2019

最优化理论与方法

研究生学位课

陈军华（副教授\主任）

- **基础知识**
  - **1、向量**
  - **2、凸集/凸组合**
  - **3、超平面**
  - **4、函数二次型与标准二次型**
  - **5、正定矩阵**
  - **6、梯度与方向**
  - **7、Hesse矩阵**
  - **8、泰勒公式**
  - **9、方向导数与最速下降方向**
  - **10、等高线**

1 向量

- 列向量

$$\boldsymbol{x} = (x_1, x_2 \dots x_n)^T \qquad \mathbf{y} = (y_1, y_2 \dots y_n)^T$$

- 向量积

$$\boldsymbol{A} = \boldsymbol{x} \cdot \boldsymbol{y} \ (\text{A为矩阵}) \qquad \boldsymbol{A} = \boldsymbol{x}^T \boldsymbol{y} \ (\text{A为数值})$$

- 向量内积

$$(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}^T \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

- 向量范数

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

$$= \sqrt{x^T x}$$

- 向量归一化

$$\boldsymbol{x_0} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$$

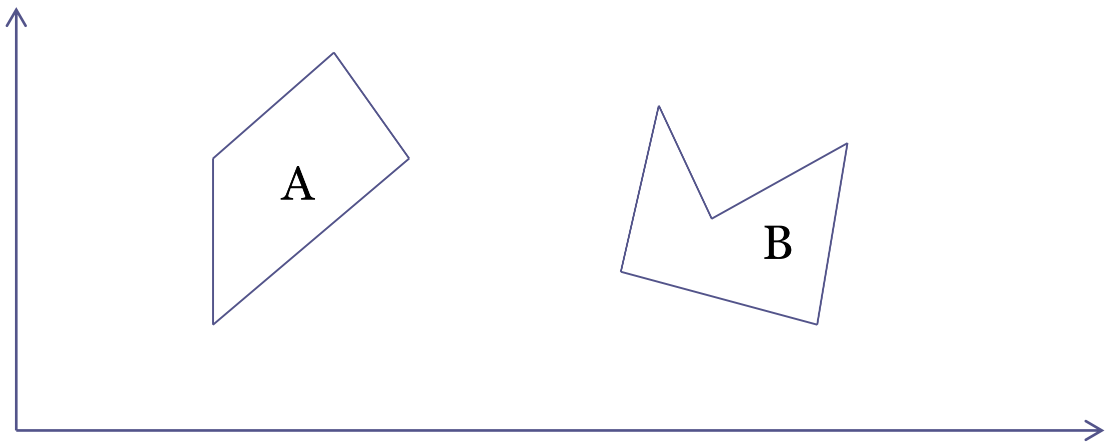
2 凸集

- 定义

$\forall \boldsymbol{x}^1, \boldsymbol{x}^2 \in \boldsymbol{R}^n, \boldsymbol{x}^1, \boldsymbol{x}^2$ 为两点连线的点，那么：

$$\boldsymbol{\lambda x}^1 + (\mathbf{1} - \boldsymbol{\lambda})\boldsymbol{x}^1 \in \boldsymbol{R}^n$$

称R为凸集



- 例： $\boldsymbol{X} = \{\boldsymbol{x} | \|\boldsymbol{x}\| \le \mathbf{5}, \boldsymbol{x} \in \boldsymbol{R}^2\}$

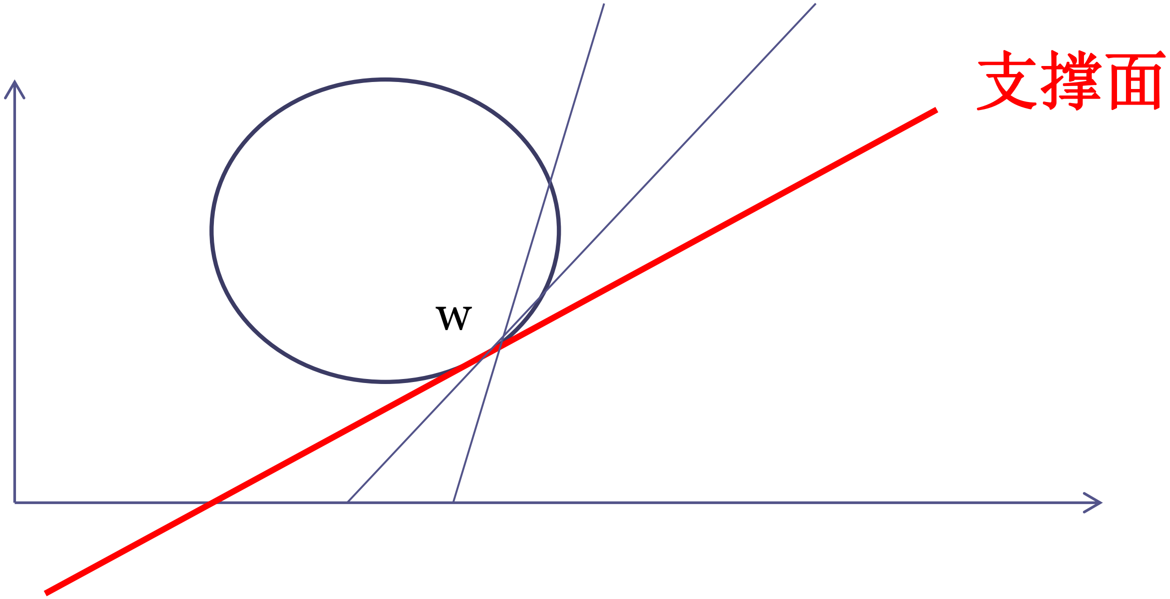
$\mathbf{Y} = \{\boldsymbol{y} | \mathbf{2} \le \|\boldsymbol{y}\| \le \mathbf{6}, \boldsymbol{y} \in \boldsymbol{R}^2\}$

