第十二章　多元函数的微分

§1　方向导数和偏导数

定义 1 (方向导数)　设 $f: D \to \mathbb{R}$ 为定义在 $\mathbb{R}^n$ 中开集 D 内的映射 (多元函数). 对于 $p \in D$ 以及 $\mathbb{R}^n$ 中单位向量 u, 极限

$$\lim_{t\to 0} \frac{f(p+tu)-t(p)}{t}$$

如果存在, 则称 f 在 p 处沿方向 u 有方向导数, 极限记为 $\dfrac{\partial f}{\partial u}(p)$, 称为 f 沿 u 的方向导数.

注　(1) 方向导数就是一元函数 $\varphi(t) = f(p+tu)$ 在 $t=0$ 处的导数. 特别地, 当 $u = e_i = (0,\cdots,0,1,0,\cdots,0)$ (第 i 个位置为 1 的单位向量) 时, 又记 $\dfrac{\partial f}{\partial u}(p)$ 为 $\dfrac{\partial f}{\partial x_i}(p)$, 称为 f 的第 i 个偏导数. 按定义, 有

$$\frac{\partial f}{\partial x_i}(p) = \lim_{t\to 0} \frac{f(p_1,\cdots,p_{i-1},p_i+t,p_{i+1},\cdots,p_n) - f(p_1,\cdots,p_n)}{t}.$$

(2) 偏导数 $\dfrac{\partial f}{\partial x_i}$ 又记为 f'_{x_i}. 如果 $f'_{x_i} = \dfrac{\partial f}{\partial x_i}$ 仍然可求偏导数, 则记 $f''_{y_i x_i} = \dfrac{\partial}{\partial y_i}\left(\dfrac{\partial f}{\partial x_i}\right)$, 称为 2 阶偏导数. 类似地可以定义 $f''_{x_i y_i} = \dfrac{\partial}{\partial x_i}\left(\dfrac{\partial f}{\partial y_i}\right)$ 以及高阶偏导数. 如果 f 存在直到 k 阶的连续偏导数, 则称 f 为 C^k 函数. 我们也使用形如这样的记号:

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right),\ \frac{\partial^2 f}{\partial x_i \partial y_i} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial y_i}\right),\ \cdots$$

例 1　求 $f(x,y) = xy$ 的 1 阶, 2 阶偏导数.

解　$f'_x(x,y) = \dfrac{\partial f}{\partial x} = \dfrac{\partial}{\partial x}(xy) = y$, $f'_y = x$, $f''_{yx} = 1$, $f''_{xy} = 1$, $f''_{xx} = f''_{yy} = 0$.

例 2　求 $f(x,y) = \sqrt{x^2+y^2} + xy$ 在 $(x_0,y_0) = (1,2)$ 处的偏导数.

解　$\dfrac{\partial f}{\partial x}(x,y) = \dfrac{x}{\sqrt{x^2+y^2}} + y$, $\dfrac{\partial f}{\partial y}(x,y) = \dfrac{y}{\sqrt{x^2+y^2}} + x \Rightarrow$

$$\frac{\partial f}{\partial x}(1,2) = \frac{\sqrt{5}}{5} + 2,\ \frac{\partial f}{\partial y}(1,2) = \frac{2}{5}\sqrt{5} + 1.$$

例 3 设 $(x_0, y_0, z_0) \in \mathbb{R}^3$, 求

$$f(x,y,z) = \left[(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2\right]^{-\frac{1}{2}}$$

的偏导数.

解 记

$$r = \left[(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2\right]^{\frac{1}{2}}$$

则

$$\frac{\partial f}{\partial x} = -\frac{1}{r^2} \cdot \frac{\partial r}{\partial x} = -\frac{1}{r^2}\frac{x-x_0}{r} = -\frac{x-x_0}{r^3}.$$

同理,

$$\frac{\partial f}{\partial y} = -\frac{y-y_0}{r^3}, \qquad \frac{\partial f}{\partial z} = -\frac{z-z_0}{r^3}.$$

和一元函数不同的是, 偏导数的存在不能保证多元函数的连续性, 这是因为偏导数只反映函数沿特定方向的性质.

例 4 设

$$f(x,y) = \begin{cases} 0, & \text{当 } x \cdot y = 0, \\ 1, & \text{当 } x \cdot y \neq 0, \end{cases}$$

则

$$\frac{\partial f}{\partial x}(x,0) = \lim_{x \to 0} \frac{f(x,0) - f(0,0)}{x} = 0,$$
$$\frac{\partial f}{\partial y}(0,0) = \lim_{y \to 0} \frac{f(0,y) - f(0,0)}{y} = 0.$$

显然, f 在 $(0,0)$ 处不连续.

定理 1 (复合求导) 设 f 如定义 1, $x^0 \in D$. 假设 f'_{x_i} $(1 \leq i \leq n)$ 在 x^0 附近连续, 则

(1) f 在 x^0 处连续;

(2) 如果 $x_i = x_i(t)$ 在 t_0 处可导, $x^0 = (x_1(t_0), \cdots, x_0(t_0))$, 则 $f(x(t))$ 在 $t = t_0$ 处可导, 且

$$\left.\frac{dt}{dt}\right|_{t=t_0} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x(t_0)) \cdot \frac{dx_i}{dt}(t_0).$$

证明 (1) 利用微分中值定理, 有

$$\begin{aligned} f(x)-f(x^0) &= \sum_{i=1}^{n}\left[f(x_1^0,\cdots,x_{i-1}^0,x_i,\cdots,x_n)-f(x_1^0,\cdots,x_{i-1}^0,x_i^0,x_{i+1},\cdots,x_n)\right] \\ &= \sum_{i=1}^{n} f'_{x_i}(x_1^0,\cdots,x_{i-1}^0,\xi_i,x_{i+1},\cdots,x_n)\cdot(x_i-x_i^0) \end{aligned}$$

当 $x\to x^0$ 时, $\xi_i\to x_i^0$, 由 f'_{x_i} 在 x^0 处连续知 $\lim\limits_{x\to x^0}(f(x)-f(x^0))=0$.

(2) 由 (1) 的证明, 有

$$\begin{aligned} \left.\frac{df}{dt}\right|_{t=t_0} &= \lim_{t\to t_0}\frac{f(x(t))-f(x(t_0))}{t-t_0} \\ &= \lim_{t\to t_0}\sum_{i=1}^{n} f'_{x_i}(x_1^0,\cdots,x_{i-1}^0,\xi_i,x_{i+1},\cdots,x_n)\cdot\frac{x(t)-x_i(t_0)}{t-t_0} \\ &= \sum_{i=1}^{n} f'_{x_i}(x^0)\cdot x_i'(t_0). \end{aligned}$$

推论 在定理 1 的条件下, 如果 $u=\sum\limits_{i=1}^{n}u_i\cdot e_i=(u_1,u_2,\cdots,u_n)$, 则

$$\frac{\partial f}{\partial u}(p)=\frac{d}{dt}|_{t=0}f(p+tu)=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(p)\cdot u_i.$$

定理 2 (求导次序的可交换性) 设 $f:D\to\mathbb{R}$ 为二元函数, $(x_0,y_0)\in D$. 如果 f''_{xy} 和 f''_{yx} 在 (x_0,y_0) 处连续, 则

$$f''_{xy}(x_0,y_0)=f''_{yx}(x_0,y_0).$$

证明 对于充分小的 $k\neq 0$, $h\neq 0$ 分别考虑函数

$$\begin{aligned} \varphi(y) &= f(x_0+h,y)-f(x_0,y), \\ \psi(x) &= f(x,y_0+k)-f(x,y_0), \end{aligned}$$

由 Lagrange 中值定理, 有

$$\begin{aligned} \varphi(y_0+k)-\varphi(y_0) &= \varphi_y'(y_0+\theta_1 k)k \quad (|\theta_1|\le 1) \\ &= \left[f_y'(x_0+h,y_0+\theta_k)-f_y'(x_0,y_1+\theta_1 k)\right]k \\ &= f''_{xy}(x_0+\theta_2\cdot h,y_0+\theta_1 k)\cdot kh \quad (|\theta_2|\le 1). \end{aligned}$$

同理,

$$\psi(x_0+h)-\psi(x_0)=f''_{yx}(x_0+\theta_3 h, y_0+\theta_4 k)hk.$$

易见,

$$\varphi(y_0+k)-\varphi(y_0)=\psi(x_0+h)-\psi(x_0),$$

故

$$f''_{xy}(x_0+\theta_2 h, y_0+\theta_1 h)=f''_{yx}(x_0+\theta_3 h, y_0+\theta_4 k).$$

令 $k, h\to 0$, 由 f''_{xy}, f''_{yx} 在 (x_0, y_0) 处连续即得欲证等式.

推论 多元函数的各阶偏导数如果连续, 则其值与求导次序无关.

例 5

$$f(x,y)=\begin{cases} xy\dfrac{x^2-y^2}{x^2+y^2}, & (x,y)\neq(0,0),\\ 0, & (x,y)=(0,0).\end{cases}$$

则 $f''_{xy}(0,0)=1$, $f''_{yx}(0,0)=-1$, 这说明定理 2 中连续性假设是不可缺少的.

§2 切线和切面

设 $\sigma:[\alpha,\beta]\to\mathbb{R}^n$ 为 $\mathbb{R}^n$ 中一条连续曲线, 记 $\sigma(t)=(x_1(t),\cdots,x_n(t))$. 如果 $x_i(t)(1\leq i\leq n)$ 在 $t=t_0$ 处均可导, 则称 σ 在 t_0 处可导, 记

$$\sigma'(t_0)=\frac{d\sigma}{dt}(t_0)=\left.\frac{d\sigma}{dt}\right|_{t=t_0}=(x'_1(t_0),\cdots,x'_n(t_0))$$

称 $\sigma'(t_0)$ 为 σ 在 t_0 处的切向量.

当 $\sigma'(t_0)\neq 0$ 时, 称 $\{\sigma(t_0)+\sigma'(t_0)u|u\in\mathbb{R}\}$ 为 σ 在 t_0 处的切线, 其方程可写为

$$P-\sigma(t_0)=u\cdot\sigma'(t_0)$$

或

$$\frac{x-x_0}{x'_1(t_0)}=\frac{y-y_0}{x'_2(t_0)}=\frac{z-z_0}{x'_3(t_0)}.$$

经过 $\sigma(t_0)$ 且与切线正交的超平面称为法面, 其方程为

$$(q-\sigma(t_0))\cdot\sigma'(t_0)=0.$$

例 1 设 f 为一元可微函数, 令

$$\sigma(t)=(t,f(t))$$

则 $\sigma'(t_0)=(1,f'(t_0))$, σ 在 t_0 处切线方程为

$$\frac{x-t_0}{1}=\frac{y-f(t_0)}{f'(t_0)}$$

即

$$y=f(t_0)+f'(t_0)\cdot(x-t_0).$$

例 2 求螺旋线

$$\sigma(t)=(a\cos t,\ a\sin t,\ t),\ t\in\mathbb{R}$$

的切线和法面方程.

