# 数学基础的复习

翟卫欣　副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院

# 目录

# 目录

# 向量与矩阵

$n$维列向量定义为含有$n$个数的数组，记为

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$a_i$表示向量$\boldsymbol{a}$的第$i$个元素。定义$\mathbb{R}$为全体实数组成的集合，那么由实数组成的$n$维列向量可表示为$\mathbb{R}^n$，称为$n$维实数向量空间。通常将$\mathbb{R}^n$的元素用小写粗体字母表示（如$\boldsymbol{x}$）。向量$\boldsymbol{x} \in \mathbb{R}^n$中元素记为$x_1, \cdots, x_n$。

# 向量与矩阵

$n$维行向量记为$[a_1, a_2, \cdots, a_n]$，行向量$\boldsymbol{a}$的转置记为$\boldsymbol{a}^T$。比如，如果

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

那么

$$\boldsymbol{a}^T = [a_1, a_2, \cdots, a_n]$$

相应的，可以记为$\boldsymbol{a} = [a_1, a_2, \cdots, a_n]^T$。

# 线性相关与线性无关

- 如果方程
$$\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k=\mathbf{0}$$
只有在所有系数$\alpha_i(i=1,\cdots,k)$都等于零的前提下才成立，那么称向量集$\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$是线性无关的，否则称向量集是线性相关的。
- 如果集合中只包括一个向量$\mathbf{0}$，由于对于任意$\alpha\neq 0$，都有$\alpha\mathbf{0}=\mathbf{0}$，因此，该集合是线性相关的。实际上，所有包含$\mathbf{0}$向量的集合都是线性相关的。
- 如果集合中只包括单个非零向量$\boldsymbol{a}\neq\mathbf{0}$，只有$\alpha=0$时，才有$\alpha\boldsymbol{a}=\mathbf{0}$成立，因此，该集合是线性无关的。
- 给定向量$\boldsymbol{a}$，如果存在标量$\alpha_1,\cdots,\alpha_k$,使得
$$\boldsymbol{a}=\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k$$
那么称向量$\boldsymbol{a}$为$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k$的线性组合。

# 线性相关与线性无关

- 向量集$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$是线性相关的，当且仅当集合中的一个向量可以表示为其他向量的线性组合。
- 证明：
- 必要性。如果$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$是线性相关的，那么有

$$\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \cdots + \alpha_k\boldsymbol{a}_k = \mathbf{0}$$

其中至少存在一个标量$\alpha_i \neq 0$，从而有

$$\boldsymbol{a}_i = -\frac{\alpha_1}{\alpha_i}\boldsymbol{a}_1 - \frac{\alpha_2}{\alpha_i}\boldsymbol{a}_2 - \cdots - \frac{\alpha_k}{\alpha_i}\boldsymbol{a}_k$$

# 线性相关与线性无关

- 充分性。假定向量$\boldsymbol{a}_1$可以被表示为其他向量的线性组合：

$$\boldsymbol{a}_1 = \alpha_2\boldsymbol{a}_2 + \alpha_3\boldsymbol{a}_3 + \cdots + \alpha_k\boldsymbol{a}_k$$

那么有

$$(-1)\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \cdots + \alpha_k\boldsymbol{a}_k = \boldsymbol{0}$$

因为第1个标量非零，所以向量集$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$是线性相关的。类似的，如果将$\boldsymbol{a}_i, i = 2, \cdots, k$表示为其他向量的线性组合，也可以得到同样的结论。

中国农业大学 China Agricultural University

# 线性相关与线性无关

- 令$\mathcal{V}$表示$\mathbb{R}^n$的一个子集，如果$\mathcal{V}$在向量加和运算及标量乘积运算下是封闭的，那么称$\mathcal{V}$为$\mathbb{R}^n$的子空间。
- 每个子空间都包含零向量$\mathbf{0}$。
- 假定$\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$是$\mathbb{R}^n$中的任意向量，它们所有线性组合的集合称为$\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$张成的子空间。记为，

$$span\left[\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\right] = \left\{\sum_{i=1}^{k} \alpha_i \boldsymbol{a}_i : \alpha_1, \cdots, \alpha_k \in \mathbb{R}\right\}$$

  任意向量集合都能张成一个子空间。
- 给定子空间$\mathcal{V}$，如果存在线性无关的向量集合$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\} \subset \mathcal{V}$使得$\mathcal{V} = span\left[\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\right]$，那么称$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$是子空间$\mathcal{V}$的一组基。子空间$\mathcal{V}$中的所有基都包含相同数量的向量，这一数量称为$\mathcal{V}$的维数，记为$dim\mathcal{V}$。

# 线性相关与线性无关

如果$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$是$\mathcal{V}$的一组基，那么$\mathcal{V}$中的任一向量$\boldsymbol{a}$可以唯一地表示为

$$\boldsymbol{a} = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k$$

其中，$\alpha_i \in \mathbb{R}, i = 1, 2, \cdots, k$。

# 标准基

给定$\mathcal{V}$的一组基$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$和向量$\boldsymbol{a} \in \mathcal{V}$，如果

$$\boldsymbol{a} = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k$$

那么系数$\alpha_i, i = 1, \cdots, k$称为$\boldsymbol{a}$对应于基$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$的坐标。

$\mathbb{R}^n$的标准基为

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

在标准基下，向量$\boldsymbol{x}$可表示为

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T = x_1 \boldsymbol{e}_1 + x_2 \boldsymbol{e}_2 + \cdots + x_n \boldsymbol{e}_n$$

# 初等变换

对矩阵进行以下三种变换的称为行初等变换

- 对换两行（对换$i, j$两行，记作$r_i \leftrightarrow r_j$);
- 以数$k \neq 0$乘某一行中所有的元素（第$i$行乘$k$，记作$r_i \times k$）；
- 把某一行所有元的$k$倍，加到另一行对应的元上去（第$j$行的$k$倍加到第$i$行上，记作$r_i + kr_j$）.

