数学分析讲义: 第八章 多元微分学

讲课教材: 《数学分析讲义》陈天权编著, 共三册, 北京大学出版社

参考书: 《数学分析》(第4版) 第一卷、第二卷, V.A. Zorich (卓里奇) 编著, 中译本, 高等教育出版社; 《数学分析》徐森林、薛春华编著, 共三册，清华大学出版社; 《数学分析教程》上下册, 常庚哲、史济怀编著, 中科大出版社

清华大学数学科学系 2016 级数学分析教学 卢旭光 Feb. 2017

数学分析讲义总目录

第八章　多元微分学

§8.1 线性变换和矩阵范数

【说明: 线性变换是指从一个线性空间X到X自身的线性映射, 而一般的线性映射是指从一个线性空间X到另一个线性空间Y的线性映射. 但为了兼顾习惯说法, 我们有时也把线性映射叫做线性变换.】

回忆一元函数的微分或切线概念: 几何上看, 切线是一条直线. 对二元函数来说, "切线" 将变成切平面. 一般地, 对多元函数来说, "切线" 将变成超平面. 把坐标原点移到切点处, 则这样的超平面便称为切空间. 它是一个线性子空间. 切空间是由所考虑的多元函数在切点处产生的线性映射唯一决定的. 因此不同的函数决定的切空间就可能不同, 特别是复合函数决定的切空间是线性子空间之间的线性变换的结果. 为弄清楚这些概念和性质, 我们就需要了解欧空间上的线性变换.

【定义】设X, Y 是同一数域$\mathbb{K}$ 上的线性空间. 称映射$\mathcal{L}: X \to Y$ 是线性的如果$\mathcal{L}$ 保持加法和数乘:

$$\mathcal{L}(x+y)=\mathcal{L}(x)+\mathcal{L}(y), \quad \mathcal{L}(\lambda x)=\lambda\mathcal{L}(x), \quad x,y\in X, \quad \lambda\in\mathbb{K}.$$

令$\mathscr{L}(X,Y)$ 是从X 到Y 的线性映射的全体. 在$\mathscr{L}(X,Y)$上定义加法和数乘为

$$(\mathcal{L}_1+\mathcal{L}_2)(x)=\mathcal{L}_1(x)+\mathcal{L}_2(x), \quad (\alpha\mathcal{L})(x)=\alpha\mathcal{L}(x), \quad x\in X, \quad \alpha\in\mathbb{K}.$$

按此运算易见$\mathscr{L}(X,Y)$ 也是数域$\mathbb{K}$上的一个线性空间 (也即若$\mathcal{L}_1, \mathcal{L}_2 \in \mathscr{L}(X,Y)$, $\alpha, \beta \in \mathbb{K}$, 则$\alpha\mathcal{L}_1+\beta\mathcal{L}_2$ 也保持加法和数乘). □

以下我们取$X=\mathbb{R}^n, Y=\mathbb{R}^m$和数域$\mathbb{K}=\mathbb{R}$, 也即考虑$\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$. 相应地, 令$\mathbb{R}^{m\times n}$ 表示m行、n 列的实矩阵的全体.

【对矩阵与向量乘法的约定】

按照惯例, 当矩阵$A=(a_{ij})_{1\le i\le m, 1\le j\le n}\in\mathbb{R}^{m\times n}$ 左乘向量$x=(x_1,x_2,...,x_n)\in\mathbb{R}^n$时, 向量$x$将自动转换为列向量, 也即$x=(x_1,x_2,...,x_n)^\tau\in\mathbb{R}^{n\times 1}\equiv\mathbb{R}^n$:

$$Ax=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn}\end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n\end{pmatrix}=\begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j\end{pmatrix}.$$

令

$$\mathbf{e}_1=\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\quad \mathbf{e}_2=\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix},\quad ...,\quad \mathbf{e}_n=\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}.$$

以下我们用花写的$\mathcal{A},\mathcal{B}$ 等表示$\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$的元素, 即$\mathbb{R}^n$ 到$\mathbb{R}^m$ 的线性映射; 而用对A,B 等表示$\mathcal{A},\mathcal{B}$ 对应的矩阵(详见下面).