解 在 $t=t_0$ 处,

$$\sigma'(t_0)=(-a\sin t_0,a\cos t_0,1)$$

故切线方程为

$$\frac{x-a\cos t_0}{-a\sin t_0}=\frac{y-a\sin t_0}{a\cos t_0}=\frac{z-t_0}{1},$$

法面方程为

$$-(x-a\cos t_0)a\sin t_0+(y-a\sin t_0)a\cos t_0+(z-t_0)\cdot 1=0.$$

设 D 为 $\mathbb{R}^m$ 中开集, 我们称连续映射 $\Sigma:D\to\mathbb{R}^n$ $(n>m)$ 为 $\mathbb{R}^n$ 中的一个参数曲面. 设 $u^0=(u_1^0,\cdots,u_m^0)\in D$, 则

$$u\mapsto\Sigma(u_1^0,\cdots,u_{i-1}^0,u,u_{i+1}^0,\cdots,u_m^0)$$

为 $\mathbb{R}^m$ 中曲线, 称为 Σ 上的 u_i 曲线. 如果 u_i 曲线在 u_i^0 处可导, 则记 $\dfrac{\partial\Sigma}{\partial u_i}(u^0)=\Sigma'_{u_i}(u^0)$, 它是该曲线在 u^0 处的切向量. 如果 $\{\Sigma'_{u_i}(u^0)|1\le i\le m\}$ 线性无关 (此时称 u^0 为 Σ 的正则点), 则称由这些切向量张成的、经过 $\Sigma(u^0)$ 的子空间为切空间, 切空间的正交补称为法空间, 法空间中的元素称为法向量.

当 $m=n-1$ 时, 称 Σ 为超曲面, 此时法空间维数为 1. 特别地, 对于 $\mathbb{R}^3$ 中的 (超) 曲面 Σ,

$$\Sigma(u,v)=(x(u,v),y(u,v),z(u,v))$$

其切向量为

$$\Sigma'_u(u_0,v_0)=(x'_u(u_0,v_0),y'_u(u_0,v_0),z'_u(u_0,v_0)$$
$$\Sigma'_v(u_0,v_0)=(x'_v(u_0,v_0),y'_v(u_0,v_0),z'_v(u_0,v_0))$$

如果 $\Sigma'_u(u_0,v_0),\Sigma'_v(u_0,v_0)$ 线性无关, 则

$$\begin{aligned}\vec{n} &= \Sigma'_u(u_0,v_0)\times\Sigma'_v(u_0,v_0)\\ &= (y'_u\cdot z'_v-z'_u\cdot y'_v,\ z'_u\cdot x'_v-x'_u\cdot z'_v,\ x'_u\cdot y'_v-y'_u\cdot x'_v)\neq 0\end{aligned}$$

$\vec{n}$ 为法向量, 从而 Σ 的切平面方程为

$$(P-\Sigma(u_0,v_0))\cdot\vec{n}=0$$

或改写为

$$\begin{vmatrix} x-x(u_0,v_0) & y-y(u_0,v_0) & z-z(u_0,v_0)\\ x'_u(u_0,v_0) & y'_u(u_0,v_0) & z'_u(u_0,v_0)\\ x'_v(u_0,v_0) & z'_v(u_0,v_0) & z'_v(u_0,v_0)\end{vmatrix}=0.$$

例 3 求球面 $S^2=\{(x,y,z)\in\mathbb{R}^3\,|\,x^2+y^2+z^2=1\}$ 的切面.

解 球面可写成参数曲面

$$x=\sin\theta\cos\varphi,\ y=\sin\theta\sin\varphi,\ z=\cos\theta,\quad 0\le\theta\le\pi,\quad 0\le\varphi\le 2\pi,$$

其法向量为

$$\vec{n}=(\cos\theta\cos\varphi,\cos\theta\sin\varphi,-\sin\theta)\times(-\sin\theta\sin\varphi,\sin\theta\cos\varphi,0)=\sin\theta\cdot(x,y,z)$$

故在 (x_0,y_0,z_0) 处切平面方程为

$$(x-x_0)\cdot x_0+(y-y_0)\cdot y_0+(z-z_0)\cdot z_0=0.$$

§3 映射的微分

我们回忆一下, 对于一元函数而言, 可微是指该函数可以被线性函数一阶逼近. 对于多元函数, 我们也可以通过线性逼近来定义可微性.

设 $D \subset \mathbb{R}^n$ 为开集, 我们把映射 $f: D \to \mathbb{R}^n$ 称为多元向量值函数, 写成分量形式为

$$f(x_1, \cdots, x_n) = (f_1(x_1, \cdots, x_n), \cdots, f_m(x_1, \cdots, x_n))$$

为方便起见, 以下把欧氏空间中的向量以列向量来表示.

定义 1 (微分) 设 f 如上, $x^0 = (x_1^0, \cdots, x_n^0)^T \in D$. 如果存在 $m \times n$ 阶的矩阵 $A = (a_{ij})_{m\times n}$, 使得对于 x^0 附近的点 x, 有

$$\|f(x) - [f(x^0) + A \cdot (x - x^0)]\| = o(\|x - x^0\|),\ x \to x_0,$$

则称 f 在 x^0 处可微, 线性映射

$$\begin{aligned} df(x^0):\ & \mathbb{R}^n \to \mathbb{R}^m \\ & v \mapsto A \cdot v \end{aligned}$$

称为 f 在 x^0 处的微分.

命题 1 (可微 $\Rightarrow$ 可导) 如果 $f: D \to \mathbb{R}^m$ 在 x^0 处可微, 则其分量 f_i $(1 \le i \le n)$ 在 x^0 处存在方向导数, 并且

$$A = \left(\frac{\partial f_i}{\partial x_j}(x^0)\right)_{m\times n}.$$

证明 从微分的定义可以看出, 如果 f 在 x^0 处可微, 则 f 在 x^0 处连续. 下面以 $m = 1$ 为例说明方向导数的存在性.

为此, 取单位向量 u, 由定义, 我们有

$$\begin{aligned} f(x^0 + tu) - f(x^0) &= A(x^0 + tu - x^0) + o(\|x^0 + tu - x^0\|) \\ &= t \cdot Au + o(|t|) \end{aligned}$$

这说明 $\dfrac{\partial f}{\partial u}(x^0) = A \cdot u$, 即方向导数存在. 特别地,

$$\frac{\partial f}{\partial x_i}(x^0) = A \cdot e_i \quad \Rightarrow \quad A = \left(\frac{\partial f}{\partial x_1}(x^0), \cdots, \frac{\partial f}{\partial x^n}(x^0)\right).$$

例 1 设

$$f(x, y) = \begin{cases} \dfrac{x^2 y}{x^2 + y^2}, & (x, y) \ne (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

因为 $|f(x,y)| \le \frac{1}{2}|x|$, 故 f 在 $(0,0)$ 处连续, 且 $f'_x(0,0) = f'_y(0,0) = 0$. 如果 $u=(u_1,u_2)$ 为单位向量, 则

$$\frac{\partial f}{\partial u}(0,0) = \lim_{t\to 0}\frac{f(tu_1,tu_2)}{t} = \lim_{t\to 0}\frac{t^3u_1^2u_2}{(u_1^2+u_2^2)t^3} = \frac{u_1^2u_2}{u_1^2+u_2^2}.$$

然而 f 在 (0,0) 处不可微 (why?).

如果 f_i $(1\le i\le m)$ 的偏导数均存在, 则记 $Jf = \left(\frac{\partial f_i}{\partial x_j}\right)_{m\times n}$, 称为 f 的 Jacobian. Jf 在每一点的值构成一个映射 $Jf: D\to \mathbb{R}^{m\cdot n}$, 这里我们把 $m\times n$ 阶矩阵视为 $\mathbb{R}^{m\cdot n}$ 中的点.

定理 1 (可微的充分条件) 如果 Jf 在 D 中存在, 且它作为映射在 x^0 处连续, 则 f 在 x^0 处可微.

证明 仍以 $m=1$ 为例来证明. 由条件, f'_{x_i} 在 x^0 处连续, $i=1,2,\cdots,n$. 根据微分中值定理, 有

$$\begin{aligned} f(x)-f(x^0) &= \sum_{i=1}^n \left[f(x_1^0,\cdots,x_{i-1}^0,x_i,x_{i+1},\cdots,x_n) - f(x_1^0,\cdots,x_i^0,x_{i+1},\cdots,x_n)\right] \\ &= \sum_{i=1}^n f'_{x_i}(x_1^0,\cdots,x_{i-1}^0,x_i^0+\theta\cdot(x_i-x_i^0),x_{i+1},\cdots,x_n)\cdot(x_i-x_i^0) \\ &= \sum_{i=1}^n f'_{x_i}(x^0)\cdot(x_i-x_i^0) + \sum_{i=1}^n \alpha_i\cdot(x_i-x_i^0) \end{aligned}$$

其中

$$\alpha_i = f'_{x_i}(x_1^0,\cdots,x_{i-1}^0,x_i^0+\theta(x_i-x_i^0),x_{i+1},\cdots,x_n) - f'_{x_i}(x_1^0,\cdots,x_n^0)\to 0,\quad (x_i\to x_i^0)$$

从而

$$\begin{aligned} \left\|f(x)-\left[f(x^0)+\sum_{i=1}^n f'_{x_i}(x^0)\cdot(x_i-x_i^0)\right]\right\| &\le \left(\sum_{i=1}^n \alpha_i^2\right)^{\frac{1}{2}}\cdot\|x_i-x_i^0\| \\ &= o(\|x-x^0\|) \end{aligned}$$

即 f 在 x^0 处可微.

如果我们把 $m\times n$ 的矩阵视为 $\mathbb{R}^{mn}$ 中的点, 则矩阵也可定义自然的范数. 即, 如果 $A=(a_{ij})_{m\times n}$, 则其范数定义为

$$\|A\| = \left[\sum_{\substack{1\le i\le m\\ 1\le j\le n}} a_{ij}^2\right]^{\frac{1}{2}}.$$

由 Schwarz 不等式, 有

$$\|A\cdot v\| \le \|A\|\cdot\|v\|, \qquad \forall v\in\mathbb{R}^n.$$

定理 2(复合求导) 设 Δ 为 $\mathbb{R}^l$ 中开集, D 为 $\mathbb{R}^m$ 中开集, $g:\Delta\to D$ 及 $f:D\to\mathbb{R}^n$ 为映射. 如果 g 在 $u^0\in\Delta$ 处可微, f 在 $x^0=g(u^0)$ 处可微, 则复合映射 $h=f\circ g:\Delta\to\mathbb{R}^n$ 在 u^0 处可微, 且

$$Jh(u^0)=Jf(x^0)\cdot Jg(u^0).$$

证明 因为 g 在 u^0 处可微, 故

$$g(u)-g(u^0)=Jg(u^0)\cdot(u-u^0)+R_g(u,u^0) \tag{1}$$

其中 $R_g(u,u^0)=o(\|u-u^0\|)$. 同理, 因为 f 在 $x^0=g(u^0)$ 处可微, 故

$$f(x)-f(x^0)=Jf(x^0)\cdot(x-x^0)+R_f(x,x^0) \tag{2}$$

其中 $R_f(x,x^0)=0(\|x-x^0\|)$.