矩阵的行初等变换与列初等变换，统称为矩阵的初等变换。

对一个矩阵每进行一次初等变换相当于为这个矩阵乘了一个初等矩阵。

# 初等变换

三种初等变换对应三种初等矩阵

(i)$r_i \leftrightarrow r_j$对应的的初等矩阵

$$
\boldsymbol{E}(i,j) = \begin{array}{cl}
\begin{pmatrix}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & 0 & & \cdots & & 1 & & & \\
 & & & & 1 & & & & & & \\
 & & & \vdots & & \ddots & & \vdots & & & \\
 & & & & & & 1 & & & & \\
 & & & 1 & & \cdots & & 0 & & & \\
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{pmatrix}
&
\begin{array}{l}
\\ \\ \\ \leftarrow \text{第i行} \\ \\ \\ \\ \leftarrow \text{第j行} \\ \\ \\ \\
\end{array}
\end{array}
$$

# 初等变换

(ii)$r_i \times k$对应的的初等矩阵

$$\boldsymbol{E}(i(k)) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & k & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \leftarrow \text{第i行}$$

# 初等变换

(iii)$r_i + kr_j$对应的的初等矩阵

$$\boldsymbol{E}(ij(k)) = \begin{array}{c} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & k & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \end{array} \begin{array}{l} \\ \\ \leftarrow 第i行 \\ \\ \leftarrow 第j行 \\ \\ \\ \end{array}$$

# 矩阵的秩

- 矩阵$\boldsymbol{A}$中线性无关列的最大数目称为$\boldsymbol{A}$的秩，记为$rank\boldsymbol{A}$。
  矩阵$\boldsymbol{A}$的秩等于它的非零子式的最高阶数。

- 在以下运算中，矩阵$\boldsymbol{A}$的秩保持不变：
  1.矩阵$\boldsymbol{A}$的某个（些）列乘以非零标量。
  2.矩阵内部交换列次序。
  3.在矩阵中加入一列，该列是其他列的线性组合。

# 矩阵的秩

- 如果矩阵$\boldsymbol{A}$的行数等于列数，那么该矩阵称为方阵。行列式是与每个方阵相对应的一个标量，记为$det\boldsymbol{A}$或$|\boldsymbol{A}|$。
- 如果一个$m \times n\,(m \geq n)$矩阵$\boldsymbol{A}$具有非零的$n$阶子式，那么$\boldsymbol{A}$的各列是线性无关的，即$rank\boldsymbol{A} = n$。

# 内积和范数

对于$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,定义欧式内积为

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

定义向量$\boldsymbol{x}$的欧氏范数为

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\boldsymbol{x}^T \boldsymbol{x}}$$

向量$\boldsymbol{x}$的欧氏范数$\|\boldsymbol{x}\|$具有如下性质：

1.非负性：$\boldsymbol{x}$的欧氏范数$\|\boldsymbol{x}\| \geq 0$,当且仅当$\boldsymbol{x} = 0$时，$\|\boldsymbol{x}\| = 0$

2.齐次性:$\|r\boldsymbol{x}\| = |r|\|\boldsymbol{x}\| \geq 0, r \in \mathbb{R}$

3.三角不等式：$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

# 内积和范数

$\mathbb{R}^n$中向量范数有很多种不同的定义方式，如1范数:

$$\|\boldsymbol{x}\|_1 = |x_1| + \cdots + |x_n|$$

$\infty$范数:

$$\|\boldsymbol{x}\|_\infty = max_i |x_i|$$

$max_i$表示对于所有$i$,取向量元素中最大的一项。
以上范数都是$p$范数的特例，$p$范数为

$$\|\boldsymbol{x}\|_p = \begin{cases} (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}, & 1 \le p < \infty \\ max\{|x_1|, \cdots, |x_n|\}, & p = \infty \end{cases}$$

# 如何求方阵的逆

设$\boldsymbol{A}$为矩阵，如果存在$n$阶方阵$\boldsymbol{B}$，使得

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$$

则称$\boldsymbol{A}$是可逆矩阵，$\boldsymbol{B}$是$\boldsymbol{A}$的逆矩阵。

- 定理1　如果$\boldsymbol{A}$是一个$n$阶可逆矩阵，则$\boldsymbol{A}$的逆矩阵是唯一的。
- 定理2　$n$阶方阵$\boldsymbol{A}$可逆的充分必要条件是行列式$|\boldsymbol{A}| \neq 0$

# 如何求方阵的逆

当$|\boldsymbol{A}| \neq 0$时，

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\boldsymbol{A}^* = \frac{1}{|\boldsymbol{A}|}\begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & Ann \end{bmatrix}$$

这是$A^{-1}$的计算公式，其中$\boldsymbol{A}^*$是$\boldsymbol{A}$的伴随阵，$A_{ij}$是$a_{ij}$的代数余子式，请注意伴随矩阵的行列关系。

# 求方阵的逆：伴随矩阵法

设$\boldsymbol{A}=\begin{bmatrix}1&2&3\\2&2&1\\3&4&3\end{bmatrix}$,求$\boldsymbol{A}^{-1}$

**解：**

$$|\boldsymbol{A}|=\begin{bmatrix}1&2&3\\2&2&1\\3&4&3\end{bmatrix}=2$$

$$A_{11}=det\begin{bmatrix}2&1\\4&3\end{bmatrix}=2\quad A_{21}=-det\begin{bmatrix}2&3\\4&3\end{bmatrix}=6\quad A_{31}=det\begin{bmatrix}2&3\\2&1\end{bmatrix}=-4$$

$$A_{12}=-det\begin{bmatrix}2&1\\3&3\end{bmatrix}=-3\quad A_{22}=det\begin{bmatrix}1&3\\3&3\end{bmatrix}=-6\quad A_{32}=-det\begin{bmatrix}1&3\\2&1\end{bmatrix}=5$$

$$A_{13}=det\begin{bmatrix}2&2\\3&4\end{bmatrix}=2\quad A_{23}=-det\begin{bmatrix}1&2\\3&4\end{bmatrix}=2\quad A_{33}=det\begin{bmatrix}1&2\\2&2\end{bmatrix}=-2$$

$$\boldsymbol{A}^{-1}=\frac{1}{|\boldsymbol{A}|}\boldsymbol{A}^{*}=\frac{1}{|\boldsymbol{A}|}\begin{bmatrix}A_{11}&A_{21}&A_{31}\\A_{12}&A_{22}&A_{32}\\A_{13}&A_{23}&A_{33}\end{bmatrix}=\begin{bmatrix}1&3&-2\\-\frac{3}{2}&-3&\frac{5}{2}\\1&1&-1\end{bmatrix}$$

