第2章 数学基础

2.1 向量与矩阵

1.向量 由线性代数知，n个有序的数 $x_1, x_2, \cdots, x_n$

组成的数组称n 維向量。

n 維向量写成一列时称列向量，记作 X

写成一行时称行向量，记作 X^T

即

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad X^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

2.矩阵

$n \times m$ 个有序的数 $a_{ij}\ (i=1,2,\cdots m\ ;\ j=1,2,\cdots n)$，

排成的 m 行 n 列数表，称 $m \times n$ 阶矩阵。

用大写字母 A 或 $A_{m\times n}$ 表示。

即

$$A = A_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

3.矩阵的运算

向量和矩阵之间可以进行各种运算，如

加减法运算；

数乘运算外；

一般的乘法运算。

4.矩阵的乘法运算

令

$$X=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix},\quad C=\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix},\quad A=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

则有

- 向量与向量相乘

$$C^TX=[c_1,c_2,...,c_n]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=c_1x_1+c_2x_2+\cdots+c_nx_n \quad （数）$$

$$CX^T=\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}[x_1,x_2,...,x_n]=\begin{bmatrix}c_1x_1&c_1x_2&\cdots&c_1x_n\\c_2x_1&c_2x_2&\cdots&c_2x_n\\\vdots&\vdots&\ddots&\vdots\\c_nx_1&c_nx_2&\cdots&c_nx_n\end{bmatrix} \quad (\text{矩阵})$$

●向量与向量相乘

$$X^TA = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$= [x_1a_{11} + x_2a_{21} + x_3a_{31}, \quad x_1a_{12} + x_2a_{22} + x_3a_{32},$$

$$x_1a_{13} + x_2a_{23} + x_3a_{33}] \qquad \text{(行向量)}$$

●向量与矩阵相乘

$$AX = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} a_{11}c_1 + a_{12}c_2 + a_{13}c_3 \\ a_{21}c_1 + a_{22}c_2 + a_{23}c_3 \\ a_{31}c_1 + a_{32}c_2 + a_{33}c_3 \end{bmatrix} \qquad \text{(列向量)}$$

$$X^T AX = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= (x_1 a_{11} + x_2 a_{21} + x_3 a_{31}) y_1 + (x_1 a_{12} + x_2 a_{22} + x_3 a_{32}) y_2$$

$$+ (x_1 a_{13} + x_2 a_{23} + x_3 a_{33}) y_3 \qquad \text{（数）}$$

即

$$\begin{bmatrix} 行 & 向 & 量 \end{bmatrix} \begin{bmatrix} 矩 & & 阵 \\ & & \end{bmatrix} = \begin{bmatrix} 行 & 向 & 量 \end{bmatrix}$$

$$\begin{bmatrix} & & \\ 矩 & & 阵 \\ & & \end{bmatrix} \begin{bmatrix} 列 \\ 向 \\ 量 \end{bmatrix} = \begin{bmatrix} 列 \\ 向 \\ 量 \end{bmatrix}$$

$$\begin{bmatrix} 行 & 向 & 量 \end{bmatrix} \begin{bmatrix} & & \\ 矩 & & 阵 \\ & & \end{bmatrix} \begin{bmatrix} 列 \\ 向 \\ 量 \end{bmatrix} = 数$$

2.2 方向导数与梯度

(1) 导数是函数在某点的变化率的数学描述

$$\left.\frac{df}{dx}\right|_{x=x_k}$$

一元函数$f(x)$在点x_k的一阶导数

$$\frac{\partial f(X^k)}{\partial x_1}, \frac{\partial f(X^k)}{\partial x_2}, \cdots$$

