



#### 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



最优化方法

## 目录

向量与矩阵

线段与超平面

导数矩阵

微分法则

水平集与梯度

泰勒级数





### 目录

#### 向量与矩阵

线段与超平面

导数矩阵

微分法则

水平集与梯度

泰勒级数



## 向量与矩阵

n维列向量定义为含有n个数的数组。记为

$$m{a} = egin{bmatrix} a_1 \ a_2 \ dots \ a_n \end{bmatrix}$$

 $a_i$ 表示向量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x)。向量 $x \in \mathbb{R}^n$ 中元素记为 $x_1, \dots, x_n$ 。

<ロ > < 回 > < 回 > < 巨 > < 巨 > 三 り < で

4 / 70

最优化方法 Optimization Methods



#### 向量与矩阵

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

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

那么

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

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

4□ > 4回 > 4 = > 4 = > 9 < 0</p>

最优化方法



• 如果方程

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

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 $\{a_1,a_2,\cdots,a_k\}$ 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**0**,由于对于任意 $\alpha \neq 0$ ,都有 $\alpha$ **0** = **0**, 因此,该集合是线性相关的。实际上,所有包含**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 $a \neq 0$ ,只有 $\alpha = 0$ 时,才有 $\alpha a = 0$ 0成立,因此,该集合是线性无关的。
- 给定向量 $\mathbf{a}$ , 如果存在标量 $\alpha_1, \dots, \alpha_k$ ,使得

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \dots + \alpha_k \mathbf{a}_k$$

那么称向量a为 $a_1, a_2, \cdots, a_k$ 的线性组合。

- 向量集 $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 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 $\{\alpha_1, \alpha_2, \cdots, \alpha_k\}$ 是线性相关的,那么有

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

其中至少存在一个标量 $\alpha_i \neq 0$ ,从而有

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

4日 > 4目 > 4目 > 4目 > 4目 > 9 < 0</p>

• 充分性。假定向量 $a_1$ 可以被表示为其他向量的线性组合:

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

那么有

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

因为第1个标量非零,所以向量集 $\{a_1, a_2, \cdots, a_k\}$ 是线性相关的。 类似的,如果将 $a_i, i = 2, \cdots, k$ 表示为其他向量的线性组合,也可以得到同样的结论。

◆□▶ ◆□▶ ◆■▶ ◆■ ● かへぐ

- ◆ 令ン表示ℝ"的一个子集,如果ン在向量加和运算及标量乘积运算下是封闭的,那么称以为ℝ"的子空间。
- 每个子空间都包含零向量0。
- 假定 $a_1, a_2, \dots, a_k$ 是 $\mathbb{R}^n$ 中的任意向量,它们所有线性组合的集合称为 $a_1, a_2, \dots, a_k$ 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 $\{a_1,a_2,\cdots,a_k\}\subset\mathcal{V}$  使得 $\mathcal{V}=span\left[a_1,a_2,\cdots,a_k\right]$ ,那么称 $\{a_1,a_2,\cdots,a_k\}$  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 $dim\mathcal{V}$ 。

◆ロト ◆昼 ▶ ◆ 豊 ト 豊 り 9 ○ ○

CHINA AGRICUL

如果 $\{a_1, a_2, \cdots, a_k\}$ 是 $\mathcal{V}$ 的一组基,那么 $\mathcal{V}$ 中的任一向量a可以唯一地表示为

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

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



#### 标准基

给定V的一组基 $\{a_1, a_2, \cdots, a_k\}$ 和向量 $a \in V$ ,如果

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

那么系数 $\alpha_i$ ,  $i=1,\cdots,k$ 称为a对应于基 $\{a_1,a_2,\cdots,a_k\}$ 的坐标。 $\mathbb{R}^n$ 的标准基为

在标准基下,向量x可表示为

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



#### 对矩阵进行一下三种变换的称为行初等变换

- 对换两行(对换i,i两行,记作 $r_i \leftrightarrow r_i$ );
- 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 $r_i + kr_i$ ).

