第七章多变量函数和它的极限与连续性

§1. 度量空间和它的重要子集类

[1° 度量空间] 在分析学中，人们关注所研究对象的逼近程度。通常这种逼近程度用所谓的⌊距离⌉来刻画。我们引入一般距离（度量）的概念，它是人们日常生活中两点之间距离的抽象和概括。

定义 1 设集合X上的一个函数：$d: X \times X \to \mathbb{R}$。如果满足如下四个条件：

(1) （非负性）对于任意的$u, v \in X$，$d(u,v) \geq 0$。

(2) （正性）$d(u,v) = 0 \Leftrightarrow (u = v)$。

(3) （对称性）对于任意的$u, v \in X$，$d(u,v) = d(v,u)$。

(4) （三角不等式）对于任意的$u,\ v,\ w \in X$，成立$d(u,w) \leq d(u,v) + d(v,w)$。

则称函数$d(\cdot,\cdot)$为X上的一个距离或度量。集合连同它上面的的距离一起，记作(X,d)，称为度量空间或距离空间。度量空间中的元素通常也称为向量或点。

例1. 在任意的集合X上，对任意的$x, y \in X$，定义函数：

$$d_{disc}(x,y) := \begin{cases} 0, & x = y; \\ 1, & x \neq y. \end{cases}$$

这个函数定义了X上的一个度量，称为X上的离散度量。这个度量的行为比较奇怪。

我们举一些常用度量空间的例子。

例2. 连续函数空间$C^0[a,b]$上，对任意两个连续函数f和g，定义：

$$d_C(f,g) = \max_{x\in[a,b]} |f(x) - g(x)|.$$

这是连续函数空间上的一种度量或距离，它反映两个函数之间一致的接近程度。

例3. Riemann可积函数空间$\mathcal{R}[a,b]$上，对任意两个Riemann可积函数f和g，定义：

$$d_p(f,g) = \Big(\int_a^b |f(x) - g(x)|^p dx \Big)^{\frac{1}{p}}, \quad 1 \leq p < +\infty.$$

这是Riemann可积函数空间上的一种度量或距离，它反映两个函数之间之间一种平均意义的接近程度。

[2° 度量空间$\mathbb{R}^m$] 本章中我们主要讨论具有有限个分量的集合$\mathbb{R}^m$。

定义 2 用$\mathbb{R}^m$表示由m个实数$x^i \in \mathbb{R}$ $(i = 1, 2, \cdots .m)$ 组成的所有有序数组的集合：

$$\mathbb{R}^m = \big\{ \boldsymbol{x} = (x^1, \cdots, x^m) \big| \ x^i \in \mathbb{R}, \quad 1 \leq i \leq m. \big\}$$

其上距离定义为：

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sqrt{\sum_{i=1}^{m} |x_1^i - x_2^i|^2}.$$

我们将有序数组用一个列向量表示，黑体字母$\boldsymbol{x}$表示一个向量，x^i表示向量$\boldsymbol{x}$的第i个坐标。我们采用上标表示向量的第几个分量。下标表示向量的可能的标号。

从距离的定义看出

$$\max_{1\le j\le m}|x_1^j-x_2^j|\le d(\boldsymbol{x}_1,\boldsymbol{x}_2)\le\sqrt{m}\max_{1\le j\le m}|x_1^j-x_2^j|.$$

也就是说，在$\mathbb{R}^m$两点$\boldsymbol{x}_1$和$\boldsymbol{x}_2$之间的距离很小，当且仅当对应的坐标之差很小。

例1. $m=1$时，$\mathbb{R}^1=\mathbb{R}$即为实数集合。在$\mathbb{R}^1$上，$d(x,y)=|x-y|$ 是一个度量。$m=2$时，$\mathbb{R}^2$ 是二维平面集合。$m=3$时，$\mathbb{R}^3$ 为三维立体空间。$m\ge 4$即为抽象立体空间。

例2. 在$\mathbb{R}^m$上可以定义无限多度量。对于任意$1\le p<+\infty$，我们定义：

$$d_p(\boldsymbol{x},\boldsymbol{y})=\Big(\sum_{i=1}^m|x^i-y^i|^p\Big)^{\frac{1}{p}}$$

和

$$d_\infty(\boldsymbol{x},\boldsymbol{y})=\max_{1\le i\le m}|x^i-y^i|.$$

上述函数均构成$\mathbb{R}^m$上的度量。

[**3° 度量空间中的开集与闭集**] 我们在一般度量空间中引入开集和闭集的概念。

定义 3 设(X,d)是一个度量空间。对任意$a\in X$和$\delta>0$，称集合

$$B(a,\delta)=\{x\in X\mid d(a,x)<\delta\}$$

为以a为心，δ为半径的球，或点$a\in X$的δ邻域。

定义 4 设(X,d)是一个度量空间。集合$G\subset X$称为X中的开集，如果对任意$a\in G$，存在一个邻域$B(a,\delta)$，使得$B(a,\delta)\subset G$。

例1. 设(X,d)是一个度量空间。X是X中的开集。

例2. 设(X,d)是一个度量空间。空集$\emptyset$不包含任意点，可以视为满足开集定义，因此空集$\emptyset$是X中的开集。

例3. 设(X,d)是一个度量空间。X中的球$B(a,r)$是X中的开集，因此称它为开球。

例4. 设(X,d)是一个度量空间。集合$G=\{x\in X\mid d(a,x)>r\}$，即$X$中与点$a$距离大于$r$的所有点构成的集合，是$X$中的开集。

定义 5 设(X,d)是一个度量空间。集合$F\subset X$称为X中的闭集，如果F在X 中的余集$X\backslash G$是X中的开集。

例5. 设(X,d)是一个度量空间。$\bar{B}(a,r)=\{x\in X\mid d(a,x)\le r\}$，即$X$中与点$a$距离不大于$r$的所有点构成的集合，是$X$中的闭集。

例6. 设(X,d)是一个度量空间。X和空集$\emptyset$是闭集。这是两个特殊的既是开集也是闭集的集合。

例7. 在度量空间(X,d_{disc})中，对任意$a\in X$，单点集合$\{a\}$既是开集也是闭集。

命题 1 设(X,d)是一个度量空间。下面结论成立：

(a) 如果$\{G_\alpha;\alpha\in A\}$是X中任意多个开集构成的集合族，则它的并集$\bigcup\limits_{\alpha\in A}G_\alpha$是$X$中的开集。

(b) X中有限多个开集$\{G_i\}_{i=1}^n$的交集$\bigcap\limits_{i=1}^n G_i$是$X$中的开集。

(c) 如果$\{F_\alpha; \alpha\in A\}$是$X$中任意多个闭集构成的集合族，则它的交集$\bigcap\limits_{\alpha\in A} G_\alpha$是$X$中的闭集。

(d) X中有限多个闭集$\{F_i\}_{i=1}^n$的并集$\bigcup\limits_{i=1}^n F_i$是$X$中的闭集。

证明：(a) 如果$a\in\bigcup\limits_{\alpha\in A} G_\alpha$，存在$\alpha_0\in A$ 使得$a\in G_{\alpha_0}$。由于G_{α_0}是开集，存在a的一个δ邻域$B(a,\delta)\subset G_{\alpha_0}$。于是$B(a,\delta)\subset\bigcup\limits_{\alpha\in A} G_\alpha$。所以，$\bigcup\limits_{\alpha\in A} G_\alpha$ 是开集。

(b) 如果$a\in\bigcap\limits_{i=1}^n G_i$，则$a\in G_i$, $i=1,2,\cdots,n$。由于每一G_i是开集，存在a的一个δ邻域$B(a,\delta_i)\subset G_i$, $i=1,2,\cdots,n$。取$\delta=\min\{\delta_1,\delta_2,\cdots,\delta_n\}>0$，则$B(a,\delta)\subset G_i$, $i=1,2,\cdots,n$。于是$B(a,\delta)\subset\bigcap\limits_{i=1}^n G_i$。所以，$\bigcup\limits_{i=1}^n G_i$是开集。

(c) 由闭集定义，只需证明$\bigcap\limits_{\alpha\in A} F_\alpha$的余集是开集。由De Morgan律，

$$X\backslash\bigcap_{\alpha\in A} F_\alpha=\bigcup_{\alpha\in A}(X\backslash F_\alpha).$$

即然F_α是闭集，于是$X\backslash F_\alpha$是开集。所以根据(a)知$\bigcup\limits_{\alpha\in A}(X\backslash F_\alpha)$是开集。

(d) 由闭集定义，只需证明$\bigcup\limits_{i=1}^n F_i$的余集是开集。由De Morgan 律，

$$X\backslash\bigcup_{i=1}^n F_i=\bigcap_{i=1}^n(X\backslash F_i).$$

即然F_i是闭集，于是$X\backslash F_i$是开集。所以根据(b)知$\bigcap\limits_{i=1}^n(X\backslash F_i)$是开集。证毕

命题 2 设(X,d)是一个度量空间，F是X中的一个集合。F是X中闭集的充分必要条件是$F=\bar F$。换句话说，F是X中闭集当且仅当F含有它的所有极限点。

证明：（必要性）设F是X中闭集，则其余集$X\backslash F$是开集。于是，对任意$a\notin F$，存在邻域$B(a,\delta)\subset X\backslash F$，此邻域不含$F$中任何点，所以$a$不是$F$的极限点。因此$\bar F\subset F$。又由定义$F\subset\bar F$，只能是$F=\bar F$。

（充分性）设$F=\bar F$。由定义，只需证$G=X\backslash F$是X中开集。对任意$a\in G$，则$a\notin F$。由于$F=\bar F$。因此，a不会是极限点。存在点a的一个邻域$O(a)$，仅包含F中的有限个点$\{x_i\}_{i=1}^n\subset F\cap O(a)$。由于$a\neq x_i$使得$d(a,x_i)>0$，存在$a$的$\delta$邻域$B(a,\delta_i)\subset O(a)$ $(0<\delta_i<d(a,x_i))$，使得$x_i\notin B(a,\delta_i)$。取$\delta=\min\{\delta_1,\delta_2,\cdots,\delta_n\}>0$。则$B(a,\delta)\subset O(a)$不含$F$中任何点，即$B(a,\delta)\subset X\backslash F$。这证明了$G$ 是开集。证毕

§2. 多变量函数的极限

[**4° 多变量函数的极限定义**] 本节中我们考察定义在集合X上而值域在$\mathbb{R}^n$上的函数$f:X\to\mathbb{R}$的极限理论。我们从一般定义开始。

定义 6 设X是一个集合，$\mathcal{B}$是X中的一族基。点$A\in\mathbb{R}^n$称为映射$f:X\to\mathbb{R}^n$关于基$\mathcal{B}$的极限，如果对于A的任一邻域$V(A)$，存在$B\in\mathcal{B}$，使得$f(B)\subset V(A)$。简言之，

$$\lim_{\mathcal{B}} f(x)=A:=(\forall V(A)\exists B\in\mathcal{B}(f(B)\subset V(A)).$$

或写作

$$(\lim_{\mathcal{B}} f(x) = A \in \mathbb{R}^n) := (\forall \epsilon > 0 \exists B \in \mathcal{B}(\forall x \in B(d(f(x), A) < \epsilon))).$$

或写作

$$(\lim_{\mathcal{B}} f(x) = A \in \mathbb{R}^n) := (\lim_{\mathcal{B}} d(f(x), A) = 0).$$

定义 7 映射$f : X \to \mathbb{R}^n$称为有界映射，如果$f(X) \subset \mathbb{R}^n$在$\mathbb{R}^n$中有界。

定义 8 设集合$\mathcal{B}$是集合X中的基，映射$f : X \to \mathbb{R}^n$ 称为是对基$\mathcal{B}$最终有界映射，如果存在$B \in \mathcal{B}$，使得f在B上有界。

命题 3 设$\mathcal{B}$是X的基和函数$f : X \to \mathbb{R}^n$。成立结论：

(a) 函数f关于基$\mathcal{B}$至多有一个极限。

(b) 若函数f关于基$\mathcal{B}$有极限，则f对于基$\mathcal{B}$是最终有界的。

[**5° 映射$\boldsymbol{f} : X \to \mathbb{R}^n$的特征**] 映射映射$\boldsymbol{f} : X \to \mathbb{R}^n$本质上是定义在集合$X$上的向量值函数：

$$\boldsymbol{f}(x) = (f^1(x), f^2(x), \cdots, f^n(x)) \qquad f^i(x) : X \to \mathbb{R}, \quad 1 \le i \le n.$$

从中看出，

$$\lim_{\mathcal{B}} \boldsymbol{f}(x) = \boldsymbol{A} \Leftrightarrow \lim_{\mathcal{B}} f^i(x) = A^i \quad (i = 1, 2, \cdots, n).$$

也就是说，函数$\boldsymbol{f} : X \to \mathbb{R}^n$在$\mathbb{R}$的收敛是按坐标收敛。

[**6° Cauchy收敛准则**] 设$X = \mathbb{N}$，函数$\boldsymbol{f} : \mathbb{N} \to \mathbb{R}^n$ 为：

$$\boldsymbol{f} = (y_k^1, y_k^2, \cdots, y_k^n) \qquad k = 1, 2, \cdots\cdots.$$

函数$\boldsymbol{f} : \mathbb{N} \to \mathbb{R}^n$也可以视为$\mathbb{R}^n$中的点列$\{\boldsymbol{y}_k\}_{k=1}^{\infty}$。

定义 9 $\mathbb{R}^n$中点列$\{\boldsymbol{y}_k\}_{k=1}^{\infty}$称为基本点列（也称为*Cauchy* 点列），如果对任意$\epsilon > 0$，存在$N \in \mathbb{N}$，使得对任意$k_1, k_2 \ge N$，必有$d(\boldsymbol{y}_{k_1}, \boldsymbol{y}_{k_2}) < \epsilon$。

显而易见，$\mathbb{R}^n$中点列$\{\boldsymbol{y}_k\}_{k=1}^{\infty}$时基本点列当且仅当它的每一坐标列$\{y_k^i\}_{k=1}^{\infty}$ $(1 \le i \le n)$是基本列。由数列的Cauchy收敛准则既得$\mathbb{R}^n$中的点列收敛当且仅当它是$\mathbb{R}^n$中的基本列（或Cauchy点列）。换句话说，Cauchy收敛准则在$\mathbb{R}$中是正确的。

定义 10 一个度量空间，如果其中每个基本列都在该空间有极限，则称这个度量空间为完备度量空间。

命题 4 对任意$n \in \mathbb{N}$，度量空间$\mathbb{R}^n$是完备度量空间。

定义 11 设$E \subset X$和映射$\boldsymbol{f} : X \to \mathbb{R}^n$，量

$$\omega(f; E) := d(\boldsymbol{f}(E))$$

称为函数$\boldsymbol{f}$在E上的振幅，其中$d(\boldsymbol{f}(E))$是集合$\boldsymbol{f}(E) \subset \mathbb{R}^n$的直径。

与$\mathbb{R}^n$的完备性相关，成立下面的Cauchy收敛准则。

定理 1 设X是一个集合，$\mathcal{B}$是X中的基，函数$\boldsymbol{f}: X \to \mathbb{R}^n$关于$\mathcal{B}$有极限，当且仅当对于任意$\epsilon > 0$，存在$B \in \mathcal{B}$，使得$\omega(\boldsymbol{f}; B) < \epsilon$，也就是

$$\exists \lim_{\mathcal{B}} \boldsymbol{f}(x) \Leftrightarrow \forall \epsilon > 0 \exists B \in \mathcal{B}(\omega(\boldsymbol{f}; B) < \epsilon).$$

对于函数值在$\mathbb{R}^n$中的函数，关于复合函数极限的重要定理也成立。

定理 2 设Y是一个集合，$\mathcal{B}_Y$是Y中的基，映射$g: Y \to \mathbb{R}^n$关于$\mathcal{B}_Y$有极限。

设X是又一个集合，$\mathcal{B}_X$是X中的基，$f: X \to Y$是从X到Y 的映射，而且对于任意的$B_Y \in \mathcal{B}_Y$，存在$B_X \in \mathcal{B}_X$，使得$f(B_X) \subset B_Y$。

在这些条件下，映射f和$\boldsymbol{g}$的复合映射$\boldsymbol{g} \circ f: X \to \mathbb{R}^n$关于$\mathcal{B}$有极限，且

$$\boxed{\lim_{\mathcal{B}_X} (\boldsymbol{g} \circ f)(x) = \lim_{\mathcal{B}_Y} \boldsymbol{g}(y).} \tag{1}$$

[**7° 函数**$f: \mathbb{R}^m \to \mathbb{R}^n$**的极限**] 我们首先感兴趣的是$X$为空间$\mathbb{R}^m$的子空间的情形。

我们约定：

$U(a)$ —— 点$a \in \mathbb{R}^m$的邻域；

$\mathring{U}(a)$ —— 点$a \in \mathbb{R}^m$的去心邻域，即$\mathring{U}(a) := U(a) \backslash a$；

$U_E(a)$ —— 点a在E的邻域，即$U_E(a) := U(a) \cap E$；

$\mathring{U}_E(a)$ —— 点a在E的去心邻域，即$\mathring{U}_E(a) := \mathring{U}(a) \cap E$；

$x \to a$ —— 点a的去心邻域组成的$\mathbb{R}^m$中的基。

$x \to \infty$ —— 无穷远点邻域组成的基，即由集合$\mathbb{R}^m \backslash B(0, r)$组成的基。

$x \to a$，$x \in E$或$(E \ni x \to a)$ —— 点a在E 中的去心邻域组成的基，且a是E的极限点。

$x \to \infty$，$x \in E$或$(E \ni x \to \infty)$ —— 无穷远点在E中的邻域组成的基，即由集合$E \backslash B(0, r)$组成的基，且E是无界集。

利用上面符号，可使各类极限具体化。

$$(\lim_{E \ni x \to a} f(x) = A) := (\forall \epsilon > 0 \exists \mathring{U}(a) \forall x \in \mathring{U}_E(a)(d(f(x), A) < \epsilon).$$

等价于

$$(\lim_{E \ni x \to a} f(x) = A) := (\forall \epsilon > 0 \exists \delta > 0 \forall x \in E(0 < d(x, a) < \delta \Rightarrow d(f(x), A) < \epsilon)),$$

其中$d(x, a)$和$d(f(x), A)$分别表示$\mathbb{R}^m$和$\mathbb{R}^n$的度量。

$$(\lim_{x \to \infty} f(x) = A) := (\forall \epsilon > 0 \exists B(0, r) \forall x \in \mathbb{R}^m \backslash B(0, r)(d(f(x), A) < \epsilon)).$$

约定

$$(\lim_{\mathcal{B}} f(x) = \infty) := (\forall B(0, r) \exists B \in \mathcal{B}(f(B) \subset \mathbb{R}^n \backslash B(0, r))).$$

设映射$\pi: \mathbb{R}^m \to \mathbb{R}$，$\pi^(x) = x^i$，$i = 1, 2, \cdots, m$。如果$\boldsymbol{x} = (x^1, x^2, \cdots, x^m) \to \boldsymbol{a} = (a^1, a^2, \cdots, a^m)$，则$\pi^i(\boldsymbol{x}) \to a^i$，$i = 1, 2, \cdots, m$。如果$m > 1$，当$\boldsymbol{x} \to \infty$ 时，有

$$\sum_{i=1}^m (\pi^i(\boldsymbol{x}))^2 \to \infty,$$

但$\pi^i(\boldsymbol{x})$可能无极限，也可能不趋于无限远。

[**8° 重极限与累次极限**] 对于多元函数而言，由于自变量维度的增加，其极限的情况非常复杂。首先自变量趋于某个极限的方式和路径多样化。一个最简单的方式是自变量按某种顺序依次取极限，记为

$$\lim_{z^{i_1}\to x^{i_1}}\lim_{z^{i_2}\to x^{i_2}}\cdots\lim_{z^{i_m}\to x^{i_m}} f(z^1,\cdots,z^m),$$

这里$i_1,i_2,\cdots,i_m$是$1,2,\cdots,m$的某个排列。这种极限方式，我们称之为累次极限，与此对应，原始的极限称为重极限，记为

$$\lim_{(z^1,\cdots,z^m)\to(x^1,\cdots,x^m)} f(z^1,\cdots,z^m),$$

我们通过以下例子来看取极限的方式不同，可能产生的结果是否相同。

例1. 设函数$f:\mathbb{R}^2\to\mathbb{R}$为

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

则$f(0,y)=f(x,0)=0$，即$\lim\limits_{y\to0}\lim\limits_{x\to0} f(x,y)=\lim\limits_{x\to0}\lim\limits_{y\to0} f(x,y)=0$。而当$y=ax\neq0$时，$f(x,ax)=\dfrac{a}{1+a^2}$。于是当$(x,y)\to(0,0)$时，这个函数的两个累次极限存在且相等，但二重极限不存在。