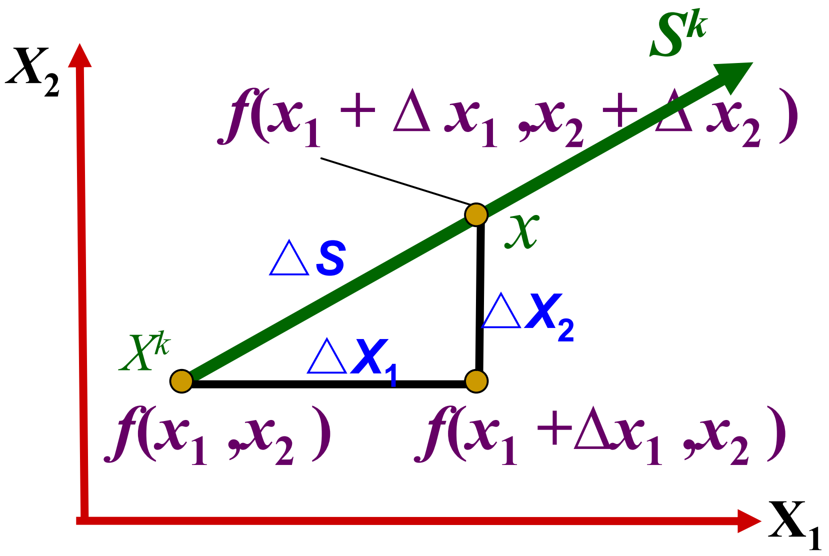
多元函数$f(X)$在点 X^k 的一阶偏导数

$$\frac{\partial f(X^k)}{\partial S}$$

多元函数$f(X)$在点 X^k 沿任意方向S的一阶偏导数，称方向导数

对于二元函数，如图所示

方向导数可根据定义写作



$$\frac{\partial f(X^k)}{\partial S}=\lim_{\Delta S\to 0}\frac{f(X^k+\Delta S)-f(X^k)}{\Delta S}$$

$$=\lim_{\Delta x_1\to 0}\frac{f(x_1+\Delta x_1,x_2)-f(x_1,x_2)}{\Delta x_1}\cdot\frac{\Delta x_1}{\Delta S}$$

$$+\lim_{\Delta x_2\to 0}\frac{f(x_1+\Delta x_1,x_2+\Delta x_2)-f(x_1+\Delta x_1,x_2)}{\Delta x_2}\cdot\frac{\Delta x_2}{\Delta S}$$

$$=\frac{\partial f(X^k)}{\partial x_1}\cos\alpha_1+\frac{\partial f(X^k)}{\partial x_2}\cos\alpha_2$$

$$=\left[\frac{\partial f(X^k)}{\partial x_1},\quad \frac{\partial f(X^k)}{\partial x_2}\right]\begin{bmatrix}\cos\alpha_1\\ \cos\alpha_2\end{bmatrix}$$

同理，对于一般 n 元函数有

$$\frac{\partial f\left(X^k\right)}{\partial S}=\frac{\partial f\left(X^k\right)}{\partial x_1}\cos\alpha_1+\frac{\partial f\left(X^k\right)}{\partial x_2}\cos\alpha_2+\ldots+\frac{\partial f\left(X^k\right)}{\partial x_n}\cos\alpha_n$$

$$=\left[\frac{\partial f\left(X^k\right)}{\partial x_1},\frac{\partial f\left(X^k\right)}{\partial x_2},\ldots,\frac{\partial f\left(X^k\right)}{\partial x_n}\right]\begin{bmatrix}\cos\alpha_1\\ \cos\alpha_2\\ \cdots\cdots\\ \cos\alpha_n\end{bmatrix}$$

$$=\left[\nabla f\left(X^k\right)\right]^T S^0$$

式中 $\nabla f\left(X^k\right)$ 称函数 $f\left(X\right)$ 在点 X^k 的梯度。

(2) 梯度是一个向量

$$\nabla f\left(X^k\right)=\left[\frac{\partial f\left(X^k\right)}{\partial x_1},\frac{\partial f\left(X^k\right)}{\partial x_2},\ldots,\frac{\partial f\left(X^k\right)}{\partial x_n}\right]^T$$