• 3 超平面

- 定义

$X = \{x | c^T x = z\} x \neq 0, z$ 为常数，那么：　称X为超平面

- 超平面将空间分成两部分：$c^T x \geq z, c^T x \leq z$



- **4 函数二次型与标准型**
  - 二次函数一般表示

$$f(x) = a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + b_1x_1 + b_2x_2 + c$$

  - 二次函数矩阵表

$$f(x) = \frac{1}{2}x^TAx + B^Tx + c$$

  - 例：$f(x) = 4x_1^2 + 3x_2^2 + 6x_1x_2 + 3x_1 + x_2 + 1$

$$f(x) = \frac{1}{2}(x_1, x_2)\begin{pmatrix} 8 & 6 \\ 6 & 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (3,1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + 1$$

- **5 正定矩阵**

$x \neq 0$

$x^T A x > 0$ 称 A 为正定矩阵 记为 $A > 0$

$x^T A x \geq 0$ 称 A 为半正定矩阵 记为 $A \geq 0$

$x^T A x < 0$ 称 A 为负定矩阵 记为 $A < 0$

$x^T A x \leq 0$ 称 A 为半负定矩阵 记为 $A \leq 0$

- **5 正定矩阵**
  - 关于正定、半正定的判定定理

(1) $\boldsymbol{A}>\mathbf{0}$ 的充要条件是：A的各阶顺序主子式全大于零

$$a_{11}>0 \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}>0 \quad ,\cdots, \quad \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}>0$$

(2) $\boldsymbol{A}\geq\mathbf{0}$ 的充分条件是：A的各阶主子式全大于等于零

- **5 正定矩阵**
  - 三阶顺序主子式的值计算

$$\boldsymbol{A}=\begin{pmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{d} & \boldsymbol{e} & \boldsymbol{f} \\ \boldsymbol{g} & \boldsymbol{h} & \boldsymbol{i} \end{pmatrix}$$

$$|\boldsymbol{A}| = \boldsymbol{a} * (\boldsymbol{ei} - \boldsymbol{hf}) + \boldsymbol{b} * (\boldsymbol{fg} - \boldsymbol{di}) + \boldsymbol{c} * (\boldsymbol{dh} - \boldsymbol{eg})$$

5 正定矩阵

- 例：判定矩阵$A=\begin{pmatrix}1 & 2 & 2\\ 2 & 5 & 0\\ 2 & 0 & 20\end{pmatrix}$的正定性。

解： A 为对称阵。

A 的一阶主子式为1>0;

二阶主子式为5-4=1>0;

三阶主子式为:

$1\times(5\times 20-0)+2\times(0\times 2-2\times 20)+2\times(2\times 0-2\times 5)$

$=100+(-80)+(-20)$

$=0$

故A为半正定矩阵。

- **6 梯度**

$f(x)$在x 点处的梯度

$$\nabla f(x) = g(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n} \right]^T$$

常见的公式

$$\nabla c = 0$$

$$\nabla (b^T x) = b$$

$$\nabla (x^T A x) = 2Ax \quad \text{其中} \quad A^T = A$$

$$\nabla (x^T x) = 2x$$

- **6 梯度**

例1 $z = 2x_1 + 3x_2 + 6x_3$

例2 $z = 2x_1^{\ 2} + 3x_2^{\ 2} + 4x_1 x_2$

- 7 Hesse阵

$$\nabla^2 f(\boldsymbol{x}) = H(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1{}^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2{}^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial^2 x_n} \end{bmatrix}$$

$$\nabla^2 c = 0$$

$$\nabla^2 (\boldsymbol{b}^T \boldsymbol{x}) = 0$$

$$\nabla^2 (\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}) = 2\boldsymbol{A} \quad （其中 \boldsymbol{A}^T = \boldsymbol{A}）$$