设$\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$. 则可写

$$\mathcal{A}(x)=\begin{pmatrix}\mathcal{A}_1(x)\\\mathcal{A}_2(x)\\\vdots\\\mathcal{A}_m(x)\end{pmatrix},\qquad x=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\sum_{j=1}^n x_j\mathbf{e}_j.$$

$\mathcal{A}$的线性性蕴含$\mathcal{A}$的每个坐标函数$\mathcal{A}_i$都是线性的(反之亦然). 于是有

$$\mathcal{A}_i(x)=\sum_{j=1}^n x_j\mathcal{A}_i(\mathbf{e}_j),\quad i=1,2,...,m.$$

$\Longrightarrow$

$$\mathcal{A}(x)=\begin{pmatrix}\sum_{j=1}^n x_j\mathcal{A}_1(\mathbf{e}_j)\\\sum_{j=1}^n x_j\mathcal{A}_2(\mathbf{e}_j)\\\vdots\\\sum_{j=1}^n x_j\mathcal{A}_m(\mathbf{e}_j)\end{pmatrix}=\begin{pmatrix}\mathcal{A}_1(\mathbf{e}_1) & \mathcal{A}_1(\mathbf{e}_2) & ... & \mathcal{A}_1(\mathbf{e}_n)\\\mathcal{A}_2(\mathbf{e}_1) & \mathcal{A}_2(\mathbf{e}_2) & ... & \mathcal{A}_2(\mathbf{e}_n)\\\multicolumn{4}{c}{\dotfill}\\\mathcal{A}_m(\mathbf{e}_1) & \mathcal{A}_m(\mathbf{e}_2) & ... & \mathcal{A}_m(\mathbf{e}_n)\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}.$$

定义从$\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 到$\mathbb{R}^{m\times n}$ 的映射如下：

$$\mathcal{A}\mapsto A:=(\mathcal{A}_i(\mathbf{e}_j))_{m\times n},\quad \mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m).$$

我们称$A=(\mathcal{A}_i(\mathbf{e}_j))_{m\times n}$ 是$\mathcal{A}$的**表示矩阵**.

回忆线性代数: $Ax\equiv Bx\ (x\in\mathbb{R}^n)\iff A=B$.

于是从上面的导出过程易见

$$A\in\mathbb{R}^{m\times n}\text{ 是 }\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)\text{ 的表示矩阵}\iff \mathcal{A}(x)\equiv Ax,\quad x\in\mathbb{R}^n.$$

设$\mathcal{A},\mathcal{B}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m),\alpha,\beta\in\mathbb{R}$, 则有

$$(\alpha\mathcal{A}+\beta\mathcal{B})(x)=\alpha\mathcal{A}(x)+\beta\mathcal{B}(x)=(\alpha A+\beta B)x,\quad x\in\mathbb{R}^m.$$

这表明$\alpha\mathcal{A}+\beta\mathcal{B}$ 的表示矩阵是$\alpha A+\beta B$.

设$\mathcal{B}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m),\mathcal{A}\in\mathscr{L}(\mathbb{R}^m,\mathbb{R}^l)$, 则$\mathcal{A}\circ\mathcal{B}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^l)$ 并有

$$(\mathcal{A}\circ\mathcal{B})(x)=\mathcal{A}(\mathcal{B}(x))=A\mathcal{B}(x)=ABx,\quad x\in\mathbb{R}^n.$$

这表明$\mathcal{A}\circ\mathcal{B}$ 的表示矩阵等于AB.

设$m=n$. 设$\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^n)$. 假设$\mathcal{A}$ 为单射. 则$\mathcal{A}(x)=0\Longrightarrow x=0$, 也即$Ax=0\Longrightarrow x=0$. 这等价于方阵$A$ 可逆, 即$\det A\neq 0$. 此时由线代数可知对任意$y\in\mathbb{R}^m$, 方程$Ax=y$ 有且有唯一解$x=A^{-1}y$. 因此$y=Ax=\mathcal{A}(x)\in\mathcal{A}(\mathbb{R}^n)$. 这表明$\mathcal{A}$ 是单射蕴含$\mathcal{A}$ 是满射. 反之设$\mathcal{A}$为满射. 则存在$b_j\in\mathbb{R}^n$使得$\mathbf{e}_j=\mathcal{A}(b_j)=Ab_j, j=1,2,...,n$. 令$B=(b_1,b_2,...,b_n)$. 则单位矩阵$\mathrm{I}=AB$. 这蕴含$\det A\neq 0$. 因此$\mathcal{A}$ 也是单射.

以上说明: 对于线性变换$\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^n)$ 而言, $\mathcal{A}$ 是满射$\Longleftrightarrow$ $\mathcal{A}$ 是单射.