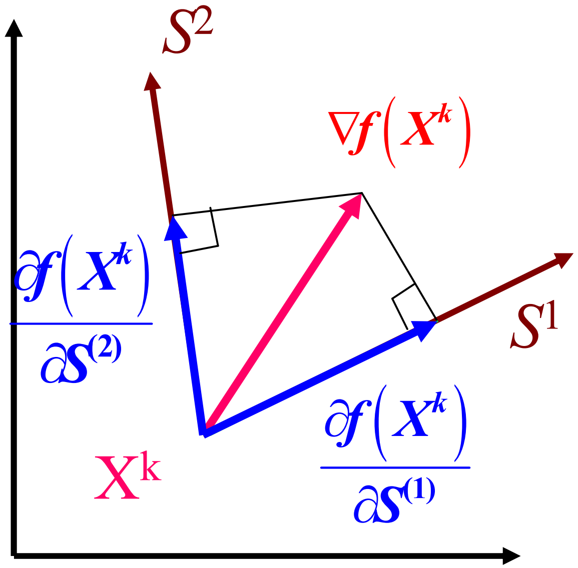
方向导数与梯度的关系

$$\frac{\partial f\left(X^k\right)}{\partial S}=\left[\nabla f\left(X^k\right)\right]^T\cdot S^0$$

$$=\|\nabla f\left(X^k\right)\|\cdot\|S^0\|\cdot\cos\left\langle\nabla f\left(X^k\right),S\right\rangle$$

$$=\|\nabla f\left(X^k\right)\|\cdot\cos\left\langle\nabla f\left(X^k\right),S\right\rangle$$

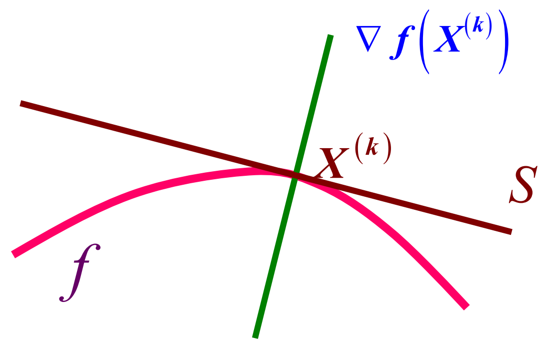
此式表明，函数在某点沿方向**S**的方向导数等于该点的梯度在方向**S**上的投影。



(3) 梯度的方向

1) 当方向S与梯度垂直时

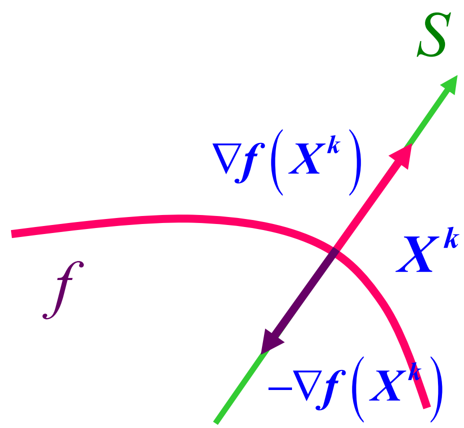
$$\frac{\partial f\left(X^k\right)}{\partial S}=\left[\nabla f\left(X^k\right)\right]^T\cdot S=0$$



说明函数的梯度方向是在一点等值线(面)的法线方向

2）当方向S与梯度的夹角为零时

$$\frac{\partial f\left(X^k\right)}{\partial S}=\left[\nabla f\left(X^k\right)\right]^T\cdot S=\|\nabla f\left(X^k\right)\|$$

