第四节

多元函数的Taylor公式与极值问题

4.1 多元函数的Taylor公式

4.2 无约束极值、最大值与最小值

4.3 有约束极值，Lagrange乘数法

本节中，我们把一元函数的Taylor公式、极值与最大最小值的问题推广到多元函数的情形。另外，本节中的向量都写成列向量。

4.1 多元函数的Taylor公式

首先，复习一元函数的Taylor公式：

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x)$$

其中 $R_n(x)=\dfrac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}$ （ξ在x_0与x之间)

Taylor公式，是用$(x\text{-}x_0)$的n次多项式对$f(x)$进行（近似）逼近。

对于n元函数$f(\boldsymbol{x})$，我们也可以用$(\boldsymbol{x}-\boldsymbol{x}_0)$的分量所构成的多项式来逼近.作为铺垫, 先介绍 $C^{(m)}$ 类函数的概念.

定义4.1 设 $f(\boldsymbol{x})$ 是定义在区域$\Omega\subset \mathrm{R}^n$内的$n$元函数, 若$f$在$\Omega$ 内连续, 则称f是Ω 上的$C^{(0)}$类函数, 记为 $f\in C^{(0)}(\Omega)$, 或 $f\in C(\Omega)$； 若 f 在Ω内有连续的m阶偏导数,则称 f 是 Ω 上的$C^{(m)}$ 类函数, 记为$f\in C^{(m)}(\Omega)$.

下面定理为多元函数的Taylor公式的一阶形式：

定理4.1 设n元函数 $f\in C^{(2)}(U(\boldsymbol{x}_0))$, $\boldsymbol{x}_0+\Delta\boldsymbol{x}\in U(\boldsymbol{x}_0)$, 其中, $\boldsymbol{x}_0=(x_{0,1},\cdots,x_{0,n})^{\mathrm{T}}\in \boldsymbol{R}^n$, $\Delta\boldsymbol{x}=(\Delta x_1,\cdots,\Delta x_n)^{\mathrm{T}}$. 则$\exists\,\theta\in(0,1)$, 使得

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})=f(\boldsymbol{x}_0)+\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{x}_0)}{\partial x_i}\Delta x_i+R_1, \qquad (4.1)$$

其中 $R_1=\dfrac{1}{2!}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f(\boldsymbol{x}_0+\theta\Delta\boldsymbol{x})}{\partial x_i\partial x_j}\Delta x_i\Delta x_j$ 称为Lagrange余项.

证 考虑一元函数

$$\varphi(t)=f(\boldsymbol{x}_0+t\Delta\boldsymbol{x}),\ 0\le t\le 1,$$

则

$$\varphi(0)=f(\boldsymbol{x}_0),\varphi(1)=f(\boldsymbol{x}_0+\Delta\boldsymbol{x}).$$

由于 $f\in C^{(2)}(U(\boldsymbol{x}_0))$, 从而复合函数

$$\varphi(t)=f(\boldsymbol{x}_0+t\Delta\boldsymbol{x})=f(x_{0,1}+t\Delta x_1,x_{0,2}+t\Delta x_2,\cdots,x_{0,n}+t\Delta x_n)$$

在 $t=0$ 的邻域内对 t 有连续的二阶导数. 根据一元函数的 Taylor 公式可得

$$\varphi(t)=\varphi(0)+\varphi'(0)t+\frac{1}{2!}\varphi''(\theta t)t^2,0<\theta<1, \qquad (4.2)$$

由于

$$\varphi'(t)=\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{x}_0+t\Delta\boldsymbol{x})}{\partial x_i}\Delta x_i\ ,$$

$$\varphi''(t)=\sum_{i=1}^{n}\frac{d}{dt}\left(\frac{\partial f(\boldsymbol{x}_0+t\Delta\boldsymbol{x})}{\partial x_i}\right)\Delta x_i=\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial^2 f(\boldsymbol{x}_0+t\Delta\boldsymbol{x})}{\partial x_i\partial x_j}\right)\Delta x_i\Delta x_j,$$