例2. 设函数$f:\mathbb{R}^2\to\mathbb{R}$为

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

有

$$\lim_{y\to0}\lim_{x\to0} f(x,y)=-1,\quad \lim_{x\to0}\lim_{y\to0} f(x,y)=1.$$

而当$y=ax\neq0$时，$f(x,ax)=\dfrac{1-a^2}{1+a^2}$。于是当$(x,y)\to(0,0)$时，这个函数的两个累次极限存在但不相等，二重极限也不存在。

例3. 设函数$f:\mathbb{R}^2\to\mathbb{R}$为

$$f(x,y)=\begin{cases}x+y\sin\dfrac{1}{x}, & x\neq 0;\\ 0, & x=0.\end{cases}$$

有

$$\lim_{(x,y)\to(0,0)}=0,\quad \lim_{x\to0}\lim_{y\to0} f(x,y)=0.$$

于是当$(x,y)\to(0,0)$时，这个函数的二重极限和累次极限$\lim\limits_{x\to0}\lim\limits_{y\to0} f(x,y)$存在，但累次极限$\lim\limits_{y\to0}\lim\limits_{x\to0} f(x,y)$不存在。

例4. 设函数$f:\mathbb{R}^2\to\mathbb{R}$为

$$f(x,y)=\begin{cases}x\sin\dfrac{1}{y}+y\sin\dfrac{1}{x}, & xy\neq 0;\\ 0, & xy=0.\end{cases}$$

既然

$$\left|x\sin\frac{1}{y}+y\sin\frac{1}{x}\right|\le|x|+|y|,$$

于是成立

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

所以当$(x,y)\to(0,0)$时，这个函数的二重极限存在，但两个累次极限均不存在。

上面例子表明，在一般情形，多元函数的重极限与累次极限没有确定关系。但在一定条件下，多元函数的重极限和累次极限存在一定的关联性。

定理 3 若函数$f:\mathbb{R}^m\to\mathbb{R}$在$(x^1,\cdots,x^m)$存在重极限

$$\lim_{(z^1,\cdots,z^m)\to(x^1,\cdots,x^m)} f(z^1,\cdots,z^m)$$

与累次极限

$$\lim_{z^{i_1}\to x^{i_1}}\lim_{z^{i_2}\to x^{i_2}}\cdots\lim_{z^{i_m}\to x^{i_m}} f(z^1,\cdots,z^m),$$

这里$i_1,i_2,\cdots,i_m$是$1,2,\cdots,m$的某个排列，则重极限与累次极限必相等。

证明：设$\lim\limits_{(z^1,\cdots,z^m)\to(x^1,\cdots,x^m)} f(z^1,\cdots,z^m)=A$。对任意$\epsilon>0$，存在$\delta>0$，只要$0<\sqrt{\sum\limits_{i=1}^{m}|z^i-x^i|^2}<\delta$，必有

$$|f(z^1,\cdots,z^m)-A|<\epsilon.$$

由定理条件$\lim\limits_{z^{i_1}\to x^{i_1}}\lim\limits_{z^{i_2}\to x^{i_2}}\cdots\lim\limits_{z^{i_m}\to x^{i_m}} f(z^1,\cdots,z^m)$存在，则依次取极限得

$$|\lim_{z^{i_1}\to x^{i_1}}\lim_{z^{i_2}\to x^{i_2}}\cdots\lim_{z^{i_m}\to x^{i_m}} f(z^1,\cdots,z^m)-A|\le\epsilon.$$

由ϵ的任意性，证明了$\lim\limits_{z^{i_1}\to x^{i_1}}\lim\limits_{z^{i_2}\to x^{i_2}}\cdots\lim\limits_{z^{i_m}\to x^{i_m}} f(z^1,\cdots,z^m)=A$。证毕

推论 1 设函数f在点$(x^1,\cdots,x^m)$的重极限存在。如果在该点两个累次极限存在，则这两个累次极限相等。

特别，在重极限和所有累次极限都存在的条件下，累次极限的求极限次序可交换。这对于极限的计算是非常有用的，有时对于某种顺序的累次极限容易计算。

推论 2 若函数f在点$(x^1,\cdots,x^m)$的两个不同的累次极限均存在但不相等，则函数f 在该点的重极限必不存在。

本推论可用来判断重极限的不存在性。

定理3及其推论告诉我们，如果重极限和累次极限同时存在，则它们必相等。但不能由重极限存在，确定累次极限存在。反之亦然。一般讲，这是一个困难的问题。下面的定理在二元函数情形给出了一些有关结果。

定理 4 如果函数f在点(x_0,y_0)处的二重极限存在，且在y_0的某个邻域内，$\lim\limits_{x\to x_0} f(x,y)=\varphi(y)$存在，则累次极限$\lim\limits_{y\to y_0}\lim\limits_{x\to x_0} f(x,y)$存在，且$\lim\limits_{y\to y_0}\lim\limits_{x\to x_0} f(x,y)=\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$.

反之，如果在y_0的某个邻域内，$f(x,y)$一致收敛到$\varphi(y)$，且累次极限$\lim\limits_{y\to y_0}\lim\limits_{x\to x_0} f(x,y)$存在，则二重极限$\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$存在，且$\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)=\lim\limits_{y\to y_0}\lim\limits_{x\to x_0} f(x,y)$。

证明：设$\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)=A$。对任意$\epsilon>0$，存在$\delta>0$，只要$0<\sqrt{|x-x_0|^2+|y-y_0|^2}<\delta$，必有

$$|f(x,y)-A|<\frac{\epsilon}{2}.$$

由定理条件$\lim\limits_{x\to x_0} f(x,y)=\varphi(y)$存在，则令$x\to x_0$，

$$|\varphi(y)-A|\le\frac{\epsilon}{2}<\epsilon.$$

这表明：$\lim\limits_{y\to y_0}\lim\limits_{x\to x_0} f(x,y)=A$。

反之，设$f(x,y)$在y_0的某个邻域内，一致收敛到$\varphi(y)$，且累次极限$\lim\limits_{y\to y_0}\lim\limits_{x\to x_0} f(x,y)=A$，则对任意$\epsilon>0$，存在$y_0$的一个邻域$B(y_0,b)$ $(b>0)$和$\delta_1>0$，对任意$y\in B(y_0,b)$ 和$0<|x-x_0|<\delta$，必有$|f(x,y)-\varphi(y)|<\epsilon/2$和存在$\delta_2>0$，只要$0<|y-y_0|<\delta_2$，必有$|\varphi(y)-A|<\epsilon/2$。取$\delta=\min\{\delta_1,\delta_2,b\}$，则当$0<\sqrt{|x-x_0|^2+|y-y_0|^2}<\delta$时，

$$|f(x,y)-A|\le|f(x,y)-\varphi(y)|+|\varphi(y)-A|<\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

这证明了$\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)=A$。证毕

[9° 多变量连续函数及其性质] 设E是$\mathbb{R}^m$中的集合。

定义 12 函数$\boldsymbol{f}:E\to\mathbb{R}^n$称为在点$\boldsymbol{a}\in E$是连续的，如果对于函数值$\boldsymbol{f}(\boldsymbol{a})$的任意邻域$V(\boldsymbol{f}(\boldsymbol{a})$，存在点$\boldsymbol{a}$在$E$中的邻域$U_E(\boldsymbol{a})$，使得$\boldsymbol{f}(U_E(\boldsymbol{a}))\subset V(\boldsymbol{f}(\boldsymbol{a}))$，即

$$(\boldsymbol{f}:E\to\mathbb{R}^n\text{在}\boldsymbol{a}\in E\text{连续}):=(\forall V(\boldsymbol{f}(\boldsymbol{a}))\exists U_E(\boldsymbol{a})(\boldsymbol{f}(U_E(\boldsymbol{a}))\subset V(\boldsymbol{f}(\boldsymbol{a})))),$$

等价于

$$(\boldsymbol{f}:E\to\mathbb{R}^n\text{在}\boldsymbol{a}\in E\text{连续}):=(\forall\epsilon>0\ \exists\delta>0\ \forall\ \boldsymbol{x}\in E(d(\boldsymbol{x},\boldsymbol{a})<\delta\Rightarrow d(\boldsymbol{f}(\boldsymbol{x}),\boldsymbol{f}(\boldsymbol{a}))<\epsilon)).$$

也可写成

$$(\boldsymbol{f}:E\to\mathbb{R}^n\text{在}\boldsymbol{a}\in E\text{连续}):=(\lim_{E\ni\boldsymbol{x}\to\boldsymbol{a}}\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{f}(\boldsymbol{a})).$$

命题 5 函数$\boldsymbol{f}:E\to\mathbb{R}^n$，$\boldsymbol{f}=(f^1,\cdots,f^n)$，在某点连续当且仅当每个分量函数$f^i$，$i=1,2,\cdots,n$，在该点连续。

定义 13 设$E\subset\mathbb{R}^m,\boldsymbol{a}\in E,B_E(\boldsymbol{a};r):=E\cap B(\boldsymbol{a};r)$。量

$$\omega(\boldsymbol{f};\boldsymbol{a}):=\lim_{r\to0^+}\omega(\boldsymbol{f};B_E(\boldsymbol{a};r))$$

称为函数$\boldsymbol{f}:E\to\mathbb{R}^n$在点$\boldsymbol{a}\in E$的振幅。

命题 6 （连续函数的局部性质）

(a) 函数$\boldsymbol{f}:E\to\mathbb{R}^n$在点$\boldsymbol{a}\in E$连续当且仅当$\omega(\boldsymbol{f};\boldsymbol{a})=0$.

(b) 函数$\boldsymbol{f}:E\to\mathbb{R}^n$在点$\boldsymbol{a}\in E$连续，则在此点的某个邻域$U_E(\boldsymbol{a})$上有界。

(c) 若集合$Y\subset\mathbb{R}^n$上的映射$\boldsymbol{g}:Y\to\mathbb{R}^k$在点$\boldsymbol{y}_0\in Y$连续，集合$X\subset\mathbb{R}^m$上的映射$\boldsymbol{f}:X\to Y$在点$\boldsymbol{x}_0\in X$ 连续，且$\boldsymbol{f}(\boldsymbol{x}_0)=\boldsymbol{y}_0$，则复合映射$\boldsymbol{g}\circ\boldsymbol{f}:X\to\mathbb{R}^k$存在且在点$\boldsymbol{x}_0\in X$ 连续。

(d) 函数$f:E\to\mathbb{R}$在点$\boldsymbol{a}\in E$连续且$f(\boldsymbol{a})>0$ $(f(\boldsymbol{a})<0)$，则存在点$\boldsymbol{a}$在E中的某个邻域$U_E(\boldsymbol{a})$，使得对任意$\boldsymbol{x}\in U_E(\boldsymbol{a})$，$f(\boldsymbol{x})>0$ $(f(\boldsymbol{x})<0)$。

(e) 若函数$f:E\to\mathbb{R}$和$g:E\to\mathbb{R}$均在点$\boldsymbol{a}\in E$连续，则它们的线性组合$\alpha f+\beta g:X\to\mathbb{R}$，其中$\alpha,\beta\in\mathbb{R}$，乘积$f\cdot g:X\to\mathbb{R}$和商$\dfrac{f}{g}:X\to\mathbb{R}$, 其中当$\boldsymbol{x}\in E$时$g\neq0$，均在点$\boldsymbol{a}\in E$连续。

下面转入连续函数的全局性质研究。引入几个概念。

定义 14 如果函数$\boldsymbol{f}: E \to \mathbb{R}^n$在$E$中的每一点连续，则称函数$\boldsymbol{f}$ 在E 上连续。

若函数$\boldsymbol{f}: E \to \mathbb{R}^n$在$E$上连续，$\tilde{E}$是$E$的一个子集，那么函数$\boldsymbol{f}$在子集$\tilde{E}$上的限制$\boldsymbol{f}|_{\tilde{E}}$在$\tilde{E}$上连续。

定义 15 从集合$E \subset \mathbb{R}^m$到$\mathbb{R}^n$的映射$\boldsymbol{f}: E \to \mathbb{R}^n$称为在$E$上一致连续映射，如果对任意$\epsilon > 0$，存在$\delta > 0$，对任意$\boldsymbol{x}_1, \boldsymbol{x}_2 \in E$，只要$d(\boldsymbol{x}_1, \boldsymbol{x}_2) < \delta$，必有$d(\boldsymbol{f}(\boldsymbol{x}_1), \boldsymbol{f}(\boldsymbol{x}_2)) < \epsilon$。

定义 16 设$I \subset \mathbb{R}$是区间，函数$\Gamma^1(t), \cdots, \Gamma^n(t)$在$I$ 上连续，称由

$$t \mapsto \boldsymbol{\Gamma} = (\Gamma^1(t), \cdots, \Gamma^n(t))$$

定义的映射$\boldsymbol{\Gamma}: I \to \mathbb{R}^n$为$\mathbb{R}^n$中的道路。

定义 17 集合$E \subset \mathbb{R}^n$称为路连通的，如果对于E中任意一对点$\boldsymbol{x}_0$与$\boldsymbol{x}_1$，总存在一条以$\boldsymbol{x}_0$与$\boldsymbol{x}_1$为端点的道路$\boldsymbol{\Gamma}: I \to E$，并且它的承载子在$E$ 中。换句话说，从E中任一点$\boldsymbol{x}_0$出发不超出E的范围可以到达E中另外的任一点$\boldsymbol{x}_1$。

定义 18 称空间$\mathbb{R}^m$中的路连通开集为$\mathbb{R}^m$中的区域。

例1. $\mathbb{R}^n$中的球$B(a,r)$ $(r > 0)$时区域。

例2. 由方程$(x^1)^2 + (x^2)^2 = r^2$给出的圆周（一维球面）是$\mathbb{R}^2$中的路连通子集，但不是$\mathbb{R}^2$中的区域。

命题 7 （连续函数的整体性质）设$K \subset \mathbb{R}^m$。

(a) 如果K是紧集且函数$\boldsymbol{f}: K \to \mathbb{R}^n$在$K$上连续，则它在$K$上一致连续。

(b) 如果K是紧集且函数$\boldsymbol{f}: K \to \mathbb{R}^n$在$K$上连续，则它在$K$上有界。

(c) 如果K是紧集且函数$f: K \to \mathbb{R}$在K上连续，则它在K上取到最大值和最小值。

(d) 如果K是路连通集且函数$f: K \to \mathbb{R}$在K上连续，存在$a, b \in K$使$f(a) = A$和$f(b) = B$，则对A和B之间的任意值C，存在$c \in K$，使得$f(c) = C$。

第八章多变量函数微分学

§1. $\mathbb{R}^m$的线性结构

[1° 向量空间$\mathbb{R}^m$]

在集合$\mathbb{R}^m$中对于任意两个元素$\boldsymbol{x}_1=(x_1^1,\cdots,x_1^m)$与$\boldsymbol{x}_2=(x_2^1,\cdots,x_2^m)$引入加法

$$\boldsymbol{x}_1+\boldsymbol{x}_2=(x_1^1+x_2^1,\cdots,x_1^m+x_2^m);$$

对于任意元素$\boldsymbol{x}=(x^1,\cdots,x^m)$与实数$\lambda\in\mathbb{R}$定义数乘运算：

$$\lambda\boldsymbol{x}=(\lambda x^1,\cdots,\lambda x^m).$$

那么$\mathbb{R}^m$便成为实数域$\mathbb{R}$上的线性空间。它的元素可以称为点或向量。

$\mathbb{R}^m$中向量组

$$\boldsymbol{e}_i=(0,\cdots,0,\underset{i}{1},0,\cdots,0)\quad i=1,2,\cdots m;$$

其中数1仅在第i个位置出现，组成$\mathbb{R}^m$的极大线性无关组，称为$\mathbb{R}^m$的一个基底，因此称$\mathbb{R}^m$为m 维向量空间。

$\mathbb{R}^m$中任意向量$\boldsymbol{x}$可以按照基底$\{\boldsymbol{e}_j\}_{j=1}^m$展开

$$\boldsymbol{x}=x^1\boldsymbol{e}_1+\cdots+x^m\boldsymbol{e}_m=\sum_{i=1}^m x^i\boldsymbol{e}_i=x^i\boldsymbol{e}_i,$$

这里记$\sum\limits_{i=1}^m x^i\boldsymbol{e}_i=x^i\boldsymbol{e}_i$称为Einstein约定：同一个指标在上方和下方均出现，就意味着关于这个指标在它的变化范围内求和。

[2° 线性映射] 引入一般的线性变换定义。

定义 19 设X和Y是两个集合，映射：$L:X\to Y$称为线性的，如果对任意的$x_1,x_2\in X$和$\lambda_1,\lambda_2\in\mathbb{R}$成立

$$L(\lambda_1x_1+\lambda_2x_2)=\lambda_1L(x_1)+\lambda_2L(x_2).$$

本章中，我们关心从$\mathbb{R}^m$到$\mathbb{R}^n$的线性映射$\mathcal{L}:\mathbb{R}^m\to\mathbb{R}^n$。注意到：$\mathcal{L}(\boldsymbol{x})=(L^1(\boldsymbol{x}),\cdots,L^n(\boldsymbol{x}))$，于是映射$\mathcal{L}:\mathbb{R}^m\to\mathbb{R}^n$是线性映射当且仅当每个$L^i:\mathbb{R}^m\to\mathbb{R},\ i=1,2,\cdots,n$，是线性映射。

我们考察线性映射$\mathcal{L}$ 的表示形式。设$\{\boldsymbol{e}_i\}_{i=1}^m$和$\{\tilde{\boldsymbol{e}}_j\}_{j=1}^n$分别是$\mathbb{R}^m$和$\mathbb{R}^n$的基底。设

$$\mathcal{L}(\boldsymbol{e}_i)=a_i^1\tilde{\boldsymbol{e}}_1+\cdots+a_i^n\tilde{\boldsymbol{e}}_m=a_i^j\tilde{\boldsymbol{e}}_j,\quad i=1,2,\cdots,m;$$

即已知基底$\{\boldsymbol{e}_i\}_{i=1}^m$在线性映射$\mathcal{L}:\mathbb{R}^m\to\mathbb{R}^n$ 在$\mathbb{R}^n$中的像，由映射$\mathcal{L}$的线性性，对于$\mathbb{R}^m$中任意向量$\boldsymbol{h}=(h^1,\cdots,h^m)$，成立

$$\boxed{\mathcal{L}(\boldsymbol{h})=L(h^i\boldsymbol{e}_i)=h^i\mathcal{L}(\boldsymbol{e}_i)=h^ia_i^j\tilde{\boldsymbol{e}}_j,}\tag{2}$$

写成坐标形式

$$\boxed{\mathcal{L}(\boldsymbol{h})=(L^1(\boldsymbol{h}),\cdots,L^n(\boldsymbol{h}))=(a_i^1h^i,\cdots,a_i^nh^i).}\tag{3}$$

命题 8 线性映射$\mathcal{L}:\mathbb{R}^m\to\mathbb{R}^n$ 可以表示成矩阵形式：

$$\boxed{\mathcal{L}(\boldsymbol{h})=\begin{pmatrix}L^1(\boldsymbol{h})\\ \vdots\\ L^n(\boldsymbol{h})\end{pmatrix}=\begin{pmatrix}a_1^1&\cdots&a_m^1\\ \vdots&&\vdots\\ a_1^n&\cdots&a_m^n\end{pmatrix}\begin{pmatrix}h^1\\ \vdots\\ h^m\end{pmatrix}=A\boldsymbol{h}} \tag{4}$$

这里A是$n\times m$矩阵。反之，对任意$n\times m$矩阵A，$A\boldsymbol{h}$定义了$\mathbb{R}^m$到$\mathbb{R}^n$的一个线性映射。

本命题表明，在给定$\mathbb{R}^m$和$\mathbb{R}^n$的基底后，可以建立线性映射$\mathcal{L}:\mathbb{R}^m\to\mathbb{R}^n$到$n\times m$阶矩阵的一一对应。

对任意线性映射$\mathcal{A}:\mathbb{R}^m\to\mathbb{R}^n$和$\mathcal{B}:\mathbb{R}^m\to\mathbb{R}^n$，定义线性组合：

$$(\alpha\mathcal{A}+\beta\mathcal{B})(\boldsymbol{x})=\alpha\mathcal{A}(\boldsymbol{x})+\beta\mathcal{B}(\boldsymbol{x}).$$

$\alpha\mathcal{A}+\beta\mathcal{B}$依然是$\mathbb{R}^m$到$\mathbb{R}^n$ 的线性映射。因此$\mathbb{R}^m$到$\mathbb{R}^n$的线性映射全体组成的集合也是线性空间。$\alpha\mathcal{A}+\beta\mathcal{B}$对应的矩阵恰为映射$\mathcal{A}$和$\mathcal{B}$对应矩阵$A$和$B$的线性组合$\alpha A+\beta B$。