中国农业大学 China Agricultural University

# 另一种求矩阵逆的方法

将一个矩阵通过数次初等变换得到一个单位阵的过程，等价于为这个矩阵分步乘了对应的初等矩阵，这些初等矩阵的乘积即为他的逆矩阵。

因此，若一个矩阵可以通过数次初等变换得到一个单位阵，则这个矩阵是非奇异的。同时其变换过程对单位阵改变后的结果，即为原矩阵的逆。

例:设$\boldsymbol{A}=\begin{pmatrix} 0 & -2 & 1 \\ 3 & 0 & -2 \\ -2 & 3 & 0 \end{pmatrix}$，证明$\boldsymbol{A}$可逆，并求$\boldsymbol{A}^{-1}$

解：经过初等变换后,若我们可以将$(\boldsymbol{A},\boldsymbol{I})$化成$(\boldsymbol{I},\boldsymbol{P})$，则矩阵$\boldsymbol{P}$即为$\boldsymbol{A}^{-1}$。运算如下：

$$(\boldsymbol{A},\boldsymbol{I}) = \begin{pmatrix} 0 & -2 & 1 & 1 & 0 & 0 \\ 3 & 0 & -2 & 0 & 1 & 0 \\ -2 & 3 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\underset{r_1 \leftrightarrow r_2}{\overset{\substack{r_3 \times 3 \\ r_3 + 2r_2}}{\sim}} \begin{pmatrix} 3 & 0 & -2 & 0 & 1 & 0 \\ 0 & -2 & 1 & 1 & 0 & 0 \\ 0 & 9 & -4 & 0 & 2 & 3 \end{pmatrix} \underset{r_3 + 9r_2}{\overset{r_3 \times 2}{\sim}} \begin{pmatrix} 3 & 0 & -2 & 0 & 1 & 0 \\ 0 & -2 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 9 & 4 & 6 \end{pmatrix}$$

$$\underset{r_2 - r_3}{\overset{r_1 + 2r_3}{\sim}} \begin{pmatrix} 3 & 0 & 0 & 18 & 9 & 12 \\ 0 & -2 & 0 & -8 & -4 & -6 \\ 0 & 0 & 1 & 9 & 4 & 6 \end{pmatrix} \underset{r_2/(-2)}{\overset{r_1/3}{\sim}} \begin{pmatrix} 1 & 0 & 0 & 6 & 3 & 4 \\ 0 & 1 & 0 & 4 & 2 & 3 \\ 0 & 0 & 1 & 9 & 4 & 6 \end{pmatrix}$$

因$\boldsymbol{A} \overset{r}{\sim} \boldsymbol{I}$，故$\boldsymbol{A}$可逆，且$\boldsymbol{A}^{-1} = \begin{pmatrix} 6 & 3 & 4 \\ 4 & 2 & 3 \\ 9 & 4 & 6 \end{pmatrix}$

# 特征值与特征向量

令$\boldsymbol{A}$是$n \times n$的实数方阵。存在标量$\lambda$（可能为复数）和非零向量$\boldsymbol{v}$满足等式

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$$

$\lambda$称为$\boldsymbol{A}$的特征值，$\boldsymbol{v}$称为$\boldsymbol{A}$的特征向量。

# 特征值与特征向量

已知$n$阶齐次线性方程组$(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = 0$有非零解的充分必要条件是系数行列式为0。

即矩阵$\lambda \boldsymbol{I} - \boldsymbol{A}$是奇异的，有$det[\lambda \boldsymbol{I} - \boldsymbol{A}] = 0$，于是有$n$次方程成立：

$$det[\lambda \boldsymbol{I} - \boldsymbol{A}] = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0$$

多项式$det[\lambda \boldsymbol{I} - \boldsymbol{A}]$称为矩阵$\boldsymbol{A}$的特征多项式，上面的方程称为特征方程。由代数的基本原理可知，特征方程必定有$n$个根（可能存在相同的根），即为$\boldsymbol{A}$的$n$个特征值。若$\boldsymbol{A}$有$n$个相异的特征值，那么它也有$n$个线性无关的特征向量。

# 特征值与特征向量

- 假定特征方程$det[\lambda \boldsymbol{I} - \boldsymbol{A}] = 0$存在$n$个相异的根$\lambda_1, \lambda_2, \cdots, \lambda_n$,那么存在$n$个线性无关的特征向量$\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$,使得

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, i = 1, 2, \cdots, n$$

- 证明：由$det[\lambda_i \boldsymbol{I} - \boldsymbol{A}] = 0(i = 1, \cdots, n)$可知，存在一组非零向量$\boldsymbol{v}_i(i = 1, \cdots, n)$,使得$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i(i = 1, \cdots, n)$。下面证明$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$是线性无关的。为此，令$c_1, \cdots, c_n$是满足关系式$\sum_{i=1}^{n} c_i \boldsymbol{v}_i = \boldsymbol{0}$的一组标量，只需证明$c_i = 0(i = 1, \cdots, n)$即可。

# 特征值与特征向量

考虑矩阵

$$\boldsymbol{Z} = (\lambda_2 \boldsymbol{I} - \boldsymbol{A})(\lambda_3 \boldsymbol{I} - \boldsymbol{A}) \cdots (\lambda_n \boldsymbol{I} - \boldsymbol{A})$$