所以

$$\varphi'(0)=\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{x}_0)}{\partial x_i}\Delta x_i,$$

$$\varphi''(\theta t)=\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial^2 f(\boldsymbol{x}_0+\theta t\Delta\boldsymbol{x})}{\partial x_i\partial x_j}\right)\Delta x_i\Delta x_j,$$

将 $\varphi(0),\varphi'(0),\varphi''(\theta t)$,代入(4.2)式再令 $t=1$,便得到Taylor (4.1). ▌

$$\varphi(t)=\varphi(0)+\varphi'(0)t+\frac{1}{2!}\varphi''(\theta t)t^2, 0<\theta<1, \qquad (4.2)$$

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})=f(\boldsymbol{x}_0)+\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{x}_0)}{\partial x_i}\Delta x_i+R_1, \qquad (4.1)$$

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})=f(\boldsymbol{x}_0)+\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{x}_0)}{\partial x_i}\Delta x_i+R_1, \qquad (4.1)$$

上式右端第二项是 f 在 $\mathbf{x}_0$ 的梯度

$$\nabla f(\mathbf{x}_0)=\left(\frac{\partial f(\mathbf{x}_0)}{\partial x_1},\frac{\partial f(\mathbf{x}_0)}{\partial x_2},\cdots,\frac{\partial f(\mathbf{x}_0)}{\partial x_n}\right)^{\mathrm{T}}$$

与 $\Delta\mathbf{x}=(\Delta x_1,\Delta x_2,\cdots,\Delta x_n)^{\mathrm{T}}$ 的内积.又由 $\dfrac{\partial^2 f(\mathbf{x})}{\partial x_i\partial x_j}=\dfrac{\partial^2 f(\mathbf{x})}{\partial x_j\partial x_i}$ 可知余项 R_1 是关于 $\Delta\mathbf{x}$ 的 n 元二次型.所以利用矩阵乘法记号,公式(4.1)可以改写成

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})=f(\boldsymbol{x}_0)+\langle\nabla f(\boldsymbol{x}_0),\Delta\boldsymbol{x}\rangle+\frac{1}{2!}(\Delta\boldsymbol{x})^{\mathrm{T}}\boldsymbol{H}_f(\boldsymbol{x}_0+\theta\Delta\boldsymbol{x})\Delta\boldsymbol{x},$$

其中实对称矩阵 $\boldsymbol{H}_f(\mathbf{x}_0+\theta\Delta\mathbf{x})$ 称为 $f(\mathbf{x})$ 在点 $\mathbf{x}_0+\theta\Delta\mathbf{x}$ 的 Hessian(海塞)矩阵.

$$H_f(x_0+\theta\Delta x)=\begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1\partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2\partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n\partial x_1} & \frac{\partial^2 f(x)}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}_{x=x_0+\theta\Delta x}$$

可以证明(从略)在定理4.1的条件下,余项

$$R_1=\frac{1}{2!}(\Delta x)^T H_f(x_0+\vartheta\Delta x)\Delta x=\frac{1}{2!}(\Delta x)^T H_f(x_0)\Delta x+o\left(\|\Delta x\|^2\right).$$

由此就得到了 $f(x)$ 在 x_0 处带有 Peano 余项的二阶 Taylor 公式

$$f(x_0+\Delta x)=f(x_0)+\langle\nabla f(x_0),\Delta x\rangle+\frac{1}{2!}(\Delta x)^T H_f(x_0)\Delta x+o\left(\|\Delta x\|^2\right),\tag{4.3}$$

或者,也可以写成

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x}\text{-}\boldsymbol{x}_0 \rangle + \frac{1}{2!}(\Delta \boldsymbol{x})^{\mathrm{T}} \boldsymbol{H}_f(\boldsymbol{x}_0) \Delta \boldsymbol{x} + o\left(\|\boldsymbol{x}\text{-}\boldsymbol{x}_0\|^2\right), \quad (4.4)$$

其中 $\boldsymbol{H}_f(\boldsymbol{x}_0) = \left(\dfrac{\partial^2 f(\boldsymbol{x}_0)}{\partial x_i \partial x_j} \right)_{n \times n}$

为 f 在点 $\boldsymbol{x}_0$ 处的Hessian 矩阵, $o\left(\|\Delta \boldsymbol{x}\|^2\right)$是当$\Delta x \to 0$时关于 $\|\Delta \boldsymbol{x}\|^2$的高阶无穷小.