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。

◆ロト ◆部 → ◆ き → ◆ き ・ か へ で

12 / 70



#### 三种初等变换对应三种初等矩阵

 $(i)r_i \leftrightarrow r_i$ 对应的的初等矩阵

第i行

← 第j行



#### (ii) $r_i \times k$ 对应的的初等矩阵

$$E(i(k)) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & k & & \\ & & & & 1 & \\ & & & & \ddots & \\ & & & & & 1_{2} \end{pmatrix}$$

CHINA AGRICUI

← 第i行



#### (iii) $r_i + kr_j$ 对应的的初等矩阵

CHINA AGRICU

← 第j行

← 第i行

## 矩阵的秩

- 矩阵A中线性无关列的最大数目称为A的秩,记为rankA。 矩阵A的秩等于它的非零子式的最高阶数。
- 在以下运算中,矩阵A的秩保持不变:
  - 1.矩阵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detA或|A|。
- 如果一个 $m \times n \ (m \ge n)$ 矩阵A具有非零的n阶子式,那么A的各列是线性无关的,即rank A = n。

4 D > 4 B > 4 E > 4 E > 9 Q C

最优化方法 Optimization Methods



## 内积和范数

对于 $x, y \in \mathbb{R}^n$ ,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x的欧氏范数为

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

向量x的欧氏范数||x||具有如下性质:

- 1.非负性: x的欧氏范数 $||x|| \ge 0$ ,当且仅当x = 0时, ||x|| = 0
- 2.齐次性: $||r\mathbf{x}|| = |r|||\mathbf{x}|| \ge 0, r \in \mathbb{R}$
- 3.三角不等式: ||x + y|| < ||x|| + ||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:

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

 $\infty$ 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

$$\|\mathbf{x}\|_{p} = \begin{cases} (|x_{1}|^{p} + \dots + |x_{n}|^{p})^{\frac{1}{p}}, & 1 \leq p < \infty \\ \max\{|x_{1}|, \dots, |x_{n}|\}, & p = \infty \end{cases}$$



# 如何求方阵的逆

设A为矩阵,如果存在n阶方阵B,使得

$$AB = BA = I$$

则称A是可逆矩阵,B是A的逆矩阵。

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

20 / 70



## 如何求方阵的逆

当 $|A| \neq 0$ 时,

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

这是 $A^{-1}$ 的计算公式,其中 $A^*$ 是A的伴随阵。



最优化方法



## 求方阵的逆: 伴随矩阵法

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

解:

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

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

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

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

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

最优化方法 Optimization Methods 22 / 70



23 / 70

### 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 的说。

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

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

Optimization Methods





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

$$\begin{array}{c} r_3 \times 3 \\ r_1 + 2r_2 \\ r_1 \leftrightarrow r_2 \end{array} \begin{pmatrix} 3 & 0 & -2 & 0 & 1 & 0 \\ 0 & -2 & 1 & 1 & 0 & 0 \\ 0 & 9 & -4 & 0 & 2 & 3 \end{pmatrix} \xrightarrow{r_3 \times 2} \begin{pmatrix} 3 & 0 & -2 & 0 & 1 & 0 \\ 0 & -2 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 9 & 4 & 6 \end{pmatrix}$$

$$\frac{r_{1}+2r_{3}}{r_{2}-r_{3}} \begin{pmatrix} 3 & 0 & 0 & 18 & 9 & 12 \\ 0 & -2 & 0 & -8 & -4 & -6 \\ 0 & 0 & 1 & 9 & 4 & 6 \end{pmatrix} \underbrace{r_{1}/3}_{r_{2}/(-2)} \begin{pmatrix} 1 & 0 & 0 & 6 & 3 & 4 \\ 0 & 1 & 0 & 4 & 2 & 3 \\ 0 & 0 & 1 & 9 & 4 & 6 \end{pmatrix}$$