- **7 Hesse阵**
  - 例1： $z = 2x_1 + 3x_2 + 6x_3$

- **7 Hesse阵**
  - 例1： $z = 2x_1 + 3x_2 + 6x_3$

$$\nabla z = (2,3,6)^T$$

$$\nabla^2 z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- **7 Hesse阵**
  - 例2： $z = 2x_1^2 + 3x_2^2 + 4x_1x_2$

- **7 Hesse阵**
  - 例2： $z = 2x_1^2 + 3x_2^2 + 4x_1x_2$

$$\nabla z = (4x_1 + 4x_2, 6x_2 + 4x_1)^T$$

$$\nabla^2 z = \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix}$$

常量矩阵

- 7 Hesse阵
  - 例3：$z = 2x_1^3 + 3x_2^2 + 4x_1x_2$

- 7 Hesse阵
  - 例3： $z = 2x_1^3 + 3x_2^2 + 4x_1x_2$

$$\nabla z = (6x_1^2 + 4x_2, 6x_2 + 4x_1)^T$$

$$\nabla^2 z = \begin{pmatrix} 12x_1 & 4 \\ 4 & 6 \end{pmatrix}$$

函数矩阵

- **7 Hesse阵**

已知 $f(\boldsymbol{u})$ 二阶可导，$\boldsymbol{u}, \boldsymbol{x} \in S \subset R^n$ ，$\boldsymbol{p} \in R^n, t \in R^1$

$\varphi(t) = f(\boldsymbol{x} + t\boldsymbol{p})$，求 $\varphi'(t), \varphi''(t)$

$$\varphi'(t) = \nabla f(\boldsymbol{x} + t\boldsymbol{p})^T \boldsymbol{p}$$

$$\varphi''(t) = \boldsymbol{p}^T \nabla^2 f(\boldsymbol{x} + t\boldsymbol{p}) \boldsymbol{p}$$

- **7 Hesse阵**

$$\varphi'(t) = \nabla f(\boldsymbol{x} + t\boldsymbol{p})^T \boldsymbol{p}$$
$$\varphi''(t) = \boldsymbol{p}^T \nabla^2 f(\boldsymbol{x} + t\boldsymbol{p}) \boldsymbol{p}$$

解 设 $\boldsymbol{u} = (u_1, u_2, \cdots, u_n)^T$， $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$，

$$\boldsymbol{p} = (p_1, p_2, \cdots, p_n)^T$$

$$\varphi(t) = f(\boldsymbol{x} + t\boldsymbol{p}) = f(x_1 + tp_1, x_2 + tp_2, \cdots, x_n + tp_n)$$

$$\varphi'(t) = \frac{\partial f(\boldsymbol{x} + t\boldsymbol{p})}{\partial u_1} \frac{d(x_1 + tp_1)}{dt} + \cdots + \frac{\partial f(\boldsymbol{x} + t\boldsymbol{p})}{\partial u_n} \frac{d(x_n + tp_n)}{dt}$$
$$= \frac{\partial f(\boldsymbol{x} + t\boldsymbol{p})}{\partial u_1} p_1 + \frac{\partial f(\boldsymbol{x} + t\boldsymbol{p})}{\partial u_2} p_2 + \cdots + \frac{\partial f(\boldsymbol{x} + t\boldsymbol{p})}{\partial u_n} p_n$$

$$= \left[\frac{\partial f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_1}, \frac{\partial f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_2}, \cdots, \frac{\partial f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_n} \right] \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}$$

$$= \nabla f(\boldsymbol{x}+t\boldsymbol{p})^{\mathrm{T}} \boldsymbol{p}$$

$$\varphi''(t) = \frac{d}{dt}(\varphi'(t)) = \frac{d}{dt}\sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_i} p_i$$

$$= \sum_{i=1}^{n} \frac{d}{dt}\left[\frac{\partial f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_i} p_i \right] = \sum_{i=1}^{n} \frac{d}{dt}\left[\frac{\partial f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_i} \right] p_i$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial^2 f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_i \partial u_j} p_j \right) p_i = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(\boldsymbol{x}+t\boldsymbol{p})}{\partial u_i \partial u_j} p_i p_j$$

$$= \boldsymbol{p}^T \nabla^2 f(\boldsymbol{x}+t\boldsymbol{p}) \boldsymbol{p}$$

8 泰勒公式

定理1（一阶泰勒公式）设n元函数$f(x)(x\in R^n)$在$U(x^0)$内连续可微，则 $\forall x\in U(x^0)$ 有

$$f(x)=f(x^0)+\nabla f(x^0)^T(x-x^0)+o(\|x-x^0\|)$$

或 $$f(x)=f(x^0)+\nabla f(\xi)^T(x-x^0)$$

其中 $\xi=x^0+\theta(x-x^0).\quad (0<\theta<1)$

定理2（二阶泰勒公式）设 n 元函数$f(x)(x\in R^n)$ 在 $U(x^0)$ 有内二阶连续可微，则 $\forall x\in U(x^0)$ 有

$$f(x)=f(x^0)+\nabla f(x^0)^T(x-x^0)+\frac{1}{2}(x-x^0)^T\nabla^2 f(x^0)(x-x^0)+o(\|x-x^0\|^2)$$