特别对于二元函数$f(x,y)$, 公式(4.4)可以写成

$$f(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) +$$

$$\frac{1}{2!}(x-x_0, y-y_0)\begin{pmatrix} f_{xx}(x_0,y_0) & f_{xy}(x_0,y_0) \\ f_{yx}(x_0,y_0) & f_{yy}(x_0,y_0) \end{pmatrix}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} + o(\rho^2), \quad (4.5)$$

其中$\rho^2 = \sqrt{(x-x_0)^2 + (y-y_0)^2}$.

例4.1 设函数 $z=z(x,y)$ 是由方程 $z^3-2xz+y=0$ 所确定的隐函数，且 $z(1,1)=1$，求 $z(x,y)$ 在点 $(1,1)$ 处带有 Peano 余项的二阶 Taylor 公式。

解 由方程 $z^3-2xz+y=0$ 两端求一阶全微分，得

$$3z^2dz-2(zdx+xdz)+dy=0,$$

故

$$dz=\frac{2z}{3z^2-2x}dx-\frac{1}{3z^2-2x}dy,$$

从而

$$z_x=\frac{2z}{3z^2-2x},\quad z_y=\frac{-1}{3z^2-2x},\qquad (4.6)$$

$z(1,1)=1$，代入可得

$$z_x|_{(1,1)}=2,\quad z_y|_{(1,1)}=-1.\qquad (4.7)$$

由(4.6)式可得

$$z_{xx}=\frac{-2\left(3z^2+2x\right)z_x+4z}{\left(3z^2-2x\right)^2},\ z_{xy}=\frac{-2\left(3z^2+2x\right)z_y}{\left(3z^2-2x\right)^2},\ z_{yy}=\frac{6zz_y}{\left(3z^2-2x\right)^2}.$$

将(4.7)式代入,可得

$$z_{xx}\big|_{(1,1,1)}=-16,\ z_{xy}\big|_{(1,1,1)}=z_{yx}\big|_{(1,1,1)}=10,\ z_{yy}\big|_{(1,1,1)}=-6.\qquad(4.8)$$

把(4.8)式代入(4.5),便得到函数$z(x,y)$在点$(1,1)$处带有Peano余项的二阶Taylor公式:

$$z(x,y)=1+2(x-1)-(y-1)+\frac{1}{2!}\left[-16(x-1)^2+20(x-1)(y-1)-6(y-1)^2\right]+o\left(\rho^2\right),$$

其中$\rho^2=(x-1)^2+(y-1)^2$.

4.2 无约束极值、最大值与最小值

1. 无约束极值

生产实践中, 我们总是希望用料最省、时间最短、效益最大、质量最好等等，这类问题往往可以归结为多元函数的极值问题. 为此，我们先把一元函数的极值的概念推广到多元函数中来，然后解决多元函数的极值问题.

定义4.2 设 $f: U(x_0) \subseteq R^n \to R$, 若 $\forall x \in U(x_0)$, 恒成立不等式 $f(x) \le f(x_0)$ $\left(f(x) \ge f(x_0)\right)$,