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

CHINA AGRICU

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

$$Av = \lambda v$$

 $\lambda$ 称为A的特征值, $\nu$ 称为A的特征向量。

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

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

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

多项式 $det[\lambda I - A]$ 称为矩阵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A的n个特征值。若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。

◆ロ → ◆園 → ◆ 恵 → ◆ 恵 → りへで

• 假定特征方程 $det[\lambda I - A] = 0$ 存在n个相异的根 $\lambda_1, \lambda_2, \dots, \lambda_n$ ,那么存在n个线性无关的特征向量 $v_1, v_2, \dots, v_n$ ,使得

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

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

27 / 70



考虑矩阵

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

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

$$Z\mathbf{v}_{n} = (\lambda_{2}\mathbf{I} - \mathbf{A})(\lambda_{3}\mathbf{I} - \mathbf{A}) \cdots (\lambda_{n-1}\mathbf{I} - \mathbf{A})(\lambda_{n}\mathbf{I} - \mathbf{A})\mathbf{v}_{n}$$
$$= (\lambda_{2}\mathbf{I} - \mathbf{A})(\lambda_{3}\mathbf{I} - \mathbf{A}) \cdots (\lambda_{n-1}\mathbf{I} - \mathbf{A})(\lambda_{n}\mathbf{v}_{n} - \mathbf{A}\mathbf{v}_{n})$$
$$= \mathbf{0}$$

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

$$\mathbf{Z}\mathbf{v}_{k}=\mathbf{0}, \quad k=2,3,\cdots,n$$



但是,

$$Z\mathbf{v}_{1} = (\lambda_{2}\mathbf{I} - \mathbf{A})(\lambda_{3}\mathbf{I} - \mathbf{A}) \cdots (\lambda_{n-1}\mathbf{I} - \mathbf{A})(\lambda_{n}\mathbf{I} - \mathbf{A})\mathbf{v}_{1}$$

$$= (\lambda_{2}\mathbf{I} - \mathbf{A})(\lambda_{3}\mathbf{I} - \mathbf{A}) \cdots (\lambda_{n-1}\mathbf{v}_{1} - \mathbf{A}\mathbf{v}_{1})(\lambda_{n} - \lambda_{1})$$

$$\vdots$$

$$= (\lambda_{2}\mathbf{I} - \mathbf{A})(\lambda_{3}\mathbf{I} - \mathbf{A})\mathbf{v}_{1} \cdots (\lambda_{n-1} - \lambda_{n})(\lambda_{n} - \lambda_{1})$$

$$= (\lambda_{2} - \lambda_{1})(\lambda_{3} - \lambda_{1}) \cdots (\lambda_{n-1} - \lambda_{1})(\lambda_{n} - \lambda_{1})\mathbf{v}_{1}$$

由上面等式可以发现

$$\mathbf{Z}\left(\sum_{i=1}^{n} c_{i} \mathbf{v}_{i}\right) = \sum_{i=1}^{n} c_{i} \mathbf{Z} \mathbf{v}_{i}$$

$$= c_{1} \mathbf{Z} \mathbf{v}_{i}$$

$$= c_{1}(\lambda_{2} - \lambda_{1})(\lambda_{3} - \lambda_{1}) \cdots (\lambda_{n} - \lambda_{1}) \mathbf{v}_{1} = \mathbf{0}$$

最优化方法 Optimization Methods

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 $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 是线性无关的。

考虑特征向量 $\{v_1, v_2, \cdots, v_n\}$ 构成的一组线性无关基。在这一组基下,可对矩阵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

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

最优化方法



则有

$$TAT^{-1} = TA [v_1, v_2, \cdots, v_n]$$

$$= T [Av_1, Av_2, \cdots, Av_n]$$

$$= T [\lambda_1 v_1, \lambda_2 v_2, \cdots, \lambda_n v_n]$$

$$= TT^{-1} \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & \ddots \\ 0 & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & \ddots \\ 0 & \lambda_n \end{bmatrix}$$