由 (1) 知, 当 $u\to u^0)$ 时, $g(u)\to g(u^0)=x^0$. 以 $x=g(u)$ 代入 (2), 得

$$\begin{aligned} f\circ g(u)-f\circ g(u^0) &= Jf(x^0)(g(u)-g(u^0))+R_f(g(u),g(u^0)) \\ &= Jf(x^0)\cdot Jg(u^0)\cdot(u-u^0)+R_{f\circ g}(u,u^0) \end{aligned} \tag{3}$$

其中

$$R_{f\circ g}(u,u^0)=Jf(x^0)\cdot R_g(u,u^0)+R_f(g(u),g(u^0))$$

从而有如下估计

$$\begin{aligned} \|R_{f\circ g}(u,u^0)\| &\le \|Jf(x^0)\cdot R_g(u,u^0)\|+\|R_f(g(u),g(u^0)\| \\ &\le \|Jf(x^0)\|\cdot\|R_g(u,u^0)\|+o(\|g(u)-g(u^0)\|) \\ &= o(\|u-u^0\|)+o(O(\|u-u^0\|))=o(\|u-u^0\|). \end{aligned}$$

从而由 (3) 及微分的定义知 $f\circ g$ 在 u^0 处可微, 且 $J(\circ g)(u^0)=Jf(x^0)\cdot J_g(u^0)$.

如果把 f, g 分别表示成分量形式

$$\begin{aligned} y_i &= f_i(x_1,\cdots,x_n), \qquad i=1,\cdots,m, \\ x_j &= g_j(u_1,\cdots,u_l), \qquad j=1,\cdots,n. \end{aligned}$$

则 $J(f\circ g)(u^0)=Jf(x^0)\cdot J_g(u^0)$ 可改写为

$$\begin{pmatrix} \frac{\partial y_1}{\partial u_1}(u^0) & \cdots & \frac{\partial y_1}{\partial u_l}(u^0) \\ & \cdots & \\ \frac{\partial y_n}{\partial u_1}(u^0) & \cdots & \frac{\partial y_n}{\partial u_l}(u^0) \end{pmatrix}_{n\times l} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1}(x^0) & \cdots & \frac{\partial y_1}{\partial x_m}(x^0) \\ & \cdots & \\ \frac{\partial y_n}{\partial x_1}(x^0) & \cdots & \frac{\partial y_n}{\partial x^m}(x^0) \end{pmatrix}_{n\times m} \cdot \begin{pmatrix} \frac{\partial x_1}{\partial u_1}(u^0) & \cdots & \frac{\partial x_1}{\partial u_l}(u^0) \\ & \cdots & \\ \frac{\partial x_m}{\partial u_1}(u^0) & \cdots & \frac{\partial x_m}{\partial u^l}(u^0) \end{pmatrix}_{m\times l}$$

即

$$\frac{\partial y_i}{\partial u_j}(u^0)=\sum_{s=1}^{n}\frac{\partial y_i}{\partial x_s}(g(u^0))\cdot\frac{\partial x_s}{\partial u_j}(u^0).$$

这也就是所谓的链规则.

例 2 设 $f(x,y)$ 可微, $\varphi(x)$ 可微, 求 $u=f(x,\varphi(x))$ 关于 x 的导数.

解 由链规则

$$\begin{aligned} u'_x &= f'_x(x,\varphi(x))\cdot x'_x+f'_y(x,\varphi(x))\cdot\varphi'(x) \\ &= f'_x(x,\varphi(x))+f'_y(x,\varphi(x))\cdot\varphi'(x). \end{aligned}$$

例 3 设 $u=f(x,y)$ 可微, $x=r\cos\theta, y=r\sin\theta$, 证明

$$\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2=\left(\frac{\partial u}{\partial r}\right)^2+\frac{1}{r^2}\left(\frac{\partial u}{\partial \theta}\right)^2.$$

证明 由链规则,

$$\begin{aligned} \frac{\partial u}{\partial r} &= \frac{\partial u}{\partial x}\cdot\frac{\partial x}{\partial r}+\frac{\partial u}{\partial y}\cdot\frac{\partial y}{\partial r}=\frac{\partial u}{\partial x}\cos\theta+\frac{\partial u}{\partial y}\cdot\sin\theta \\ \frac{\partial u}{\partial \theta} &= \frac{\partial u}{\partial x}\cdot\frac{\partial x}{\partial \theta}+\frac{\partial u}{\partial y}\cdot\frac{\partial y}{\partial \theta}=-r\cdot\frac{\partial u}{\partial x}\cdot\sin\theta+r\frac{\partial u}{\partial y}\cos\theta \end{aligned}$$

这说明

$$\begin{aligned} \left(\frac{\partial u}{\partial r}\right)^2+\frac{1}{r^2}\left(\frac{\partial u}{\partial \theta}\right)^2 &= \left(\frac{\partial u}{\partial x}\cos\theta+\frac{\partial u}{\partial y}\sin\theta\right)^2+\left(-\frac{\partial u}{\partial x}\sin\theta+\frac{\partial u}{\partial y}\cos\theta\right)^2 \\ &= \left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2. \end{aligned}$$

例 4 设 $z=f(u,v,w), v=\varphi(u,s),\ s=\psi(u,w)$, 求 $\frac{\partial z}{\partial u}, \frac{\partial z}{\partial w}$.

解

$$z = f(u,v,w) = f(u,\varphi(u,s),w) = f(u,\varphi(u,\psi(u,w)),w).$$

由链规则,

$$\begin{aligned}
\frac{\partial z}{\partial u} &= \frac{\partial f}{\partial u} + \frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial u} = \frac{\partial f}{\partial u} + \frac{\partial f}{\partial v}\cdot\left(\frac{\partial \varphi}{\partial u} + \frac{\partial \varphi}{\partial s}\cdot\frac{\partial s}{\partial u}\right) \\
&= \frac{\partial f}{\partial u} + \frac{\partial f}{\partial v}\cdot\frac{\partial \varphi}{\partial u} + \frac{\partial f}{\partial v}\cdot\frac{\partial \varphi}{\partial s}\cdot\frac{\partial \psi}{\partial u} \\
\frac{\partial z}{\partial w} &= \frac{\partial f}{\partial w} + \frac{\partial f}{\partial v}\cdot\frac{\partial v}{\partial w} = \frac{\partial f}{\partial w} + \frac{\partial f}{\partial v}\left(\frac{\partial \varphi}{\partial s}\cdot\frac{\partial s}{\partial w}\right) = \frac{\partial f}{\partial w} + \frac{\partial f}{\partial v}\cdot\frac{\partial \varphi}{\partial s}\cdot\frac{\partial \psi}{\partial w}.
\end{aligned}$$

最后, 我们简单地介绍全微分 (形式微分) 的概念. 设 $f: D \to \mathbb{R}$ 为可微的多元函数, 由定义, f 在 x 处的微分 $df(x)$ 是一个线性映射

$$\begin{aligned}
df(x): \quad & \mathbb{R}^m \to \mathbb{R} \\
& v \mapsto \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}\cdot v_i
\end{aligned}$$

我们把映射 $x \mapsto df(x)$ 称为 f 的全微分, 记为 df. 由于

$$d(\lambda f + \mu g)(x) = \lambda df(x) + \mu dg(x),\ \ \lambda,\mu \in \mathbb{R},$$

因此, 全微分之间也可以定义加法和数乘运算, 在这个意义下, 有

$$\begin{aligned}
df &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\cdot dx^i \quad (*) \\
d(\lambda f + \mu g) &= \lambda df + \mu dg, \\
d(f,g) &= f\cdot dg + g df, \\
d(\frac{f}{g}) &= \frac{g df - f dg}{g^2} \quad (g \neq 0).
\end{aligned}$$

如果将 (*) 写成矩阵形式:

$$df = Jf(dx_1,\cdots,dx_n)^T$$

则复合映射的链规则可写为

$$\begin{aligned}
d(f\circ g) &= J(f\circ g)\cdot(du_1,\cdots,du_m)^T, \\
&= Jf(x)\cdot Jg(u)(du_1,\cdots,du_m)^T \qquad (x = g(u)) \\
&= Jf(d(u))\cdot(dg_1,\cdots,dg_n)^T
\end{aligned}$$

这个等式称为全微分的形式不变性.

例 5 $u=\log\dfrac{z^2}{\sqrt{x^2+y^2}}$, 求 du 及 u 的偏导数.

解

$$\begin{aligned} du=d\,\log\frac{z^2}{x^2+y^2} &= d\left(\log z^2-\ln(x^2+y^2)\right)\\ &= \frac{2}{z}dz-\frac{1}{x^2+y^2}d(x^2+y^2)\\ &= \frac{-2x}{x^2+y^2}dx-\frac{2y}{x^2+y^2}dy+\frac{2}{z}dz. \end{aligned}$$

这说明

$$\frac{\partial u}{\partial x}=-\frac{\partial x}{x^2+y^2},\quad \frac{\partial u}{\partial y}=-\frac{2y}{x^2+y^2},\quad \frac{\partial u}{\partial z}=\frac{2}{z}.$$

§4 中值公式与 Taylor 公式

设 $p,q\in\mathbb{R}^n$, 令

$$\sigma(t)=(1-t)\cdot p+t\cdot q,\qquad t\in\mathbb{R}.$$

我们称 $\sigma:[0,1]\to\mathbb{R}^n$ 为 $\mathbb{R}^n$ 中连接 p,q 的直线段. 设 A 为 $\mathbb{R}^n$ 中的子集, 如果 $\forall a_1,a_2\in A$, 连接 a_1,a_2 的直线段仍包含于 A, 则称 A 为凸集. 特别地, 凸集都是道路连通的. 我们把开的凸集称为凸域.

例 1 $\mathbb{R}^n$ 中开球 $B_r(x)$ 都是凸集.

定理 1 (微分中值定理) 设 D 为 $\mathbb{R}^n$ 中凸域, $f:D\to\mathbb{R}$ 可微, 则 $\forall\ a,b\in D$, $\exists\ \xi\in D$, 使得

$$f(b)-f(a)=Jf(\xi)\cdot(b-a),$$

其中 $\xi=a+\theta(b-a)$, $\theta\in(0,1)$, ξ 位于连接 a,b 的直线段上.