首先证明$c_1 = 0$。因为有$\lambda_n \boldsymbol{v}_n - \boldsymbol{A}\boldsymbol{v}_n = 0$，故有

$$\begin{aligned} \boldsymbol{Z}\boldsymbol{v}_n &= (\lambda_2 \boldsymbol{I} - \boldsymbol{A})(\lambda_3 \boldsymbol{I} - \boldsymbol{A}) \cdots (\lambda_{n-1} \boldsymbol{I} - \boldsymbol{A})(\lambda_n \boldsymbol{I} - \boldsymbol{A})\boldsymbol{v}_n \\ &= (\lambda_2 \boldsymbol{I} - \boldsymbol{A})(\lambda_3 \boldsymbol{I} - \boldsymbol{A}) \cdots (\lambda_{n-1} \boldsymbol{I} - \boldsymbol{A})(\lambda_n \boldsymbol{v}_n - \boldsymbol{A}\boldsymbol{v}_n) \\ &= \boldsymbol{0} \end{aligned}$$

考虑到$(\lambda_n \boldsymbol{I} - \boldsymbol{A})\boldsymbol{v}_k = (\lambda_n - \lambda_k)\boldsymbol{v}_k$ 重复这一过程，可得

$$\boldsymbol{Z}\boldsymbol{v}_k = \boldsymbol{0}, \quad k = 2, 3, \cdots, n$$

# 特征值与特征向量

但是，

$$
\begin{aligned}
\boldsymbol{Z}\boldsymbol{v}_1 &= (\lambda_2\boldsymbol{I}-\boldsymbol{A})(\lambda_3\boldsymbol{I}-\boldsymbol{A})\cdots(\lambda_{n-1}\boldsymbol{I}-\boldsymbol{A})(\lambda_n\boldsymbol{I}-\boldsymbol{A})\boldsymbol{v}_1 \\
&= (\lambda_2\boldsymbol{I}-\boldsymbol{A})(\lambda_3\boldsymbol{I}-\boldsymbol{A})\cdots(\lambda_{n-1}\boldsymbol{v}_1-\boldsymbol{A}\boldsymbol{v}_1)(\lambda_n-\lambda_1) \\
&\ \ \vdots \\
&= (\lambda_2\boldsymbol{I}-\boldsymbol{A})(\lambda_3\boldsymbol{I}-\boldsymbol{A})\boldsymbol{v}_1\cdots(\lambda_{n-1}-\lambda_1)(\lambda_n-\lambda_1) \\
&= (\lambda_2-\lambda_1)(\lambda_3-\lambda_1)\cdots(\lambda_{n-1}-\lambda_1)(\lambda_n-\lambda_1)\boldsymbol{v}_1
\end{aligned}
$$

由上面等式可以发现

$$
\begin{aligned}
\boldsymbol{Z}\left(\sum_{i=1}^{n} c_i\boldsymbol{v}_i\right) &= \sum_{i=1}^{n} c_i\boldsymbol{Z}\boldsymbol{v}_i \\
&= c_1\boldsymbol{Z}\boldsymbol{v}_1 \\
&= c_1(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)\cdots(\lambda_n-\lambda_1)\boldsymbol{v}_1 = \boldsymbol{0}
\end{aligned}
$$

# 特征值与特征向量

因为$\lambda_i$是唯一的，必定有$c_1 = 0$。依此类推，可以证明所有的$c_i$都必定为零，因此特征向量$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 是线性无关的。

考虑特征向量$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$构成的一组线性无关基。在这一组基下，可对矩阵$\boldsymbol{A}$进行对角化，即对所有的$i \neq j$,对角矩阵的第$(i,j)$个元素$a_{ij} = 0$。令

$$\boldsymbol{T} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n]^{-1}$$

则有

$$
\begin{aligned}
\boldsymbol{TAT}^{-1} &= \boldsymbol{TA}\left[\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\right] \\
&= \boldsymbol{T}\left[\boldsymbol{Av}_1, \boldsymbol{Av}_2, \cdots, \boldsymbol{Av}_n\right] \\
&= \boldsymbol{T}\left[\lambda_1\boldsymbol{v}_1, \lambda_2\boldsymbol{v}_2, \cdots, \lambda_n\boldsymbol{v}_n\right] \\
&= \boldsymbol{TT}^{-1}\begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}
\end{aligned}
$$

上面用到了关系式$\boldsymbol{TT}^{-1} = \boldsymbol{I}$。

对于矩阵$\boldsymbol{A}$,若$\boldsymbol{A} = \boldsymbol{A}^T$,则称$\boldsymbol{A}$为对称矩阵。

# 特征值与特征向量

对于任意$n \times n$实对称矩阵，存在$n$个相互正交的特征向量。

证明（$n$个特征值各不相同的情况）：

假定$\boldsymbol{A}\boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1$,$\boldsymbol{A}\boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_2$,其中$\lambda_1 \neq \lambda_2$,那么有

$$\boldsymbol{v}_1^T \boldsymbol{A}\boldsymbol{v}_2 = \boldsymbol{v}_1^T(\boldsymbol{A}\boldsymbol{v}_2) = \lambda_2(\boldsymbol{v}_1^T \boldsymbol{v}_2)$$

根据$\boldsymbol{A} = \boldsymbol{A}^T$,有

$$\boldsymbol{v}_1^T \boldsymbol{A}\boldsymbol{v}_2 = (\boldsymbol{v}_1^T \boldsymbol{A}^T)\boldsymbol{v}_2 = (\boldsymbol{A}\boldsymbol{v}_1)^T \boldsymbol{v}_2 = \lambda_1(\boldsymbol{v}_1^T \boldsymbol{v}_2)$$

因此，

$$\lambda_1(\boldsymbol{v}_1^T \boldsymbol{v}_2) = \lambda_2(\boldsymbol{v}_1^T \boldsymbol{v}_2)$$