设$\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^n)$ 为单满射. 则自然有$\mathcal{A}^{-1}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^n)$. 设$B$ 是$\mathcal{A}^{-1}$的表示矩阵, 则有

$$x\equiv\mathcal{A}\circ\mathcal{A}^{-1}(x)=ABx,\qquad x\in\mathbb{R}^n$$

从而有

$$AB=\mathrm{I}\qquad\text{即}\quad B=A^{-1}.$$

这表明$\mathcal{A}^{-1}$的表示矩阵等于$\mathcal{A}$的表示矩阵的逆, 即等于A^{-1}.

概括起来我们有如下结论:

【命题8.1.】

(a) 记$A=(\mathcal{A}_i(\mathbf{e}_j))_{m\times n}$ 为$\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的表示矩阵. 则映射$\mathcal{A}\mapsto A$ 是从$\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$到$\mathbb{R}^{m\times n}$的线性同构映射, 因此$\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 与$\mathbb{R}^{m\times n}$同构. 此外有: $A\in\mathbb{R}^{m\times n}$是$\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$的表示矩阵当且仅当$\mathcal{A}(x)=Ax$ for all $x\in\mathbb{R}^n$.

(b) 映射$\mathcal{A}\mapsto A=(\mathcal{A}_i(\mathbf{e}_j))_{m\times n}$ 把线性变换的复合乘法变为矩阵的乘法: 即若$\mathcal{B}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m),\mathcal{A}\in\mathscr{L}(\mathbb{R}^m,\mathbb{R}^l)$, 则$\mathcal{A}\circ\mathcal{B}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^l)$ 且

$$\mathcal{A}\circ\mathcal{B}(x)=ABx\quad\text{for all}\, x\in\mathbb{R}^n.$$

(c) 设$m=n$. 则: $\mathcal{A}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^n)$ 为单满射 $\Longleftrightarrow$ A 为可逆矩阵.

而当$\mathcal{A}$ 为单满射时, $\mathcal{A}^{-1}$ 的表示矩阵等于A^{-1}. $\square$

根据上面的分析和结果可见

如果把矩阵$A \in \mathbb{R}^{m\times n}$ 也看成线性变换$A : \mathbb{R}^n \to \mathbb{R}^m, A(x) := Ax$, 则

线性变换A 与矩阵A、复合乘法$A \circ B$ 与矩阵乘法AB、逆变换A^{-1}与逆矩阵A^{-1}, 等等, 便没有实质区别.

换言之, 如果把矩阵也看成是线性变换, 则有$\mathscr{L}(\mathbb{R}^n, \mathbb{R}^m) = \mathbb{R}^{m\times n}$. 因此, 对于有限维线性空间来说, 研究线性变换等价于研究矩阵.

【矩阵范数的定义】我们称非负函数

$$\|\cdot\| : \bigcup_{m,n\in\mathbb{N}} \mathbb{R}^{m\times n} \longrightarrow [0, +\infty)$$

是一个矩阵范数, 如果$\|\cdot\|$满足下面(i),(ii):

(**i**) 对任意$m, n \in \mathbb{N}$, $\|\cdot\|$是$\mathbb{R}^{m\times n}$上的范数, 即满足

正定性：$\|A\| \ge 0$; $\quad \|A\| = 0 \iff A = 0.$

正齐次性：$\|\alpha A\| = |\alpha|\|A\|, \quad \alpha \in \mathbb{R};$

三角不等式: $\|A + B\| \le \|A\| + \|B\|.$

(**ii**) $\|\cdot\|$ 满足矩阵乘积的范数不等式: 对任意$l, m, n \in \mathbb{N}$

$$\|AB\| \le \|A\|\|B\| \qquad \forall\, A \in \mathbb{R}^{l\times m},\ \forall\, B \in \mathbb{R}^{m\times n}. \qquad \Box$$

【注】由于当$(m, n) \ne (m', n')$ 时$\mathbb{R}^{m\times n}$与$\mathbb{R}^{m'\times n'}$ 不相交, 故上述范数$\|\cdot\|$是良好定义的(well defined).

与欧空间情形相同, 矩阵空间上的任意两个范数彼此等价, 并且矩阵按范数收敛等价于矩阵的每个元素收敛. 准确的说法是

【命题8.2.】设$\|\cdot\|, \|\cdot\|_*$ 是任意两个矩阵范数. 则对任意$m, n \in \mathbb{N}$ 存在只依赖于mn的常数$0 < a < b < +\infty$ 使得

$$a\|A\| \le \|A\|_* \le b\|A\| \qquad \forall\, A \in \mathbb{R}^{m\times n}.$$

此外$\|A\|$ 与$A=(a_{ij})_{m\times n}$的元素a_{ij}之间可以相互控制: 存在一个只依赖于mn的常数$0<C<+\infty$ 使得对所有$A=(a_{ij})_{m\times n}$有

$$|a_{ij}|\le C\|A\|,\quad 1\le i\le m,\ 1\le j\le n;\qquad \|A\|\le C\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|.$$

因此在给定的矩阵空间$\mathbb{R}^{m\times n}$上, 按矩阵范数的收敛:

$B\to A$ (即$\|B-A\|\to 0$)等价于按矩阵每个元素的收敛: $b_{ij}\to a_{ij}$, $i=1,2,...,m; j=1,2,...,n$.