证明 设 $\sigma:[0,1]\to D$ 为连接 a,b 的直线段, 则复合函数 $\varphi(t)=f\circ\sigma(t)$ 是可微的一元函数. 由一元函数的微分中值定理, $\exists\ \theta\in(0,1)$ 使得

$$\varphi(1)-\varphi(0)=\varphi'(\theta)\cdot(1-0).$$

上式即

$$f(b)-f(a)=Jf(\xi)\cdot(b-a).$$

其中 $\xi=\sigma(\theta)=a+\theta(b-a)$.

推论 1 设 D 为 $\mathbb{R}^n$ 中区域 (道路连通开集), $f:D\to\mathbb{R}$ 可微. 如果 $Jf\equiv 0$, 则 f 为常数.

证明 如果 D 为凸域, 则欲证结论从定理 1 立即得到. 一般地, 任取 $p,q\in D$, 设 $\tau:[0,1]\to D$ 是连接 p,q 的连结曲线. 令

$$T=\sup\{t\,|\,\text{当 } 0\le s\le t \text{ 时, } f\circ\tau(s)=f(p)\}.$$

因为 D 为开集, 故 $\exists\,\varepsilon>0$ 使得 $B_{\varepsilon_0}(p)\subset D$. 由 τ 的连续性知, $\exists\,\delta>0$, 使得 $\tau([0,\delta))\subset B_{\varepsilon_0}(p)$. 由于 $B_{\varepsilon_0}(p)$ 为凸域, f 在 $B_{\varepsilon_0}(p)$ 中为常值函数, 故 $T\ge\delta>0$.

下面证明 $T=1$ (反证法). 如果不然, 则在 $\tau(T)\in D$ 处作和上面在 $\tau(0)=p$ 处一样的讨论知, $\exists\,\delta'>0$ 使得

$$[0,T+\delta')\subset\{s\in[0,1]\,|\,f(s)=f(p)\}.$$

这和 T 的定义相矛盾! 从而 $T=1$. 由连续性立知 $f(q)=f(1)=f(T)=f(p)$. 因为 p,q 是任取的, 故 f 为常数.

例 2 考虑向量值函数 $f:\mathbb{R}\to R^2$, $f(t)=(t^2,t^3)$, 则

$$Jf(t)=\begin{pmatrix}2t\\3t^2\end{pmatrix}.$$

如果 $\exists\,\theta\in(0,1)$ 使得

$$f(1)-f(0)=Jf(\theta)\cdot(1-0)$$

则

$$\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}2t\\3t^2\end{pmatrix}.\qquad\text{(无解)}$$

这个例子说明定理 1 对向量值函数不再成立.

定理 2 (拟微分平均值定理) 设 D 为 $\mathbb{R}^n$ 中凸域, $f:D\to\mathbb{R}^m$ 可微, 则 $\forall\,a,b\in D$, $\exists\,\xi\in D$ 使得

$$\|f(b)-f(a)\|\le\|Jf(\xi)\|\cdot\|b-a\|.$$

其中 $\xi=a+\theta\cdot(b-a)$, $\theta\in(0,1)$.

证明 记 $\sigma:[0,1]\to D$ 为连接 a,b 的直线段. 考虑复合函数

$$\begin{aligned}&\varphi:[0,1]\to\mathbb{R}\\&\varphi(t)=\langle f(b)-f(a),\ f\circ\sigma(t)\rangle\end{aligned}$$

则 φ 为可微一元函数, 且

$$\varphi'(t)=\langle f(b)-f(a),\ Jf(\sigma(t))\cdot(b-a)\rangle$$

由一元函数的微分中值定理, $\exists\ \theta\in(0,1)$ 使得

$$\begin{aligned}|\varphi(1)-\varphi(0)| &= |\varphi'(\theta)|\le\|f(b)-f(a)\|\cdot\|Jf(\sigma(\theta))\cdot(b-a)\|\\&\le \|f(b)-f(a)\|\cdot\|Jf(\xi)\|\cdot\|b-a\|\end{aligned}$$

因为

$$\begin{aligned}|\varphi(1)-\varphi(0)| &= |\langle f(b)-f(a),\ f(b)\rangle-\langle f(b)-f(a),\ f(a)\rangle|\\&= \|f(b)-f(a)\|^2\end{aligned}$$

结合两个不等式就得到了定理的证明.

推论 2 设 D 为 $\mathbb{R}^n$ 中区域, $f:D\to\mathbb{R}^m$ 可微. 如果 $Jf\equiv 0$, 则 f 为常值映射.

证明 和推论 1 的证明类似, 略.

根据定义, 如果一个多元函数可微, 则它可以用线性函数逼近, 利用这一点我们可以做近似计算.

例 3 求 $\dfrac{1.03^2}{\sqrt{0.98}\sqrt[3]{1.06}}$ 的近似值.

解 考虑函数 $f(x,y,z)=\dfrac{x^2}{\sqrt{y}\sqrt[3]{z}}$, $x_0=(1,1,1)$, $h=(0.03,-0.02,0.06)$, 则

$$\begin{aligned}\frac{1.03^2}{\sqrt{0.98}\sqrt[3]{1.06}} &= f(1.03,0.98,1.06)\\&= f(x_0+h)\approx f(x_0)+Jf(x_0)\cdot h\\&= 1+(2,-\frac{1}{2},-\frac{1}{3})\begin{pmatrix}0.03\\-0.02\\0.06\end{pmatrix}=1.05.\end{aligned}$$

为了更好地求逼近, 我们要用多元多项式来逼近多元函数. 为此先引进一些记号. 设 $\alpha_i \in \mathbb{Z}(1 \le i \le n)$, 记 $\alpha = (\alpha_1, \cdots, \alpha_n)$, 称为多重指标. 记

$$|\alpha| = \sum_{i=1}^{n} \alpha_i, \quad \alpha! = \alpha_1! \cdot \alpha_2! \cdots \alpha_n!.$$

如果 $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, 则记

$$x^\alpha = x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdots x_n^{\alpha_n}.$$

对于多元函数 f, 还用下面的记号表示 $|\alpha|$ 阶偏导数:

$$D^\alpha f(x^0) = \frac{\partial^{|\alpha|} f}{\partial x_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n}}(x_0).$$

定理 3 (Taylor 公式) 设 D 为 $\mathbb{R}^n$ 中凸域, $f \in C^{m+1}(D)$(即 f 具有 $m+1$ 阶连续偏导数), $a = (a_1, \cdots, a_m) \in D$. 则 $\forall\, x \in D$, $\exists\, \theta \in (0,1)$ 使得

$$f(x) = \sum_{k=0}^{m} \sum_{|\alpha|=k} \frac{D^\alpha f(a)}{\alpha!} \cdot (x-a)^\alpha + \sum_{|\alpha|=m+1} \frac{D^\alpha f(a+\theta(x-a))}{\alpha!} \cdot (x-a)^\alpha.$$

证明 考虑一元函数 $\varphi(t) = f(a + t\cdot(x-a))$, $t \in [0,1]$. φ 具有 $m+1$ 阶连续导数, 故由一元函数的 Taylor 公式, 有

$$(*) \quad \varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2!}\varphi''(0) + \cdots + \frac{1}{m!}\varphi^{(m)}(0) + \frac{1}{(m+1)!}\varphi^{(m+1)}(\theta),\ \theta \in (0,1)$$

利用归纳法不难证明

$$\varphi^{(k)}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(a + t(x-a)) \cdot (x-a)^\alpha$$

特别地, $t=0$ 时, 有

$$\varphi^{(k)}(0) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} D^\alpha f(a) \cdot (x-a)^\alpha,$$

上式代入 $(*)$ 即得欲证公式.

记

$$R_m = f(x) - \sum_{k=0}^{m} \sum_{|\alpha|=k} \frac{D^\alpha f(a)}{\alpha!} \cdot (x-a)^\alpha,$$

由定理 3, $f \in C^{m+1}(D)$ 时

$$R_m = \sum_{|\alpha|=m+1} \frac{D^\alpha f(a+\theta(x-a))}{\alpha!} \cdot (x-a)^\alpha. \qquad \text{(Lagrange余项)}$$

推论 3 在定理 3 的条件下, 当 $\|x-a\|$ 充分小时, $R_m = O(\|x-a\|^{m+1})$.

证明 取 $\delta > 0$, 使得 $\bar{B}_\delta(a) \subset D$. 由于 $f \in C^{m+1}(D)$, $\bar{B}_\delta(a)$ 紧致, 故 $\exists\, M > 0$ 使得

$$|D^\alpha f(x)| \le M,\ \forall x \in \bar{B}_\delta(a),\ |\alpha| \le m+1.$$

因此

$$\begin{aligned} |R_m| &\le M \cdot \sum_{|\alpha|=m+1} |(x-a)^\alpha| \\ &\le M \cdot \sum_{|\alpha|=m+1} \|x-a\|^{|\alpha|} = M \cdot n^{m+1} \cdot \|x-a\|^{m+1}. \end{aligned}$$

注 (1) 如果 $f \in C^m(D)$, 则 Taylor 公式可写为

$$\begin{aligned} f(x) &= \sum_{k=0}^{m-1} \sum_{|\alpha|=k} \frac{D^\alpha f(a)}{\alpha!} \cdot (x-a)^\alpha + \sum_{|\alpha|=m} \frac{D^\alpha f(a+\theta(x-a))}{\alpha!} \cdot (x-a)^\alpha \\ &= \sum_{k=0}^{m} \sum_{|\alpha|=k} \frac{D^\alpha f(a)}{\alpha!} (x-a)^\alpha + R_m \end{aligned}$$

其中

$$R_m = \sum_{|\alpha|=m} \frac{1}{\alpha!} \left[D^\alpha f(a+\theta(x-a) - D^\alpha f(a)\right] \cdot (x-a)^\alpha$$

用推论 3 的证明方法可得如下估计:

$$R_m = o(\|x-a\|^m),\ x \to a. \quad \text{(Peano余项)}.$$

(2) 多元函数 Taylor 展开的前三项为

$$f(x) = f(a) + Jf(a) \cdot (x-a) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \cdot (x_i - a_i) \cdot (x_j - a_j) + \cdots$$

记 $\mathrm{Hess}(f) = \left[\dfrac{\partial^2 f}{\partial x_i \partial x_j}(a)\right]_{n \times n}$, 称为 f 在 a 处的 Hessian.

(3) Taylor 展开式的系数由 f 惟一确定.

定义 1 (凸函数) 设 D 为 $\mathbb{R}^n$ 中的凸域, $f: D \to \mathbb{R}$ 为多元函数. 如果 $\forall\, x, y \in D$, 均有

$$f(tx + (1-t)y) \le t \cdot f(x) + (1-t)f(y),\ \forall\, t \in [0,1]$$

则称 f 为 D 上的凸函数. 上式中 "$\le$" 换成 "$<$" 时, 称为严格凸函数.