上面用到了关系式 $TT^{-1} = I$ 。

对于矩阵A,若 $A = A^T$ ,则称A为对称矩阵。





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 $Av_1 = \lambda_1 v_1 Av_2 = \lambda_2 v_2$ ,其中 $\lambda_1 \neq \lambda_2$ ,那么有

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

根据 $\mathbf{A} = \mathbf{A}^T$ ,有

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

因此,

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

由于 $\lambda_1 \neq \lambda_2$ ,可以推出

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

(ロ) (部) (注) (注) 注 り(0)



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。

最优化方法

如果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 $\{v_1, v_2, \cdots, v_n\}$ 进行标准化,使得每个向量 $v_i$ 的范数都为1,那么可以定义矩阵

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

该矩阵满足

$$T^TT = I$$

从而有

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

如果一个矩阵的转置等于它的逆,那么该矩阵为正交矩阵。

◆□▶ ◆□▶ ◆■▶ ◆■ ◆9<0</p>



#### 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:

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

其中Q是一个 $n \times n$ 实数矩阵。

$$f(\mathbf{x}) = q_{11}x_1^2 + q_{12}x_1x_2 + \dots + q_{1n}x_1x_n + q_{21}x_2x_1 + q_{22}x_2^2 + \dots + q_{2n}x_2x_n + \dots + q_{n1}x_nx_1 + q_{n2}x_2^2 + \dots + q_{nn}x_n^2$$

$$= [x_1, x_2, \dots, x_n] \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ q_{n1} & q_{n2} & \dots & q_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}^T \mathbf{Q} \mathbf{x}$$





#### 二次型函数

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

$$f(\mathbf{x}) = q_{11}x_1^2 + q_{22}x_2^2 + \dots + q_{nn}x_n^2$$
  
+  $2q_{12}x_1x_2 + q_{13}x_1x_3 + \dots + q_{n-1,n}x_{n-1}x_n$ 

若Q非对称,也可以用对称阵代替

$$oldsymbol{\mathcal{Q}}_0 = oldsymbol{\mathcal{Q}}_0^T = rac{1}{2} \left( oldsymbol{\mathcal{Q}} + oldsymbol{\mathcal{Q}}^T 
ight)$$

可以看出

$$x^{T}Qx = x^{T}Q_{0}x = \frac{1}{2}x^{T}(Q + Q^{T})x$$

## 顺序主子式

当对于任一非零向量x,都有 $x^TQx>0$ ,则二次型 $x^TQx$ , $Q=Q^T$ 都是正定的,若 $x^TQx\geq0$ 则此二次型是半正定。类似的,  $x^TQx<0$ ,或者 $x^TQx\leq0$ 则说明二次型是负定或半负定的。

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

$$\det \begin{bmatrix} q_{i_1i_1} & q_{i_1i_2} & \cdots & q_{i_1i_p} \\ q_{i_2i_1} & q_{i_2i_2} & \cdots & q_{i_2i_p} \\ \vdots & \vdots & & \vdots \\ q_{i_pi_1} & q_{i_pi_2} & \cdots & q_{i_pi_p} \end{bmatrix}, 1 \leq i_1 < \cdots < i_p \leq n, \quad p = 1, 2, \cdots, n$$

4 D > 4 B > 4 B > 4 B > 4 C

37 / 70

最优化方法 Optimization Methods

## 顺序主子式

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

$$\Delta_1 = q_{11}, \qquad \Delta_2 = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$
 $\Delta_3 = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{21} & q_{22} & q_{23} \end{bmatrix}, \cdots, \Delta_n = \det \boldsymbol{Q}$ 

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

4 D > 4 B > 4 E > 4 E > 9 Q C

38 / 70

最优化方法 Optimization Methods

## 西尔维斯特准则

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

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

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

最优化方法



## 利用特征值判断二次型函数正定



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

4日 > 4目 > 4目 > 4目 > 4目 > 990

最优化方法 Optimization Methods

## 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Q具有对称的半正定平方根 $Q^{1/2}$ ,满足 $Q^{1/2}=Q$ ,采用算子T,可将 $Q^{1/2}$ 表示为

$$oldsymbol{\mathcal{Q}}^{1/2} = oldsymbol{T} egin{bmatrix} \lambda_1^{1/2} & & & 0 \ & \lambda_2^{1/2} & & \ & & \ddots & \ 0 & & & \lambda_n^{1/2} \end{bmatrix} oldsymbol{T}^T$$