由于$\lambda_1 \neq \lambda_2$,可以推出

$$(\boldsymbol{v}_1^T \boldsymbol{v}_2) = 0$$

# 特征值与特征向量

实际上，每一个特征值$\lambda$都对应一个重数，$k$重特征值实际上对应的是一个$k$维子空间，$k=1$时就是特征向量例如$n$阶单位矩阵只有一个特征值1，该特征值的重数为$n$，对应一个$n$维子空间。

# 特征值与特征向量

如果$\boldsymbol{A}$是对称矩阵，那么它的特征向量集合构成了$\mathbb{R}^n$空间的正交基。如果对基$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$进行标准化，使得每个向量$\boldsymbol{v}_i$的范数都为1，那么可以定义矩阵

$$\boldsymbol{T} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n]$$

该矩阵满足

$$\boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{I}$$

从而有

$$\boldsymbol{T}^T = \boldsymbol{T}^{-1}$$

如果一个矩阵的转置等于它的逆，那么该矩阵为正交矩阵。

# 二次型函数

设二次型函数$f:\mathbb{R}^n \to \mathbb{R}$ 定义为具有如下形式的函数:

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$$

其中$\boldsymbol{Q}$是一个$n \times n$实数矩阵。

$$\begin{aligned}
f(\boldsymbol{x}) &= q_{11}x_1^2 + q_{12}x_1x_2 + \cdots + q_{1n}x_1x_n \\
&+ q_{21}x_2x_1 + q_{22}x_2^2 + \cdots + q_{2n}x_2x_n \\
&+ \cdots \\
&+ q_{n1}x_nx_1 + q_{n2}x_2^2 + \cdots + q_{nn}x_n^2 \\
&= [x_1, x_2, \cdots, x_n] \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}
\end{aligned}$$

# 二次型函数

不妨设$\boldsymbol{Q}$是对称阵，即$\boldsymbol{Q}=\boldsymbol{Q}^T$。则

$$\begin{aligned} f(\boldsymbol{x}) &= q_{11}x_1^2 + q_{22}x_2^2 + \cdots + q_{nn}x_n^2 \\ &\quad + 2q_{12}x_1x_2 + q_{13}x_1x_3 + \cdots + q_{n-1,n}x_{n-1}x_n \end{aligned}$$

若$\boldsymbol{Q}$非对称，也可以用对称阵代替

$$\boldsymbol{Q}_0 = \boldsymbol{Q}_0^T = \frac{1}{2}\left(\boldsymbol{Q} + \boldsymbol{Q}^T\right)$$

可以看出

$$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{Q}_0\boldsymbol{x} = \frac{1}{2}\boldsymbol{x}^T\left(\boldsymbol{Q} + \boldsymbol{Q}^T\right)\boldsymbol{x}$$

# 顺序主子式

当对于任一非零向量$\boldsymbol{x}$,都有$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} > 0$, 则二次型$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$，$\boldsymbol{Q} = \boldsymbol{Q}^T$都是正定的, 若$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \geq 0$则此二次型是半正定。类似的， $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} < 0$，或者$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \leq 0$ 则说明二次型是负定或半负定的。

$\boldsymbol{Q}$的主子式是包括$\det \boldsymbol{Q}$和$\boldsymbol{Q}$移除第$i$行和第$i$列获得的其他子式。可表示为

$$\det \begin{bmatrix} q_{i_1 i_1} & q_{i_1 i_2} & \cdots & q_{i_1 i_p} \\ q_{i_2 i_1} & q_{i_2 i_2} & \cdots & q_{i_2 i_p} \\ \vdots & \vdots & & \vdots \\ q_{i_p i_1} & q_{i_p i_2} & \cdots & q_{i_p i_p} \end{bmatrix}, 1 \leq i_1 < \cdots < i_p \leq n, \quad p = 1, 2, \cdots, n$$

# 顺序主子式

矩阵$\boldsymbol{Q}$的顺序主子式为$\det \boldsymbol{Q}$自身以及从矩阵$\boldsymbol{Q}$ 中依次移除最后一行和最后一列获得的所有子式，即

$$\Delta_1 = q_{11}, \qquad \Delta_2 = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

$$\Delta_3 = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}, \cdots, \Delta_n = \det \boldsymbol{Q}$$

我们可以用西尔维斯特准则证明，根据$\boldsymbol{Q}$的顺序主子式判定二次型$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$是否正定。

# 西尔维斯特准则

- 给定的二次型$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$，其中$\boldsymbol{Q}=\boldsymbol{Q}^T$，该二次型是正定的，当且仅当的$\boldsymbol{Q}$顺序主子式是正定的。
- 举例，判断$\boldsymbol{Q}$对应的二次型$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$是否正定

$$\boldsymbol{Q}=\begin{bmatrix}5 & -2 & 3\\ -2 & 1 & 0\\ 3 & 0 & 1\end{bmatrix}$$

由于$\Delta_1=5>0,\Delta_2=1>0,\Delta_3=-8<0$，所以非正定。

# 利用特征值判断二次型函数正定

- 对称矩阵$\boldsymbol{Q}$是正定（半正定）的，当且仅当$\boldsymbol{Q}$的所有特征值是正的（非负的）。
- 证明：对于任意向量$\boldsymbol{x}$，令$\boldsymbol{y} = \boldsymbol{T}^{-1}\boldsymbol{x} = \boldsymbol{T}^T\boldsymbol{x}$, 其中$\boldsymbol{T}$是一个正交矩阵，各列就是矩阵$\boldsymbol{Q}$的特征向量。那么，有$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{T}^T\boldsymbol{Q}\boldsymbol{T}\boldsymbol{y} = \sum_{i=1}^n \lambda_i y_i^2$。由此可推知定理成立。