对任意线性映射$\mathcal{A}:\mathbb{R}^m\to\mathbb{R}^n$和$\mathcal{B}_2:\mathbb{R}^n\to\mathbb{R}^k$，定义复合映射：

$$(\mathcal{B}\circ\mathcal{A})(\boldsymbol{x})=\mathcal{B}(\mathcal{A}(\boldsymbol{x})).$$

它是$\mathcal{A}:\mathbb{R}^m$到$\mathbb{R}^k$的线性映射，对应矩阵为$C=BA$，这里BA是矩阵B 与A的乘积。

[3° 赋范线性空间]

定义 20 设集合V是一个线性空间，一个函数$\|\cdot\|:V\to\mathbb{R}$，如果满足下面四个条件：

(1) （非负性）对任意$v\in V$，$\|v\|\geq 0$;

(2) （正定性）$(\|v\|=0)\Leftrightarrow(v=0)$;

(3) （齐次性）对任意$v\in V$和$\lambda\in\mathbb{R}$，$\|\lambda v\|=|\lambda|\|v\|$;

(4) （三角不等式）对任意$u,\ v\in V$，$\|u+v\|\leq\|u\|+\|v\|$;

则称$\|\cdot\|$为V上的一个范数。线性空间V连同其上的范数$\|\cdot\|$一起，$(V,\|\cdot\|)$，称为一个赋范线性空间。

三角不等式可以推广到任意有限个向量相加的情形：

$$\|v_1+\cdots+v_k\|\leq\|v_1\|+\cdots+\|v_k\|.$$

设$(V,\|\cdot\|)$是一个赋范线性空间，定义

$$d(\boldsymbol{x},\boldsymbol{y}):=\|\boldsymbol{x}-\boldsymbol{y}\|.$$

则$d(\cdot,\cdot)$满足度量的定义，称为由范数导出的度量或距离。在此意义下，一个赋范线性空间必是一个度量空间。

在$\mathbb{R}^m$上，定义函数

$$\|\boldsymbol{x}\|=\sqrt{(x^1)^2+\cdots+(x^m)^2},\ \ \forall\ \boldsymbol{x}=(x^1,\cdots,x^m)\in\mathbb{R}^m.$$

则$\|\boldsymbol{x}\|$是$\mathbb{R}^m$上的一个范数。在这个范数意义下，$\mathbb{R}^m$是一个赋范线性空间。

上述范数依赖于空间的定义，为区别或明确空间的差别，在范数的记号右下角标记空间的记号，例如$\|\cdot\|_{R^m}$。但在空间及其范数的含义在行文中非常明确时，一般不这样做。

对任意$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^m$，定义$d(\boldsymbol{x},\boldsymbol{y}):=\|\boldsymbol{x}-\boldsymbol{y}\|$。恰为$\mathbb{R}^m$中两点之间的距离。

当$\boldsymbol{x}\to\boldsymbol{x}_0$时，条件$d(\boldsymbol{x},\boldsymbol{x}_0)\to 0$与条件$\|\boldsymbol{x}-\boldsymbol{x}_0\|\to 0$等价。特别地，$\|\boldsymbol{x}\|=d(\boldsymbol{0},\boldsymbol{x})$。

范数的存在，可以比较函数之间的值。

定义 21 设函数$f:X\to\mathbb{R}^m$与函数$g:X\to\mathbb{R}^n$。$\mathcal{B}$ 是X中的基。如果

$$\lim_{\mathcal{B}}\frac{\|f(x)\|_{R^m}}{\|g(x)\|_{R^n}}=0.$$

即在$\mathcal{B}$下$\|f(x)\|_{R^m}=o(\|g(x)\|_{R^n})$，则记作$f(x)=o(g(x))$，或$f=o(g)$。

定义 22 设函数$f:X\to\mathbb{R}^m$与函数$g:X\to\mathbb{R}^n$。$\mathcal{B}$ 是X中的基。如果$\|f(x)\|_{R^m}/\|g(x)\|_{R^n}$对于基$\mathcal{B}$最终有界，则记作$f(x)=O(g(x))$，或$f=O(g)$。

如果将映射$f:X\to\mathbb{R}^m$写成坐标形式$f(x)=(f^1(x),\cdots,f^m(x))$，则有结论：

$$\boxed{(\text{对于基}\mathcal{B},\ f=o(g))\Leftrightarrow(\text{对于基}\mathcal{B},\ f^i=o(g),\ i=1,\cdots,m)}\tag{5}$$

$$\boxed{(\text{对于基}\mathcal{B},\ f=O(g))\Leftrightarrow(\text{对于基}\mathcal{B},\ f^i=O(g),\ i=1,\cdots,m)}\tag{6}$$

命题 9 设$\mathcal{L}:\mathbb{R}^m\to\mathbb{R}^n$是线性映射，$\boldsymbol{h}=x^1\boldsymbol{e}_1+\cdots+x^m\boldsymbol{e}_m$是$\mathbb{R}^m$中任一向量，则

$$\boxed{\|\mathcal{L}(\boldsymbol{h})\|_{R^n}\le\sum_{i=1}^m\|\mathcal{L}(\boldsymbol{e}_i)\||h^i|\le\Big(\sum_{i=1}^m\|\mathcal{L}(\boldsymbol{e}_i)\|\Big)\|\boldsymbol{h}\|.}\tag{7}$$

命题 10 线性映射$\mathcal{L}:\mathbb{R}^m\to\mathbb{R}^n$在任一点$\boldsymbol{x}_0\in\mathbb{R}^m$是一致连续的。

证明：对任意$\boldsymbol{x},\boldsymbol{x}_0\in\mathbb{R}^m$，由上面命题得：

$$\|\mathcal{L}(\boldsymbol{x})-\mathcal{L}(\boldsymbol{x}_0)\|=\|\mathcal{L}(\boldsymbol{x}-\boldsymbol{x}_0)\|\le\Big(\sum_{i=1}^m\|\mathcal{L}(\boldsymbol{e}_i)\|\Big)\|\boldsymbol{x}-\boldsymbol{x}_0\|.$$

对任意$\epsilon>0$，取$\delta=\epsilon/(1+\sum\limits_{i=1}^m\|\mathcal{L}(\boldsymbol{e}_i)\|)$，只要$\|\boldsymbol{x}-\boldsymbol{x}_0\|<\delta$，必有$\|\mathcal{L}(\boldsymbol{x})-\mathcal{L}(\boldsymbol{x}_0)\|<\epsilon$。证毕

[4° 内积空间] 本段介绍内积空间的一些基础知识。

定义 23 设集合H是一个线性空间，一个函数$\langle\cdot,\cdot\rangle:H\times H\to\mathbb{R}$，如果满足下面四个条件：

(1) （非负性）对任意$v\in H$，$\langle v,v\rangle\ge 0$;

(2) （正定性）$(\langle v,v\rangle=0)\Leftrightarrow(v=0)$;

(3) （对称性）对任意$u,\ v\in H$，$\langle u,v\rangle=\langle v,u\rangle$;

(4) （线性性）对任意$u,\ v,\ w\in H$和$\alpha,\beta\in\mathbb{R}$，$\langle\alpha u+\beta v,w\rangle=\alpha\langle u,w\rangle+\beta\langle v,w\rangle$;

则称$\langle\cdot,\cdot\rangle$为$H$上的一个内积。线性空间$H$连同其上的内积$\langle\cdot,\cdot\rangle$一起，$(H,\langle\cdot,\cdot\rangle)$，称为一个内积空间。

利用对称性和线性性，可得内积关于第二个元素的线性性：

(4') （线性性）对任意$u,\ v,\ w\in H$和$\alpha,\beta\in\mathbb{R}$，$\langle w,\alpha u+\beta v\rangle=\alpha\langle w,u\rangle+\beta\langle w,v\rangle$。

定理 5 （*Cauchy-Schwarz*不等式）设$\langle\cdot,\cdot\rangle$是线性空间$H$上的一个内积，则成立不等式：

$$\boxed{\langle u,v\rangle^2\le\langle u,u\rangle\langle v,v\rangle.} \tag{8}$$

等号成立的充分必要条件是：存在$\lambda\in\mathbb{R}$，使得$u=\lambda v$或$v=\lambda u$ 。

证明：利用内积的线性性和非负性，我们得到

$$0\le\langle\lambda u+v,\lambda u+v\rangle=\lambda^2\langle u,u\rangle+2\lambda\langle u,v\rangle+\langle v,v\rangle.$$

于是，除非所有系数为零（此时不等式自然成立），相应的一元二次方程$\lambda^2\langle u,u\rangle+2\lambda\langle u,v\rangle+\langle v,v\rangle=0$不可能有两个不同的实根，因此，必有$\langle u,v\rangle^2-\langle u,u\rangle\langle v,v\rangle\le 0$。不等式得证。

接下来证明等号的情形。（必要性）设$\langle u,v\rangle^2=\langle u,u\rangle\langle v,v\rangle$。如果$u=0$，则$u=0v$，且$\langle u,v\rangle^2=0\langle v,v\rangle^2=0$。如果$u\ne 0$，则对任意$\lambda\in\mathbb{R}$，

$$0\le\langle\lambda u+v,\lambda u+v\rangle=\lambda^2(u,u)\pm 2\lambda\sqrt{\langle u,u\rangle\langle v,v\rangle}+\langle v,v\rangle=(\lambda\sqrt{\langle u,u\rangle}\pm\sqrt{\langle v,v\rangle})^2.$$

取$\lambda=\mp\sqrt{\langle v,v\rangle}/\sqrt{\langle u,u\rangle}$，则$\langle u+\lambda v,u+\lambda v\rangle=0$。由内积的正性，$v=-\lambda u$。

（充分性）设$u=\lambda v$，则$\langle u,v\rangle^2=\lambda^2\langle v,v\rangle^2=\langle\lambda v,\lambda v\rangle\langle v,v\rangle=\langle u,u\rangle\langle v,v\rangle$。即等式成立。证毕

利用Cauchy-Schwarz不等式，我们得到下面命题。

命题 11 设$(H,\langle\cdot,\cdot\rangle)$是一个内积空间，则函数$\|v\|=\sqrt{\langle v,v\rangle}$ 定义了一个范数，称为内积诱导出的范数。在这个范数意义下，$(H,\|\cdot\|)$是一个赋范线性空间。*Cauchy-Schwarz*不等式变形为

$$\boxed{|\langle u,v\rangle|\le\|u\|\|v\|.} \tag{9}$$

证明：我们验证$\|v\|=\sqrt{\langle v,v\rangle}$满足范数条件。显然$\|v\|\ge 0$ 且$\|v\|=0$ 的充要条件是$v=0$。又$\|\lambda v\|=\sqrt{\langle\lambda v,\lambda v\rangle}=|\lambda|\sqrt{\langle v,v\rangle}=|\lambda|\|v\|$。最后验证三角不等式：

$$\|u+v\|^2=\|u\|^2+2\langle u,v\rangle+\|v\|^2\le\|u\|^2+2\|u\|\|v\|+\|v\|^2=(\|u\|+\|v\|)^2.$$

即三角不等式成立。证毕

命题 12 设$(H,\langle\cdot,\cdot\rangle)$是一个内积空间。对任意$u,v\in H$，存在$\varphi\in[0,\pi]$，使得

$$\boxed{\langle u,v\rangle=\|u\|\|v\|\cos\varphi,} \tag{10}$$

φ称为向量u和v的夹角。

定义 24 设$(H,\langle\cdot,\cdot\rangle)$是一个内积空间。$H$中两个向量$u$和$v$ 称为是正交的，记为$u\perp v$，如果$\langle u,v\rangle=0$。

命题 13 设$(H,\langle\cdot,\cdot\rangle)$是一个内积空间，则

$$\boxed{\|u+v\|^2=\|u\|^2+\|v\|^2,\quad\forall\ u,v\in H\text{且}\ \langle u,v\rangle=0.} \tag{11}$$

例. 在Riemann可积函数类$\mathcal{R}[a,b]$上，定义函数

$$\langle f,g\rangle=\int_a^b f(x)g(x)dx,\quad\forall\ f,g\in\mathcal{R}[a,b].$$

构成$\mathcal{R}[a,b]$上的一个内积，诱导出的范数为$\|f\| = \sqrt{\int_a^b f^2(x)dx}$。函数$f$和$g$正交意味着$\int_a^b f(x)g(x)dx = 0$。Cauchy-Schwarz不等式为：

$$\Big|\int_a^b f(x)g(x)dx\Big|^2 \leq \int_a^b f^2(x)dx \int_a^b g^2(x)dx.$$

[**5° $\mathbb{R}^m$的Euclid结构**] 本段介绍$\mathbb{R}^m$上的内积。代数中，在向量空间$\mathbb{R}^m$中引入了数量积的运算

$$\boldsymbol{x} \cdot \boldsymbol{y} := \boldsymbol{x}\boldsymbol{y}^T = \sum_{i=1}^m x^i y^i, \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m,$$

其中$\boldsymbol{x}\boldsymbol{y}^T$是矩阵的乘积。容易验证数量积定义的从$\mathbb{R}^m \times \mathbb{R}^m$到$\mathbb{R}$的函数满足内积定义。

定义 25 在$\mathbb{R}^m$上定义内积：对$\mathbb{R}^m$中任意点$\boldsymbol{x} = (x^1, \cdots, x^m)$和$\boldsymbol{y} = (y^1, \cdots, y^m)$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^m x^i y^i.$$

$\mathbb{R}^m$连同上述内积，$(\mathbb{R}^m, \langle \cdot, \cdot \rangle)$，称为一个*Euclid*空间。

容易验证，上述定义的函数$\langle \cdot, \cdot \rangle : \mathbb{R}^m \to \mathbb{R}$满足内积条件，因而是一个内积。诱导出的范数为：

$$\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^m (x^i)^2}.$$

恰为上节引入的$\mathbb{R}^m$上的范数。Cauchy-Schwarz不等式为：

$$\Big|\sum_{i=1}^m x^i y^i\Big|^2 \leq \Big(\sum_{i=1}^m (x^i)^2\Big)\Big(\sum_{i=1}^m (y^i)^2\Big).$$

在Euclid空间中，任何一个线性函数$L : \mathbb{R}^m \to \mathbb{R}$都能表达成

$$L(\boldsymbol{x}) = \langle \boldsymbol{\xi}, \boldsymbol{x} \rangle, \quad \forall\ \boldsymbol{x} \in H$$

的形式，其中$\boldsymbol{\xi} \in \mathbb{R}^m$是由函数$L$唯一确定的向量。

线性空间$\mathbb{R}^m$中向量可视为在m维空间中向量在某个基底上的坐标组成的向量。上述内积是由坐标定义的内积，实际问题中，同一个点在不同的基底上的坐标是不同的，自然导致一个问题，基于不同的坐标系，观察同一事物，结果会否不同？我们研究在不同基底下的内积表达式。设$\{\boldsymbol{a}_i\}_{i=1}^m$是$\mathbb{R}^m$的一组基底，$\boldsymbol{x}$和$\boldsymbol{y}$是基于$\{\boldsymbol{a}_i\}_{i=1}^m$观察获得的坐标，则

$$\bar{\boldsymbol{x}} = x^1\boldsymbol{a}_1 + \cdots + x^m\boldsymbol{a}_m, \quad \bar{\boldsymbol{y}} = y^1\boldsymbol{a}_1 + \cdots + y^m\boldsymbol{a}_m,$$

于是

$$\langle \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \rangle = \sum_{i,j=1}^m \langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle x^i y^j$$

定义 26 基底$\{\boldsymbol{a}_i\}_{i=1}^m$称为规范正交基，如果

$$\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle = \delta_{ij} = \begin{cases} 1, & i = j; \\ 0, & \imath \neq j. \end{cases}$$

如果$\{\boldsymbol{a}_i\}_{i=1}^m$是规范正交基，则
$$\langle \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle.$$
上式表明：如果选择的基底是规范正交基，则由不同基底观察得到的内积形式和结果都相同。当基底不是规范正交基时，得到的结论将会不同。

在规范正交系下，内积具有相同的形式：$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$。反之亦然。具有这种内积形式的坐标系称为Descartes坐标系，相应坐标称为Descartes坐标。

可以定义一般形式的内积：
$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_A = \langle A\boldsymbol{x}, A\boldsymbol{y} \rangle,$$
其中，A是任意一个非奇异的$m \times m$矩阵。容易验证，$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_A$构成内积。在这个内积意义下，$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_A = 0$不一定意味着$\boldsymbol{x}$和$\boldsymbol{y}$ 在通常意义下的垂直。建立在这个内积意义下的几何学将不是Euclid几何学。

§2. 多变量函数的微分

[6° 多元可微函数和微分]

定义 27 设$E \subset \mathbb{R}^m$，$\boldsymbol{x} \in E$是极限点。称函数$\boldsymbol{f} : E \to \mathbb{R}^n$在$\boldsymbol{x}$可微，如果
$$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x})(\boldsymbol{h}) + \boldsymbol{\alpha}(\boldsymbol{x};\boldsymbol{h}), \quad \forall\ \boldsymbol{x}+\boldsymbol{h} \in E,$$
其中：$\mathcal{L}(\boldsymbol{x}) : \mathbb{R}^m \to \mathbb{R}^n$是依赖于$\boldsymbol{x}$的关于$\boldsymbol{h}$ 的线性函数；当$\boldsymbol{h} \to \boldsymbol{0}$时，$\boldsymbol{\alpha}(\boldsymbol{x};\boldsymbol{h}) = \boldsymbol{o}(\boldsymbol{h})$。线性函数$\mathcal{L}(\boldsymbol{x}) : \mathbb{R}^m \to \mathbb{R}^n$称为函数$\boldsymbol{f}$在点$\boldsymbol{x}$处的全微分，也称为切映射。

引入记号：

自变量增量：	$\Delta\boldsymbol{x}(\boldsymbol{h}) := (\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{x} = \boldsymbol{h}$,	简记为：	$\Delta\boldsymbol{x}$;
函数增量：	$\Delta\boldsymbol{f}(\boldsymbol{x};\boldsymbol{h}) := \boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x})$,	简记为：	$\Delta\boldsymbol{f}(\boldsymbol{x})$;
函数全微分：	$d\boldsymbol{f}(\boldsymbol{x})(\boldsymbol{h}) := \mathcal{L}(\boldsymbol{x})(\boldsymbol{h})$,	简记为：	$d\boldsymbol{f}(\boldsymbol{x})$.