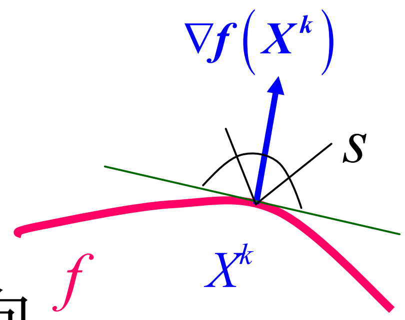


说明方向导数达到最大值，故梯度方向是函数在一点上方向导数最大的方向 。

与梯度的相反的方向称为负梯度方向，记作 $-\nabla f\left(X^k\right)$ ，

它是函数在一点上函数值下降最快的方向。

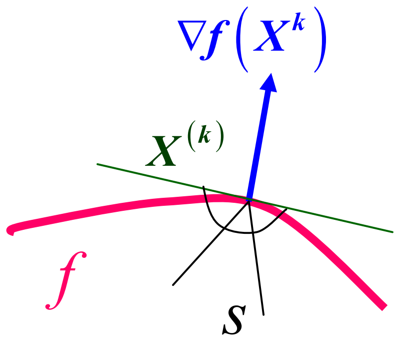
3）当方向S与梯度方向的夹角为锐角时



$$\frac{\partial f\left(X^k\right)}{\partial S}=\left[\nabla f\left(X^k\right)\right]^T\cdot S>0$$

说明与梯度成锐角的方向是函数值上升的方向。

4）当方向S与梯度方向的夹角为钝角时



$$\frac{\partial f\left(X^k\right)}{\partial S}=\left[\nabla f\left(X^k\right)\right]^T\cdot S<0$$

说明与梯度成钝角的方向是函数值下降的方向

可见函数的梯度具有以下性质：

（1）梯度是一个向量。

（2）梯度方向是等值线(面)的外法线方向。

（3）梯度是函数在一点邻域内局部性态的描述。

2.3 多元函数的泰勒展开

一元函数f（x）若在点的邻域内 n 阶可导，则函数可在该点的邻域内可作如下展开：

$$f(x)=f(x_k)+f'(x_k)\cdot(x-x_k)+f''(x_k)\cdot(x-x_k)^2+\cdots+R_n$$

多元函数 $f(X)$在某点 X^k 也可泰勒展开，展开式一般取三项

$$f(X)\approx f\left(X^k\right)+\left[\nabla f\left(X^k\right)\right]^T\left[X-X^k\right]$$
$$+\frac{1}{2}\left[X-X^k\right]^T\nabla^2 f\left(X^k\right)\left[X-X^k\right]$$

其中 $\nabla^2 f\left(X^k\right)$ 称二阶导数矩阵。

二阶导数矩阵的组成

$$\nabla^2 f\left(X^k\right)=\begin{bmatrix} \dfrac{\partial^2 f\left(X^k\right)}{\partial x_1{}^2} & \dfrac{\partial^2 f\left(X^k\right)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f\left(X^k\right)}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f\left(X^k\right)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f\left(X^k\right)}{\partial x_2{}^2} & \cdots & \dfrac{\partial^2 f\left(X^k\right)}{\partial x_2 x_n} \\ \dfrac{\partial^2 f\left(X^k\right)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f\left(X^k\right)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f\left(X^k\right)}{\partial x_n{}^2} \end{bmatrix}$$

可以看出，函数的二阶导数矩阵是一个$n\times n$ 阶对称矩阵。

矩阵的正定性

矩阵有正定、负定和不定之分：

对于任意非零向量***X***

（1）若有

$X^T HX > 0$ 则称矩阵***H***是正定矩阵；

（2）若有

$X^T HX < 0$ 则称矩阵***H***是负定矩阵；

（3）若有时 $X^T HX > 0$ 有时 $X^T HX < 0$

则称矩阵***H***是不定矩阵。

矩阵正定性的判定（主子式法）

(1) 如果矩阵***H***的各阶主子式的值均大于零，即

一阶主子式 $\left|h_{11}\right| > 0$

二阶主子式 $\begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} > 0$

三阶主子式 $\begin{vmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{vmatrix} > 0$