# 正定的二次型函数

通过对角化，可以证明一个对称的半正定矩阵$\boldsymbol{Q}$具有对称的半正定平方根$\boldsymbol{Q}^{1/2}$, 满足$\boldsymbol{Q}^{1/2}\boldsymbol{Q}^{1/2}=\boldsymbol{Q}$，采用算子$\boldsymbol{T}$，可将$\boldsymbol{Q}^{1/2}$表示为

$$\boldsymbol{Q}^{1/2}=\boldsymbol{T}\begin{bmatrix}\lambda_1^{1/2} & & & 0\\ & \lambda_2^{1/2} & & \\ & & \ddots & \\ 0 & & & \lambda_n^{1/2}\end{bmatrix}\boldsymbol{T}^T$$

易证$\boldsymbol{Q}^{1/2}$是半正定的对称阵。于是$\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}$ 可以表示为$||\boldsymbol{Q}^{1/2}\boldsymbol{x}||^2$。

以上给出了判定二次型函数和对称矩阵正定性的两类判断依据。

# 目录

# 线段

对于$n$维向量$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$, $\boldsymbol{y} = [y_1, y_2, \cdots, y_n]^T$，两点之间的所有点的集合称为两点之间的线段。如果$\boldsymbol{z}$在这条线段上，那么有

$$\boldsymbol{z} - \boldsymbol{y} = \alpha(\boldsymbol{x} - \boldsymbol{y})$$

其中，$\alpha$是位于区间$[0, 1]$的实数。或者可以写成 $\boldsymbol{z} = \alpha\boldsymbol{x}(1 - \alpha)\boldsymbol{y}$。因此，$\boldsymbol{x}, \boldsymbol{y}$之间的线段可以表示为

$$\{\alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y} : \alpha \in [0, 1]\}$$

# 超平面

令$u_1, u_2, \cdots, u_n, v \in \mathbb{R}$，其中至少存在一个不为零的$u_i$。由所有满足线性方程

$$u_1x_1 + u_2x_2 + \cdots + u_nx_n = v$$

的点$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$组成的集合称为空间$\mathbb{R}^n$的**超平面**。超平面可以写为

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T\boldsymbol{x} = v\}$$

其中，

$$\boldsymbol{u} = [u_1, u_2, \cdots, u_n]^T$$

注意：超平面不一定是$\mathbb{R}^n$的子空间，因为超平面通常不包含原点。而且超平面不一定是一个平面，在二维空间中就是一条直线，而三维空间中是一些普通平面。

如图，我们可以通过一些变换使其经过原点。

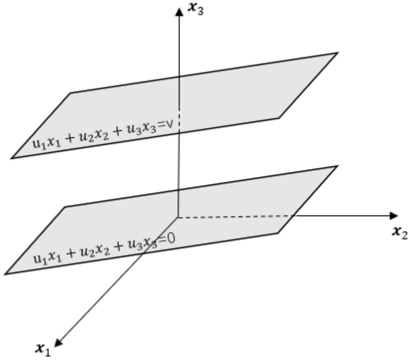


图 1: 超平面的变换

# 超平面

把超平面$H = \{\boldsymbol{x} : u_1x_1 + \cdots + u_nx_n = v\}$可以换一种形式去表述。
在超平面内一点$\boldsymbol{a}$和任一点$\boldsymbol{x}$，将满足$\boldsymbol{u}^T(\boldsymbol{x}-\boldsymbol{a}) = 0$ 于是有$H = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T(\boldsymbol{x}-\boldsymbol{a}) = 0\}$。

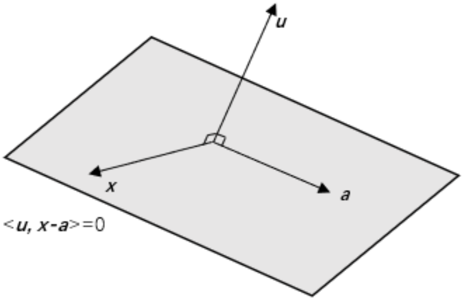


图 2: 超平面$H = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T(\boldsymbol{x}-\boldsymbol{a}) = 0\}$

# 目录

# 导数矩阵

任意从$\mathbb{R}^n$到$\mathbb{R}^m$的线性变换，特别是$\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^m$的导数$\mathcal{L}$，都可以表示一个$m\times n$的矩阵。为了确定可微函数$\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^m$的导数$\mathcal{L}$ 所对应的矩阵表示$\boldsymbol{L}$，引入$\mathbb{R}^n$空间的的标准基$\{\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n\}$。考虑向量

$$\boldsymbol{x}_j=\boldsymbol{x}_0+t\boldsymbol{e}_j,\quad j=1,\cdots,n$$

根据导数的定义，有

$$\lim_{t\to 0}\frac{\boldsymbol{f}(\boldsymbol{x}_j)-(t\boldsymbol{L}\boldsymbol{e}_j+\boldsymbol{f}(\boldsymbol{x}_0))}{t}=\boldsymbol{0}$$

这意味着，对于$j=1,\cdots,n$有

$$\lim_{t\to 0}\frac{\boldsymbol{f}(\boldsymbol{x}_j)-\boldsymbol{f}(\boldsymbol{x}_0)}{t}=\boldsymbol{L}\boldsymbol{e}_j$$

$\boldsymbol{L}\boldsymbol{e}_j$是矩阵$\boldsymbol{L}$的第$j$列，该列有$m$个元素。

# 导数矩阵

向量$\boldsymbol{x}_j$与$\boldsymbol{x}_0$仅仅在第$j$个元素上有差异，该元素上的差值为$t$。因此，上式的左边等于偏导数

$$\frac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{x}_0)$$

可以通过向量的每一个元素求极限的方式来计算向量极限，因此，如果

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$$

那么有

$$\frac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(\boldsymbol{x}_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(\boldsymbol{x}_0) \end{bmatrix}$$