定理 4 (1) 如果 f 在凸域 D 上有连续一阶偏导数, 则 f 为凸函数 $\Leftrightarrow$

$$f(y) \ge f(x) + Jf(x) \cdot (y-x),\ \forall\, x, y \in D.$$

(2) 如果 f 有连续二阶偏导数, 则 f 为凸函数 $\Leftrightarrow$ $\mathrm{Hess}(f) \ge 0$ (半正定矩阵).

证明 (1) "$\Rightarrow$" $\forall\, x, y \in D,\ t \in [0,1]$, 有

$$\begin{aligned} t \cdot (f(y) - f(x)) &\ge f(x + t(y-x)) - f(x) \\ &= Jf(x) \cdot t(y-x) + o(t\|y-x\|) \end{aligned}$$

上式两边除以 t, 然后令 $t \to 0$ 即得

$$f(y) \ge f(x) + Jf(x) \cdot (y-x).$$

"$\Leftarrow$" $\forall\, x, y \in D,\ t \in [0,1]$, 记 $z = tx + (1-t) \cdot y$, 则

$$\begin{aligned} f(x) &\ge f(z) + Jf(z)(x-z) \\ f(y) &\ge f(z) + Jf(z)(y-z) \end{aligned}$$

这说明

$$t \cdot f(x) + (1-t)f(y) \ge f(z) + Jf(z)\,[t(x-z) + (1-t)(y-z)] = f(z).$$

(2) "$\Leftarrow$" 设 f 有连续的二阶偏导数, 且 $\mathrm{Hess}(f)$ 半正定, 则由 Taylor 公式, $\forall\, x, y \in D$, $\exists\, \xi = x + \theta \cdot (y-x)$, $\theta \in (0,1)$ 使得

$$\begin{aligned} f(y) &= f(x) + Jf(x) \cdot (y-x) + \frac{1}{2}(y-x)^T \cdot \mathrm{Hess}(f)(\xi) \cdot (y-x) \\ &\ge f(x) + Jf(x) \cdot (y-x). \end{aligned}$$

由 (1) 知 f 为凸函数.

“⇒” (反证法). 如果 Hess(f) 不是半正定的, 则 $\exists\, x\in D$, 以及 $h=(h_1,\cdots,h_n)\in R^n$ 使得

$$(h_1,\cdots,h_n)\cdot \mathrm{Hess}(f)(x)\cdot(h_1,\cdots,h_n)^T<0.$$

由 Taylor 公式及其注记, 当 $\varepsilon\to 0$ 时, 有

$$\begin{aligned} f(x+\varepsilon\cdot h) &= f(x)+\varepsilon\cdot Jf(x)\cdot h+\frac{1}{2}\varepsilon^2\cdot h^T\cdot \mathrm{Hess}(f)\cdot h+o(\|\varepsilon h\|^2)\\ &= f(x)+\varepsilon\cdot Jf(x)\cdot h+\varepsilon^2\left[\frac{1}{2}h^T\cdot \mathrm{Hess}(f)(x)\cdot h+o(1)\right]\end{aligned}$$

当 $\varepsilon\neq 0$ 充分小时, 上式第三项 <0. 此时

$$f(x+\varepsilon h)<f(x)+Jf(x)\cdot\varepsilon h.$$

这与 f 为凸函数相矛盾 (用到 (1)).

§5 逆映射定理和隐映射定理

回忆一下: 对于一元函数, 如果它可微且导数处处非零, 则该函数可逆且其逆仍可微. 这里, 可微性和导数非零保证了函数在局部上与可逆线性函数有好的逼近, 因而也是 (局部) 可逆的. 下面我们考虑多元向量值函数类似的问题.

例 1 我们先考虑线性映射. 由线性代数, 线性映射 $A:\mathbb{R}^n\to\mathbb{R}^n$ 可逆 $\Leftrightarrow$ $\det A\neq 0\Leftrightarrow A$ 为单射. 我们有如下观察: 如果 $B:\mathbb{R}^n\to\mathbb{R}^n$ 为线性映射, $\|B\|<1$, 则 I_n-B 可逆.

事实上, 设 $(I_n-B)v=0$, 则

$$\|v\|=\|I_nv\|=\|Bv\|\le\|B\|\cdot\|v\|$$

这说明 $v=0$.

即, 把恒同映射这样一个可逆映射作一个小的挠动以后仍然为可逆映射. 一般地, 任何可逆映射在微扰下仍为可逆映射.

定理 1 (逆映射定理) 设 W 为 $\mathbb{R}^n$ 中开集, $f:W\to\mathbb{R}^n$ 为 $C^k(k\ge 1)$ 映射, $x^0\in W$. 如果 $\det Jf(x^0)\neq 0$, 则存在 x^0 的开邻域 $U\subset W$ 以及 $y^0=f(x^0)$ 的开邻域 $V\subset\mathbb{R}^n$, 使得 $f|_U:U\to V$ 是可逆映射, 且其逆仍为 C^k 映射.

证明 不失一般性, 可设 $x^0=0$, $y^0=f(x^0)=0$. 以 A 记 f 在 $x^0=0$ 处的微分, 则 A 可逆, 且 $f\circ A^{-1}$ 在 0 处微分为恒同映射. 如果欲证的结论对 $f\circ A^{-1}$ 成立, 则对 f 也成立. 因此, 不妨从一开始就假设 $Jf(x^0)=I_n$.

定义映射

$$\begin{aligned} g:\ \ & W\to\mathbb{R}^n \\ & x\mapsto f(x)-x \end{aligned}$$

则 g 为 C^k 映射, 且 $Jg(0)=0$. 因此, $\exists\ \varepsilon_0>0$ 使得

$$\|Jg(x)\|\le\frac{1}{2},\ \forall x\in\bar{B}_{\varepsilon_0}(0)\subset W.$$

由拟微分平均值定理,

$$\begin{aligned} \|g(x_1)-g(x_2)\| &\le \|Jg(\xi)\|\cdot\|x_1-x_2\| \\ &\le \frac{1}{2}\cdot\|x_1-x_2\|,\quad \forall\ x_1,x_2\in\bar{B}_{\varepsilon_0}(0). \end{aligned}$$

设 $y\in B_{\frac{\varepsilon_0}{2}}(0)$, 我们来解方程

$$f(x)=y,\qquad x\in B_{\varepsilon_0}(0). \tag{5.1}$$

这等价于在 $B_{\varepsilon_0}(0)$ 中寻找 $g_y(x)=x+y-f(x)=y-g(x)$ 的不动点. 我们利用压缩映像原理来找这样的不动点. 首先,

$$\begin{aligned} \|g_y(x)\| &= \|y-g(x)\|\le\|y\|+\|g(x)\| \\ &< \frac{\varepsilon_0}{2}+\frac{1}{2}\|x\|\le\frac{\varepsilon_0}{2}+\frac{\varepsilon_0}{2}=\varepsilon_0,\quad \forall x\in\bar{B}_{\varepsilon_0}(0) \end{aligned} \tag{5.2}$$

其次, $g_y:\bar{B}_{\varepsilon_0}(0)\to B_{\varepsilon_0}(0)\subset\bar{B}_{\varepsilon_0}(0)$ 是压缩映射:

$$\|g_y(x_1)-g_y(x_2)\|=\|g(x_2)-g(x_1)\|\le\frac{1}{2}\|x_1-x_2\|,\ \forall x_1,x_2\in\bar{B}_{\varepsilon_0}(0).$$

从而 (5.1) 在 $\bar{B}_{\varepsilon_0}(0)$ 中有惟一解, 记为 x_y. 由 (5.2), $x_y\in B_{\varepsilon_0}(0)$. 记

$$U=f^{-1}(B_{\frac{\varepsilon_0}{2}}(0))\cap B_{\varepsilon_0}(0),\ V=B_{\frac{\varepsilon_0}{2}}(0),$$

则我们已经证明了 $f|_U:U\to V$ 是一一的 C^k 映射, 其逆映射 $h(y)=x_y$ 满足

$$y-g(h(y))=h(y). \tag{5.3}$$

(1) $h:V\to U$ 是连续映射: 当 $y_1,\,y_2\in V$ 时

$$\begin{aligned}\|h(y_1)-h(y_2)\| &\le \|y_1-y_2\|+\|g(h(y_1))-g(h(y_2))\|\\ &\le \|y_1-y_2\|+\frac{1}{2}\|h(y_1)-h(y_2)\|\end{aligned}$$

这说明

$$\|h(y_1)-h(y_2)\|\le 2\cdot\|y_1-y_2\|,\ \ y_1,y_2\in V.$$

(2) $h:V\to U$ 是可微映射: 设 $y_0\in V$, 则 $\forall\ y\in V$, 有

$$\begin{aligned}h(y)-h(y_0) &= (y-y_0)-[g(h(y))-g(h(y_0))]\\ &= (y-y_0)-Jg(h(y_0))\cdot(h(y-h(y_0))+o(\|h(y)-h(y_0)\|)\end{aligned}$$

上式可改写为

$$[I_n+Jg(h(y_0))]\cdot(h(y)-h(y_0))=(y-y_0)+o(\|y-y_0)\|).$$

因而

$$h(y)-h(y_0)=[I_n+(Jg)(h(y_0))]^{-1}\cdot(y-y_0)+o(\|y-y_0\|).$$

即 h 在 y_0 处可微.

(3) $h:V\to U$ 为 C^k 映射.

由 (2) 的证明知

$$Jh(y)=[I_n+Jg(h(y))]^{-1}=[Jf(h(y))]^{-1},\ \forall y\in V. \tag{5.4}$$

由 $f\in C^k\Rightarrow Jf\in C^{k-1}$. 由 (2) 及 (5.4) $\Rightarrow Jh\in C^0$, 即 $h\in C^1$. 再由 $Jf\in C^{k-1}$, $h\in C^1$ 及 (5.4) $\Rightarrow Jh\in C^1$ 即 $h\in C^2$. 依次类推, 最后我们就得到 $h\in C^k$.

注 从证明可以看出, 如果 $f:W\to\mathbb{R}^n$ 的 Jacobian 非退化, 则 $f(W)$ 为开集.

例 2 考虑 $f:\mathbb{R}^2\to\mathbb{R}^2$, $f(x,y)=(e^x\cdot\cos y,\ e^x\cdot\sin y)$. 显然, f 不是单射, 但

$$\det Jf(x,y)=\det\begin{pmatrix} e^x\cos y & -e^x\sin y\\ e^x\sin y & e^x\cos y\end{pmatrix}=e^{2x}\ne 0.$$

这说明定理 1 的结论只能局部地成立.

例 3 考虑 $f:\mathbb{R}^2\to\mathbb{R}^2$, $f(x,y)=(x^3,y^3)$, 则 f 为光滑单射, 也是满射, 但 $Jf(0,0)=0$, f 的逆映射不可微. 这说明 Jacobian 非退化的条件不能去掉.