…

则矩阵***H***是正定的；

(2) 如果矩阵***H***的各阶主子式的值负正相间，即

奇数阶主子式小于零，偶数阶主子式大于零时，

矩阵***H***负定；

(3) 其他情况下***H***不定。

例2-2 用泰勒展开的方法将函数 $f(X)=x_1^5+x_2^4$ 在点 $X^1=[1，1]^T$ 简化成线性函数和二次函数。

解：分别求函数在点X_1的函数值、梯度和二阶导数矩阵为

$$f\left(X^1\right)=2$$

$$\nabla f\left(X^1\right)=\begin{bmatrix}5x_1^4\\4x_2^3\end{bmatrix}_{\begin{bmatrix}1\\1\end{bmatrix}}=\begin{bmatrix}5\\4\end{bmatrix}$$

$$\nabla^2 f\left(X^1\right)=\begin{bmatrix}20x_1^3 & 0\\0 & 12x_2^2\end{bmatrix}_{\begin{bmatrix}1\\1\end{bmatrix}}=\begin{bmatrix}20 & 0\\0 & 12\end{bmatrix}$$

$$X-X^1=\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}x_1-1\\x_2-1\end{bmatrix}$$

代入泰勒展开式得简化的线性函数

$$f(X) \approx f(X^1) + \left[\nabla f(X^1)\right]^T \left[X - X^1\right]$$

$$= 2 + \begin{bmatrix} 5 & 4 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$

$$= 5x_1 + 4x_2 - 7$$

和展开式的二次项 $\frac{1}{2}\left[X - X^1\right]^T \nabla^2 f(X^1)\left[X - X^1\right]$

$$= \frac{1}{2}\left[x_1 - 1,\ x_2 - 1\right] \begin{bmatrix} 20 & 0 \\ 0 & 12 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$

$$= 10(x_1 - 1)^2 + 6(x_2 - 1)^2$$

将上面的结果相加得简化后的二次函数

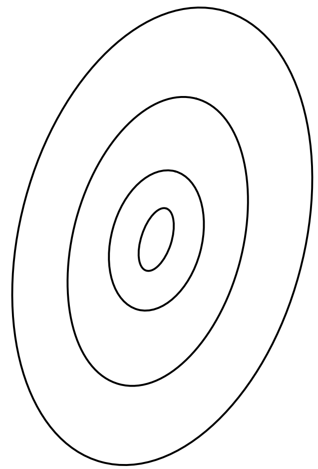
$$f(X) \approx 10(x_1 - 1)^2 + 6(x_2 - 1)^2 + 5x_1 + 4x_2 - 7$$