根据命题8，对于线性函数$\mathcal{L}(\boldsymbol{x})(\boldsymbol{h})$，存在$n \times m$矩阵$A(\boldsymbol{x})$使得$\mathcal{L}(\boldsymbol{x})(\boldsymbol{h}) = A(\boldsymbol{x})\boldsymbol{h}$。利用这些记号，我们有
$$\Delta\boldsymbol{f}(\boldsymbol{x};\boldsymbol{h}) = d\boldsymbol{f}(\boldsymbol{x})(\boldsymbol{h}) + \boldsymbol{\alpha}(\boldsymbol{x};\boldsymbol{h}) = A(\boldsymbol{x})\boldsymbol{h} + \boldsymbol{\alpha}(\boldsymbol{x};\boldsymbol{h}).$$
确全定微分的关键是求出矩阵$A(\boldsymbol{x})$。

与一元函数的微分意义相同，$\boldsymbol{f}$在$\boldsymbol{x} \in E$点的全微分是函数$\boldsymbol{f}$在$\boldsymbol{x}$点近旁的增量$\Delta\boldsymbol{f}(\boldsymbol{x})$的线性主要部分：

线性—— $d\boldsymbol{f}(\boldsymbol{x})(\boldsymbol{h}) = A(\boldsymbol{x})\boldsymbol{h}$关于$\boldsymbol{h}$是线性的，易于计算。

主要部分—— $\Delta\boldsymbol{f}(\boldsymbol{x};\boldsymbol{h}) - A(\boldsymbol{x})\boldsymbol{h}$精确到比$\boldsymbol{h} \to \boldsymbol{0}$ 高阶的小量。

注意到$\boldsymbol{h} = (\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{x}$是一个起点在$\boldsymbol{x}$，终点在$\boldsymbol{x}+\boldsymbol{h}$ 的向量，可以解释为津贴在点$\boldsymbol{x} \in \mathbb{R}^m$的向量全体，记其全体为$\mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}$，称为紧贴点$\boldsymbol{x}$的切空间。

$A(\boldsymbol{x})\boldsymbol{h}$是在点$\boldsymbol{x}$近旁与函数$\boldsymbol{f}$紧密贴合的超平面，后面将看到它是$\boldsymbol{f}$在$\boldsymbol{x}$的切平面，记$\mathrm{T}\mathbb{R}^n_{\boldsymbol{f}(\boldsymbol{x})} = \{A(\boldsymbol{x})\boldsymbol{h};\ \boldsymbol{h} \in \mathbb{R}^m\}$，称为$\mathbb{R}^n$中函数$\boldsymbol{f}$在$\boldsymbol{f}(\boldsymbol{x})$处的切空间。

[7° 实值函数的偏导数和微分] 我们开始推导多元函数全微分的表达式。

命题 14 设$E \subset \mathbb{R}^m$。函数$\boldsymbol{f} = (f^1, \cdots, f^n) :\to \mathbb{R}^n$在点$\boldsymbol{x} \in E$可微，当且仅当每个坐标函数$f^i : E \to \mathbb{R}$，$i = 1, 2, \cdots, n$，在$\boldsymbol{x} \in E$可微，且
$$\boxed{d\boldsymbol{f}(\boldsymbol{x}) = (df^1(\boldsymbol{x}), \cdots, df^n(\boldsymbol{x})).} \tag{12}$$

基于以上命题，我们先考虑实值函数的微分。设$f: E \to \mathbb{R}$，求

$$df(\boldsymbol{x})(\boldsymbol{h}) = a_1(x)h^1 + \cdots + a_m h^m,$$

其中$a_i(\boldsymbol{x})$是依赖于$\boldsymbol{x}$的函数。取$\boldsymbol{h} = (0, \cdots, 0, h^i, 0, \cdots, 0)$，即$\boldsymbol{h} = h^i \boldsymbol{e}_i$，则

$$f(\boldsymbol{x} + h^i \boldsymbol{e}_i) - f(\boldsymbol{x}) = a_i(x)h^i + \alpha_i(\boldsymbol{x}; \boldsymbol{h}).$$

于是，

$$\lim_{h^i \to 0} \frac{f(\boldsymbol{x} + h^i \boldsymbol{e}_i) - f(\boldsymbol{x})}{h^i} = a_i(\boldsymbol{x}).$$

定义 28 设$f: E \to \mathbb{R}$。对$1 \leq i \leq m$，如果极限

$$\lim_{h^i \to 0} \frac{f(\boldsymbol{x} + h^i \boldsymbol{e}_i) - f(\boldsymbol{x})}{h^i}$$

存在有限极限，则称函数$\boldsymbol{f}$在点$\boldsymbol{x}$关于变量x^i可导，极限称为函数$\boldsymbol{f}$在点$\boldsymbol{x}$关于变量x^i的偏导数，记为：

$$\lim_{h^i \to 0} \frac{f(\boldsymbol{x} + h^i \boldsymbol{e}_i) - f(\boldsymbol{x})}{h^i} = \frac{\partial f(\boldsymbol{x})}{\partial x^i}.$$

下面记号都表示偏导数：

$$\frac{\partial f}{\partial x^i}(\boldsymbol{x}),\ \partial_i f(\boldsymbol{x}),\quad D_i f(\boldsymbol{x}),\quad f'_{x^i}(\boldsymbol{x}).$$

偏导数为仅视一个变量变化，其它变量为常量时多元函数f的变化率或导数，因此关于一元函数的求导法则和结果对于偏导数都成立。

例1. 设函数$f(u, v) = \dfrac{u^3}{v^2} + v^2 \sin u$。则

$$\frac{\partial f}{\partial u} = \frac{3u^2}{v^2} + v^2 \cos u, \qquad \frac{\partial f}{\partial v} = -\frac{2u^3}{v^3} + 2v \sin u.$$

根据上面的分析，我们得到下面结论。

命题 15 设$E \subset \mathbb{R}^m$，函数$f: E \to \mathbb{R}$在E的内点可微，则f 在点$\boldsymbol{x}$关于所有变量的偏导数存在，并且其微分由偏导数唯一确定：

$$\boxed{df(\boldsymbol{x})(\boldsymbol{h}) = \frac{\partial f(\boldsymbol{x})}{\partial x^1} h^1 + \cdots + \frac{\partial f(\boldsymbol{x})}{\partial x^m} h^m.} \tag{13}$$

[**7° 多元映射的Jacobi矩阵和微分**] 本段推导多元映射的微分。

定义 29 设$\boldsymbol{f}: E \to \mathbb{R}^n$在$\boldsymbol{x} \in E$的各方向偏导数存在，由其偏导数组成的矩阵，

$$\begin{pmatrix} \partial_1 f^1(\boldsymbol{x}) & \cdots & \partial_m f^1(\boldsymbol{x}) \\ \vdots & & \vdots \\ \partial_1 f^n(\boldsymbol{x}) & \cdots & \partial_m f^n(\boldsymbol{x}) \end{pmatrix}$$

称为$\boldsymbol{f}$在点$\boldsymbol{x}$的$Jacobi$矩阵，记为$D\boldsymbol{f}$，$\boldsymbol{f}'$。

根据微分的定义，对$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$，我们有

$$\Delta \boldsymbol{x}(\boldsymbol{h}) = \boldsymbol{h} = d\boldsymbol{x}$$

定理 6 设$E \subset \mathbb{R}^m$，函数$\boldsymbol{f}: E \to \mathbb{R}^n$在$E$的内点可微，则$\boldsymbol{f}$在点$\boldsymbol{x}$的*Jacobi*矩阵存在，并且其微分$d\boldsymbol{f}(\boldsymbol{x}): \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}} \to \mathrm{T}\mathbb{R}^n_{\boldsymbol{f}(\boldsymbol{x})}$由偏导数唯一确定：

$$\boxed{d\boldsymbol{f}(\boldsymbol{x})(\boldsymbol{h}) = \begin{pmatrix} \partial_1 f^1 & \cdots & \partial_m f^1 \\ \vdots & & \vdots \\ \partial_1 f^n & \cdots & \partial_m f^n \end{pmatrix} \begin{pmatrix} h^1 \\ \vdots \\ h^m \end{pmatrix}} \tag{14}$$

简记为

$$\boxed{d\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}.} \tag{15}$$

在上面的表达式中，多元函数的微分具有与一元函数相同的形式，区别在于$\boldsymbol{f}'$是Jacobi矩阵，$d\boldsymbol{x}$是一个向量。

[**8° 多元函数的连续性、可导性和可微性**] 在一元函数情形，函数可导必连续，可微与可导是等价的。但在多元函数情形，情况是复杂的。

命题 16 设$E \subset \mathbb{R}^m$。函数$\boldsymbol{f}(\boldsymbol{x}): E \to \mathbb{R}^n$在点$\boldsymbol{x}$可微，则函数$\boldsymbol{f}$在点$\boldsymbol{x}$连续。

证明：设函数$\boldsymbol{f}(\boldsymbol{x}): E \to \mathbb{R}^n$在点$\boldsymbol{x}$可微，则

$$\boldsymbol{f}(\boldsymbol{x}) - (\boldsymbol{x}_0) = \boldsymbol{f}'(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \boldsymbol{\alpha}(\boldsymbol{x}_0; \boldsymbol{x} - \boldsymbol{x}_0) \to \boldsymbol{0}, \quad \boldsymbol{x} \to \boldsymbol{x}_0.$$

即$\boldsymbol{f}$在$\boldsymbol{x}$连续。证毕

类似于一元函数情形，多元函数也是可微必连续。但在多元函数情形，许多其它性质并不相同。

例2. 考虑函数

$$f(x,y) = \begin{cases} 0, & xy = 0; \\ \\ 1, & xy \neq 0. \end{cases}$$

注意到：$f(0,y) = 0$和$f(x,0) = 0$，所以$\partial_x f(0,0) = \partial_y f(0,0) = 0$。但是函数$f$在点$(0,0)$间断，从而也不可微。这个例子表明，对于多元函数而言，可导与可微不再等价，并且可导也不保证连续。其原因在于，偏导数仅反映多元函数沿坐标轴方向的变化情况，而多元函数的连续性和可微性还取决于函数沿各方向的变化情况，仅沿坐标轴方向的变化不能控制其它方向的变化趋势。

[**9° 多变量函数可微性的充分条件**] 对于多元函数，可导不一定可微。本段讨论断定可微性的充分条件。

定理 7 设$\boldsymbol{x} \in \mathbb{R}^m$和$U(\boldsymbol{x}) \subset \mathbb{R}^m$是点$\boldsymbol{x}$的一个邻域。函数$f: U(\boldsymbol{x}) \to \mathbb{R}$在$U(\boldsymbol{x})$的每一点偏导数$\partial_1 f, \cdots, \partial_m f$都存在且连续。则$f$在点$\boldsymbol{x}$可微。

证明：由于$\boldsymbol{x}$是内点，存在开球$B(\boldsymbol{x}, \delta) \subset U(\boldsymbol{x})$，设$\boldsymbol{x} + \boldsymbol{h} \in B(\boldsymbol{x}, \delta)$，则根据一元函数中值定理，

$$\begin{aligned} & f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x}) \\ & \quad = \sum_{i=1}^m (f(x^1, \cdots, x^{i-1}, x^i + h^i, \cdots, x^m + h^m) - f(x^1, \cdots, x^i, x^{i+1} + h^{i+1}, \cdots, x^m + h^m)) \\ & \quad = \sum_{i=1}^m \partial_i f(x^1, \cdots, x^m) h^i + \sum_{i=1}^m (\partial_i (f(x^1, \cdots, x^{i-1}, x^i + \theta_i h^i, \cdots, x^m + h^m) - \partial_i f(x^1, \cdots, x^m)) h^i \end{aligned}$$

由偏导数的连续性知：

$$\sum_{i=1}^m (\partial_i (f(x^1, \cdots, x^{i-1}, x^i + \theta_i h^i, \cdots, x^m + h^m) - \partial_i f(x^1, \cdots, x^m)) h^i = o(\boldsymbol{h}), \quad \boldsymbol{h} \to 0.$$

这证明了：

$$f(\boldsymbol{x}+\boldsymbol{h})-f(\boldsymbol{x})=\sum_{i=1}^{m}\partial_i f(x^1,\cdots,x^m)h^i+o(\boldsymbol{h}).$$

即f在点$\boldsymbol{x}$可微。证毕

§3. 微分法的基本定律

[10° 函数四则运算的微分]

定理 8 设$E\subset\mathbb{R}^m$。函数$\boldsymbol{f}:E\to\mathbb{R}^n$和$\boldsymbol{g}:E\to\mathbb{R}^n$在点$\boldsymbol{x}\in E$可微，则它们的线性组合$\alpha\boldsymbol{f}+\beta\boldsymbol{g}:E\to\mathbb{R}^n$在点$\boldsymbol{x}\in E$可微，且有

$$\boxed{d(\alpha\boldsymbol{f}+\beta\boldsymbol{g})(\boldsymbol{x})=\alpha d\boldsymbol{f}(\boldsymbol{x})+\beta d\boldsymbol{g}(\boldsymbol{x}),}\tag{16}$$

等价地

$$\boxed{(\alpha\boldsymbol{f}+\beta\boldsymbol{g})'(\boldsymbol{x})=\alpha\boldsymbol{f}'(\boldsymbol{x})+\beta\boldsymbol{g}'(\boldsymbol{x}).}\tag{17}$$

证明：由微分的定义：

$$\begin{aligned}&(\alpha\boldsymbol{f}+\beta\boldsymbol{g})(\boldsymbol{x}+\boldsymbol{h})-(\alpha\boldsymbol{f}+\beta\boldsymbol{g})(\boldsymbol{x})=\alpha(\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h})-\boldsymbol{f}(\boldsymbol{x}))+\beta(\boldsymbol{g}(\boldsymbol{x}+\boldsymbol{h}))-\beta\boldsymbol{g}(\boldsymbol{x}))\\&\quad=\alpha(d\boldsymbol{f}(\boldsymbol{x})+\boldsymbol{a}(\boldsymbol{x};\boldsymbol{h}))+\beta(d\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{b}(\boldsymbol{x};\boldsymbol{h}))=\alpha d\boldsymbol{f}(\boldsymbol{x})+\beta d\boldsymbol{g}(\boldsymbol{x})+\alpha\boldsymbol{a}(\boldsymbol{x};\boldsymbol{h})+\beta\boldsymbol{b}(\boldsymbol{x};\boldsymbol{h}).\end{aligned}$$

(16)获证。另一方面，由

$$\begin{pmatrix}\partial_1(\alpha f^1+\beta g^1)&\cdots&\partial_m(\alpha f^1+\beta g^1)\\\vdots&&\vdots\\\partial_1(\alpha f^n+\beta g^n)&\cdots&\partial_m(\alpha f^n+\beta g^n)\end{pmatrix}=\alpha\begin{pmatrix}\partial_1 f^1&\cdots&\partial_m f^1\\\vdots&&\vdots\\\partial_1 f^n&\cdots&\partial_m f^n\end{pmatrix}+\beta\begin{pmatrix}\partial_1 g^1&\cdots&\partial_m g^1\\\vdots&&\vdots\\\partial_1 g^n&\cdots&\partial_m g^n\end{pmatrix}.$$

(17)获证。证毕

定理 9 设$E\subset\mathbb{R}^m$。函数$f:E\to\mathbb{R}$和$g:E\to\mathbb{R}$在点$\boldsymbol{x}\in E$可微。

(a) 它们的乘积$fg:E\to\mathbb{R}$ 在点$\boldsymbol{x}\in E$可微，且有

$$\boxed{d(fg)(\boldsymbol{x})=g(\boldsymbol{x})df(\boldsymbol{x})+f(\boldsymbol{x})dg(\boldsymbol{x}),}\tag{18}$$

等价地

$$\boxed{(fg)'(\boldsymbol{x})=g(\boldsymbol{x})f'(\boldsymbol{x})+f(\boldsymbol{x})g'(\boldsymbol{x}).}\tag{19}$$

(b) 当$g(\boldsymbol{x})\neq0$时，它们的商$\dfrac{f}{g}:E\to\mathbb{R}$ 在点$\boldsymbol{x}\in E$ 可微，且有

$$\boxed{d\Big(\frac{f}{g}\Big)(\boldsymbol{x})=\frac{1}{g^2(\boldsymbol{x})}(g(\boldsymbol{x})df(\boldsymbol{x})-f(\boldsymbol{x})dg(\boldsymbol{x})).}\tag{20}$$

等价地

$$\boxed{\Big(\frac{f}{g}\Big)'(\boldsymbol{x})=\frac{1}{g^2(\boldsymbol{x})}(g(\boldsymbol{x})f'(\boldsymbol{x})-f(\boldsymbol{x})g'(\boldsymbol{x})).}\tag{21}$$

[11° 复合函数的偏导数和微分] 我们首先考虑单个多变量函数的复合函数的偏导数。

定理 10 设$X\subset\mathbb{R}^m$和$Y\subset\mathbb{R}^n$。如果映射$f:Y\to\mathbb{R}$在$\boldsymbol{y}=\boldsymbol{g}(\boldsymbol{x})$可微和映射$\boldsymbol{g}:X\to Y$在点$\boldsymbol{x}\in X$可导，则复合映射$f\circ\boldsymbol{g}:X\to\mathbb{R}$在点$\boldsymbol{x}$可导，且

$$\boxed{\frac{\partial f(\boldsymbol{x})}{\partial x^i}=\sum_{j=1}^{n}\frac{\partial f(\boldsymbol{y})}{\partial y^j}\Big|_{\boldsymbol{y}=\boldsymbol{g}(\boldsymbol{x})}\frac{\partial g^j(\boldsymbol{x})}{\partial x^i}=\partial_j f(\boldsymbol{g}(\boldsymbol{x})\partial_i g^j(\boldsymbol{x}),\quad i=1,2,\cdots,m.}\tag{22}$$

证明：根据微分定义，我们得到：

$$f(\boldsymbol{g}(\boldsymbol{x}+h\boldsymbol{e}_i))-f(\boldsymbol{g}(\boldsymbol{x}))=\sum_{j=1}^{n}\frac{\partial f(\boldsymbol{y})}{\partial y^j}\Big|_{\boldsymbol{y}=\boldsymbol{g}(\boldsymbol{x})}(g^j(\boldsymbol{x}+h\boldsymbol{e}_i)-g^j(\boldsymbol{x}))+\boldsymbol{\alpha}(\boldsymbol{x};\boldsymbol{g}(\boldsymbol{x}+h\boldsymbol{e}_i)-\boldsymbol{g}(\boldsymbol{x})).$$

又根据偏导数定义，我们有

$$\frac{\partial f(\boldsymbol{x})}{\partial x^i}=\lim_{h\to 0}\frac{f(\boldsymbol{g}(\boldsymbol{x}+h\boldsymbol{e}_i))-f(\boldsymbol{g}(\boldsymbol{x}))}{h}=\sum_{j=1}^{n}\frac{\partial f(\boldsymbol{y})}{\partial y^j}\Big|_{\boldsymbol{y}=\boldsymbol{g}(\boldsymbol{x})}\lim_{h\to 0}\frac{g^j(\boldsymbol{x}+h\boldsymbol{e}_i)-g^j(\boldsymbol{x})}{h}$$

这即为(22)。证毕

公式(22)给出了多变量函数求导得链式法则。在上面定理中，函数f的可微性是需要的。

定理 11 设$X\subset\mathbb{R}^m$和$Y\subset\mathbb{R}^n$。如果映射$f:Y\to\mathbb{R}^k$ 在$\boldsymbol{y}=\boldsymbol{g}(\boldsymbol{x})$可微和映射$\boldsymbol{g}:X\to Y$在点$\boldsymbol{x}\in X$可导，则复合映射$f\circ\boldsymbol{g}:X\to\mathbb{R}^k$在点$\boldsymbol{x}$可导，且

$$\boxed{(\boldsymbol{f}\circ\boldsymbol{g})'(\boldsymbol{x})=\boldsymbol{f}'(\boldsymbol{g}(\boldsymbol{x}))\boldsymbol{g}'(\boldsymbol{x}),}\tag{23}$$

即两个函数的复合函数的$Jacobi$矩阵等于两个函数的$Jacobi$矩阵的乘积。

证明：根据Jacobi矩阵的定义，我们得到：

$$\begin{aligned}(\boldsymbol{f}\circ\boldsymbol{g})'&=\begin{pmatrix}\partial_1(f^1\circ\boldsymbol{g}) & \cdots & \partial_m(f^1\circ\boldsymbol{g})\\ \vdots & & \vdots\\ \partial_1(f^k\circ\boldsymbol{g}) & \cdots & \partial_m(f^k\circ\boldsymbol{g})\end{pmatrix}\\
&=\begin{pmatrix}\sum\limits_{j=1}^{n}\partial_j f^1(\boldsymbol{g}(\boldsymbol{x}))\partial_1 g^j(\boldsymbol{x}) & \cdots & \sum\limits_{j=1}^{n}\partial_j f^1(\boldsymbol{g}(\boldsymbol{x}))\partial_m g^j(\boldsymbol{x})\\ \vdots & & \vdots\\ \sum\limits_{j=1}^{n}\partial_j f^k(\boldsymbol{g}(\boldsymbol{x}))\partial_1 g^j(\boldsymbol{x}) & \cdots & \sum\limits_{j=1}^{n}\partial_j f^k(\boldsymbol{g}(\boldsymbol{x}))\partial_m g^j(\boldsymbol{x})\end{pmatrix}\\
&=\begin{pmatrix}\partial_1 f^1(\boldsymbol{g}(\boldsymbol{x})) & \cdots & \partial_n f^1(\boldsymbol{g}(\boldsymbol{x}))\\ \vdots & & \vdots\\ \partial_1 f^k(\boldsymbol{g}(\boldsymbol{x})) & \cdots & \partial_n f^k(\boldsymbol{g}(\boldsymbol{x}))\end{pmatrix}\begin{pmatrix}\partial_1 g^1(\boldsymbol{x}) & \cdots & \partial_m g^1(\boldsymbol{x})\\ \vdots & & \vdots\\ \partial_1 g^n(\boldsymbol{x}) & \cdots & \partial_m g^n(\boldsymbol{x})\end{pmatrix}.\end{aligned}$$

定理获证。证毕

例1. 设$z=ln(u^2+v)$, $u=e^{x+y^2}$, $v=x^2+y$. 求偏导数$\partial_x z$和$\partial_y z$。

解：采用复合函数求导法则。

$$\frac{\partial z}{\partial u}=\frac{2u}{u^2+v},\quad \frac{\partial z}{\partial v}=\frac{1}{u^2+v},$$

$$\frac{\partial u}{\partial x}=e^{x+y^2},\quad \frac{\partial u}{\partial y}=2ye^{x+y^2},\quad \frac{\partial v}{\partial x}=2x,\quad \frac{\partial v}{\partial y}=1.$$

于是

$$\frac{\partial z}{\partial x}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial x}=\frac{2ue^{x+y^2}}{u^2+v}+\frac{2x}{u^2+v}=\frac{2(e^{2(x+y^2)}+x)}{e^{2(x+y^2)}+x^2+y},$$

$$\frac{\partial z}{\partial y}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial y}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial y}=\frac{4uye^{x+y^2}}{u^2+v}+\frac{1}{u^2+v}=\frac{4ye^{2(x+y^2)}+1}{e^{2(x+y^2)}+x^2+y}.$$