例 4 设 $f:\mathbb{R}^2\to\mathbb{R}$ 为 $C^k(k\geq 1)$ 映射, $f(x^0,y^0)=0, \dfrac{\partial f}{\partial y}(x^0,y^0)\neq 0$, 在 (x^0,y^0) 附近解方程

$$f(x,y)=0.$$

解 令 $F:\mathbb{R}^2\to\mathbb{R}^2$, $F(x,y)=(x,f(x,y))$, 则 $F(x^0,y^0)=(x^0,0)$, 且

$$\det JF(x^0,y^0)=\det\begin{pmatrix} 1 & 0 \\ \dfrac{\partial f}{\partial x}(x^0,y^0) & \dfrac{\partial f}{\partial y}(x^0,y^0)\end{pmatrix}=\frac{\partial f}{\partial y}(x^0,y^0)\neq 0.$$

由逆映射定理, 在 (x^0,y^0) 附近, F 为可逆映射. 于是当 x 在 x^0 附近时, y^0 附近存在 $g(x)$, 使得 $F(x,y(x))=(x,0)$, 即 $f(x,y(x))=0$. 对 x 求导 $\Rightarrow$

$$\frac{\partial f}{\partial x}(x,g(x))+\frac{\partial f}{\partial y}(x,g(x))\cdot g'(x)=0$$

从而

$$g'(x)=-\frac{\frac{\partial f}{\partial x}(x,g(x))}{\frac{\partial f}{\partial y}(x,g(x))}.$$

$y=g(x)$ 称为由 $f(x,y)=0$ 决定的隐函数. 这个例子的高维推广称为隐映射定理.

定理 2 (隐映射定理) 设 W 为 $\mathbb{R}^{n+m}$ 中开集, W 中的点用 (x,y) 表示, $x=(x_1,\cdots,x_n)$, $y=(y_1,\cdots,y_m)$. $f:W\to\mathbb{R}^m$ 为 C^k 映射,

$$f(x,y)=(f_1(x,y),f_2(x,y),\cdots,f_m(x,y)).$$

设 $(x^0,y^0)\in W$, $f(x^0,y^0)=0$ 且 $\det Jf_y(x^0,y^0)\neq 0$, 其中

$$Jf_y(x,y)=\left(\frac{\partial f_i}{\partial y_j}(x,y)\right)_{m\times m}$$

则存在 x^0 的开邻域 $V\subset\mathbb{R}^n$ 以及惟一的 C^k 映射 $g:V\to\mathbb{R}^m$, 使得

(1) $y^0=g(x^0)$, $f(x,g(x))=0$, $\forall\ x\in V$.

(2) $Jg(x)=-[Jf_y(x,g(x))]^{-1}Jf_x(x,g(x))$,

其中

$$Jf_x(x,y)=\left(\frac{\partial f_i}{\partial x_j}(x,y)\right)_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}.$$

证明 令 $F: W \to \mathbb{R}^{n+m}$ 为

$$F(x,y) = (x, f(x,y))$$

在 (x^0, y^0) 处利用逆映射定理即可, 因为证明和例 4 类似, 故略.

注 $g(x)$ 称为由 $f(x,y)=0$ 决定的隐映射.

例 5 设 $x^2+2y^2+3z^2+xy-z-9=0$, 求 $x=1$, $y=-2$, $z=1$ 时 $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$, $\frac{\partial^2 z}{\partial x \partial y}$ 的值.

解 令 $F(x,y,z)=x^2+2y^2+3z^2+xy-z-9$, 则 $F(1,-2,1)=0$, $F_z(1,-2,1)=(6z-1)|_{z=1}=5\neq 0$, 故 z 可局部地表示为 x,y 的函数, 记为 $z=z(x,y)$. 由定理 2, 在 $(1,-2,1)$ 处, 有

$$\begin{aligned}(z_x, z_y) &= -F_z^{-1}\dot{}(F_x, F_y) \\ &= -\frac{1}{5}(2x+y, 4y+x)|_{(x,y,z)=(1,-2,1)} = (0, \frac{7}{5}).\end{aligned}$$

又因为

$$z_y = -F_z^{-1} \cdot F_y = -\frac{1}{6z-1}(4y+x)$$

故

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{6}{(6z-1)^2} \cdot z_y(4y+x) - \frac{1}{6z-1},$$

从而

$$\frac{\partial^2 z}{\partial x \partial y}(1,-2,1) = -\frac{1}{5}.$$

例 6 平面上直角坐标 (x,y) 和极坐标 (r,θ) 之间的转换公式为

$$x = r\cos\theta,\ y = r\sin\theta.$$

求 r 和 θ 关于 x,y 的偏导数.

解 由逆映射定理

$$\begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix}^{-1} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

从而有
$$\frac{\partial r}{\partial x}=\cos\theta,\quad \frac{\partial r}{\partial y}=\sin\theta,\quad \frac{\partial\theta}{\partial x}=-\frac{\sin\theta}{r},\quad \frac{\partial\theta}{\partial y}=\frac{\cos\theta}{r}.$$

例 7 (隐式曲面) 设 $F:\mathbb{R}^3\to\mathbb{R}$ 为 $C^k(k\geq 1)$ 的多元函数. 令

$$S=\{(x,y,z)\in\mathbb{R}^3\,|\,F(x,y,z)=0\}.$$

如果对于 $\forall\,(x,y,z)\in S$, F'_x, F'_y, F'_z 不全为 0, 则称 S 为由 F 决定的隐式曲面.

设 $m_0=(x_0,y_0,z_0)\in S$, 不妨设 $F'_z(x_0,y_0,z_0)\neq 0$. 则由定理 2, 在 (x_0,y_0,z_0) 附近, S 可用参数曲面 $z=z(x,y)$ 表示:

$$F(x,y,z(x,y))=0.$$

在 (x^0,y^0,z^0) 处的法向量为

$$\begin{aligned}\vec{n} &= (1,0,z_x(x^0,y^0)\times(0,1,z_y(x^0,y^0))\\ &= (-z_x(x^0,y^0),-z_y(x^0,y^0),1)\\ &= \frac{1}{F'_z}(F'_x(x^0,y^0,z^0),F'_y(x^0,y^0,z^0),F'_z(x^0,y^0,z^0))\end{aligned}$$

S 在 (x^0,y^0,z^0) 处的切平面方程为

$$F'_x\cdot(x-x_0)+F'_y\cdot(y-y_0)+F'_z(z-z_0)=0.$$

在 (x^0,y^0,z^0) 处的法线方程为

$$\frac{x-x_0}{F'_x(x_0,y_0,z_0)}=\frac{y-y_0}{F'_y(x_0,y_0,z_0)}=\frac{z-z_0}{F'_z(x_0,y_0,z_0)}.$$

例 8 (隐式曲线) 设 $F:\mathbb{R}^3\to\mathbb{R}$, $G:\mathbb{R}^3\to\mathbb{R}$ 为 $C^k(k\geq 1)$ 的映射, 令

$$l=\{(x,y,z)\in\mathbb{R}^3\,|\,F(x,y,z)=G(x,y,z)=0\},$$

如果对 $\forall\,(x,y,z)\in l$,

$$\frac{\partial(F,G)}{\partial(y,z)}(x,y,z)=\det\begin{pmatrix}F'_y & F'_z\\ G'_y & G'_z\end{pmatrix}\neq 0,$$

则称 l 为由 F, G 决定的隐式曲线.

设 $m_0=(x_0,y_0,z_0)\in l$, 由隐映射定理, 在 m_0 附近 l 可由参数曲线

$$y=y(x),\ z=z(x)$$

表示, 其中

$$F(x,y(x),z(x)=G(x,y(x),z(x))=0,$$

且

$$\begin{pmatrix} y'(x) \\ z'(x) \end{pmatrix} = -\begin{pmatrix} F'_y & F'_z \\ G'_y & G'_z \end{pmatrix}^{-1}\cdot\begin{pmatrix} F'_x \\ G'_x \end{pmatrix}$$

或

$$\begin{cases} F'_x+F'_y\cdot y'+F'_z\cdot z'=0 \\ G'_x+G'_y\cdot y'+G'_z\cdot z'=0 \end{cases}$$

这说明 $(1,y',z')$ 与 (F'_x,F'_y,F'_z) 和 (G'_x,G'_y,G'_z) 均正交. 因而 m_0 处的切线方程为

$$\frac{x-x_0}{\begin{vmatrix} F'_y & F'_z \\ G'_y & G'_z \end{vmatrix}}=\frac{y-y_0}{\begin{vmatrix} F'_z & F'_x \\ G'_z & G'_x \end{vmatrix}}=\frac{z-z_0}{\begin{vmatrix} F'_x & F'_y \\ G'_x & G'_y \end{vmatrix}}$$

法面方程为

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ F'_x & F'_y & F'_z \\ G'_x & G'_y & G'_z \end{vmatrix}=0.$$

§6 无条件极值

定义 1 (极值) 设 $f:U\to\mathbb{R}$ 为多元函数, U 为 $\mathbb{R}^n$ 中开集, $x^0=(x_1^0,\cdots,x_n^0)\in U$. 如果 $\exists\,\delta>0$ 使得 $B_\delta(x^0)\subset U$, 且

$$f(x^0)\ge f(x)\quad (\text{或}f(x^0)\le f(x)),\ \forall x\in B_\delta(x^0),$$

则称 x^0 为 f 的极大 (小) 值点, $f(x^0)$ 为 f 的极大 (小值). 当上式中 "$\ge$"("$\le$") 换成 "$>$"("$<$") 时, 相应地把 x_0 称为严格极值点, $f(x_0)$ 为严格极值.

命题 1 (达到极值的必要条件) 如果 x^0 为 f 的极值点, 且 f 在 x^0 处存在一阶偏导数, 则

$$f'_{x_1}(x^0) = f'_{x_2}(x^0) = \cdots = f'_{x_n}(x^0) = 0.$$

证明 以 f'_{x_1} 为例, 考虑一元函数

$$\varphi(x) = f(x, x_2^0, \cdots, x_n^0)$$

φ 可导, 且以 x_1^0 为极值点, 由 Fermat 定理,

$$\varphi'(x_1^0) = 0$$

即 $f'_{x_1}(x^0) = 0$.

注 满足条件 $f'_{x_1}(x^0) = \cdots = f'_{x_n}(x^0) = 0$ 的点 x^0 称为 f 的驻点.

定理 1 设 U 为 $\mathbb{R}^n$ 中开集, $f: U \to \mathbb{R}$ 具有二阶连续偏导数, x^0 为 f 的驻点. 则

(1) 如果 x^0 为 f 的极小 (大) 值点, 则 $\mathrm{Hess}(f)(x^0)$ 为半正 (负) 定方阵;

(2) 如果 $\mathrm{Hess}(f)(x^0)$ 为正 (负) 定方阵, 则 $x^0)$ 为 f 的严格极小 (大) 值点;

(3) 如果 $\mathrm{Hess}(f)(x^0)$ 为不定方阵, 则 $x^0)$ 不是 f 的极值点.