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

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。

◆□ > ◆□ > ◆ = > ◆ = > ◆ へ ○

41 / 70

最优化方法 Optimization Methods

## 目录

向量与矩阵

#### 线段与超平面

导数矩阵

微分法则

水平集与梯度

泰勒级数



# 线段

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

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

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

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

◆□▶ ◆□▶ ◆■▶ ◆■ ● かへぐ

## 超平面

 $\phi u_1, u_2, \cdots, u_n, v \in \mathbb{R}$ , 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 $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T$ 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R"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。

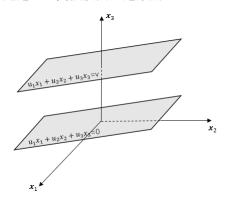


图 1: 超平面的变换



#### 超平面

把超平面 $H = \{ \boldsymbol{x} : u_1 x_1 + \dots + u_n x_n = v \}$ 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 $H = \{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T(\boldsymbol{x} - \boldsymbol{a}) = 0 \}$ 。

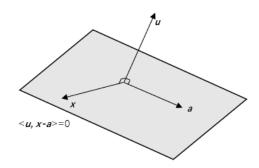


图 2: 超平面 $H = \{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^T (\boldsymbol{x} - \boldsymbol{a}) = 0 \}$ 

## 目录

向量与矩阵

线段与超平面

#### 导数矩阵

微分法则

水平集与梯度

泰勒级数



47 / 70

任意从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换,特别是 $f: \mathbb{R}^n \to \mathbb{R}^m$ 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 $f: \mathbb{R}^n \to \mathbb{R}^m$ 的导数 $\mathcal{L}$  所对应的矩阵表示L,引入 $\mathbb{R}^n$ 空间的的标准基 $\{e_1, \cdots, e_n\}$ 。考虑向量

$$\mathbf{x}_{i} = \mathbf{x}_{0} + t\mathbf{e}_{i}, \quad j = 1, \cdots, n$$

根据导数的定义,有

$$\lim_{t\to 0} \frac{f(\mathbf{x}_j) - (t\mathbf{Le}_j + f(\mathbf{x}_0))}{t} = \mathbf{0}$$

这意味着,对于 $i=1,\dots,n$ 有

$$\lim_{t\to 0} \frac{f(x_j) - f(x_0)}{t} = Le_j$$

 $Le_{i}$ 是矩阵L的第i列,该列有m个元素。

**◆□▶◆□▶◆豆▶◆豆> 豆 か**900

向量 $x_j$ 与 $x_0$ 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

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

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

$$m{f}(m{x}) = egin{bmatrix} f_1(m{x}) \\ \vdots \\ f_m(m{x}) \end{bmatrix}$$

那么有

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

矩阵 $L = [Le_1, Le_2, \cdots, Le_n]$ 可写为

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

矩阵L称为在f在点 $x_0$ 的<mark>雅克比矩阵或导数矩阵</mark>,记作 $Df(x_0)$ 。我们常用 $Df(x_0)$ 表示f在点 $x_0$ 处的导数。另外,如果f(x) = Ax,则Df = A。如果 $f: \mathbb{R}^n \to \mathbb{R}$ 是可微的,那么函数

$$abla f(oldsymbol{x}) = egin{bmatrix} rac{\partial f}{\partial x_1}(oldsymbol{x}) \ dots \ rac{\partial f}{\partial x_n}(oldsymbol{x}) \end{bmatrix} = Df(oldsymbol{x})^T$$

称为f的<mark>梯度</mark>。梯度是从一个从 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 的函数。如果绘制梯度向量 $\nabla f(x)$ ,其起点是x,箭头表示方向,以此表示向量场。