【证】给定$m,n\in\mathbb{N}$. 我们把证明转到欧空间$\mathbb{R}^{mn}$ 上. 做映射$\Phi:\mathbb{R}^{m\times n}\to\mathbb{R}^{mn}$如下: 对于$A=(a_{ij})_{1\le i\le m,1\le j\le n}\in\mathbb{R}^{m\times n}$, 定义

$$\Phi(A)=(a_{11},a_{12},...,a_{1n},a_{21},a_{22},...,a_{2n},......,a_{n1},a_{n2},...,a_{mn}).$$

Φ 显然是线性映射且为单射. 满射也是显然的: 对任意

$$x=(x_1,x_2,...,x_n,...,x_{(j-1)n+1},x_{(j-1)n+2},...,x_{jn},...,x_{(m-1)n+1},x_{(m-1)n+2},...,x_{mn})\in\mathbb{R}^{mn}$$

令

$$a_{ij}=x_{(i-1)n+j},\quad i=1,2,...,m;\quad j=1,2,...,n.$$

则$x=\Phi(A)$. 于是任意矩阵$A\in\mathbb{R}^{m\times n}$ 可表为$A=\Phi^{-1}(x), x\in\mathbb{R}^{nm}$.

设$\|\cdot\|$ 是任意矩阵范数. 来证明$\|\Phi^{-1}(\cdot)\|$ 是欧空间$\mathbb{R}^{mn}$ 上的范数. 证明将用到$\Phi^{-1}:\mathbb{R}^{mn}\to\mathbb{R}^{m\times n}$的线性性.

正定性: 显然$\|\Phi^{-1}(x)\|\ge 0$, 而当$\|\Phi^{-1}(x)\|=0$ 时有$\Phi^{-1}(x)=0$ 因此$x=0$.

正齐次性:

$$\|\Phi^{-1}(\alpha x)\|=\|\alpha\Phi^{-1}(x)\|=|\alpha|\|\Phi^{-1}(x)\|.$$

三角不等式:

$$\|\Phi^{-1}(x+y)\|=\|\Phi^{-1}(x)+\Phi^{-1}(y)\|\le\|\Phi^{-1}(x)\|+\|\Phi^{-1}(y)\|.$$

所以$\|\Phi^{-1}(\cdot)\|$ 是欧空间$\mathbb{R}^{mn}$ 上的范数.

设$\|\cdot\|,\|\cdot\|_*$为任意两个矩阵范数. 则$\|\Phi^{-1}(\cdot)\|$, $\|\Phi^{-1}(\cdot)\|_*$便是欧空间$\mathbb{R}^{mn}$ 上的两个范数. 因此存在常数$0<a<b<+\infty$ 使得

$$a\|\Phi^{-1}(x)\|\le\|\Phi^{-1}(x)\|_*\le b\|\Phi^{-1}(x)\|\qquad\forall\, x\in\mathbb{R}^{mn}.$$

但任意矩阵$A \in \mathbb{R}^{m\times n}$都可表为$A=\Phi^{-1}(x), x\in\mathbb{R}^{mn}$, 这就证明了矩阵范数$\|\cdot\|$ 与$\|\cdot\|_*$在同一空间$\mathbb{R}^{m\times n}$上等价.

最后因为$\|\Phi^{-1}(\cdot)\|$和$|\cdot|$都是$\mathbb{R}^{mn}$上的范数, 故存在只依赖于mn的常数$0<C<+\infty$ 使得

$$|x|\le C\|\Phi^{-1}(x)\|,\quad \|\Phi^{-1}(x)\|\le C|x| \qquad \forall\, x\in\mathbb{R}^{mn}.$$

于是对任意$A=(a_{ij})_{m\times n}\in\mathbb{R}^{m\times n}$, 取$x=\Phi(A)$ 有

$$|a_{ij}|=|x_{(i-1)m+j}|\le |x|\le C\|\Phi^{-1}(x)\|=C\|A\|,\quad i=1,2,...,m;\quad j=1,2,...,n.$$

$$\|A\|=\|\Phi^{-1}(x)\|\le C|x|=C\Big(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\Big)^{1/2}\le C\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|. \qquad \square$$

下面的命题是研究矩阵空间拓扑性质的基础。

【命题8.3.】给定$m,n\in\mathbb{N}$. 设$\|\cdot\|$ 是$\mathbb{R}^{m\times n}$上的任意一种矩阵范数. 则$(\mathbb{R}^{m\times n},\|\cdot-\cdot\|)$是完备的度量空间.