证明 证明的思想是在 x^0 附近利用 f 的 Taylor 展开:

$$\begin{aligned} f(x^0+h) - f(x^0) &= Jf(x^0)\cdot h + \frac{1}{2}h^T \cdot \mathrm{Hess}(f)(x^0)\cdot h + o(\|h\|^2) \\ &= \frac{1}{2}h^T \cdot \mathrm{Hess}(f)(x^0)\cdot h + o(\|h\|^2) \end{aligned} \tag{6.1}$$

(1) 设 x^0 为 f 的极小值点, v 为任意固定的单位向量, 在 (6.1) 中取 $h = t\cdot v$, $t \in \mathbb{R}$, 则 $|t|$ 充分小时,

$$0 \le f(x^0+h) - f(x^0) = t^2 \left[\frac{1}{2}\cdot v^T \cdot \mathrm{Hess}(f)(x^0)\cdot v + o(1)\right]$$

两边除以 t^2, 然后令 $t \to 0$ 知

$$0 \le v^T \cdot \mathrm{Hess}(f)(x^0)\cdot v,\ \forall\ v \in \mathbb{R}^n,\ \|v\| = 1.$$

这说明 $\mathrm{Hess}(f)(x^0)$ 为半正定方阵. x^0 为极大值点时证明完全类似.

(2) 如果 $\mathrm{Hess}(f)(x^0)$ 正定, 则

$$\lambda_0 = \min_{\|v\|=1} v^T \cdot \mathrm{Hess}(f)(x^0) \cdot v > 0.$$

从而

$$\begin{aligned} f(x^0+h) - f(x^0) &\geq \frac{1}{2}\|h\|^2 \cdot \lambda_0 + o(\|h\|^2) \\ &= \|h\|^2(\frac{1}{2}\lambda_0 + o(1)) \geq \frac{1}{4}\lambda_0 \cdot \|h\|^2, \quad (\|h\| \text{ 充分小}) \end{aligned}$$

即 x^0 为严格极小值点. $\mathrm{Hess}(f)(x^0)$ 负定的情形类似.

(3) 证明与 (1), (2) 类似, 略.

例 1 求 $f(x,y) = x^4 + y^4 - (x+y)^2$ 的极值.

解 先求驻点:

$$0 = \frac{\partial f}{\partial x} = 4x^3 - 2x - 2y,\ 0 = \frac{\partial f}{\partial y} = 4y^3 - 2x - 2y$$

有三个解

$$m_0 = (0,0), \quad m_1 = (1,1), \quad m_2 = (-1,-1).$$

为了判极大, 极小值, 再在驻点处求 Hessian:

$$\mathrm{Hess}(f)(x,y) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 12x^2 - 2 & -2 \\ -2 & 12y^2 - 2 \end{pmatrix}$$

在 m_1, m_2 处, Hessian 正定, 从而 m_1, m_2 为极小值点; 在 m_0 处 $\mathrm{Hess}(f)$ 退化, 用 $\mathrm{Hess}(f)$ 无法判别. 但对于 $0 < x < 1$, 令 $y = x$, $f(x,y) = 2x^4 - 4x^2 < 0$; 令 $y = -x$, $f(x,y) = 2x^4 > 0$, 故 m_0 不是极值点.

例 2 (最小二乘法) 设 $(x_1,y_1), \cdots, (x_n,y_n)$ 为平面 $\mathbb{R}^2$ 上 n 个点, 求一条直线 $y = ax + b$, 使得

$$F(a,b) = \sum_{i=1}^{n}(ax_i + b - y_i)^2$$

最小.

解 函数 $F(a,b)$ 是关于 (a,b) 的光滑函数. 先求驻点:

$$0 = \frac{\partial F}{\partial a}(a,b) = 2\sum_{i=1}^{n}(ax_i + b - y_i)\cdot x_i$$
$$0 = \frac{\partial F}{\partial b}(a,b) = 2\sum_{i=1}^{n}(ax_i + b - y_i)$$

当

$$n\cdot\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2 = \frac{1}{2}\sum_{i\neq j}(x_i - x_j)^2 \neq 0$$

时, 方程有唯一的解, 在此驻点处, F 的 Hessian 为

$$2\begin{pmatrix} \sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i \\ \sum_{i=1}^{n}x_i & n \end{pmatrix}$$

它是正定方阵, 故该驻点为极小值点. 由于 $F(a,b)\to+\infty$, $(a,b)\to\infty$, 故该驻点为惟一的最小值点 (见下面的引理 1).

例 3 求 $f(x,y) = x^2 + y^3 - 3y$ 的极值.

解 先求驻点:

$$0 = \frac{\partial f}{\partial x} = 2x,\ 0 = \frac{\partial f}{\partial y} = 3y^2 - 3.$$

驻点为 $(0,1)$, $(0,-1)$. 再计算 Hessian:

$$\mathrm{Hess}(f)(0,1) = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}, \qquad \mathrm{Hess}(f)(0,-1) = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$$

这说明 $(0,1)$ 为极小值点, $f(0,1) = -2$; $(0,-1)$ 不是极值点.

注意 $(0,1)$ 虽然是 f 的唯一极小值点, 但不是最小值点!

引理 1 设 $f:\mathbb{R}^n\to\mathbb{R}$ 为连续映射, 则

(1) 如果 $\lim\limits_{\|x\|\to\infty} f(x) = -\infty$, 则 f 在 $\mathbb{R}^n$ 上达到最大值;

(2) 如果 $\lim\limits_{\|x\|\to\infty} f(x) = +\infty$, 则 f 在 $\mathbb{R}^n$ 上达到最小值.

证明 以 (1) 为例, 不妨设 $f \not\equiv 0$, 取 $x^0 \in \mathbb{R}^n$, 由已知条件, $\exists R > 0$ 使得 $\|x\| \geq R$ 时,

$$f(x) < f(x^0).$$

则 f 在 $\bar{B}_R(0)$ 上的最大值即为 f 为 $\mathbb{R}^n$ 上的最大值.

例 4 求 $f(x,y) = x^2 + y^2 - 2x - y$ 在 $D: x \geq 0, y \geq 0, 2x + y \leq 4$ 上的最大值, 最小值.

解 本例须考虑 f 在 D 上的边界上的值. 先求 D 内驻点:

$$0 = \frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} \Rightarrow (x,y) = (1, \frac{1}{2}),\ f(1, \frac{1}{2} = -\frac{5}{4};$$

在边界上:

(1) $y = 0,\ 0 \leq x \leq 2$: $f(x,0) = x^2 - 2x$, 其驻点为 $(1,0)$, $f(1,0) = -1$.

(2) $x = 0,\ 0 \leq y \leq 4$: $f(0,y) = y^2 - y$, 驻点为 $(0, \frac{1}{2})$, $f(0, \frac{1}{2}) = -\frac{1}{4}$;

(3) $x \geq 0,\ y \geq 0,\ 2x + y = 4$: $f(x,y) = 5x^2 - 16x + 12$, 驻点为 $(\frac{8}{5}, \frac{4}{5})$, $f(\frac{8}{5}, \frac{4}{5}) = -\frac{4}{5}$.

最后, 在三个端点上:

$$f(0,0) = 0, \quad f(2,0) = 0, \quad f(0,4) = 12$$

$\Rightarrow f$ 的最大值为 12, 最小值为 $-\frac{5}{4}$.

§7 二次型与极值

在利用 $\mathrm{Hess}(f)$ 判别极值时, 需要判别其正定或负定性. 我们下面给出判别一个对称方阵为正定阵的方法.

设 A 为 n 阶对称实方阵, A 可以视为线性映射 $A: \mathbb{R}^n \to \mathbb{R}^n$. 另一方面, A 也诱导出函数 $Q: \mathbb{R}^n \to \mathbb{R}$,

$$Q(x) = \langle x, Ax \rangle = \sum_{i,j}^{n} a_{ij} x_i \cdot x_j,\ A = (a_{ij})_{n \times n}.$$

因为 $n-1$ 维球面

$$S^{n-1} = \{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i^2 = 1\}$$

为有界闭集, 故 Q 在 S^{n-1} 可以取到最小值和最大值.

引理 1 $Q(x)$ 在 S^{n-1} 上的最小值必为线性映射 A 的特征值.

证明 记 $\lambda_0 = \min\limits_{S^{n-1}}\{Q(x)\}$, 则 $\exists\, x^0 \in S^{n-1}$, 使得

$$\lambda_0 = Q(x^0).$$

因此有

$$Q(x) \geq \lambda_0 \cdot \|x\|^2,\ \forall x \in \mathbb{R}^n.$$

特别地, 对 $\forall\, y \in \mathbb{R}^n$, $t \in \mathbb{R}$, 有

$$\varphi(t) = Q(x^0 + ty, x^0 + ty) - \lambda_0 \|x^0 + ty\|^2 \geq 0.$$

$\varphi(t)$ 关于 t 为光滑函数, $t = 0$ 时取到最小值 0, 故 $\varphi'(0) = 0$, 简单的计算表明

$$\langle y, Ax^0\rangle + \langle x^0, Ay\rangle - \lambda_0(\langle x^0, y\rangle + \langle y, x^0\rangle) = 0.$$

由 A 为对称方阵 $\Rightarrow$

$$\langle y, Ax^0 - \lambda_0 \cdot x^0\rangle = 0.$$

取 $y = Ax^0 - \lambda_0 \cdot x^0$, $\Rightarrow$

$$Ax^0 - \lambda_0 \cdot x^0 = 0.$$

引理 1 的证明可以推广如下: 设 $V \subset \mathbb{R}^n$ 为子向量空间, 如果 $AV \subset V$, 则称 V 为 A 的一个不变子空间. 令

$$\mu = \inf\{Q(x) | x \in V, \|x\| = 1\}$$

则完全类似的证明可以推出 μ 为 A 的特征值.

如果 V 为不变子空间, 则其正定补

$$V^\perp = \{y \in \mathbb{R}^n \,|\, \langle y, x\rangle = 0,\ \forall x \in V\}$$

也是 A 的不变子空间; 而如果 μ 为特征值, 则特征子空间

$$V(\mu) = \{x \in \mathbb{R}^n \,|\, Ax = \mu x\}$$

是不变子空间. 因此, 如果重复上面的证明过程, 我们就可以得到 A 的所有特征值

$$\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_r, \quad r \leq n.$$

其中

$$\lambda_i = \min\{Q(x)\,|\,x \perp V(\lambda_0)\oplus V(\lambda_1)\oplus\cdots\oplus V(\lambda_{i-1}),\ \|x\|=1\},\quad i\geq 1.$$

从而有

引理 2 设 A 为 n 阶实对称方阵, 则

(1) A 的特征值全为实数;

(2) A 为正定矩阵 $\Leftrightarrow$ 其特征值都是正实数.