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

$$D(\mathbf{A}x) = \mathbf{A}$$
$$D(\alpha \mathbf{x}) = \alpha \mathbf{I}$$
$$\frac{d(\alpha \mathbf{x})}{d\alpha \mathbf{x}} = \mathbf{x}$$

 $d\alpha$ 



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

$$D^{2}f = \begin{bmatrix} \frac{\partial^{2}f}{\partial x_{1}^{2}} & \frac{\partial^{2}f}{\partial x_{2}\partial x_{1}} & \cdots & \frac{\partial^{2}f}{\partial x_{n}\partial x_{1}} \\ \frac{\partial^{2}f}{\partial x_{1}\partial x_{2}} & \frac{\partial^{2}f}{\partial x_{2}^{2}} & \cdots & \frac{\partial^{2}f}{\partial x_{n}\partial x_{2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^{2}f}{\partial x_{1}\partial x_{n}} & \frac{\partial^{2}f}{\partial x_{2}\partial x_{n}} & \cdots & \frac{\partial^{2}f}{\partial x_{n}^{2}} \end{bmatrix}$$

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

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



CHINA AGRICU

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。

53 / 70

最优化方法 Optimization Methods



#### 一个简单的例子:

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

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





## 目录

线段与超平面

导数矩阵

#### 微分法则

水平集与梯度

泰勒级数





## 微分法则

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

$$h^{'}(t) = D\mathbf{g}(\mathbf{f}(t))D\mathbf{f}(t) = \nabla \mathbf{g}(\mathbf{f}(t))^{T}D\mathbf{f}(t) = \nabla \mathbf{g}(\mathbf{f}(t))^{T}\begin{bmatrix} \mathbf{f}_{1}^{'}(t) \\ \vdots \\ \mathbf{f}_{n}^{'}(t) \end{bmatrix}$$

## 微分法则

CHINA AGRICUI

乘法法则: 设定 $\mathbf{f}:\mathbb{R}^n \to \mathbb{R}^m$  和 $\mathbf{g}:\mathbb{R}^n \to \mathbb{R}^m$ 表示两个可微函数,函

数 $h: \mathbb{R}^n \to \mathbb{R}$  定义为 $h(x) = f(x)^T g(x)$ , 那么h也是可微的, 且

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



## 微分法则

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

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

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

如果 $\mathbf{y} \in \mathbb{R}^n$ . 那么

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

如果Q是对矩阵,那么

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

特别的, 当Q = I时, 有

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

## 目录

向量与矩阵

线段与超平面

导数矩阵

微分法则

水平集与梯度

泰勒级数





## 水平集

函数 $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}$ 在水平c上的水平集定义为

$$S = \{ \boldsymbol{x} : f(\boldsymbol{x}) = c \}$$

对于 $f: \mathbb{R}^2 \to \mathbb{R}$ 水平集S是一条曲线,能够直接观察;