$$= 10x_1^{\ 2} + 6x_2^2 - 15x_1 - 8x_2 + 9$$

2.4 正定二次函数

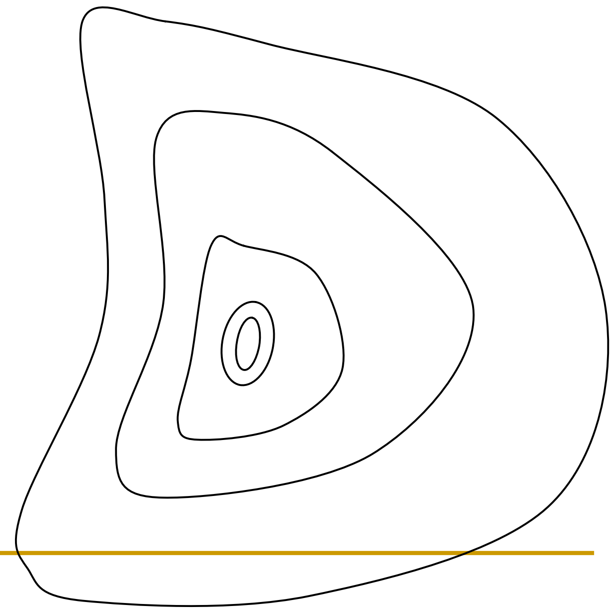
二次函数可以写成以下向量形式：

$$f(X)=\frac{1}{2}X^T HX+B^T X+c$$



如果H 是正定的，则函数 f（X）称正定二次函数。

正定二次函数具有以下性质：



（1）正定二次函数的等值线（面）是一族 同心椭圆（球），其中心就是该二次函数的极小点。

（2）非正定二次函数在极小点附近的等值线（面）近似于椭圆（球）。

2.5 多元函数的极值条件

2.5.1 无约束问题的极值条件

一元函数f（x）在点 x_k取得极值的

必要条件 $f'(x_k)=0$

充分条件 $f''(x_k)\neq 0$

当 $f''(x_k)>0$ 时，取极小值。

当 $f''(x_k)<0$ 时，取极大值。

多元函数f(X)在点 X^k 取得极值的

必要条件 $\nabla f\left(X^k\right)=0$

把函数在点 X^k 展开成泰勒二次式，将必要条件代入，有

$$f(X)-f(X^k)=\frac{1}{2}\left[X-X^k\right]^T\nabla^2 f(X^k)\left[X-X^k\right]$$

当 X^k 为函数的极小点时，有

$$\left[X-X^k\right]^T\nabla^2 f(X^k)\left[X-X^k\right]>0$$

这说明 $\nabla^2 f(X^k)$ 是正定的

同理，当 X^k 为函数的极大点时，有

$$\left[X-X^k\right]^T\nabla^2 f(X^k)\left[X-X^k\right]<0$$

这说明 $\nabla^2 f(X^k)$ 是负定的

可见，多元函数极值的充分条件是：

当$\nabla^2 f\left(X^k\right)$ 正定时$\rightarrow$在X^k取极小值

当$\nabla^2 f\left(X^k\right)$ 负定时$\rightarrow$在X^k取极大值

当$\nabla^2 f\left(X^k\right)$ 不定时$\rightarrow$在X^k无极值

例**2-3** 求函数 $f(X)=x_1^3+x_2^3+3x_1^2+2x_2^2-9x_1-4x_2$ 的极值。

解：由极值的必要条件

$$\nabla f(X)=\begin{bmatrix}3x_1^2+6x_1-9\\3x_2^2+4x_2-4\end{bmatrix}=0$$

解得以下四个驻点：

$$X^1=\begin{bmatrix}1\\2/3\end{bmatrix},\quad X^2=\begin{bmatrix}1\\-2\end{bmatrix}$$

$$X^3=\begin{bmatrix}-3\\2/3\end{bmatrix},\quad X^4=\begin{bmatrix}-3\\-2\end{bmatrix}$$

由极值的充分条件得

$$\nabla^2 f(X^1)=\begin{bmatrix}6x_1+6 & 0\\ 0 & 6x_2+4\end{bmatrix}_{\begin{bmatrix}1\\2/3\end{bmatrix}}=\begin{bmatrix}12 & 0\\ 0 & 8\end{bmatrix}$$ 矩阵正定

$$\nabla^2 f(X^2)=\begin{bmatrix}6x_1+6 & 0\\ 0 & 6x_2+4\end{bmatrix}_{\begin{bmatrix}1\\-2\end{bmatrix}}=\begin{bmatrix}12 & 0\\ 0 & -8\end{bmatrix}$$ 矩阵不定

$$\nabla^2 f(X^3)=\begin{bmatrix}6x_1+6 & 0\\ 0 & 6x_2+4\end{bmatrix}_{\begin{bmatrix}-3\\2/3\end{bmatrix}}=\begin{bmatrix}-12 & 0\\ 0 & 8\end{bmatrix}$$ 矩阵不定

$$\nabla^2 f(X^4)=\begin{bmatrix}6x_1+6 & 0\\ 0 & 6x_2+4\end{bmatrix}_{\begin{bmatrix}-3\\-2\end{bmatrix}}=\begin{bmatrix}-12 & 0\\ 0 & -8\end{bmatrix}$$ 矩阵负定

