I-P_0)=\operatorname{tr}(I)-\operatorname{tr}(P_0)=n-1$，且：

$$M_0x=x-\frac{1}{n}\iota\iota'x=\begin{pmatrix}x_1-\bar{x}\\ \vdots\\ x_n-\bar{x}\end{pmatrix}$$

那么：

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=(M_0x)'M_0x=x'M_0x$$

为一个二次型的形式。

5 矩阵微积分

对于一个向量 $\theta=[\theta_1,\theta_2,\ldots,\theta_n]'$，其实值函数：$f(\theta):\mathbb{R}^n\to\mathbb{R}$，我们可以定义函数 $f(\cdot)$ 对向量 θ 的导数为：

$$\frac{\partial f}{\partial\theta}=\begin{bmatrix}\frac{\partial f}{\partial\theta_1}\\ \frac{\partial f}{\partial\theta_2}\\ \vdots\\ \frac{\partial f}{\partial\theta_n}\end{bmatrix}$$

同时定义其二阶导：

$$\frac{\partial^2 f}{\partial\theta\partial\theta'}=\begin{bmatrix}\frac{\partial^2 f}{\partial\theta_1^2} & \frac{\partial^2 f}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 f}{\partial\theta_1\partial\theta_n}\\ \frac{\partial^2 f}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 f}{\partial\theta_2^2} & \cdots & \frac{\partial^2 f}{\partial\theta_2\partial\theta_d}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial^2 f}{\partial\theta_n\partial\theta_1} & \frac{\partial^2 f}{\partial\theta_n\partial\theta_2} & \cdots & \frac{\partial^2 f}{\partial\theta_n^2}\end{bmatrix}$$

比如，如果 $f(\theta)=\frac{\mu^2}{2\sigma^2}+\ln(\sigma)$，$\theta=(\mu,\sigma)'$ 那么：

$$\frac{\partial f(\theta)}{\partial\theta}=\begin{bmatrix}\frac{\mu}{\sigma^2}\\ \frac{1}{\sigma}-\frac{\mu^2}{\sigma^3}\end{bmatrix}$$

$$\frac{\partial^2 f(\theta)}{\partial\theta\partial\theta'}=\begin{bmatrix}\frac{1}{\sigma^2} & -\frac{2\mu}{\sigma^3}\\ -\frac{2\mu}{\sigma^3} & \frac{3\mu^2}{\sigma^4}-\frac{1}{\sigma^2}\end{bmatrix}$$

我们知道对于一个实值函数，$\frac{\partial^2 f}{\partial\theta_i\partial\theta_j}=\frac{\partial^2 f}{\partial\theta_j\partial\theta_i}$，因而 $\frac{\partial^2 f}{\partial\theta\partial\theta'}$ 是一个实对称矩阵。回忆极值原理，如果函数 f 可微，那么函数 f 在 θ_0 处为极值点的必要条件是 $\frac{\partial f(\theta_0)}{\partial\theta}$=0，如果 $\frac{\partial^2 f(\theta_0)}{\partial\theta\partial\theta'}$ 为正定矩阵（$f(\theta)$ 的所有特征值为正），那么 f 在 θ_0 处为极小值点，否则如果 $\frac{\partial^2 f(\theta_0)}{\partial\theta\partial\theta'}$ 为负定矩阵（$f(\theta)$ 的所有特征值为负），那么 f 在 θ_0 处为极大值点，如果 $\frac{\partial^2 f(\theta_0)}{\partial\theta\partial\theta'}$ 的特征值既有正值又有负值，那么 f 在 θ_0 处为鞍点（saddle point）。我们称 $\frac{\partial^2 f(\theta)}{\partial\theta\partial\theta'}$ 为**海塞矩阵**（**Hessian matrix**）。

例 2. 记向量 $x=(x_1,x_2)'$，效用函数为 $f(x)=\ln x_1+\ln x_2$。为了求解如下效用最大化问题：

$$\max_x f(x)$$
$$s.t. p'x=I$$

其中 $p=(p_1,p_2)'$ 为价格，I 为收入，可以构造拉格朗日函数：

$$\mathscr{L}(x,\lambda;p,I)=\ln x_1+\ln x_2+\lambda(I-p'x)$$

其一阶条件为：

$$\frac{\partial\mathscr{L}(x,\lambda;p,I)}{\partial x}=\begin{pmatrix}\frac{1}{x_1}\\ \frac{1}{x_2}\end{pmatrix}-\lambda\begin{pmatrix}p_1\\ p_2\end{pmatrix}=0$$

解得：

$$x=\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}\frac{I}{2p_1}\\ \frac{I}{2p_2}\end{pmatrix}$$

在所有向量的实值函数中，有两类函数是非常常用的，即线性型以及二次

型。对于线性型：$f(\theta)=\alpha'\theta$，其一阶导数为：

$$\frac{\partial\alpha'\theta}{\partial\theta}=\alpha$$

而二次型 $g(\theta)=\theta'A\theta$ 的一阶导数为：

$$\frac{\partial(\theta'A\theta)}{\partial\theta}=A\theta+A'\theta=(A+A')\theta$$

而其二阶导为：

$$\frac{\partial^2\theta'A\theta}{\partial\theta\partial\theta'}=A'+A$$

特别的，如果 A 为对称矩阵，那么

$$\frac{\partial(\theta'A\theta)}{\partial\theta}=2A\theta$$
$$\frac{\partial^2\theta'A\theta}{\partial\theta\partial\theta'}=2A$$

例 3. 如果存在 n 个有价证券，其期望收益率向量为 $r=[r_1,...,r_n]$，而其收益率的协方差矩阵为 Σ。现在给定一个权重向量 x，满足 $x'\iota=1$，即所有权重相加等于 1，那么根据权重向量配置 n 个有价证券，我们就得到了一个资产组合，该资产组合的期望收益率为 $x'r$，而资产组合收益率的方差为 $x'\Sigma x$。在所有的资产组合中，有一个资产组合其方差最小，为了求出这个使得方差最小的资产组合权重，我们可以解如下问题：

$$\begin{aligned} &\min_x x'\Sigma x \\ s.t.\quad & x'\iota=1 \end{aligned}$$

构建拉格朗日函数[1]：

$$\mathscr{L}(x,\lambda;r,\Sigma)=x'\Sigma x+2\lambda(1-x'\iota)$$

其一阶条件为：

$$\frac{\partial\mathscr{L}(x,\lambda;p,I)}{\partial x}=2\Sigma x-2\lambda\iota=0$$

解得：

$$x=\lambda\Sigma^{-1}\iota$$

将其带入约束中，得到：

$$\lambda\iota'\Sigma^{-1}\iota=1$$

[1]为了方便起见，不失一般性，我们令拉格朗日乘子为 2λ。

从而：

$$\lambda = \frac{1}{\iota'\Sigma^{-1}\iota}$$

最终我们得到权重向量：

$$x = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}$$

习题

1. 证明：

   (a) $Ax = 0 \Longleftrightarrow A'Ax = 0$

   (b) $AB = 0 \Longleftrightarrow A'AB = 0$

   (c) $A'AB = A'AC \Longleftrightarrow AB = AC$

2. 在例 (3) 中，如果投资者的效用函数为：

   $$U(x; r, \Sigma) = x'r - \gamma x'\Sigma x$$

   其中 $\gamma \geq 0$ 为一个标量，度量了投资者的风险厌恶程度，请求出使得效用最大化的资产组合权重 x。

3. 同样在例 (3) 中，请计算给定投资组合期望收益 $x'r = r*$ 的条件下，方差最小的投资组合权重 x。