# 导数矩阵

矩阵$\boldsymbol{L} = [\boldsymbol{L}\boldsymbol{e}_1, \boldsymbol{L}\boldsymbol{e}_2, \cdots, \boldsymbol{L}\boldsymbol{e}_n]$可写为

$$\left[\frac{\partial \boldsymbol{f}}{\partial x_1}(\boldsymbol{x}_0) \quad \cdots \quad \frac{\partial \boldsymbol{f}}{\partial x_n}(\boldsymbol{x}_0)\right] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{x}_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\boldsymbol{x}_0) \end{bmatrix}$$

矩阵$\boldsymbol{L}$称为在$\boldsymbol{f}$在点$\boldsymbol{x}_0$的雅克比矩阵或导数矩阵，记作$\boldsymbol{Df}(\boldsymbol{x}_0)$。我们常用$\boldsymbol{Df}(\boldsymbol{x}_0)$表示$\boldsymbol{f}$在点$\boldsymbol{x}_0$处的导数。另外，如果$f(\boldsymbol{x}) = \boldsymbol{A}x$，则$\boldsymbol{D}f = A$。

如果$f:\mathbb{R}^n\to\mathbb{R}$是可微的，那么函数

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{bmatrix} = Df(\boldsymbol{x})^T$$

称为$\boldsymbol{f}$的梯度。梯度是从一个从$\mathbb{R}^n$到$\mathbb{R}$的函数。如果绘制梯度向量$\nabla \boldsymbol{f}(\boldsymbol{x})$，其起点是$\boldsymbol{x}$，箭头表示方向，以此表示向量场。梯度的方向就是函数在这点增长最快的方向。

# 导数矩阵

设$\boldsymbol{x}$是$n \times 1$列向量，$\boldsymbol{A}$是$m \times n$矩阵， $\alpha$是标量

$$D(\boldsymbol{A}x) = \boldsymbol{A}$$

$$D(\alpha \boldsymbol{x}) = \alpha I$$

$$\frac{d(\alpha \boldsymbol{x})}{d\alpha} = \boldsymbol{x}$$

# 导数矩阵

- 给定函数$\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}$，如果梯度$\nabla\boldsymbol{f}$可微，则称$\boldsymbol{f}$是二次可微，$\nabla\boldsymbol{f}$的导数记为

$$D^2f=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n\partial x_1}\\ \frac{\partial^2 f}{\partial x_1\partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial^2 f}{\partial x_1\partial x_n} & \frac{\partial^2 f}{\partial x_2\partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2}\end{bmatrix}$$

  其中，$\frac{\partial^2 f}{\partial x_i\partial x_j}$表示$\boldsymbol{f}$首先对$x_i$求导，再对$x_j$求导的偏导数。矩阵$\boldsymbol{D}^2\boldsymbol{f}(\boldsymbol{x})$称为$\boldsymbol{f}$在点$\boldsymbol{x}$的黑塞矩阵，记作$\boldsymbol{F}(\boldsymbol{x})$。

- 给定函数$\boldsymbol{f}:\Omega\to\mathbb{R}^n,\Omega\subset\mathbb{R}^n$，如果该函数在$\Omega$上是可微的，且$\boldsymbol{Df}:\Omega\to\mathbb{R}^{m\times n}$是连续的，则称该函数在$\Omega$上是连续可微的，即$\boldsymbol{f}$的各个元素具有连续偏导数。满足这种条件的函数$\boldsymbol{f}$，将其记作$\boldsymbol{f}\in\mathcal{C}^1$。如果$\boldsymbol{f}$中的各元素具有$p$阶连续偏导数，记作$\boldsymbol{f}\in\mathcal{C}^p$。

# 导数矩阵

如果$f:\mathbb{R}^n\to\mathbb{R}$在点$x$是二次连续可微的，那么$f$在点$x$的黑塞矩阵是对称的。称作克莱罗定理或施瓦茨定理。然而，$f$的二次偏导数不连续，就不能保证黑塞矩阵是对称的。

# 导数矩阵

一个简单的例子：

$$f(\boldsymbol{x}) = x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6$$

$$\boldsymbol{F} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 6 \end{bmatrix}$$

# 目录

# 微分法则

利用函数$\boldsymbol{f}:\mathbb{R}\to\mathbb{R}^n$和函数$\boldsymbol{g}:\mathbb{R}^n\to\mathbb{R}^m$可构成复合函数$\boldsymbol{h}(t)=\boldsymbol{g}(\boldsymbol{f}(t))$,对其微分需要用到链式法则。

$m=1$时,

$$\boldsymbol{h}'(t)=D\boldsymbol{g}(\boldsymbol{f}(t))D\boldsymbol{f}(t)=\nabla\boldsymbol{g}(\boldsymbol{f}(t))^TD\boldsymbol{f}(t)=\nabla\boldsymbol{g}(\boldsymbol{f}(t))^T\begin{bmatrix}\boldsymbol{f}_1'(t)\\ \vdots\\ \boldsymbol{f}_n'(t)\end{bmatrix}$$

例如

$$\boldsymbol{h}(t)=\boldsymbol{g}(\boldsymbol{f}(t)),\boldsymbol{g}(\boldsymbol{f})=\begin{bmatrix}f_1^2+f_2^2\\ 2f_1f_2\end{bmatrix},\boldsymbol{f}(t)=\begin{bmatrix}2t\\ t\end{bmatrix}$$

则

$$\boldsymbol{h}'(t)=\begin{bmatrix}2f_1 & 2f_2\\ 2f_2 & 2f_1\end{bmatrix}\begin{bmatrix}2\\ 1\end{bmatrix}=\begin{bmatrix}10t\\ 8t\end{bmatrix}$$

# 微分法则

乘法法则：设定$\boldsymbol{f}:\mathbb{R}^n\to\mathbb{R}^m$和$\boldsymbol{g}:\mathbb{R}^m\to\mathbb{R}$表示两个可微函数，函数$\boldsymbol{h}:\mathbb{R}^n\to\mathbb{R}$定义为$h(\boldsymbol{x})=f(\boldsymbol{x})^T g(\boldsymbol{x})$，那么$h$也是可微的，且

$$Dh(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{x})^T D\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{g}(\boldsymbol{x})^T D\boldsymbol{f}(\boldsymbol{x})$$