由此知

X^1 是函数的极小值点， X^4是函数的极大值点，

X^2 和 X^3 均为非极值点。

2.5.2 约束问题的极值条件

（1）等式约束问题的极值条件

由高等数学可知，对于等式约束优化问题

$$\min \ f(X)$$
$$s.t. \ h_v(X)=0 \ (v=1,2,\cdots,m)$$

可以建立如下拉格朗日函数极小化问题

$$\min L(X,\lambda)=f(X)+\sum_{v=1}^{m}\lambda_v h_v(X)$$

两问题同解

令▽L（X*， λ）=0，得

$$\nabla f(X^*)+\sum_{v=1}^{m}\lambda_v \nabla h_v(X^*)=0$$

λ_v不全为零

这就是等式约束问题极值的必要条件。简称k-t条件。

k−*t*条件可以描述为：目标函数的负梯度等于起作用约束函数梯度的非零线性组合。

（2）不等式约束问题的极值条件

对于不等式约束优化问题

$$\min\ f(X)$$
$$s.t.\ g_u(X) \le 0\ \ (u=1,2,\cdots,p)$$

引入m个松驰变量 $x_{n+u} \geqslant 0$，将不等式约束变成等式约束

$$\min f(X)$$
$$s.t.\ \ g_u(X)+x_{n+u}^2=0\ \ \ (u=1,2,\cdots,p)$$

建立拉格朗日函数

$$L(X,\lambda,\overline{X})=f(X)+\sum_{u=1}^{p}\lambda_u\left[g_u(X)+x_{n+u}^2\right]$$

令 $\nabla L\left(X,\lambda,\overline{X}\right)=0$ ，有

$$\nabla f\left(X^{*}\right)+\sum_{i\in I_k}\lambda_i\nabla g_i\left(X^{*}\right)=0$$