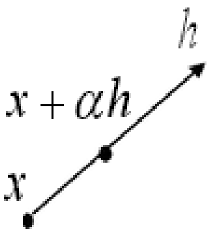
或

$$f(x)=f(x^0)+\nabla f(x^0)^T(x-x^0)+\frac{1}{2}(x-x^0)^T\nabla^2 f(\xi)(x-x^0)$$

• 9 方向导数与最速下降方向

设有单位向量 $\boldsymbol{h}=(h_1,h_2,\cdots,h_n)^T\in R^n$ 可微函数 $f(\boldsymbol{x})$

在 $\boldsymbol{x}$ 点沿 $\boldsymbol{h}$ 方向的方向导数定义为



$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{h}}=\lim_{\alpha\to 0^+}\frac{f(\boldsymbol{x}+\alpha h)-f(\boldsymbol{x})}{\alpha}$$

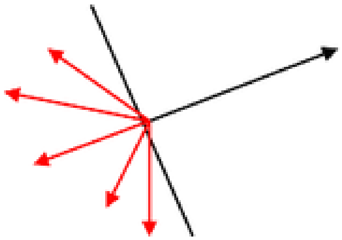
$$=\lim_{\alpha\to 0^+}\frac{\nabla f(\boldsymbol{x})^T(\alpha\boldsymbol{h})+o(\|\alpha\boldsymbol{h}\|)}{\alpha}$$

$$=\nabla f(\boldsymbol{x})^T\boldsymbol{h}+\lim_{\alpha\to 0^+}\frac{o(\|\alpha\boldsymbol{h}\|)}{\alpha}$$

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

$$=\|\nabla f(\boldsymbol{x})\|\cos(\nabla f(\boldsymbol{x}),\boldsymbol{h})$$

- **9 方向导数与最速下降方向**
  - 对于方向导数有以下结论



（1）若 $\dfrac{\partial f(x)}{\partial h} > 0$ ，则 h 为 $f(x)$ 在点的上升方向

（2）若 $\dfrac{\partial f(x)}{\partial h} < 0$ ，则 h 为 $f(x)$ 在点的下降方向

（3）若 $\nabla f(x) = 0$ 　则对任何方向 h ，有 $\dfrac{\partial f(x)}{\partial h} = 0$

（4）若 $\nabla f(x) \neq 0$ ，则

当 $h = \dfrac{\nabla f(x)}{\|\nabla f(x)\|}$ 时，$\dfrac{\partial f(x)}{\partial h}$ 取得最大值，此时 $\dfrac{\partial f(x)}{\partial h} = \|\nabla f(x)\|$

当 $h = -\dfrac{\nabla f(x)}{\|\nabla f(x)\|}$ 时，$\dfrac{\partial f(x)}{\partial h}$ 取得最小值，此时 $\dfrac{\partial f(x)}{\partial h} = -\|\nabla f(x)\|$

9 方向导数与最速下降方向

- 对于方向导数有以下结论

由此可知:

（1）$\nabla f(x)$方向为$f(x)$在点 x 处函数值增加最快的方向，

称为函数$f(x)$ 在点 x 处的最速上升方向;

（2）$-\nabla f(x)$方向为 $f(x)$在点x 处函数值减少最快的方向，

称为函数$f(x)$ 在点 x 处的最速下降方向;

- **9 方向导数与最速下降方向**
  - 例：求 $f(x)=x^2+2y^2$ 在点 $P(x,y)=(1,1)$沿方向 $\boldsymbol{d}=[1,3]^T$的方向导数。

- **10 等高线**

定义 设有二元函数 $z=f(x,y)$，若令$f(x,y)=c$

它代表函数值为C的点连成的曲线，故将曲线$f(x,y)=c$

称为二元函数$z=f(x,y)$的等高线或等值线。

性质:

（1）二次函数在极值点附近的等高线是准确的同心椭圆族，极值点正好是椭圆的共同中心。

非二次函数在极值点附近的等高线是近似的同心椭圆族，极值点正好是椭圆的共同中心。

因此求函数的极值，从几何上来讲，就是求等高线族中同心椭圆组的共同中心。

（2）函数在某点的梯度方向与过该点等高线在该点的切线垂直。