对于 $f: \mathbb{R}^3 \to \mathbb{R}$ , 水平集S通常是一组曲面。



# 梯度

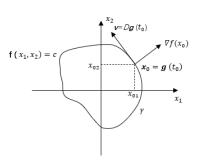


图 3: 梯度与水平集中任意曲线的切向量正交

如图, 当f连续可微时, 以下定理成立。

定理: 对于水平为 $f(\mathbf{x}) = f(\mathbf{x}_0)$ 的水平集中的任意一条经过 $\mathbf{x}_0$ 的光滑曲

线,其在点 $x_0$ 的切向量与函数f在点 $x_0$ 的梯度 $\nabla f(x_0)$ 正文。

最优化方法 Optimization Methods 61 / 70



# 梯度

- 根据定理,可以说 $\nabla f(\mathbf{x}_0)$ 在点 $\mathbf{x}_0$ 正交于水平集 $\mathbf{S}$ ,或者说 $\nabla f(\mathbf{x}_0)$ 是水平集在点 $\mathbf{x}_0$ 的法向量。同时,如果 $\nabla f(\mathbf{x}_0) \neq 0$ ,那么所有满足 $\nabla f(\mathbf{x}_0)^T(\mathbf{x} \mathbf{x}_0) = 0$ 的所有 $\mathbf{x}$ 组合的集合称为水平集 $\mathbf{S}$ 在点 $\mathbf{x}_0$ 的切平面(或切线)。
- 接下来将会发现, $\nabla f(\mathbf{x}_0)$ 是函数f在点 $\mathbf{x}_0$ 处增加速度最快的方向。由于 $\nabla f(\mathbf{x}_0)$ 正交于水平为 $\nabla f(\mathbf{x}) = \nabla f(\mathbf{x}_0)$ 的水平集,因此,可以推论:一个实值可微函数在某点增加速度最快的方向正交于函数f在该点的水平集。

62 / 70

63 / 70

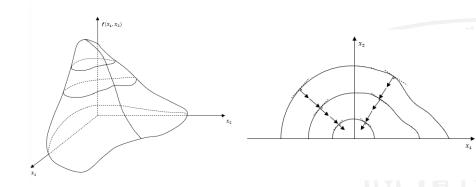


图 4: 最快上升路径

如图给出 $f: \mathbb{R}^2 \to \mathbb{R}$ 时,最速上升路径的示意图。图中,阴影曲面上自上而下的曲线具有如下性质:

曲线在 $(x_1,x_2)$ 平面上投影总是正交于轮廓线。由于该曲线指示的是函数f增长速度最快的方向,因此称为最速上升路径。

# 梯度

函数 $f: \mathbb{R}^n \to \mathbb{R}$ 的图像为集合  $\{[\mathbf{x}^T, f(\mathbf{x})]^T, \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$ ,函数f的梯度可以理解为函数图像的的一个切平面,且是一个超平面。

易证:如果f在点x0处可微,那么切线超平面可以用梯度表示:

$$z - z_0 = Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = (\mathbf{x} - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0)$$

证明过程: 切平面可以写作

$$[\mathbf{u}^T, v] \begin{bmatrix} \mathbf{x} - \mathbf{x}_0 \\ z - z_0 \end{bmatrix} = 0$$

若 $v \neq 0$ ,则可写为 $z - z_0 = d^T(x - x_0)$ 的形式,由于f(x)和z(x)在 $x_0$ 有相同的偏导数,所以 $z - z_0 = \nabla z(x)^T(x - x_0) = \nabla f(x)^T(x - x_0)$ 若v = 0. 在 $x_0$ 处的梯度为无穷。

## 目录

向量与拓荫

**华**郎 上 却 示 而

导数矩阵

微分法则

水平集与梯度

泰勒级数



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

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

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

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

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

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

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

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

最优化方法 Optimization Methods 66 / 70



(3)

## 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 $x = \bar{x}$ 时,显然成立,因此我们只考虑 $x \neq \bar{x}$ 情况,设

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

其中 $d = x - \bar{x}$ ,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

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

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

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

将以上式子带入3式,便得1式。

(ロ) (部) (注) (注) 注 の(



## 泰勒级数

(d)设

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

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

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

又有

$$\phi(a) = f(\mathbf{x}), \phi(0) = f(\overline{\mathbf{x}}),$$
  
$$\phi'(0)a = \nabla f(\overline{\mathbf{x}})^T (\mathbf{x} - \overline{\mathbf{x}}),$$
  
$$\phi''(0)a^2 = (\mathbf{x} - \overline{\mathbf{x}})^T \nabla^2 f(\overline{\mathbf{x}})^T (\mathbf{x} - \overline{\mathbf{x}})$$

将以上式子带入4式,可得到2式。

CHINA AGRICUI

(4)

(ロ) (個) (量) (量) (量) のQで

## 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

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

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

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

◆□▶ ◆□▶ ◆■▶ ◆■ ● かへぐ





# 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

4D > 4A > 4B > 4B > B 900