例2. 设$y=x^x$. 求导数。

解：令$y=u^v$，$u=x$，$v=x$。则

$$y'=\partial_u y\partial_x u+\partial_v y\partial_x v=uv^{v-1}+u^v\ln v=x^x(1+\ln x).$$

例3. 设$f(x,y)$在$\mathbb{R}^2$可微，满足方程$yf_x(x,y)=xf(x,y)$. 求证：在极坐标中，f仅是r的函数。

证明：我们要证明$\dfrac{\partial f}{\partial\theta}=0$。设$x=r\cos\theta$，$y=r\sin\theta$。则

$$\frac{\partial f}{\partial\theta}=\frac{\partial f}{\partial x}\frac{\partial x}{\partial\theta}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial\theta}=-f_x r\sin\theta+f_y r\cos\theta==-yf_x+xf_y=0.$$

证毕

定理 12 设$X\subset\mathbb{R}^m$和$Y\subset\mathbb{R}^n$。如果映射$\boldsymbol{f}:X\to Y$在点$\boldsymbol{x}\in X$可微，映射$\boldsymbol{g}:Y\to\mathbb{R}^k$在$\boldsymbol{y}=\boldsymbol{f}(\boldsymbol{x})$可微，复合映射$\boldsymbol{g}\circ\boldsymbol{f}:X\to\mathbb{R}^k$在点$\boldsymbol{x}$可微，且复合映射微分$d(\boldsymbol{g}\circ\boldsymbol{f}):\mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}\to\mathrm{T}\mathbb{R}^k_{\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))}$等于微分$d\boldsymbol{f}(\boldsymbol{x}):\mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}\to\mathrm{T}\mathbb{R}^n_{\boldsymbol{f}(\boldsymbol{x})}$和微分$d\boldsymbol{g}(\boldsymbol{y}):\mathrm{T}\mathbb{R}^n_{\boldsymbol{y}}\to\mathrm{T}\mathbb{R}^k_{\boldsymbol{g}(\boldsymbol{y})}$的复合$d\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))\circ d\boldsymbol{f}(\boldsymbol{x})$，即

$$\boxed{d(\boldsymbol{g}\circ\boldsymbol{f})(\boldsymbol{x})(\boldsymbol{h})=d\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))(d\boldsymbol{f}(\boldsymbol{x})(\boldsymbol{h})).}\tag{24}$$

等价地

$$\boxed{d(\boldsymbol{g}\circ\boldsymbol{f})(\boldsymbol{x})=\boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{x}))\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}.}\tag{25}$$

注意到$d\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}$，我们有

$$\boxed{(\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x})))'d(\boldsymbol{x})=\boldsymbol{g}'(\boldsymbol{f})d\boldsymbol{f}=\boldsymbol{g}'(\boldsymbol{f}(\boldsymbol{x}))\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}.}\tag{26}$$

即无论视函数为中间变量还是最终变量的函数，其全微分形式不变，称为全微分的形式不变性。这为全微分计算带来方便，可以先对中间变量求全微分，再对中间变量求全微分。

例4. 设$z=e^{xy}\sin(x+y)$。求z的全微分dz。

解：令$u=xy$，$v=x+y$。则$z=e^u\sin v$.

$$dz=e^u\sin v du+e^u\cos v dv,\quad du=ydx+xdy,\quad dv=dx+dy$$

于是

$$\begin{aligned}dz&=e^u\sin v(ydx+xdy)+e^u\cos v(dx+dy)=e^{xy}(\sin vydx+\sin vxdy+\cos(x+y)(dx+dy)\\&=e^{xy}((y\sin(x+y)+\cos(x+y))dx+(x\sin(x+y)+\cos(x+y))dy).\end{aligned}$$

[12° 逆映射的微分]

定理 13 设$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^m$，$U(\boldsymbol{x})\subset\mathbb{R}^m$ 和$V(\boldsymbol{y})\subset\mathbb{R}^m$分别是$\boldsymbol{x}$和$\boldsymbol{y}$的邻域。$\boldsymbol{f}:U(\boldsymbol{x})\to V(\boldsymbol{y})$是将$U(\boldsymbol{x})$映到$V(\boldsymbol{y})$的映射，且$\boldsymbol{y}=\boldsymbol{f}(\boldsymbol{x})$。$\boldsymbol{f}$ 在$\boldsymbol{x}$ 连续，且存在逆映射$\boldsymbol{f}^{-1}:V(\boldsymbol{y})\to U(\boldsymbol{x})$，而且逆映射在点$\boldsymbol{y}$连续。如果映射$\boldsymbol{f}$在点$\boldsymbol{x}$可微，其逆映射$\boldsymbol{f}^{-1}$点$\boldsymbol{y}$可导，则

$$\boxed{(\boldsymbol{f}^{-1})'(\boldsymbol{y})=[\boldsymbol{f}'(\boldsymbol{x})]^{-1},\quad \boldsymbol{y}=\boldsymbol{f}(\boldsymbol{x}).}\tag{27}$$

这里$[\boldsymbol{f}'(\boldsymbol{x})]^{-1}$是$Jacobi$矩阵$\boldsymbol{f}'(\boldsymbol{x})$的逆矩阵。并且$\boldsymbol{f}^{-1}$在点$\boldsymbol{y}$可微，成立

$$\boxed{d(\boldsymbol{f}^{-1})(\boldsymbol{y})=[\boldsymbol{f}'(\boldsymbol{x})]^{-1}d\boldsymbol{y}.}\tag{28}$$

证明：注意到对于$\boldsymbol{z} \in V(\boldsymbol{y})$，$\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{f}^{-1}(\boldsymbol{z}))$，则由复合函数求导法则：

$$\mathcal{I} = \boldsymbol{z}' = (\boldsymbol{f} \circ \boldsymbol{f}^{-1}(\boldsymbol{z}))' = \boldsymbol{f}'(\boldsymbol{f}^{-1}(\boldsymbol{z}))(\boldsymbol{f}^{-1})'(\boldsymbol{z}),$$

其中$\mathcal{I}$是单位矩阵。这表明：矩阵$\boldsymbol{f}'(\boldsymbol{f}^{-1}(\boldsymbol{y}))$与$(\boldsymbol{f}^{-1})'(\boldsymbol{y})$ 互为逆矩阵，于是在$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$处，$(\boldsymbol{f}^{-1})'(\boldsymbol{y}) = (\boldsymbol{f}'(\boldsymbol{x}))^{-1}$。(27)成立。

下证(28)。设$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$，$\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x})$和$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h}) = \boldsymbol{y}+\boldsymbol{z}$。则$\boldsymbol{f}^{-1}(\boldsymbol{y}) = \boldsymbol{x}$，$\boldsymbol{h} = \boldsymbol{f}^{-1}(\boldsymbol{y}+\boldsymbol{z}) - \boldsymbol{f}^{-1}(\boldsymbol{y})$和$\boldsymbol{f}^{-1}(\boldsymbol{y}+\boldsymbol{z}) = \boldsymbol{x}+\boldsymbol{h}$。我们考察$\boldsymbol{f}^{-1}(\boldsymbol{y}+\boldsymbol{z}) - \boldsymbol{f}^{-1}(\boldsymbol{y})$。因为$\boldsymbol{f}$在$\boldsymbol{x}$可微，所以

$$\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{h} + \boldsymbol{o}(\boldsymbol{h}) \quad (\boldsymbol{h} \to \boldsymbol{0}).$$

于是$\boldsymbol{h} = (\boldsymbol{f}'(\boldsymbol{x}))^{-1}\boldsymbol{z} + \boldsymbol{o}(\boldsymbol{h}) \quad (\boldsymbol{h} \to \boldsymbol{0})$，使得

$$\boldsymbol{f}^{-1}(\boldsymbol{y}+\boldsymbol{z}) - \boldsymbol{f}^{-1}(\boldsymbol{y}) = (\boldsymbol{f}'(\boldsymbol{x}))^{-1}\boldsymbol{z} + \boldsymbol{o}(\boldsymbol{h}).$$

另一方面，我们看到

$$\|(\boldsymbol{f}'(\boldsymbol{x}))^{-1}\boldsymbol{z}\| \geq \|\boldsymbol{h}\| - \|\boldsymbol{o}(\boldsymbol{h})\| = \|\boldsymbol{h}\|\Big(1 - \frac{\|\boldsymbol{o}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|}\Big).$$

即然$\|\boldsymbol{o}(\boldsymbol{h})\|/\|\boldsymbol{h}\| \to 0$，存在$\delta > 0$，当$\|\boldsymbol{h}\| < \delta$时，$1 - \|\boldsymbol{o}(\boldsymbol{h})\|/\|\boldsymbol{h}\| \geq 1/2$，从而

$$\|\boldsymbol{h}\| \leq 2\|(\boldsymbol{f}'(\boldsymbol{x}))^{-1}\boldsymbol{z}\| = O(\|\boldsymbol{z}\|) \quad (\|\boldsymbol{z}\| \to 0).$$

因此

$$\varlimsup_{\|\boldsymbol{z}\|\to 0} \frac{\|\boldsymbol{o}(\boldsymbol{h})\|}{\|\boldsymbol{z}\|} \leq 2 \varlimsup_{\|\boldsymbol{h}\|\to 0} \frac{\|\boldsymbol{o}(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} \varlimsup_{\|\boldsymbol{z}\|\to 0} \frac{\|(\boldsymbol{f}'(\boldsymbol{x}))^{-1}\boldsymbol{z}\|}{\|\boldsymbol{z}\|} = 0.$$

这表明

$$\boldsymbol{f}^{-1}(\boldsymbol{y}+\boldsymbol{z}) - \boldsymbol{f}^{-1}(\boldsymbol{y}) = (\boldsymbol{f}'(\boldsymbol{x}))^{-1}\boldsymbol{z} + \boldsymbol{o}(\boldsymbol{z}).$$

这证明了$\boldsymbol{f}^{-1}$在$\boldsymbol{y}$可微，且(28)成立。证毕

[13° 多元函数的方向导数和梯度] 在应用中，人们不仅关心函数沿坐标轴方向的的变化情况，即各方向偏导数，也经常关心沿某一个其它方向或向量的变化情况。

定义 30 设函数f定义在点$\boldsymbol{x} \in \mathbb{R}^m$的邻域内，$\boldsymbol{v} \in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}$是在点$\boldsymbol{x}$的向量，如果

$$\boxed{D_{\boldsymbol{v}} f(\boldsymbol{x}) := \lim_{t\to 0} \frac{f(\boldsymbol{x}+t\boldsymbol{v}) - f(\boldsymbol{x})}{t},} \tag{29}$$

极限存在，称为函数f在点$\boldsymbol{x}$沿向量$\boldsymbol{v}$的导数。

定义 31 如果$\boldsymbol{l}$是单位向量，即$\|\boldsymbol{l}\| = 1$，则称$D_{\boldsymbol{l}} f(\boldsymbol{x})$为沿给定方向$\boldsymbol{l}$ 的方向导数，记为

$$\frac{\partial f(\boldsymbol{x})}{\partial l} = D_{\boldsymbol{l}} f(\boldsymbol{x}).$$

特别，如果$\boldsymbol{l}_i = \boldsymbol{e}_i$，则$\dfrac{\partial f(\boldsymbol{x})}{\partial l_i} = \dfrac{\partial f(\boldsymbol{x})}{\partial x^i}$，$i = 1, 2, \cdots, m$。

方向导数是函数沿一个单位向量的变化率，因此是一个沿某向量的导数，但沿某向量的导数一般不是方向导数。二者的区别在于，但沿某向量的导数不但依赖于向量所指的方向，而且与该向量的长度有关，而方向导数仅与方向有关（向量为单位长度）。方向导数反映了函数沿该方向的变化率。对任意向量$\boldsymbol{v}$，相应单位向量为$\bar{\boldsymbol{v}} = \boldsymbol{v}/\|\boldsymbol{v}\|$，则$\boldsymbol{v} = \|\boldsymbol{v}\|\bar{\boldsymbol{v}}$。记$f$的$\bar{\boldsymbol{v}}$ 方向导数为$\dfrac{\partial f(\boldsymbol{x})}{\partial \bar{v}}$，则

$$\boxed{D_{\boldsymbol{v}} f(\boldsymbol{x}) = \|\boldsymbol{v}\| \frac{\partial f(\boldsymbol{x})}{\partial \bar{v}}.} \tag{30}$$

上式给出了沿向量的导数与该向量确定的方向的方向导数的关系。

多元函数有无限多个方向导数，人们最关心变化率绝对值最大的方向以及对应的变化率。由于这个方向和变化率最大绝对值在理论研究和应用中的重要性，引入梯度的概念。

定义 32 设f在点$\boldsymbol{x} \in \mathbb{R}^m$存在所有方向的方向导数。一个向量称为函数$f$在点$\boldsymbol{x}$的梯度，记为$\mathbf{grad} f(\boldsymbol{x})$，如果满足以下两个条件：

(a) $\mathbf{grad} f(\boldsymbol{x})$指向函数$f$增长最快的方向，即该方向的方向导数取所有方向导数的最大值。

(b) $\mathbf{grad} f(\boldsymbol{x})$的长度$\|\mathbf{grad} f(\boldsymbol{x})\|$恰为所有方向导数的最大值。

即，若记$\boldsymbol{g} = \mathbf{grad} f(\boldsymbol{x})/\|\mathbf{grad} f(\boldsymbol{x})\|$是沿梯度方向的单位向量，则沿$\boldsymbol{g}$的方向导数

$$\frac{\partial f(\boldsymbol{x})}{\partial g} = \|\mathbf{grad} f(\boldsymbol{x})\| = \max\left\{\left|\frac{\partial f(\boldsymbol{x})}{\partial l}\right|;\quad \|\boldsymbol{l}\| = 1,\quad \forall\ \boldsymbol{l} \in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}\right\}. \tag{31}$$

梯度方向是函数f增长最快的方向，其负方向即为函数f下降最快的方向。这在应用中是非常有用的，众多优化问题的算法与梯度有关。下面讨论方向导数和梯度的计算算法。

一般地，沿向量导数$D_{\boldsymbol{v}} f(\boldsymbol{x})$的存在性不依赖于$f$在在点$\boldsymbol{x}$可微性，情况很复杂。但如果$f$在点$\boldsymbol{x}$可微，则情况非常简单。

命题 17 如果f在点$\boldsymbol{x}$可微，则对任意$\boldsymbol{v} \in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}$，$D_{\boldsymbol{v}} f(\boldsymbol{x})$必存在，且

$$\boxed{D_{\boldsymbol{v}} f(\boldsymbol{x}) = df(\boldsymbol{x})(\boldsymbol{v}) = \partial_1 f(\boldsymbol{x}) v^1 + \cdots + \partial_m f(\boldsymbol{x}) v^m.} \tag{32}$$

证明：由可微性，$f(\boldsymbol{x} + t\boldsymbol{v}) - f(\boldsymbol{x}) = t df(\boldsymbol{x})(\boldsymbol{v}) + o(t\boldsymbol{v})$，于是

$$\lim_{t\to 0} \frac{f(\boldsymbol{x} + t\boldsymbol{v}) - f(\boldsymbol{x})}{t} = df(\boldsymbol{x})(\boldsymbol{v}).$$

证毕

推论 3 如果f在点$\boldsymbol{x}$可微，那么对任意向量$\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}$和任意的实数$\lambda_1, \lambda_2 \in \mathbb{R}$，函数$f$在点$\boldsymbol{x}_0$沿向量$\lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 \in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}$ 有导数，且

$$\boxed{D_{\lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2} f(\boldsymbol{x}) = \lambda_1 D_{\boldsymbol{v}_1} f(\boldsymbol{x}) + \lambda_2 D_{\boldsymbol{v}_2} f(\boldsymbol{x}).} \tag{33}$$

命题 18 设f在点$\boldsymbol{x} \in \mathbb{R}^m$可微，$\mathbf{gradf}(\boldsymbol{x})$ 是f 在点$\boldsymbol{x}$ 的梯度，则

$$\boxed{\mathbf{grad} f(\boldsymbol{x}) = (\partial_1 f(\boldsymbol{x}), \cdots, \partial_m f(\boldsymbol{x})).} \tag{34}$$

证明：记$\boldsymbol{\xi} = (\partial_1 f(\boldsymbol{x}), \cdots, \partial_m f(\boldsymbol{x}))$，则对任意单位向量$\boldsymbol{l}$，

$$\left|\frac{\partial f(\boldsymbol{x})}{\partial l}\right| = |D_{\boldsymbol{l}} f(\boldsymbol{x})(\boldsymbol{l})| = |\langle \boldsymbol{\xi}, \boldsymbol{l}\rangle| \le \|\boldsymbol{\xi}\| \|\boldsymbol{l}\| = \|\boldsymbol{\xi}\|,$$

另一方面，记$\bar{\boldsymbol{\xi}} = \boldsymbol{\xi}/\|\boldsymbol{\xi}\|$为$\boldsymbol{\xi}$方向的单位向量，则

$$\frac{\partial f(\boldsymbol{x})}{\partial \bar{\xi}} = \langle \boldsymbol{\xi}, \frac{\boldsymbol{\xi}}{\|\boldsymbol{\xi}\|}\rangle = \|\boldsymbol{\xi}\|.$$

于是

$$\|\boldsymbol{\xi}\| = \max\left\{\left|\frac{\partial f(\boldsymbol{x})}{\partial l}\right|;\quad \|\boldsymbol{l}\| = 1,\quad \forall\ \boldsymbol{l} \in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}\right\}.$$

由梯度的定义，$\boldsymbol{\xi}=(\partial_1 f(\boldsymbol{x}),\cdots,\partial_m f(\boldsymbol{x}))$是$f$在点$\boldsymbol{x}$的梯度。

下证唯一性。若存在另一个$\boldsymbol{\eta}$也满足梯度条件，令$\bar{\eta}=\boldsymbol{\eta}/\|\boldsymbol{\eta}\|$，则

$$\|\boldsymbol{\xi}\|=\max\left\{\left|\frac{\partial f(\boldsymbol{x})}{\partial l}\right|;\ \ \|\boldsymbol{l}\|=1,\ \ \forall\ \boldsymbol{l}\in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}\right\}=\|\boldsymbol{\eta}\|=\frac{\partial f(\boldsymbol{x})}{\partial\bar{\eta}}=\langle\boldsymbol{\xi},\frac{\boldsymbol{\eta}}{\|\boldsymbol{\eta}\|}\rangle.$$

于是

$$\langle\boldsymbol{\xi},\boldsymbol{\eta}\rangle=\|\boldsymbol{\xi}\|\|\boldsymbol{\eta}\|.$$

必有$\boldsymbol{\xi}=\boldsymbol{\eta}$。证毕

对任意单位向量$\boldsymbol{l}=(l^1,\cdots,l^m)$，由于$|l^i|\le 1$, 所以存在$\theta_i\in[0,\pi]$使得$\cos\theta_i=l^i$，称为$\boldsymbol{l}$和方向$\boldsymbol{e}_i$的方向余弦，可写$\boldsymbol{l}=(\cos\theta_1,\cdots,\cos\theta_m)$。

推论 4 设f在点$\boldsymbol{x}\in\mathbb{R}^m$可微，$\mathbf{gradf}(\boldsymbol{x})$ 是f 在点$\boldsymbol{x}$ 的梯度，则对任意$\boldsymbol{v}\in \mathrm{T}\mathbb{R}^m_{\boldsymbol{x}}$，

$$\boxed{D_{\boldsymbol{v}}f(\boldsymbol{x})=\langle\mathbf{grad}f(\boldsymbol{x}),\boldsymbol{v}\rangle=\mathbf{grad}f(\boldsymbol{x})\cdot\boldsymbol{v}.}\tag{35}$$

特别，对任意单位向量$\boldsymbol{l}=(\cos\theta_1,\cdots,\cos\theta_m)$，

$$\boxed{\frac{\partial f(\boldsymbol{x})}{\partial l}=\mathbf{grad}f(\boldsymbol{x})\cdot\boldsymbol{l}=\partial_1 f(\boldsymbol{x})\cos\theta_1+\cdots+\partial_m f(\boldsymbol{x})\cos\theta_m.}\tag{36}$$

上面的公式表明，函数沿一个向量的导数等于梯度与向量的数量积。

§4. 多变量实值函数微分学的基本事实

[14° 中值定理]

定理 14 设凸开区域$G\subset\mathbb{R}^m$，$f:G\to\mathbb{R}$在G的所有内点都可微，则对任意$\boldsymbol{x}\in G$和$\boldsymbol{x}+\boldsymbol{h}\in G$，存在$0\le\theta\le 1$，使得

$$\boxed{f(\boldsymbol{x}+\boldsymbol{h})-f(\boldsymbol{x})=\sum_{i=1}^m\partial_i f(\boldsymbol{x}+\theta\boldsymbol{h})h^i=f'(\boldsymbol{x}+\theta\boldsymbol{h})\boldsymbol{h}.}\tag{37}$$