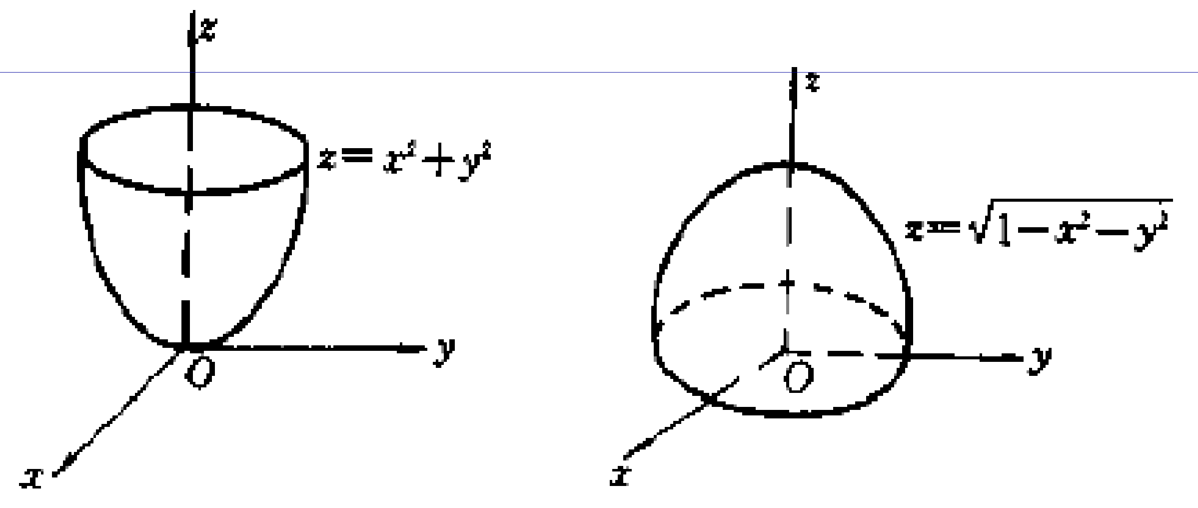
则称 f 在点 x_0 取得 **无约束极大值(无约束极小值)** $f(x_0)$,

简称为 **极大值(极小值)** $f(x_0)$. 点 x_0 称为 f 的 **极大值点(极小值点)**.

极大值与极小值统称为**极值**, 极大值点与极小值点统称为 **极值点.**

例如, 二元函数 $z=x^2+y^2$ 在点 $(0,0)$ 取得极小值 0。

$z=\sqrt{1-x^2-y^2}$ 在点 $(0,0)$ 取得极大值 1, 它们所对应的曲面在相应点分别呈现为"谷"和"峰"(图5.17)



(a) 图5.17 (b)

如果二元函数 $z=f(x,y)$ 在点 (x_0,y_0) 的偏导数存在，而且 (x_0,y_0) 为 f 的极值点，则一元函数 $f(x,y_0)$ 在 $x=x_0$ 必取得极值.由一元函数极值的必要条件,必有 $f_x(x_0,y_0)=0$,同理有 $f_y(x_0,y_0)=0$. 因此,若 f 在 (x_0,y_0) 处可微,则由梯度的定义，f 在 (x_0,y_0) 取得极值的必要条件是

$$\nabla f(x_0,y_0)=\left(f_x,f_y\right)^T\Big|_{(x_0,y_0)}=0.$$

所以有下面的定理

定理4.3 (极值的必要条件) 设 n 元函数 f 在点 x_0 可微,且 x_0 为 f 的极值点,则必有 $\nabla f(x_0)=0$.

我们称满足 $\nabla f(x_0)=0$ 的点 x_0 为 f 的**驻点.** 因此,该定理也可以说成:可微函数 f 的极值点必是驻点. 反之,不一定.

例如，二元函数 $f(x,y)=x^2-y^2$，显然点(0,0)是它的驻点，但却不是它的极值点，因为在点(0,0)的任何去心邻域内总有点 $(x,0)$ 使 $f(x,0)=x^2>0=f(0,0)$，也总有点 $(0,y)$ 使得 $f(0,y)=-y^2<0=f(0,0)$。从 f 的图像(图5.18)来看，双曲抛物面 $z=x^2-y^2$ 在原点处呈现"鞍点"，"鞍点"附近既有向上延伸的部分，也有向下延伸的部分，形如"马鞍"，因而函数 f 在该点不能取到极值。