$$\lambda_i\geq 0\ \left(i\in I_k\right)$$

这就是不等式约束问题的极值条件，称 **k-t**条件。

其中，I_k代表点 X^k 的起作用约束的下标集合，即

$$I_k=\{u\,|\,g_u(X)=0,\quad u=1,2,\ldots,p\}$$

起作用约束：即 敏感性约束。

若有 $g_u(X^k)=0$，则称 $g_u(X)\leq 0$ 为点 X^k 的起作用约束。

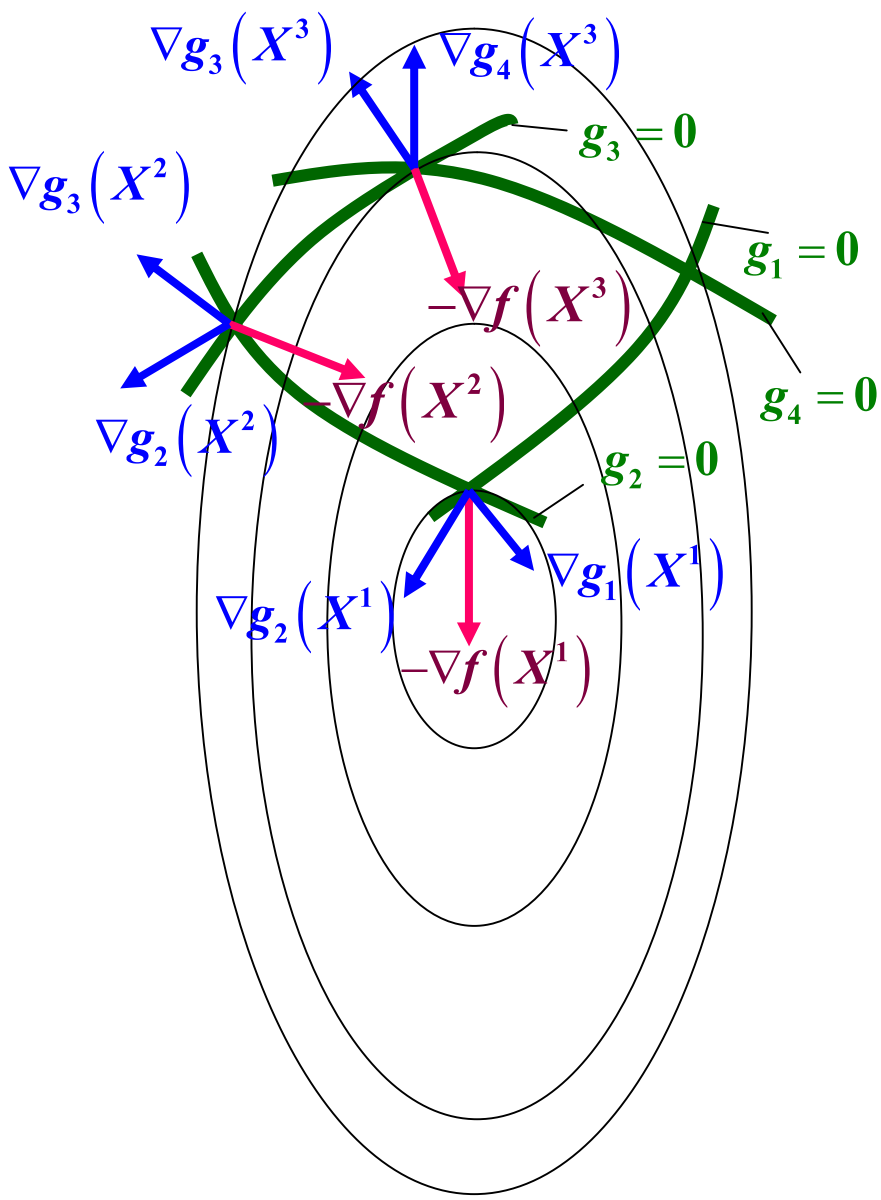
或者说，若某约束边界通过点 X^k ，则称对应的约束为该点的起作用约束。

k-t条件的意义：

目标函数的负梯度等于起作用约束梯度的非负线性组合

k-t条件的几何意义：

目标函数的负梯度位于起作用约束梯度所成夹角或锥体之内。



例2-3 用 ***k-t*** 条件判断：点 X^k=[2，0]^T 是否为下面问题的极小点。

$$\min f(X)=(x_1-3)^2+x_2^2$$
$$s.t.\ g_1(X)=x_1^2+x_2-4\le 0$$
$$g_2(X)=-x_2\le 0$$
$$g_3(X)=-x_1\le 0$$

解：因 $g_1(X^k)=2^2+0-4=0$

$$g_2(X^k)=0$$
$$g_3(X^k)=-2$$

点 X^k 的起作用约束是

和 $g_1(X)\le 0$
$g_2(X)\le 0$

在点 X^k $\nabla f(X^k)=\begin{bmatrix}2(x_1-3)\\2x_2\end{bmatrix}=\begin{bmatrix}-2\\0\end{bmatrix}$

$$\nabla g_1(X^k)=\begin{bmatrix}2x_1\\1\end{bmatrix}=\begin{bmatrix}4\\1\end{bmatrix}$$

$$\nabla g_2(X^k)=\begin{bmatrix}0\\-1\end{bmatrix}$$

将以上梯度值代入 ***k-t*** 条件

$$-\nabla f\left(X^k\right)=\lambda_1\nabla g_1\left(X^k\right)+\lambda_2\nabla g_2\left(X^k\right)$$

$$-\begin{bmatrix}-2\\0\end{bmatrix}=\lambda_1\begin{bmatrix}4\\1\end{bmatrix}+\lambda_2\begin{bmatrix}0\\-1\end{bmatrix}$$

解得 $\lambda_1=\lambda_2=0.5$，

均大于零，满足 **k-t** 条件。

说明

$X^k=[2,\ 0]^T$

就是所求最优点，如图所示。