证明：引入一维函数

$$\Phi(t)=f(\boldsymbol{x}+t\boldsymbol{h}),\ \ 0\le t\le 1.$$

则$f(\boldsymbol{x}+\boldsymbol{h})-f(\boldsymbol{x})=\Phi(1)-\Phi(0)$. 根据一元函数中值定理，存在$0\le\theta\le 1$使得

$$\Phi(1)-\Phi(0)=\Phi'(\theta).$$

利用沿向量求导公式得到：

$$\Phi'(\theta)=\sum_{i=1}^m\partial_i f(\boldsymbol{x}+\theta\boldsymbol{h})h^i.$$

(37)获证。证毕

推论 5 设区域$G\subset\mathbb{R}^m$。如果$f:G\to\mathbb{R}$在G中可微，且$\partial_i f=0$，$i=1,2,\cdots,m$，则$f(\boldsymbol{x})$在G上是常数函数。

[15° 高阶偏导数] 多元函数的偏导数仍然是多元函数，可以继续沿各坐标方向求任意阶偏导数，

$$\partial_{x^{i_1}\cdots x^{i_k}}f(\boldsymbol{x}):=\partial_{x^{i_1}}(\partial_{x^{i_2}\cdots x^{i_k}}f(\boldsymbol{x}))$$

一阶以上的偏导数统称为高阶偏导数。一般记为：

$$\partial_{x^{i_1}\cdots x^{i_k}} f(\boldsymbol{x}) = \partial_{i_1\cdots i_k} f(\boldsymbol{x}) = \frac{\partial^k f(\boldsymbol{x})}{\partial x^{i_1}\cdots \partial x^{i_k}},\quad 1 \le i_1, \cdots, i_k \le m.$$

在高阶导数中，求导顺序可能是不能任意交换的。

例1. 设函数

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

考虑偏导数。

$$f_x(x,y) = \begin{cases} \dfrac{y(x^4+4x^2y^2-y^4)}{(x^2+y^2)^2}, & x^2+y^2 \neq 0; \\ 0, & x^2+y^2 = 0. \end{cases}$$

$$f_y(x,y) = \begin{cases} \dfrac{x(x^4-4x^2y^2-y^4)}{(x^2+y^2)^2}, & x^2+y^2 \neq 0; \\ 0, & x^2+y^2 = 0. \end{cases}$$

进而在点$(0,0)$求两个不同顺序的混合导数。

$$f_{xy}(0,0) = \lim_{x\to 0} \frac{f_y(x,0)-f_y(0,0)}{x} = \lim_{x\to 0}\frac{x}{x} = 1,$$

$$f_{yx}(0,0) = \lim_{y\to 0} \frac{f_x(0,y)-f_x(0,0)}{y} = \lim_{y\to 0}\frac{-y}{y} = -1,$$

于是$f_{xy}(0,0) \neq f_{yx}(0,0)$。

下面的定理给出了高阶导数的求导顺序可交换的条件。

引理 1 设区域$G\subset \mathbb{R}^m$，函数$f: G\to\mathbb{R}$在G 上存在偏导数

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x^i \partial x^j},\quad \frac{\partial^2 f(\boldsymbol{x})}{\partial x^j \partial x^i},$$

如果它们在点$\boldsymbol{x}\in G$连续，则它们在这点是相等的。

证明：由于上述导数仅涉及两个变量x^i和x^j，不妨假设函数是二元函数$f(x,y)$。考虑函数

$$F(h,t) = f(x+h,y+t) - f(x+h,y) - f(x,y+t) + f(x,y).$$

由一元函数的中值定理，我们得到

$$\begin{aligned} F(h,t) &= (f(x+h,y+t)-f(x+h,y)) - (f(x,y+t)-f(x,y)) \\ &= h\partial_x(f(x+\xi h,y+t)-f(x+\xi h,y)) \\ &= ht\partial_{yx}f(x,y) + ht(\partial_{yx}f(x+\xi h,y+\eta t) - \partial_{yx}f(x,y)). \end{aligned}$$

另一方面

$$\begin{aligned} F(h,t) &= (f(x+h,y+t)-f(x,y+t)) - (f(x+h,y)-f(x,y)) \\ &= t\partial_y(f(x+h,y+\eta' t)-f(x,y+\eta' t)) \\ &= ht\partial_{xy}f(x,y) + ht(\partial_{xy}f(x+\xi' h,y+\eta' t) - \partial_{xy}f(x,y)). \end{aligned}$$

所以

$$\begin{aligned}&\partial_{xy}f(x,y)+(\partial_{xy}f(x+\xi'h,y+\eta't)-\partial_{xy}f(x,y))\\&\quad=\partial_{yx}f(x,y)+(\partial_{yx}f(x+\xi h,y+\eta t)-\partial_{yx}f(x,y)).\end{aligned}$$

令$h,t\to 0$，则

$$\partial_{xy}f(x,y)=\partial_{yx}f(x,y).$$

等式获证。证毕

设$G\subset\mathbb{R}^m$，定义集合$C^k(G)$为集合G上具有直到k阶连续偏导数的函数$f:G\to\mathbb{R}$全体。

定理 15 如果$f\in C^k(G)$，则偏导数$\partial_{x^{i_1}x^{i_2}\cdots x^{i_k}}f(\boldsymbol{x})$不依赖于对变量$x^{i_1},\cdots,x^{i_k}$的求导顺序，也就是说，任意排列指标$i_1,i_2,\cdots,i_k$为$j_1,j_2,\cdots,j_k$，不会改变偏导数的值，即

$$\boxed{\partial_{x^{i_1}x^{i_2}\cdots x^{i_k}}f(\boldsymbol{x})=\partial_{x^{j_1}x^{j_2}\cdots x^{j_k}}f(\boldsymbol{x}).}\tag{38}$$

证明：对高阶偏导数的阶数作数学归纳。$k=2$时，由引理1即得对二阶混合偏导数，求导次序可交换。设$1\le s\le k$，s阶偏导数$\partial_{x^{i_1}x^{i_2}\cdots x^{i_s}}f(\boldsymbol{x})$ 与求导次序无关。对于$k+1$阶偏导数$\partial_{x^{i_1}x^{i_2}\cdots x^{i_{k+1}}}f(\boldsymbol{x})$。设指标列$j_1,j_2,\cdots,j_{k+1}$是指标列$i_1,i_2,\cdots,i_{k+1}$的一个重新排列。如果$i_1=j_1$，由归纳法假设，$k$阶偏导数$\partial_{x^{i_2}\cdots x^{i_{k+1}}}f(\boldsymbol{x})=\partial_{x^{j_2}\cdots x^{j_{k+1}}}f(\boldsymbol{x})$，进而$\partial_{x^{i_1}\cdots x^{i_{k+1}}}f(\boldsymbol{x})=\partial_{x^{j_1}\cdots x^{j_{k+1}}}f(\boldsymbol{x})$。如果$i_1\ne j_1$，设$j_s=i_1$ $(2\le s\le k+1)$，由归纳法假设，k 阶偏导数$\partial_{x^{j_2}\cdots x^{j_{k+1}}}f(\boldsymbol{x})=\partial_{x^{j_s}x^{j_2}\cdots x^{j_{s-1}}x^{j_{s+1}}\cdots x^{j_{k+1}}}f(\boldsymbol{x})$。进一步

$$\begin{aligned}&\partial_{x^{j_1}x^{j_s}x^{j_2}\cdots x^{j_{s-1}}x^{j_{s+1}}\cdots x^{j_{k+1}}}f(\boldsymbol{x})=\partial_{x^{j_1}x^{j_s}}(\partial_{x^{j_2}\cdots x^{j_{s-1}}x^{j_{s+1}}\cdots x^{j_{k+1}}}f(\boldsymbol{x}))\\&\quad=\partial_{x^{j_s}x^{j_1}}(\partial_{x^{j_2}\cdots x^{j_{s-1}}x^{j_{s+1}}\cdots x^{j_{k+1}}}f(\boldsymbol{x}))=\partial_{x^{i_1}x^{j_1}x^{j_2}\cdots x^{j_{s-1}}x^{j_{s+1}}\cdots x^{j_{k+1}}}f(\boldsymbol{x}).\end{aligned}$$

回到了第$k+1$次偏导数方向相同的情形。结论对$k+1$阶偏导数也成立。根据归纳法原理，结论对任意自然数成立。证毕

[16° Taylor公式]

定理 16 *(Taylor定理)* 若函数f在点$\boldsymbol{x}=(x^1,\cdots,x^m)$的某个凸邻域$U(\boldsymbol{x})$ 上有直到$n+1$阶的连续偏导数，则对$U(\boldsymbol{x})$中任意点$\boldsymbol{x}+\boldsymbol{h}\in U(\boldsymbol{x})$，成立

$$\boxed{\begin{aligned}&f(x^1+h^1,\cdots,x^m+h^m)=f(x^1,\cdots,x^m)\\&\qquad+\sum_{k=1}^{n}\frac{1}{k!}(h^1\partial_{x^1}+\cdots+h^m\partial_{x^m})^kf(x^1,\cdots,x^m)+r_{n+1}(\boldsymbol{x};\boldsymbol{h}),\end{aligned}}\tag{39}$$

其中：

$$\boxed{r_{n+1}(\boldsymbol{x};\boldsymbol{h})=\frac{1}{n!}\int_0^1(1-t)^n(h^1\partial_{x^1}+\cdots+h^m\partial_{x^m})^{n+1}f(x^1+th^1,\cdots,x^m+th^m)dt.}\tag{40}$$

公式(43)和(40)一起称为具有积分形式余项的Taylor公式。

证明：考虑辅助函数

$$\varphi(t)=f(\boldsymbol{x}+t\boldsymbol{h}).$$

由单变量函数的Taylor公式，我们得到

$$\varphi(1)=\varphi(0)+\sum_{k=1}^{n}\frac{\varphi^{(k)}(0)}{k!}+\frac{1}{n!}\int_0^1\varphi^{(n+1)}(\tau)(1-t)^ndt.$$

另一方面，我们有

$$\varphi^{(k)}(0) = (h^1\partial_1 + \cdots + h^m\partial_m)^k f(\boldsymbol{x}), \quad i = 1, 2, \cdots, n;$$
$$\varphi^{(n+1)}(t) = (h^1\partial_1 + \cdots + h^m\partial_m)^{n+1} f(\boldsymbol{x} + t\boldsymbol{h}).$$

代入上式即得(43)和(40)。证毕

类似于一元函数，我们有其它形式的余项。

$$\boxed{\text{拉格朗日余项: } r_{n+1}(\boldsymbol{x}, \boldsymbol{h}) = \frac{1}{(n+1)!}(h^1\partial_1 + \cdots + h^m\partial_m)^{n+1} f(\boldsymbol{x} + \theta\boldsymbol{h}), \quad 0 < \theta < 1.} \tag{41}$$

$$\boxed{\text{皮亚诺余项: } r_{n+1}(\boldsymbol{x}, \boldsymbol{h}) = o(\|\boldsymbol{h}\|^n).} \tag{42}$$

为了简捷的表示高阶偏导数，我们引入多重指标记号。记$\alpha = (\alpha_1, \cdots, \alpha_m)$，它由非负整数$\alpha_i$ $(i = 1, 2, \cdots, m)$组成，称为多重指标α。又记

$$|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_m, \quad \alpha! = \alpha_1!\alpha_2! \cdots \alpha_m!.$$

若$\boldsymbol{a} = (a_1, \cdots, a_m)$，则记

$$\boldsymbol{a}^\alpha = a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_m^{\alpha_m}.$$

记高阶偏导数为：

$$D^\alpha f(\boldsymbol{x}) := \frac{\partial^{|\alpha|} f(\boldsymbol{x})}{(\partial x^1)^{\alpha_1} \cdots (\partial x^m)^{\alpha_m}}.$$

使用这些记号，我们有

$$(a_1 + \cdots + a_m)^k = \sum_{\alpha_1 + \cdots + \alpha_m = k} \frac{k!}{\alpha_1! \cdots \alpha_m!} a_1^{\alpha_1} \cdots a_m^{\alpha_m} = \sum_{|\alpha| = k} \frac{k!}{\alpha!} \boldsymbol{a}^\alpha$$

Taylor公式可写为：

$$\boxed{f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \sum_{k=1}^{n} \sum_{|\alpha|=k} \frac{1}{\alpha!} D^\alpha f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)^\alpha + \sum_{|\alpha|=n+1} \frac{1}{\alpha!} D^\alpha f((1-\theta)\boldsymbol{x}_0 + \theta\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{x}_0)^\alpha.} \tag{43}$$

§5. 多变量实值函数的极值

[17° 极值和极值点的必要条件]

定义 33 设$E \subset \mathbb{R}^m$和函数$f : E \to \mathbb{R}$。如果对于E的内点$\boldsymbol{x}_0$，存在邻域$U(\boldsymbol{x}_0) \subset E$，当$\boldsymbol{x} \in U(\boldsymbol{x}_0)$时，有$f(\boldsymbol{x} \leq f(\boldsymbol{x}_0)$ $(f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0))$，则称函数$f$在$E$的内点$\boldsymbol{x}_0$有局部最大值（局部最小值）。

如果对$\boldsymbol{x} \in U(\boldsymbol{x}_0) \backslash \boldsymbol{x}_0$时，有严格不等式$f(\boldsymbol{x} < f(\boldsymbol{x}_0)$ $(f(\boldsymbol{x}) > f(\boldsymbol{x}_0))$成立，则称函数$f$在$E$的内点$\boldsymbol{x}_0$有严格局部最大值（严格局部最小值）。

定义 34 函数的局部极小值和局部极大值通称为函数的局部极值。

定理 17 设$\boldsymbol{x}_0 \in \mathbb{R}^m$和$U(\boldsymbol{x}_0) \subset \mathbb{R}^m$是$\boldsymbol{x}_0$的一个邻域，函数$f : U(\boldsymbol{x}_0) \to \mathbb{R}$在点$\boldsymbol{x}_0$关于每个变量$x^1, \cdots, x^m$存在偏导数，如果函数$f$在点$\boldsymbol{x}_0$有局部极值，则有

$$\frac{\partial f(\boldsymbol{x}_0)}{\partial x^1} = 0, \quad \cdots, \quad \frac{\partial f(\boldsymbol{x}_0)}{\partial x^m} = 0. \tag{44}$$

定理17表明：函数在它的局部极值点或是不可导，如果可导则$f' = \mathbf{0}$，即条件(44)成立。在可导条件下，条件(44)是极值点的必要条件，但不是极值点充分条件。

定义 35 实值函数满足(44)的点称为该函数的临界点，也称稳定点。

[**18° 极值点的充分条件**] 本段研究函数取极值的充分条件。

定义 36 对任意的$x^1, \cdots, x^m$且$\|\boldsymbol{h}\| \neq 0$，

a) 如果二次型$\sum\limits_{i,j=1}^{m} a_{ij}x^i x^j > 0$，称该二次型是正定的二次型。

b) 如果二次型$\sum\limits_{i,j=1}^{m} a_{ij}x^i x^j < 0$，称该二次型是负定的二次型。

c) 如果二次型$\sum\limits_{i,j=1}^{m} a_{ij}x^i x^j \geq 0$，称该二次型是半正定的二次型。

d) 如果二次型$\sum\limits_{i,j=1}^{m} a_{ij}x^i x^j \leq 0$，称该二次型是半负定的二次型。

e) 如果二次型$\sum\limits_{i,j=1}^{m} a_{ij}x^i x^j$既不半正定，也不半负定称，即二次型可取到至少两个符号相反的非零值，称该二次型是不定的二次型。

定理 18 设$\boldsymbol{x}_0 \in \mathbb{R}^m$，$U(\boldsymbol{x}_0) \subset \mathbb{R}^m$ 是$\boldsymbol{x}_0$的一个邻域和函数$f : U(\boldsymbol{x}_0) \to \mathbb{R}$。$f \in C^{(2)}(U(\boldsymbol{x}_0))$。点$\boldsymbol{x}_0$是$f$的稳定点，如果函数$f$在点$\boldsymbol{x}_0$的$Taylor$展开式

$$f(x_0^1 + h^1, \cdots, x_0^m + h^m) = f(x_0^1, \cdots, x_0^m) + \frac{1}{2}\sum_{i,j=1}^{m} \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x^i \partial x^j} h^i h^j + o(\|\boldsymbol{h}\|^2) \tag{45}$$

中的二次型

$$\sum_{i,j=1}^{m} \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x^i \partial x^j} h^i h^j \tag{46}$$

a) 当二次型(46)正定时，函数f在$\boldsymbol{x}_0$取得严格局部极小值。

b) 当二次型(46)负定时，函数f 在$\boldsymbol{x}_0$取得严格局部极大值。

c) 当二次型(46)不定时，则函数f在$\boldsymbol{x}_0$没有极值。

证明：设$\|\boldsymbol{h}\| \neq 0$和$\boldsymbol{x}_0 + \boldsymbol{h} \in U(\boldsymbol{x}_0)$，将(45)改写为

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) = f(\boldsymbol{x}_0) + \frac{\|\boldsymbol{h}\|^2}{2}\Big(\sum_{i,j=1}^{m} \frac{\partial^2 f(\boldsymbol{x}_0)}{\partial x^i \partial x^j} \frac{h^i}{\|\boldsymbol{h}\|} \frac{h^j}{\|\boldsymbol{h}\|} + o(1)\Big) \tag{47}$$

其中$o(1)$当$\boldsymbol{h} \to \mathbf{0}$时是无穷小量。

注意到向量$(\frac{h^1}{\|\boldsymbol{h}\|}, \cdots, \frac{h^m}{\|\boldsymbol{h}\|})$ 的范数为1。二次型(47)在$\mathbb{R}^m$中的单位球面

$$S(\mathbf{0}; 1) = \{\boldsymbol{x} \in \mathbb{R}^m |\ \|\boldsymbol{x}\| = 1\}$$

上连续且有界。但球面$S(\mathbf{0}; 1)$是有界闭集，于是二次型(46)在$S(\mathbf{0}; 1)$取到最大值M和最小值m。

如果二次型(46)正定，则$M \geq m > 0$，于是存在$\delta > 0$，使得$\|\boldsymbol{h}\| \leq \delta$时，有$|o(1)| < m$，这时(47)左端的量保持正号，从而当$\|\boldsymbol{h}\| \leq \delta$ 时，$f(\boldsymbol{x}_0 + \boldsymbol{h}) > f(\boldsymbol{x}_0)$，即$f(\boldsymbol{x}_0)$是严格局部极小值，$\boldsymbol{x}_0$是$f$严格极小点。a)成立。

如果二次型(46)负定，则$m \le M < 0$，于是存在$\delta > 0$，使得$\|\boldsymbol{h}\| \le \delta$时，有$|o(1)| < |M|$，这时(47)左端的量保持负号，从而当$\|\boldsymbol{h}\| \le \delta$ 时，$f(\boldsymbol{x}_0 + \boldsymbol{h}) < f(\boldsymbol{x}_0)$，即$f(\boldsymbol{x}_0)$是严格局部极大值，$\boldsymbol{x}_0$是$f$严格局部极大点。b)成立。

若二次型(46)有不同的符号，则$m < 0 < M$，(47)左端的量具有不同的符号。$f(\boldsymbol{x}_0)$不是局部极值。c)成立。证毕

定理18表明：多元函数的局部极值点的判别，决定于二次型(46)的正定性或负定性的判定。关于二次型的性质我们有下面的Sylvester定理。

定理 19 *(Sylvester准则)* 设实系数二次型 $\sum\limits_{i,j=1}^{m} a_{ij}x^i x^j$，满足$a_{ij} = a_{ji}$ $(1 \le i, j \le m)$, 记系数为矩阵

$$\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}.$$

(a) 二次型正定，当且仅当它的矩阵的所有顺序主子式大于零，即

$$A_k = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix} > 0, \quad 1 \le k \le m.$$

(b) 二次型负定，当且仅当：$a_{11} < 0$，且每一顺序主子式与下一阶顺序主子式异号，即全部奇数阶顺序主子式小于零，而所有偶数阶顺序主子式大于零。

例. 设函数

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

求函数的极值。

解. 设$y = x^2$，则$f(x,x) = x^3\ln(x^2+x^6)$关于$x = 0$是奇函数，所以，原点不是极值点。对于$(x,y) \ne (0,0)$，解方程组

$$\begin{cases} \dfrac{\partial f(x,y)}{\partial x} = y\ln(x^2+y^2) + \dfrac{2x^2y}{x^2+y^2} = 0; \\ \dfrac{\partial f(x,y)}{\partial y} = x\ln(x^2+y^2) + \dfrac{2xy^2}{x^2+y^2} = 0. \end{cases}$$