证明 (1) 以及 (2) 的 “$\Rightarrow$” 部分已证.

(2) 的 “$\Leftarrow$”: 如果 A 的特征值 $\lambda_0,\cdots,\lambda_r$ 均为正数, 则 $\mathbb{R}^n$ 有正交分解

$$\mathbb{R}^n = V(\lambda_0)\oplus V(\lambda_1)\oplus\cdots\oplus V(\lambda_r)$$

$\forall x\in\mathbb{R}^n$, $x\neq 0$, x 有正交分解

$$x = x_0+x_1+\cdots+x_r,\quad x_i\in V(\lambda_i)$$

从而

$$Q(x)=\sum_{i=0}^{r}\lambda_i\cdot Q(x_i)=\sum_{i=0}^{n}\lambda_i\cdot\|x_i\|^2>0.$$

即 A 是正定方阵.

引理 3 设 A 为 n 阶实对称方阵, 其正特征值个数 (含重数) 为 k, 如果 V 为 $\mathbb{R}^n$ 的子线性空间, 且 $\forall\ x\neq 0\in V$, $Q(x)>0$, 则 $\dim V<k$.

证明 (**反证法**) 设 $\dim V>k$, 记 A 的正特征值对应的特征子空间的直和为 V_0, 则 $\dim V_0=k$. 考虑正交投影

$$P: V\to V_0$$

因为 $\dim V>\dim V_0$, 故 $\ker P\neq\{0\}$. 从而存在 $x\neq 0\in V$ 使得 $x\perp V_0$. 由引理 2 的证明易见, 此时必有 $Q(x)\leq 0$. 矛盾!

现在我们就得到了矩阵正定性的如下判别法:

定理 1 $A=(a_{ij})_{n\times n}$ 为 n 阶正定方阵 $\Leftrightarrow \det(a_{ij})_{\substack{1\leq i\leq k\\ 1\leq j\leq k}}>0$, $1\leq k\leq n$.

证明 “$\Rightarrow$” 如果 A 正定, 则显然 $(a_{ij})_{\substack{1\leq i\leq k\\ 1\leq j\leq k}}$ 正定, 其特征值全为正实数, 故行列式为正 (行列式为特征值之积).

"$\Leftarrow$" 对 n 用归纳法. $n=1$ 显然. 设命题对 $n-1$ 成立, 则对于 n 阶方阵, 由归纳假设, $(a_{ij})_{\substack{1\le i\le n-1\\1\le j\le n-1}}$ 正定, 这说明 A 在子向量空间 $\mathbb{R}^{n-1}=\{x\in\mathbb{R}^n\,|\,x_n=0\}$ 上正定. 由引理 3, A 至少有 $n-1$ 个正的特征值. 又因为 $\det A>0$, 故所有特征值均为正数.

注 对半正定, 负定等方阵有类似的结果.

§8 Lagrange 乘数法

在前一节中, 我们在 S^{n-1} 上求 $Q(x)$ 的极值, S^{n-1} 满足方程

$$x_1^2+\cdots+x_n^2-1=0.$$

我们把上式称为一个约束条件. 一般地, 设 U 为 $\mathbb{R}^n$ 中开集, $f:U\to\mathbb{R}$ 为 U 上多元函数, $\Phi:U\to\mathbb{R}^m(m<n)$ 为 C^1 映射, 令

$$A=\{x\in U|\Phi(x)=0\}$$

f 在 A 上的局部极值称为条件极值, 方程 $\Phi(x)=0$ 称为约束条件.

利用隐 (函数) 映射定理, 我们有如下求条件极值的方法.

定理 1 (Lagrange 乘数法) 设 $f\in C^1(U)$, $x^0\in A$ 为 f 的条件极值. 如果 $J\Phi(x^0)$ 的秩为 m, 则存在 $\lambda\in\mathbb{R}^m$, 使得

$$Jf(x^0)-\lambda\cdot J\Phi(x^0)=0.$$

证明 不妨设 $\Phi=(\varphi_1,\cdots,\varphi_m)$, 且

$$\det\begin{pmatrix}\dfrac{\partial\varphi_1}{\partial x_1} & \cdots & \dfrac{\partial\varphi_m}{\partial x_m}\\ \dfrac{\partial\varphi_m}{\partial x_1} & \cdots & \dfrac{\partial\varphi_m}{\partial x_m}\end{pmatrix}(x^0)\neq 0$$

由隐映射定理, 存在点 $z^0=(x_{m+1}^0,\cdots,x_n^0)$ 的开邻域 V 以及 C^1 映射 $g:V\to\mathbb{R}^m$ 使得

$$y^0=g(z^0),\ \Phi(g(z),z)=0,\ \forall z\in V.$$

其中 $y^0=(x_1^0,\cdots,x_n^0)$, $y=(x_1^0,\cdots,x_m)$, $z=(x_{m+1},\cdots,x_n)$, $x=(y,z)\in U$. 在 $x^0=(y^0,z^0)$ 处求导, 有

$$Jg(z^0)=-(J_y\Phi)^{-1}(x^0)\cdot J_z\Phi(x^0). \tag{8.1}$$

由于 $x^0)$ 为 f 的条件极值点, 故 z^0 为 $f(g)(z), z)$ 的极值点, 在 z_0 处求导, 得

$$J_y f(x^0) \cdot Jg(z^0) + J_z f(x^0) = 0. \tag{8.2}$$

(8.1) 代入 (8.2), 得

$$J_z f(x^0) = J_y f(x^0) \cdot (J_y \Phi)^{-1}(x^0) \cdot J_z \Phi(x^0) \tag{8.3}$$

记 $\lambda = J_y f(x^0) \cdot (J_y \Phi)^{-1}(x^0)$, 则

$$J_y f(x^0) = \lambda \cdot (J_y \Phi)(x^0) \tag{8.4}$$

(8.3) 用 λ 改写为

$$J_z f(x^0) = \lambda \cdot J_z \Phi(x^0). \tag{8.5}$$

(8.4) 和 (8.5) 结合起来 $\Rightarrow$

$$Jf(x^0) - \lambda \cdot J\Phi(x^0) = 0.$$

注 由定理 1 知, 如果 x^0 为条件极值点, 则 (x^0, λ) 为函数

$$F(x, \lambda) = f(x) - \sum_{i=1}^{n} \lambda_i \cdot \varphi_i(x)$$

的驻点.

例 1 求圆周 $(x-1)^2 + y^2 = 1$ 上的点与固定点 $(0,1)$ 的距离的最大值, 最小值.

解 在约束条件 $(x-1)^2 + y^2 - 1 = 0$ 下求函数 $d = \sqrt{x^2 + (y-1)^2}$ 的最大值和最小值. 考虑辅助函数

$$F(x, y, \lambda) = x^2 + (y-1)^2 - \lambda[(x-1)^2 + y^2 - 1]$$

求其驻点:

$$F'_x = F'_y = F'_\lambda = 0 \Leftrightarrow \begin{cases} x + \lambda(x-1) = 0 \\ y - 1 + \lambda y = 0 \\ (x-1)^2 + y^2 - 1 = 0 \end{cases} \Rightarrow \begin{cases} x = \dfrac{\lambda}{1+\lambda} \\ y = \dfrac{1}{1+\lambda} \\ \lambda = \pm\sqrt{2} - 1 \end{cases}$$

在驻点处

$$d^2=\lambda^2,\ \Rightarrow d=|\lambda|$$

由于 d 在圆周上达到最大, 最小值, 故其最大值必为 $\sqrt{2}+1$, 最小值必为 $\sqrt{2}-1$.

例 2 体积为 V 的长方体面积最小值为多少?

解 设长方体长, 宽, 高分别为 x, y, z, 则面积为

$$A=2(xy+yz+zx)$$

体积 $v=xyz$. 在约束条件 $xyz-v=0$ 之下求 A 的最小值.

令

$$F(x,y,\lambda)=2(xy+yz+zx)-\lambda(xyz-V)$$

求驻点:

$$F'_x=F'_y=F'_z=F'_x=0\Rightarrow\begin{cases}2(y+z)-\lambda yz=0\\2(z+x)-\lambda zx=0\\2(x+y)-\lambda xy=0\\xyz-v=0\end{cases}\Rightarrow x=y=z=v^{\frac{1}{3}}.$$

因为当 x, y 或 $z\to 0$ 时, $A\to+\infty$, 故 A 的最小值存在, 从而等于 $\dfrac{2}{3}V^{\frac{2}{3}}$.

例 3 设 $\alpha_i>0$, $x_i>0$, $i=1,\cdots,n$. 证明

$$x_1^{\alpha_1}\cdots x_n^{\alpha_n}\le\left(\frac{\alpha_1x_1+\cdots+\alpha_nx_n}{\alpha_1+\cdots+\alpha_n}\right)^{\alpha_1+\cdots+\alpha_n},$$

等号成立当且仅当 $x_1=\cdots=x_n$.

解 考虑函数

$$f(x_1,\cdots,x_n)=\log(x_1^{\alpha_1}\cdots x_n^{\alpha_n})=\sum_{i=1}^n\alpha_i\cdot\log x_i$$

在约束条件 $\sum\limits_{i=1}^n\alpha_ix_i=c$ $(c>0)$ 下的条件极值. 令

$$F(x,\lambda)=\sum_{i=1}^n\alpha_i\cdot\log x_i-\alpha\left(\sum_{i=1}^n\alpha_i\cdot x_i-c\right),$$

求驻点:

$$\begin{aligned} F'_{x_i} = F'_\lambda = 0 \quad &\Rightarrow \quad \frac{\alpha_i}{x_i} = \lambda \cdot \alpha_i, \quad \sum_{i=1}^{n} \alpha_i \cdot x_i - c = 0 \\ &\Rightarrow \quad x_i = \frac{c}{\sum\limits_{i=1}^{n} \alpha_i}, \quad i = 1, \cdots, n. \end{aligned}$$

因为在集合

$$D : x_i \geq 0, \ \sum_{i=1}^{n} \alpha_i \cdot x_i = c$$

的边界上, f 取值为 $-\infty$, 因此 f 在 D 的内部取到最大值, 上述唯一驻点必为最大值点, 从而

$$\log(x_1^{\alpha_1} \cdots x_n^{\alpha_n}) \leq \sum_{i=1}^{n} \alpha_i \log \frac{c}{\sum_{i=1}^{n} \alpha_i}$$

即

$$x_1^{\alpha_1} \cdots x_n^{\alpha_n} \leq \left(\frac{\alpha_1 x_1 + \cdots + \alpha_n x_n}{\alpha_1 + \cdots + \alpha_n} \right)^{\alpha_1 + \cdots + \alpha_n}.$$