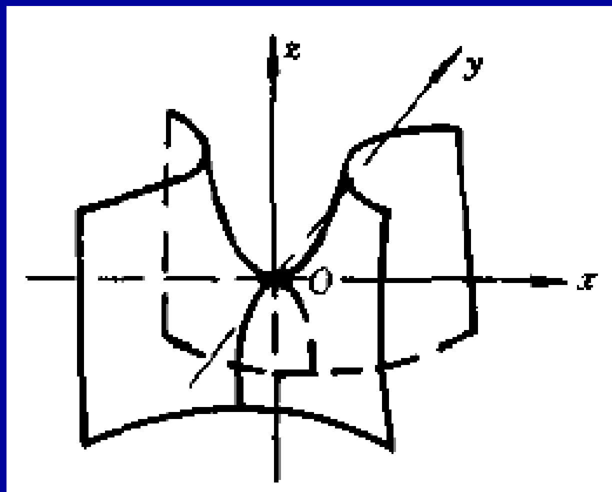


图5.18

现在我们利用Taylor 公式建立多元函数极值的充分条件.

定理4.4 (极值的充分条件) 设n元函数 $f\in C^{(2)}(U(\mathbf{x}_0))$, $\nabla f(\mathbf{x}_0)=\mathbf{0}$, $H_f(\mathbf{x}_0)$为 f 在点 $\mathbf{x}_0$ 的Hessian 矩阵. 若 $H_f(\mathbf{x}_0)$ 正定(负定), 则 $f(\mathbf{x}_0)$为 f 的极小值(极大值).

证 因为$\nabla f(\mathbf{x}_0)=\mathbf{0}$, 由Taylor 公式(4.3), 有

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})-f(\boldsymbol{x}_0)=\frac{1}{2!}(\Delta\boldsymbol{x})^{\mathrm{T}}\boldsymbol{H}_f(\boldsymbol{x}_0)\Delta\boldsymbol{x}+o\left(\|\Delta\boldsymbol{x}\|^2\right),$$

当$\boldsymbol{H}_f(\boldsymbol{x}_0)$正定时,由线性代数的知识可知, $\boldsymbol{H}_f(\boldsymbol{x}_0)$的最小特征值 $\lambda_1>0$, 而且, $\forall\Delta\mathbf{x}\neq\mathbf{0}$, 恒有

$$(\Delta\boldsymbol{x})^{\mathrm{T}}\boldsymbol{H}_f(\boldsymbol{x}_0)\Delta\boldsymbol{x}\geq\lambda_1\|\Delta\boldsymbol{x}\|^2>0,$$

于是有

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})-f(\boldsymbol{x}_0)\geq\frac{1}{2}\lambda_1\|\Delta\boldsymbol{x}\|^2+o\left(\|\Delta\boldsymbol{x}\|^2\right)=\left(\frac{1}{2}\lambda_1+o(1)\right)\|\Delta\boldsymbol{x}\|^2,$$

其中 $o(1)$ 是当 $\Delta\boldsymbol{x}\to\mathbf{0}$ 时的无穷小量，由上式可知，当 $\Delta\boldsymbol{x}\neq\mathbf{0}$ 且 $\|\Delta\boldsymbol{x}\|$ 充分小时，就有

$$f(\boldsymbol{x}_0+\Delta\boldsymbol{x})-f(\boldsymbol{x}_0)\geq 0,\text{ 或 } f(\boldsymbol{x}_0+\Delta\boldsymbol{x})\geq f(\boldsymbol{x}_0),$$

这就是说，当 $\boldsymbol{H}_f(\boldsymbol{x}_0)$ 正定时，$f(\boldsymbol{x}_0)$ 为 f 的极小值；同理可证，当 $\boldsymbol{H}_f(\boldsymbol{x}_0)$ 负定时，$f(\boldsymbol{x}_0)$ 为 f 的极大值. ▌