得到函数的所有稳定点：

$$(0,\pm 1), \quad (\pm 1, 0), \quad \Big(\pm\frac{1}{\sqrt{2e}}, \pm\frac{1}{\sqrt{2e}}\Big), \quad \Big(\pm\frac{1}{\sqrt{2e}}, \mp\frac{1}{\sqrt{2e}}\Big).$$

注意到：$f(x,\pm 1) = \pm x\ln(x^2+1)$和$f(\pm 1, y) = \pm y\ln(1+y^2)$是关于$x$或$y$的奇函数，显然不是极值点。考虑二阶导数：

$$\frac{\partial^2 f(x,y)}{\partial x^2} = \frac{6xy}{x^2+y^2} - \frac{4x^3y}{(x^2+y^2)^2},$$
$$\frac{\partial^2 f(x,y)}{\partial x \partial y} = \ln(x^2+y^2) + 2 - \frac{4x^2y^2}{(x^2+y^2)^2},$$
$$\frac{\partial^2 f(x,y)}{\partial y^2} = \frac{6xy}{x^2+y^2} - \frac{4xy^3}{(x^2+y^2)^2},$$

所以，二次型$\sum\limits_{i,j=1}^{2}\partial_{ij}f(x,y)h^ih^j$在点$\Big(\frac{1}{\sqrt{2e}},\frac{1}{\sqrt{2e}}\Big)$和$\Big(-\frac{1}{\sqrt{2e}},-\frac{1}{\sqrt{2e}}\Big)$的矩阵为

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

它是正定的，从而函数在点$\Big(\frac{1}{\sqrt{2e}},\frac{1}{\sqrt{2e}}\Big)$和$\Big(-\frac{1}{\sqrt{2e}},-\frac{1}{\sqrt{2e}}\Big)$取局部极小值$f\Big(\pm\frac{1}{\sqrt{2e}},\pm\frac{1}{\sqrt{2e}}\Big)=-\frac{1}{2e}$。

二次型$\sum\limits_{i,j=1}^{2}\partial_{ij}f(x,y)h^ih^j$在点$\Big(\frac{1}{\sqrt{2e}},-\frac{1}{\sqrt{2e}}\Big)$和$\Big(-\frac{1}{\sqrt{2e}},\frac{1}{\sqrt{2e}}\Big)$的矩阵为

$$\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

它是负定的，从而函数在点$\Big(\frac{1}{\sqrt{2e}},-\frac{1}{\sqrt{2e}}\Big)$和$\Big(-\frac{1}{\sqrt{2e}},\frac{1}{\sqrt{2e}}\Big)$取局部极大值$f\Big(\pm\frac{1}{\sqrt{2e}},\mp\frac{1}{\sqrt{2e}}\Big)=\frac{1}{2e}$。

例. （最小二乘方法）设通过观测或实验得到一列数据$\{(x_i,y_i)\}_{i=1}^{n}$。人们希望获得它们满足的精确曲线的近似解，为此猜测近似曲线的形式，例如$f(x)$可能是某种多项式函数，其系数待定，我们可以通过使得其与观测值的误差$f(x_i)-y_i$平方和最小来确定待定的系数。这种方法称为最小二乘法。我们通过取f为线性函数的情形说明方法。

解. 设所求直线方程为：$y=ax+b$，其中a和b是待定系数。考虑误差的平方和：

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

我们考虑优化问题：

$$J(\bar{a},\bar{b})=\min_{a,b}J(a,b).$$

为此，令

$$J_a'=2\sum_{i=1}^{n}x_i(ax_i+b-y_i)=0,\quad J_b'=2\sum_{i=1}^{n}(ax_i+b-y_i)=0.$$

整理得：

$$a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_iy_i,\quad a\sum_{i=1}^{n}x_i+nb=\sum_{i=1}^{n}y_i.$$

解出稳定点：

$$\bar{a}=\frac{n\sum\limits_{i=1}^{n}x_iy_i-(\sum\limits_{i=1}^{n}x_i)(\sum\limits_{i=1}^{n}y_i)}{n\sum\limits_{i=1}^{n}x_i^2-(\sum\limits_{i=1}^{n}x_i)^2},\qquad \bar{b}=\frac{(\sum\limits_{i=1}^{n}x_i^2)(\sum\limits_{i=1}^{n}y_i)-(\sum\limits_{i=1}^{n}x_iy_i)(\sum\limits_{i=1}^{n}x_i)}{n\sum\limits_{i=1}^{n}x_i^2-(\sum\limits_{i=1}^{n}x_i)^2}.$$

这里分母$n\sum\limits_{i=1}^{n}x_i^2-(\sum\limits_{i=1}^{n}x_i)^2=0$，当且仅当$x^1=x^2=\cdots=x^m$。参见本章定理5。

为确定极值的性质，计算二阶偏导数：

$$J_{aa}''=2\sum_{i=1}^{n}x_i^2>0,\quad J_{ab}''=2\sum_{i=1}^{n}x_i,\quad J_{bb}''=2\sum_{i=1}^{n}1=2n>0.$$

矩阵为：

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

一阶顺序主子式为$2\sum\limits_{i=1}^n x_i^2>0$，二阶主子式为$4n\sum\limits_{i=1}^n x_i^2-4(\sum\limits_{i=1}^n x_i)^2>0$。所以对应二次型是正定的。$J(a,b)$在$(\bar{a},\bar{b})$取到最小值。最佳的函数选择为$y=\bar{a}x+\bar{b}$。

[19° 点到平面的距离] 作为多元函数极值的一个应用，本段讨论空间一点到平面的距离。函数

$$z^{m+1}=D+A_1z^1+\cdots+A_mz^m \tag{48}$$

定义了$\mathbb{R}^{m+1}$中一个超平面Π。设$\boldsymbol{x}\in\mathbb{R}^{m+1}$. 点$\boldsymbol{x}$ 到平面Π的距离为:

$$dist(\boldsymbol{x},\Pi)=\min_{Q\in\Pi}\|P-Q\|.$$

我们有下面的结论。

命题 19 设超平面由(48)给出，则$\mathbb{R}^{m+1}$中点$(\boldsymbol{x}_0,x^{m+1})$到平面$\Pi$的距离为:

$$\boxed{dist(\boldsymbol{x},\Pi)=\frac{\left|x^{m+1}-D-A_1x^1-\cdots-A_mx^m\right|}{\sqrt{1+A_1^2+\cdots+A_m^2}}} \tag{49}$$

证明：记$P=(x^1,\cdots,x^{m+1})$是$\mathbb{R}^{m+1}$中一点，$Q=(z^1,\cdots,z^m,D+A_1z^1+\cdots+A_mz^m)$是平面$\Pi$上任意点，则

$$\|P-Q\|^2=\sum_{i=1}^m(x^i-z^i)^2+(x^{m+1}-D-\sum_{i=1}^m A_iz^i)^2$$

求稳定点:

$$\partial_{z^j}\|P-Q\|^2=-2(x^j-z^j)-2A_j(x^{m+1}-D-\sum_{i=1}^m A_iz^i)=0$$

即

$$(x^j-z^j)+A_j(x^{m+1}-D-\sum_{i=1}^m A_iz^i)=0,\quad 1\le j\le m.$$

整理得

$$x^{m+1}-D-\sum_{i=1}^m A_ix^i=(1+\sum_{i=1}^m A_i^2)(x^{m+1}-D-\sum_{i=1}^m A_iz^i).$$

回代得

$$x^i-z^i=-\frac{A_i(x^{m+1}-D-A_1x^1-\cdots-A_mx^m)}{1+A_1^2+\cdots+A_m^2},\quad 1\le i\le m.$$

考察二阶偏导数

$$\partial_{z^iz^j}\|P-Q\|^2=2\delta_{ij}+2A_iA_j$$

二次型

$$\sum_{i,j=1}^m(\delta_{ij}+A_iA_j)h^ih^j=2\|\boldsymbol{h}\|^2+2(A_1h^1+\cdots+A_mh^m)^2>0,\ \ \forall\ \|\boldsymbol{h}\|\neq 0$$

正定。稳定点为极小值点。于是

$$\min_{Q\in\Pi}\|P-Q\|=\frac{|x^{m+1}-D-A_1x^1-\cdots-A_mx^m|}{\sqrt{1+A_1^2+\cdots+A_m^2}}.$$

证毕

[20° 全微分的几何意义] 本段研究函数可微性的几何意义。

a. 超平面的法向量 在$\mathbb{R}^m$过点$\boldsymbol{x}_0$的超平面Π的一般形式为

$$A_1(x^1-x_0^1)+\cdots+A_m(x^m-x_0^m)=0.$$

记$\boldsymbol{A}=(A^1,\cdots,A^m)$，则上式可以表示为

$$\langle \boldsymbol{A},\boldsymbol{x}-\boldsymbol{x}_0\rangle=0,\ \ \forall\ \boldsymbol{x}\in\Pi.$$

这表明向量$\boldsymbol{A}$和向量$\boldsymbol{x}-\boldsymbol{x}_0$正交或垂直。当$\boldsymbol{x}\in\Pi$时，$\boldsymbol{x}-\boldsymbol{x}_0$为放置在$\Pi$上的一个向量，于是几何上，此式表明：向量$\boldsymbol{A}$正交或垂直于超平面$\Pi$，换言之，超平面是过一给定点与法向量垂直的直线构成的。在$\mathbb{R}^3$中，这和人们的常识吻合的。

定义 37 设$\mathbb{R}^m$是过点$\boldsymbol{x}_0$的超平面Π：

$$A_1(x^1-x_0^1)+\cdots+A_m(x^m-x_0^m)=0.$$

向量$\boldsymbol{A}=(A^1,\cdots,A^m)$称为超平面$\Pi$的法向量。

在$\mathbb{R}^m$给定一个非零向量$\boldsymbol{A}$作为法向量和一个位置点$\boldsymbol{x}_0$，即确定了一个过点$\boldsymbol{x}_0$的超平面。

b. 曲面的法向量 设$G\subset\mathbb{R}^m$和函数$f:G\to\mathbb{R}$。其图像构成了$\mathbb{R}^{m+1}$中的一片超曲面$S:\{(\boldsymbol{x},f(\boldsymbol{x}))|\boldsymbol{x}\in G\}$。若函数$f(\boldsymbol{x})$ 在点$\boldsymbol{x}_0$可微，考察过点$\boldsymbol{x}_0$的超平面：

$$x^{m+1}=f(\boldsymbol{x}_0)+\partial_1 f(\boldsymbol{x}_0)(x^1-x_0^1)+\cdots+\partial_m f(\boldsymbol{x}_0)(x^m-x_0^m).$$

写成标准形式：

$$\partial_1 f(\boldsymbol{x}_0)(x^1-x_0^1)+\cdots+\partial_m f(\boldsymbol{x}_0)(x^m-x_0^m)-(x^{m+1}-f(\boldsymbol{x}_0))=0. \tag{50}$$

即为$\mathbb{R}^{m+1}$中的超平面，其法向量为：

$$(\partial_1 f(\boldsymbol{x}_0),\cdots,\partial_m f(\boldsymbol{x}_0),-1) \tag{51}$$

称为曲面S在点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$处的法向量。

曲面S过点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$的法线方程为：

$$x^1=x_0^1+t\partial_1 f(\boldsymbol{x}_0),\ \ \cdots,\ \ x^m=x_0^m+t\partial_m f(\boldsymbol{x}_0),\ \ z=f(\boldsymbol{x}_0)-t,$$

或等价地

$$\frac{x^1-x_0^1}{\partial_1 f(\boldsymbol{x}_0)}=\cdots=\frac{x^m-x_0^m}{\partial_m f(\boldsymbol{x}_0)}=\frac{z-f(\boldsymbol{x}_0)}{-1}.$$

c. 曲面的切向量和切平面 如果S上的一条由点$\boldsymbol{x}_0$出发的道路$\Gamma:I\to S$，其表达式$(x^1(t),\cdots,x^{m+1}(t))$是可微函数，那么向量$(\dot{x}^1(0),\cdots,\dot{x}^{m+1}(0))$是时刻$t=0$的速度向量，由速度方向与经过点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$确定的直线在点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$与道路确定的$S$ 上曲线相切，该直线称为曲面S在点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$的一条切线。由此引入切向量定义。

定义 38 如果由向量$\boldsymbol{a}=(a^1,\cdots,a^{m+1})$和经过点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$确定的直线在点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$处与曲面$S$上过此点的某切线重合，则$\boldsymbol{a}$称为曲面$S$在点$(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$处的切向量。

我们考察切向量的性质。

命题 20 设f在点$\boldsymbol{x}_0$的某个领域内可微，S是由$(\boldsymbol{x}, f(\boldsymbol{x}))$生成的曲面。一个向量$\boldsymbol{\xi}$是$S$在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$的切向量，当且仅当：$\boldsymbol{\xi}$与$S$在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$的法向量(51)正交。

证明：设曲面S上一条由点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$出发的道路$\Gamma : I \to S$，其表达式为$(x^1(t), \cdots, x^m(t), f(x^1(t), \cdots, x^m(t)))$，则在$t=0$时刻，其切向量为：

$$\big(\dot{x}^1(0), \cdots, \dot{x}^m(0), \partial_1 f(\boldsymbol{x}_0)\dot{x}^1(0) + \cdots + \partial_m f(\boldsymbol{x}_0)\dot{x}^m(0)\big).$$

显然，这个向量与S在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$处的法向量的内积为零，即两向量正交。如果$\boldsymbol{\xi}$是$S$在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$的切向量，则过点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$沿$\boldsymbol{\xi}$ 方向的直线与S上某条道路$\boldsymbol{x}(t)$在该点生成的切线重合，从而$\boldsymbol{\xi} = \lambda \dot{\boldsymbol{x}}(0)$，其中$\lambda \in \mathbb{R}$。因此$\boldsymbol{\xi}$也与$S$在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$处的法向量正交。

反之，设向量$\boldsymbol{\xi}$与S在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$处的法向量正交，即

$$\partial_1 f(\boldsymbol{x}_0)\xi + \cdots + \partial_m f(\boldsymbol{x}_0)\xi^m - \xi^{m+1} = 0.$$

考虑过点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$沿方向$\boldsymbol{\xi}$的直线：

$$x^1 = x_0^1 + \xi^1 t, \ \ \cdots, \ \ x^m = x_0^m + \xi^m t, x^{m+1} = f(\boldsymbol{x}_0) + \xi^{m+1} t.$$

我们证明：它与过点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$的某条切线重合。取道路：

$$x^1 = x_0^1 + \xi^1 t, \ \cdots, \ x^m = x_0^m + \xi^m t, x^{m+1} = f(x_0^1 + \xi^1 t, \cdots, x_0^m + \xi^m t).$$

则

$$\dot{x}^1(0) = \xi^1, \ \cdots, \ \dot{x}^m(0) = \xi^m, \ \dot{x}^{m+1}(0) = \partial_1 f(\boldsymbol{x}_0)\xi^1 + \cdots + \partial_m f(\boldsymbol{x}_0)\xi^m.$$

它与该点处法向量正交，即

$$\partial_1 f(\boldsymbol{x}_0)\dot{x}^1(0) + \cdots + \partial_m f(\boldsymbol{x}_0)\dot{x}^m(0) - \dot{x}^{m+1}(0) = 0$$

既然$\dot{x}^1(0) = \xi^1$，$\cdots$，$\dot{x}^m(0) = \xi^m$。于是$\xi^{m+1} = \dot{x}^{m+1}(0)$。这证明了该直线为一条切线。所以$\boldsymbol{\xi}$是切向量。证毕

推论 6 设$f(\boldsymbol{x})$在点$\boldsymbol{x}_0$的某个领域内可微。则$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$处超平面(50)上的任意方向向量都是曲面$S$在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$处的切向量。

根据以上事实由(50)确定的超平面是曲面S在点$(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$处的切平面。函数在一点可微等价于在该点存在切平面。全微分的几何意义在于用该点处的切平面局部近似曲面。

下面讨论全微分的应用：近似计算和误差估计。

例. 近似计算$1.08^{3.96}$的近似值。

解. 设$f(x,y) = x^y$。令$x=1$，$y=2$，$\Delta x = 0.08$和$\Delta y = -0.04$。于是

$$\begin{aligned}1.08^{3.96} &= f(x+\Delta x, y+\Delta y) \simeq f(1,4) + f_x(1,4)\Delta x + f_y(1,4)\Delta y \\ &= x^y + yx^{y-1}\Delta x + x^y \ln x \Delta y = 1^4 + 4\cdot 1^3 \cdot 0.08 + 1^4 \cdot \ln 1 \cdot (-0.04) = 1.32\end{aligned}$$

例. 应用公式$S = \dfrac{1}{a} b \sin C$计算三角形面积。现测得$a = 12.50, b = 8.30, C = 30°$。若测量$a, b$的误差为$\pm 0.01$，$C$的测量误差为$\pm 0.1°$。求用此公式计算三角形面积时的绝对误差限与相对误差限。

解. 依据题意，测量中a,b,C的绝对误差限分别为

$$|\Delta a|=0.01,\quad |\Delta b|=0.01,\quad |\Delta C|=0.1^\circ=\frac{\pi}{1800}.$$

由于

$$\begin{aligned}|\Delta S|\simeq|dS|&=\left|\partial_a S\Delta a+\partial_b S\Delta b+\partial_C S\Delta C\right|\\&\le\frac{1}{2}\Big(|b\sin C|\Delta a|+|a\sin C||\Delta b|+|ab\cos C||\Delta C|\Big)\end{aligned}$$

绝对误差限为：

$$|\Delta S|\simeq 0.13.$$

因为$S=0.5\cdot 12.50\cdot 8.330\cdot 0.5\simeq 25.94$，所以相对误差限为

$$\left|\frac{\Delta S}{S}\right|\simeq\frac{0.13}{25.94}\approx 0.5\%.$$

§6. 隐函数定理

[**21° 隐函数的基本概念**] 具有形式$y=f(x)$的函数称为显函数，简称为函数。由一个方程式$F(x,y)=0$确定的变量对应法则，称为隐函数。我们以二元函数为例，给出隐函数的一些基本概念。

设$E\subset\mathbb{R}^2$，函数$F:E\to\mathbb{R}$。对于方程

$$F(x,y)=0\tag{52}$$

如果存在集合$I\times J\subset E$，对任意$x\in I$，有唯一确定的$y\in J$，使得$(x,y)\in E$，且满足方程(52)，则称方程(52)确定了定义在I上，值域含于J的隐函数，若将它记为

$$y=f(x),\quad x\in I,\ \ y\in J,$$

则成立恒等式：

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

隐式方程确定的函数往往是局部的。隐式方程可能确定y为自变量x的函数，也可能x为自变量y的函数。

隐式方程确定的函数可以是多解的。隐函数的存在唯一性是局部性质，即在局部限定的区域上存在唯一。例如：考虑圆方程$x^2+y^2=1$。它在区间$[-1,1]$上确定两个函数$y=\sqrt{1-x^2}$和$y=-\sqrt{1-x^2}$。但如果分别限定在$[-1,1]\times[0,1)$和$[-1,1]\times[0,-1]$上，则解唯一。

并不是任一隐式方程式都能确定出隐函数，例如方程式$x^2+y^2+c=0$，当$c>0$时，就没有实值解。研究在什么条件，隐式方程式能够确定出隐式函数，这是本节的任务。

倘若隐式方程式能够确定出隐函数，一般讲也不一定能从方程中解出$y=f(x)$，这里$f(x)$用初等函数和它们的复合表达。例如对于方程

$$y-x-\frac{1}{2}\sin y=0$$

确实存在一个定义在$(-\infty,+\infty)$上的函数，但这个函数却无法用初等函数和它们的复合表达出来。

[**22° 隐函数定理的最简单情形**] 我们首先考虑隐式方程仅含两个变量的最简单的情形。

定理 20 若函数$F(x,y)$满足如下条件：

(1) 存在点(x_0,y_0)，使得$F(x_0,y_0)=0$;

(2) $F(x,y)$在以(x_0,y_0)为内点的某个区域$D\subset\mathbb{R}^2$上连续；

(3) $F(x,y)$在D内存在连续的偏导数$F'_y(x,y)$；

(4) $F'_y(x_0,y_0)\neq 0$；

则

1° 在点(x_0,y_0)的某邻域$(x_0-\alpha,x_0+\alpha)\times(y_0-\beta,y_0+\beta)\subset D$ 上，方程$F(x,y)=0$唯一地确定了一个定义在$(x_0-\alpha,x_0+\alpha)$上，取值在$(y_0-\beta,y_0+\beta)$ 的隐函数$y=f(x)$，使得当$x\in(x_0-\alpha,x_0+\alpha)$时，$F(x,f(x))\equiv 0$和$y_0=f(x_0)$。