# 微分法则

一些不同情况下函数关于$\boldsymbol{x}$的导数：给定矩阵$\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 和向量$\boldsymbol{y} \in \mathbb{R}^{m}$，有

$$D(\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{y}^T\boldsymbol{A}$$

特别的，当m = n时，有

$$D(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}^T(\boldsymbol{A}+\boldsymbol{A}^T)$$

如果$\boldsymbol{y} \in \mathbb{R}^{n}$，那么

$$D(\boldsymbol{y}^T\boldsymbol{x}) = \boldsymbol{y}^T$$

如果$\boldsymbol{Q}$是对矩阵，那么

$$D(\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}) = 2\boldsymbol{x}^T\boldsymbol{Q}$$

特别的，当$\boldsymbol{Q} = \boldsymbol{I}$时，有

$$D(\boldsymbol{x}^T\boldsymbol{x}) = 2\boldsymbol{x}^T$$

# 目录

# 泰勒级数

关于泰勒级数，在这里给出一些结论和证明：

(a)设函数$f(\boldsymbol{x})$ ：$\mathbb{R}^n \rightarrow \mathbb{R}$。若$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$的某个邻域 $N(\bar{\boldsymbol{x}})$内一阶连续可微， 存在$\theta \in (0,1)$，使得

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}} + \theta(\boldsymbol{x} - \bar{\boldsymbol{x}}))^T(\boldsymbol{x} - \bar{\boldsymbol{x}}), \boldsymbol{x} \in N(\bar{\boldsymbol{x}})$$

(b)设函数$f(\boldsymbol{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$，若$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$的某个邻域 $N(\bar{\boldsymbol{x}})$内的一阶连续可微，则

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) + o(||\boldsymbol{x} - \bar{\boldsymbol{x}}||), \boldsymbol{x} \in N(\bar{\boldsymbol{x}})$$

(c)设函数$f(\boldsymbol{x})$ ：$\mathbb{R}^n \rightarrow \mathbb{R}$，若$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$的某个邻域 $N(\bar{\boldsymbol{x}})$内的二阶连续可微，则存在$\theta \in (0,1)$使得

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla^2 f(\bar{\boldsymbol{x}} + \theta(\boldsymbol{x} - \bar{\boldsymbol{x}}))(\boldsymbol{x} - \bar{\boldsymbol{x}}), \boldsymbol{x} \in N(\bar{\boldsymbol{x}}) \quad (1)$$

(d)设函数$f(\boldsymbol{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$，若$f(\boldsymbol{x})$在点$\bar{\boldsymbol{x}}$的某个邻域 $N(\bar{\boldsymbol{x}})$内的二阶连续可微，则

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla^2 f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) + o(||\boldsymbol{x} - \bar{\boldsymbol{x}}||^2), \boldsymbol{x} \in N(\bar{\boldsymbol{x}}) \quad (2)$$

# 泰勒级数

证明　结论(a)和(b)这里不予证明，感兴趣的同学可以自己尝试证明，下面证明结论(c)、(d)

(c)当$\boldsymbol{x}=\bar{\boldsymbol{x}}$时，显然成立，因此我们只考虑$\boldsymbol{x}\neq\bar{\boldsymbol{x}}$情况，设

$$\phi(a)=f(\bar{\boldsymbol{x}}+a\boldsymbol{d})$$

其中$\boldsymbol{d}=\boldsymbol{x}-\bar{\boldsymbol{x}}$，由一元函数的Taylor公式有

$$\phi(a)=\phi(0)+\phi'(0)a+\frac{1}{2}\phi''(\theta)a^2$$

其中$0<\theta<a$，取$a=1$得

$$\phi(1)=\phi(0)+\phi'(0)+\frac{1}{2}\phi''(\theta) \tag{3}$$

显然$\phi(1)=f(\boldsymbol{x}),\phi(0)=f(\bar{\boldsymbol{x}})$，可推得

$$\phi'(0)=\boldsymbol{d}^T\nabla f(\bar{\boldsymbol{x}}),$$
$$\phi''(\theta)=\boldsymbol{d}^T\nabla^2 f(\bar{\boldsymbol{x}}+\theta\boldsymbol{d})\boldsymbol{d}$$

将以上式子带入3式，便得1式。

# 泰勒级数

(d)设

$$\phi(a) = f(\bar{\boldsymbol{x}} + a\boldsymbol{d})$$

其中$a = ||\boldsymbol{x} - \bar{\boldsymbol{x}}||, \boldsymbol{d} = \frac{\boldsymbol{x}-\bar{\boldsymbol{x}}}{||\boldsymbol{x}-\bar{\boldsymbol{x}}||}$，由一元函数的Taylor公式有

$$\phi(a) = \phi(0) + \phi'(0)a + \frac{1}{2}\phi''(0)a^2 + o(a^2) \tag{4}$$

又有

$$\phi(a) = f(\boldsymbol{x}), \phi(0) = f(\bar{\boldsymbol{x}}),$$

$$\phi'(0)a = \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}),$$

$$\phi''(0)a^2 = (\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla^2 f(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}})$$

将以上式子带入4式，可得到2式。

# 泰勒级数

在(2)和(4)中，若略去高阶无穷小，就会有近似关系式

$$f(\boldsymbol{x}) \approx f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}), \boldsymbol{x} \in N(\bar{\boldsymbol{x}})$$

$$f(\boldsymbol{x}) \approx f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \nabla f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \boldsymbol{x} \in N(\bar{\boldsymbol{x}})$$

通常，我们把上式的右边称作函数$f(\boldsymbol{x})$再点$\bar{x}$处的线性近似（函数）和二次近似（函数）。

# Thank you for your attention!

翟卫欣 副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院