根据以上结论，我们可以得出求 $C^{(2)}$ 类函数 f 的极值的步骤：首先求出 f 的所有驻点，然后求出 f 在各个驻点处的Hessian 矩阵，最后判定Hessian 矩阵的类型，确定出 f 的极值点。设 $\boldsymbol{x}_0$ 是 f 的驻点，则当 $\boldsymbol{H}_f(\boldsymbol{x}_0)$ 正定时，$f(\boldsymbol{x}_0)$ 是 f 的极小值；当 $\boldsymbol{H}_f(\boldsymbol{x}_0)$ 负定时，$f(\boldsymbol{x}_0)$ 是 f 的极大值；当 $\boldsymbol{H}_f(\boldsymbol{x}_0)$ 不定时，因为二次型 $(\boldsymbol{x}-\boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{H}_f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)$ 不定，因而在 $\boldsymbol{x}_0$ 的邻域内 $[f(\boldsymbol{x})-f(\boldsymbol{x}_0)]$ 不定号，所以 $f(\boldsymbol{x}_0)$ 不是极值。

特别地，对于二元函数 $z=f(x,y)$，若点 P_0 为 f 的驻点，记

$$f_{xx}(P_0)=A，f_{xy}(P_0)=B，f_{yy}(P_0)=C，$$

则 f 在点 P_0 的 Hessian 矩阵为

$$H_f(P_0)=\begin{pmatrix} A & B \\ B & C \end{pmatrix}.$$

于是根据矩阵正定、负定和不定的判定方法，有：

(1) 若 $A>0, AC-B^2>0$，则 $\boldsymbol{H}_f(P_0)$ 正定，$f(P_0)$ 为极小值；

(2) 若 $A<0, AC-B^2>0$，则 $\boldsymbol{H}_f(P_0)$ 负定，$f(P_0)$ 为极大值；

(3) 若 $AC-B^2<0$，则 $\boldsymbol{H}_f(P_0)$ 不定，$f(P_0)$ 不是极值．

当 $AC-B^2=0$，称为临界情况．这时，只根据二阶 Taylor 公式不足以判断 P_0 点是不是 f 的极值点．

例4.2 求二元函数 $f(x,y)=x^3+y^3+3xy$ 的极值．

解 由
$$\begin{cases} f_x=3x^2+3y=0, \\ f_y=3y^2+3x=0, \end{cases}$$
求出 f 的驻点有两个：$M_1(0,0)$，$M_2(-1,-1)$．再求二阶偏导数，得
$$f_{xx}=6x,\quad f_{xy}=3,\quad f_{yy}=6y,$$
$$\boldsymbol{H}_f(M_1)=\begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix},\quad \boldsymbol{H}_f(M_2)=\begin{pmatrix} -6 & 3 \\ 3 & -6 \end{pmatrix}.$$

显然 $\boldsymbol{H}_f(M_1)$ 不定，$\boldsymbol{H}_f(M_2)$ 负定，故 M_1 不是 f 的极值点，$f(M_2)=1$ 是 f 的极大值．▌

例4.3 求函数 $f(x,y)=2x^2-3xy^2+y^4$ 的极值点．

解 由

$$\begin{cases} f_x=4x-3y^2=0, \\ f_y=2y(2y^2-3x)=0. \end{cases}$$

可求出 f 有唯一的驻点 $(0,0)$． f 在 $(0,0)$ 点的二阶偏导数为

$$f_{xx}(0,0)=4,\quad f_{xy}(0,0)=0,\quad f_{yy}(0,0)=0,$$

所以

$$\boldsymbol{H}_f(0,0)=\begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}.$$

因为 $\boldsymbol{H}_f(0,0)$ 的行列式为零，所以点 $(0,0)$ 到底是不

[illegible]

例4.3 求函数 $f(x,y)=2x^2-3xy^2+y^4$ 的极值点．

事实上，当 $(x,y)\neq(0,0)$ 时，从

$$f(x,y)-f(0,0)=2x^2-3xy^2+y^4=(2x-y^2)(x-y^2)$$

可以看出，[illegible]

$$当\ \frac{1}{2}y^2<x<y^2\ 时，f(x,y)<f(0,0)．$$

因此 $(0,0)$ 不是 f 的极值点．从而知 f 没有极值点． ▌

注意：上例的Hessian 矩阵 $H_f(0,0)=\begin{pmatrix}4&0\\0&0\end{pmatrix}$ 半正定．

由此可见，驻点处的Hessian 矩阵半正定时，驻点却未必是极值点．