2° $f(x)$在$(x_0-\alpha,x_0+\alpha)$上连续。

证明：先证隐函数f的存在唯一性。由条件(4)，不妨设$F'_y(x_0,y_0)>0$（若$F'_y(x_0,y_0)<0$，则考虑方程$-F(x,y)=0$)。在由条件(3)，$F'_y(x,y)$在D上连续，由连续函数的局部保号性，存在(x_0,y_0)的一个方邻域$(x_0-\alpha',x_0+\alpha')\times(y_0-\beta,y_0+\beta)\subset D$，在其上$F'_y(x,y)>0$。对$(x_0-\alpha',x_0+\alpha')$中每一个$x$，函数$F(x,y)$关于$y$在$(y_0-\beta,y_0+\beta)$是严格单调增函数。由条件(1)，我们有

$$F(x_0,y_0-\beta)<F(x_0,y_0)=0<F(x_0,y_0+\beta)$$

再次利用F在D上的连续性，存在$0<\alpha\leq\alpha'$，使得

$$F(x,y_0-\beta)<0<F(x,y_0+\beta),\ \ \forall\ x\in(x_0-\alpha,x_0+\alpha).$$

对每一个$x\in(x_0-\alpha,x_0+\alpha)$，利用关于$y$的一元函数$F(x,y)$的连续性和介值定理，存在$y_0-\beta<y<y_0+\beta$，使得$F(x,y)=0$。又由$F(x,y)$关于$y$的严格单调增性质，在$[y_0-\beta,y_0+\beta]$ 中，这样的(x,y)是唯一的。这个唯一确定的关系记为$y=f(x)$，它的定义域为$(x_0-\alpha,x_0+\alpha)$，值域包含在$(y_0-\beta,y_0+\beta)$。则邻域$(x_0-\alpha,x_0+\alpha)\times(y_0-\beta,y_0+\beta)$满足$1^\circ$ 所有结论。

再证明隐函数f的连续性。设$\bar{x}\in(x_0-\alpha,x_0+\alpha)$。对任意$\epsilon>0$，且满足

$$y_0-\beta\leq\bar{y}-\epsilon<\bar{y}<\bar{y}+\epsilon\leq y_0+\beta.$$

由$F(\bar{x},y)$的严格单调增性，

$$F(\bar{x},\bar{y}-\epsilon)<F(\bar{x},\bar{y})=0<F(\bar{x},\bar{y}+\epsilon).$$

由F的连续性，存在$\delta>0$，使得当$x\in(\bar{x}-\delta,\bar{x}+\delta)$时，

$$F(x,\bar{y}-\epsilon)<0<F(x,\bar{y}+\epsilon).$$

再由F的关于y的连续性和介值定理，存在$\bar{y}-\epsilon\leq y\leq\bar{y}+\epsilon$，使得$F(x,y)=0$。但由$F$的关于$y$的局部严格单调增性，在$(y_0-\beta,y_0+\beta)$中满足$F(x,y)=0$的$y$是唯一的，而$F(x,f(x))=0$且$f(x)\in(y_0-\beta,y_0+\beta)$，于是必有$y=f(x)$，这表明：

$$|f(x)-f(\bar{x})|=|y-\bar{y}|<\epsilon,\ \ \forall\ x\in(\bar{x}-\delta,\bar{x}+\delta).$$

这证明了f在点$\bar{x}$的连续性。再由$\bar{x}$的任意性，得证f在$(x_0-\alpha,x_0+\alpha)$的连续性。证毕

注1 条件(3)和(4)仅仅用来保证在点(x_0,y_0)的某个邻域内函数$F(x,y)$关于y是严格单调增的，因此这两个条件可以减弱为"F在点(x_0,y_0)的某个邻域内关于y是严格单调增的"。

注2 如果条件(3)和(4)改为$F_x(x,y)$连续和$F_x(x_0,y_0)\neq 0$，则结论为存在唯一的局部连续函数$x=g(y)$。

定理 21 设定理*20*的条件成立，又设在D上还存在连续的偏导数$F_x(x,y)$，则由方程式$F(x,y)=0$确定的隐函数$y=f(x)$在其定义域$(x_0-\alpha,x_0+\alpha)$上有连续的导数，且

$$\boxed{f'(x)=-\frac{F'_x(x,f(x))}{F'_y(x,f(x))}.} \tag{53}$$

进一步，如果$F\in C^{(p)}(D)$，其中$p\geq 1$，则隐函数$f\in C^{(p)}(x_0-\alpha,x_0+\alpha)$。例如

$$\boxed{\begin{aligned} f''(x)&=-\frac{F''_{xx}+2F''_{xy}f'(x)+F''_{yy}(f'(x))^2}{F'_y}\\ &=\frac{F'_x(F''_{xy}+F''_{yy}f'(x))-F'_y(F''_{xx}+F''_{xy}f'(x))}{(F'_y)^2}\\ &=\frac{2F'_xF'_yF''_{xy}-(F'_x)^2F''_{yy}-(F'_y)^2F''_{xx}}{(F'_y)^3},\end{aligned}} \tag{54}$$

其中$F'_x,F'_y,F''_{xx},F''_{xy},F''_{yy}$是在点$(x,f(x))$的值。

证明：对任意x，$x+\Delta x\in(x_0-\alpha,x_0+\alpha)$，注意到

$$F(x,f(x))=F(x+\Delta x,f(x+\Delta x)=0.$$

由多元函数的中值定理，得

$$\begin{aligned}0=F(x+\Delta x,f(x+\Delta x))-F(x,f(x))=F'_x(x+\theta\Delta x,f(x)+\theta(f(x+\Delta x)-f(x))\Delta x\\ +F'_y(x+\theta\Delta x,f(x)+\theta(f(x+\Delta x)-f(x))(f(x+\Delta x)-f(x)),\quad 0<\theta<1.\end{aligned}$$

于是

$$\frac{f(x+\Delta x)-f(x)}{\Delta x}=-\frac{F'_x(x+\theta\Delta x,f(x)+\theta(f(x+\Delta x)-f(x)))}{F'_y(x+\theta\Delta x,f(x)+\theta(f(x+\Delta x)-f(x)))},$$

进而

$$f'(x)=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}=-\frac{F'_x(x,f(x))}{F'_y(x,f(x))}.$$

对等式

$$F'_x(x,f(x))+F'_y(x,f(x))f'(x)=0$$

上式两端关于x利用复合函数求导得链式法则再求导，得

$$F''_{xx}(x,f(x))+2F''_{xy}(x,f(x))f'(x)+F''_{yy}(x,f(x))(f'(x))^2+F'_y(x,f(x))f''(x)=0$$

整理即得二阶导数:

$$f''(x)=-\frac{F''_{xx}+2F''_{xy}f'(x)+F''_{yy}(f'(x))^2}{F'_y}$$

依次类推，可得更高阶导数。证毕

注3 定理的条件(3)和(4)仅仅是充分的。例如方程$y^3-x^3=0$在点$(0,0)$，$F_y(0,0)=0$，但仍能确定唯一的连续函数$y=x$。但$F_y(0,0)=0$往往导致定理结论不成立。例如双纽线方程

$$F(x,y)=(x^2+y^2)^2-x^2+y^2=0.$$

在原点$(0,0)$处$F'_x(0,0)=F'_y(0,0)=0$，且在原点$(0,0)$处不可能存在唯一的连续可微的隐函数，因为过原点，双纽线有两个交叉的分枝，无论上下还是左右分开，在原点都不可导。

[**23° $F(x^1,\cdots,x^m,y)=0$情形的隐函数定理**] 很容易将隐函数定理推广到$F(x^1,\cdots,x^m,y)=0$的情况。

定理 22 若函数$F(x^1,\cdots,x^m,y)$满足如下条件：

(1) 存在$(x_0^1,\cdots,x_0^m,y_0)$，使得$F(x_0^1,\cdots,x_0^m,y_0)=0$；

(2) $F(x^1,\cdots,x^m,y)$在以$(x_0^1,\cdots,x_0^m,y_0)$为内点的某个区域$D\subset\mathbb{R}^{m+1}$上连续；

(3) $F(x^1,\cdots,x^m,y)$在D内存在连续的偏导数$F'_y(x^1,\cdots,x^m,y)$；

(4) $F'_y(x_0^1,\cdots,x_0^m,y_0)\neq 0$；

则存在$m+1$维区间$I=I_x^m\times I_y$，其中

$$I_x^m=\{(x^1,\cdots,x^m)\in\mathbb{R}^m\big|\ |x^i-x_0^i|<\alpha^i,\ i=1,2,\cdots,m\},$$

$$I_y=\{y\in\mathbb{R}\big|\ |y-y_0|<\beta\},$$

满足：$I\subset D$，且存在函数$y=f(x^1,\cdots,x^m)$，对任意点$(x^1,\cdots,x^m,y)\in I_x^m\times I_y$，成立

$$\boxed{F(x^1,\cdots,x^m,y)=0\quad\Leftrightarrow\quad y=f(x^1,\cdots,x^m).}\tag{55}$$

并且$y=f(x^1,\cdots,x^m)$在I_x^m上连续和$y_0=f(x_0^1,\cdots,x_0^m)$。同时，$y=f(x^1,\cdots,x^m)$的偏导数可按下式计算

$$\boxed{\frac{\partial f(x^1,\cdots,x^m)}{\partial x^i}=-\frac{F'_{x^i}(x^1,\cdots,x^m,f(x^1,\cdots,x^m))}{F'_y(x^1,\cdots,x^m,f(x^1,\cdots,x^m))},\quad i=1,2\cdots,m.}\tag{56}$$

进一步，如果$F(x^1,\cdots,x^m,y)\in C^{(p)}(D)$，$(p\geq 1)$，则$f\in C^{(p)}(I_x^m)$。

[24° 单个函数隐函数定理的应用] 我们给出在单个实值函数情形的隐函数定理的一些应用。

a. 隐函数求导

例1. 设方程

$$F(x,y)=y-x-\frac{1}{2}\sin y=0.$$

函数F及其偏导数F_x和F_y在平面$\mathbb{R}^2$上任意点连续，且$F(0,0)=0$和$F_y(x,y)=1-\dfrac{1}{2}\cos y\geq\dfrac{1}{2}>0$，所以方程至少在原点$(0,0)$的近旁确定了一个连续可导的函数$y=f(x)$。进一步可将这个函数关系开拓到这个$(-\infty,+\infty)$。其导数为

$$f'(x)=-\frac{F'_x(x,y)}{F'_y(x,y)}=\frac{2}{2-\cos y}.$$

在这个例子中，我们无法利用初等函数的复合显示地表示出$y=f(x)$。

例2. 设函数$z=z(x,y)$由方程$F(x-z,y-z)=0$确定，其中F具有直到二阶的连续偏导数。求证：

$$z''_{xx}+2z''_{xy}+z''_{yy}=0.$$

证明：注意到$\tilde{F}(x,y,z)=F(u,v)$，其中$u=x-z$和$v=y-z$。利用复合函数求导公式：

$$\tilde{F}'_x=F'_u,\quad \tilde{F}'_y=F'_v,\quad \tilde{F}'_z=-(F'_u+F'_v).$$

再利用隐函数求导公式：

$$z'_x=-\frac{\tilde{F}'_x}{\tilde{F}'_z}=\frac{F'_u}{F'_u+F'_v},\quad z'_y=-\frac{\tilde{F}'_y}{\tilde{F}'_z}=\frac{F'_v}{F'_u+F'_v},$$

所以$z'_x+z'_y=1$。于是

$$z''_{xx}+z''_{xy}=0,\quad z''_{xy}+z''_{yy}=0.$$

二式相加得到：

$$z''_{xx}+2z''_{xy}+z''_{yy}=0.$$

例3. 求由方程

$$F(x,y,z)=xyz^3+x^2+y^3-z=0$$

在点$(01,1)$附近所确定的二元函数$z=z(x,y)$在$(01,1)$点处的全微分。

解：由于$F(0,1,1)=0$，$F'_x=yz^3+2x$，$F'_y=xz^3+3y^2$和$F'_z=3xy^2-1$，$F'x$，F'_y和F'_z连续，且$F'_z(0,1,1)=-1<0$，根据隐函数存在定理，方程$F(x,y,z)=0$在点$(0,1,1)$近旁唯一确定连续可微隐函数$z=z(x,y)$，其偏导数为：

$$\frac{\partial z}{\partial x}=-\frac{F'_x}{F'_z}=\frac{yz^3+2x}{1-3xyz^2},\qquad \frac{\partial z}{\partial x}=-\frac{F'_y}{F'_z}=\frac{xz^3+3y^2}{1-3xyz^2}.$$

在点$(0,1,1)$处，$\partial_x z(0,1)=1$和$\partial_y z(0,1)=3$。所以，隐函数$z=z(x,y)$在点$(0,1)$近旁的全微分为

$$dz\big|_{(0,1)}=dx+3dy.$$

b. 隐函数的极值问题 作为一个例子，我们考虑下面问题。

例4. 讨论Decartes叶形线

$$x^3+y^3-3axy=0\ \ (a>0)$$

所确定的隐函数$y=f(x)$的极值问题。

解：令 $F(x,y)=x^3+y^3-3axy$。则

$$F'_x=3(x^2-ay),\quad F'_y=3(y^2-3x).$$

在点$(0,0)$和$(\sqrt[3]{4}a,\sqrt[3]{2}a)$处，$F'(0,0)=F'(\sqrt[3]{4}a,\sqrt[3]{2}a)=0$。除这两点外，在其它点近旁，方程$F(x,y)=0$ 均可确定隐函数 $y=f(x)$，且其一阶导数为：

$$y'=-\frac{F'_x}{F'_y}=\frac{ay-x^2}{y^2-ax}.$$

又由$F''_{xx}=6x$，$F''_{xy}=-3a$和$F''_{yy}=6y$使得

$$2F'_xF'_yF''_{xy}=-54a(x^2-ay)(y^2-ax),\ \ (F'_x)^2F''_{yy}=54y(x^2-ay)^2,\ \ (F'_y)^2F''_{xx}=54x(y^2-ax)^2.$$

于是

$$y''=\frac{2F'_xF'_yF''_{xy}-(F'_x)^2F''_{yy}-(F'_y)^2F''_{xx}}{(F'_y)^3}=-\frac{2a^3xy}{(y^2-ax)^3}.$$

联立方程

$$x^3+y^3-3axy=0,\quad \frac{ay-x^2}{y^2-ax}=0,$$

解出稳定点$A=(\sqrt[3]{2}a,\sqrt[3]{4}a)$。在稳定点，$y''|_A=-4/(\sqrt[3]{2}a)<0$。所以隐函数$y=f(x)$在点$(\sqrt[3]{2}a,\sqrt[3]{4}a)$取局部极大值$\sqrt[3]{4}a$。

d. 反函数的存在性及其导数 我们应用多元方程的隐函数定理研究反函数的存在性及其导数。设函数$y=f(x)$在点x_0的某个邻域内有连续的导数$f'(x)$且$y_0=f(x_0)$。考虑方程：

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

由于$F(x_0,y_0)=0$，$F'_y(x,y)=1$和$F'_x(x,y)=-f(x)$。所以，只要$f'(x_0)\neq 0$, 根据隐函数定理，在(x_0,y_0)的某个邻域内$(x_0-\alpha,x_0+\alpha)\times(y_0-\beta,y_0+\beta)$上存在连续可微的反函数$x=g(y)$，且反函数的导数是：

$$g'(y)=-\frac{F'_y}{F'_x}=\frac{1}{f'(g(y))}.$$

此即一元函数的反函数求导公式。

c. 隐函数确定的平面曲线的切线与法线 设曲线由方程

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

给出，我们考察平面曲线的切线方程和法线方程。设$F(x,y)$在点(x_0,y_0)的某个邻域内满足隐函数存在定理。不妨设$F'_y(x_0,y_0)\neq 0$。于是方程$F(x,y)=0$在(x_0,y_0)的某个邻域内唯一确定连续可微函数$y=f(x)$，使得$F(x,f(x))=0$，$y_0=f(x_0)$且

$$\frac{dy}{dx}=-\frac{F'_x(x,y)}{F'_y(x.y)}.$$

在(x_0,y_0)点近旁，平面曲线L为$(x,y(x))$，其切线方程为：

$$\boxed{\text{切线方程：}y-y_0=-\frac{F'_x(x_0,y_0)}{F'_z(x_0.y_0)}(x-x_0).} \tag{57}$$

写成对称的形式

$$\boxed{\text{切线方程：}\quad F'_x(x_0,y_0)(x-x_0)+F'_y(x_0,y_0)(y-y_0)=0.} \tag{58}$$

法线方程为：

$$\boxed{\text{法线方程：}\quad y-y_0=\frac{F'_y(x_0,y_0)}{F'_x(x_0,y_0)}(x-x_0).} \tag{59}$$

写成对称的形式

$$\boxed{\text{法线方程：}\quad \frac{x-x_0}{F'_x(x_0,y_0)}=\frac{y-y_0}{F'_y(x_0,y_0)}.} \tag{60}$$

d. 隐函数确定的曲面的切平面与法线 设曲面由方程

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

给出，我们考察曲面的法线方程和切面方程。设$F(x,y,z)$在点(x_0,y_0,z_0)的某个邻域内满足隐函数存在定理。不妨设$F'_z(x_0,y_0,z_0)\neq 0$。于是方程$F(x,y,z)=0$在(x_0,y_0,z_0)的某个邻域内唯一确定连续可微函数$z=z(x,y)$，使得$F(x,y,z(x,y))=0$且

$$\frac{\partial z}{\partial x}=-\frac{F'_x(x,y,z)}{F'_z(x.y,z)},\quad \frac{\partial z}{\partial y}=-\frac{F'_y(x,y,z)}{F'_z(x.y,z)}.$$

在(x_0,y_0,z_0)点近旁，曲面S为$(x,y,z(x,y))$，其切平面方程为：

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

写成对称的形式

$$\boxed{\text{切平面方程：}F'_x(x_0,y_0,z_0)(x-x_0)+F'_y(x_0,y_0,z_0)(y-y_0)+F'_z(x_0.y_0,z_0)(z-z_0)=0.} \tag{61}$$

法线方程为：

$$\boxed{\text{法线方程：}\quad \frac{x-x_0}{F'_x(x_0,y_0,z_0)}=\frac{y-y_0}{F'_y(x_0,y_0,z_0)}=\frac{z-z_0}{F'_z(x_0.y_0,z_0)}} \tag{62}$$

或

$$\boxed{x=x_0+tF'_x(x_0,y_0,z_0),\quad y=y_0+tF'_y(x_0,y_0,z_0),\quad z=z_0+tF'_z(x_0.y_0,z_0),\quad -\infty<t\leq+\infty.} \tag{63}$$

例5. 求椭球面$x^2+2y^2+3z^2=6$在点$(1,1,1)$处的切平面和法线方程。

解：令$F(x,y,z)=x^2+2y^2+3z^2-6$，则$F'_x=2x$，$F'_y=4y$和$F'_z=6z$。它们在全空间处处连续。在点$(1,1,1)$处，$F(1,1,1)=0$，且$F'_x(1,1,1)=2$，$F'_y(1,1,1)=4$和$F'_z(1,1,1)=6$，其切平面为

$$2(x-1)+4(y-1)+6(z-1)=0$$

即

$$x+2y+3z=6.$$

法线方向为：

$$\frac{x-1}{1}=\frac{y-1}{2}=\frac{z-1}{3},$$

即

$$x=1+t,\quad y=1+2t,\quad z=1+3t,\quad -\infty<t<+\infty.$$

例6. 证明：曲面$f(\dfrac{x-a}{z-c},\dfrac{y-b}{z-c})=0$的任意切平面都过某个定点，其中$f$ 是连续可微函数。

证明：令$F(x,y,z)=f(\dfrac{x-a}{z-c},\dfrac{y-b}{z-c})$。记$f=f(u,v)$，其中$u=\dfrac{x-a}{z-c},v=\dfrac{y-b}{z-c}$。设在点$P_0=(x_0,y_0,z_0)$，函数$F(P_0)=0$，且

$$F'_xP_0)=\frac{f'_u(P_0)}{z_0-c},\quad F'_y=\frac{f'_v(P_0)}{z_0-c},\quad F'_z=-\frac{(x_0-a)f'_u(P_0)+(y_0-b)f'_v(P_0)}{(z_0-c)^2},$$

切平面方程为：

$$(x-x_0)f'_u(P_0)+(y-y_0)f'_v(P_0)-(z-z_0)\frac{(x_0-a)f'_u(P_0)+(y_0-b)f'_v(P_0)}{z_0-c}=0.$$

显然，取点$(x,y,z)=(a,b,c)$，代入上式即得恒等式：

$$(a-x_0)f'_u(P_0)+(b-y_0)f'_v(P_0)-(c-z_0)\frac{(x_0-a)f'_u(P_0)+(y_0-b)f'_v(P_0)}{z_0-c}\equiv 0.$$

这表明：任意点的切平面都过点(a,b,c)。证毕