【证】易见$(\mathbb{R}^{m\times n},\|\cdot\|)$是一个实的赋范线性空间. 因$\|\cdot-\cdot\|$ 是由$\|\cdot\|$ 诱导的度量, 故$(\mathbb{R}^{m\times n},\|\cdot-\cdot\|)$是一个度量空间。

设$\{A_p\}_{p=1}^{\infty}$是$(\mathbb{R}^{m\times n},\|\cdot-\cdot\|)$中的Cauchy 列, 即

$$\lim_{p>q\to\infty}\|A_p-A_q\|=0.$$

写$A_p=(a_{ij}^{(p)})_{m\times n}$. 则由上一命题知存在一个只与$mn$ 有关的常数$0<C<+\infty$ 使得

$$|a_{ij}^{(p)}-a_{ij}^{(q)}|\le C\|A_p-A_q\|,\quad i=1,2,...,m; j=1,2,...,n;\quad p>q\ge 1$$

因此对每一对(i,j), 数列$\{a_{ij}^{(p)}\}_{p=1}^{\infty}$ 是$\mathbb{R}$中的Cauchy 列从而在$\mathbb{R}$中收敛. 令

$$a_{ij}=\lim_{p\to\infty}a_{ij}^{(p)},\quad i=1,2,...,m; j=1,2,...,n,\quad A=(a_{ij})_{m\times n}.$$

则$A\in\mathbb{R}^{m\times n}$ 并再由上面的不等式有

$$\|A_p-A\|\le C\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}^{(p)}-a_{ij}|\to 0\quad (p\to\infty).$$

这表明$\{A_p\}_{p=1}^{\infty}$在$(\mathbb{R}^{m\times n},\|\cdot-\cdot\|)$中收敛. 所以$(\mathbb{R}^{m\times n},\|\cdot-\cdot\|)$是完备的. $\square$

矩阵乘积的连续性也是常用的性质:

设$A_p, A \in \mathbb{R}^{l\times m}, B_p, B \in \mathbb{R}^{m\times n}, p=1,2,3,...$ 且$\|A_p - A\| \to 0, \|B_p - B\| \to 0\,(p\to\infty)$. 则有$\|A_pB_p - AB\| \to 0\,(p\to\infty)$.

事实上我们有

$$\|A_pB_p - AB\| = \|(A_p - A)B_p + A(B_p - B)\| \le \|(A_p - A)B_p\| + \|A(B_p - B)\|$$
$$\le \|A_p - A\|\|B_p\| + \|A\|\|B_p - B\|$$
$$\le \|A_p - A\|(\|B_p - B\| + \|B\|) + \|A\|\|B_p - B\| \to 0 \quad (p\to\infty).$$

【常用的两种矩阵范数】

♠ 矩阵的2-范数(或欧几里德范数)：

$$\|A\| = \|A\|_2 := \Big(\sum_{i=1}^m\sum_{j=1}^n |a_{ij}|^2\Big)^{1/2}, \qquad A = (a_{ij})_{m\times n} \in \mathbb{R}^{m\times n}.$$

我们证明$\|\cdot\|$确为矩阵范数.

【证】对每个$\mathbb{R}^{m\times n}$, $\|\cdot\|$在$\mathbb{R}^{m\times n}$上显然满足正定性和正齐次性. 为证三角不等式, 如上我们把$A = (a_{ij})_{m\times n} \in \mathbb{R}^{m\times n}$ 的元素按固定方式排成一个mn 维的向量:

$$\Phi(A) = (a_{11}, a_{12}, ..., a_{1n}, a_{21}, a_{22}, ..., a_{2n}, ..., a_{n1}, a_{n2}, ..., a_{mn}).$$

上面已证$A \mapsto \Phi(A)$ 是线性的: $\Phi(A+B) = \Phi(A) + \Phi(B)$, etc. 记$|\Phi(A)|$ 为向量$\Phi(A)$的欧几里德范数. 则显然有$\|A\| = |\Phi(A)|$ 从而有

$$\|A+B\| = |\Phi(A+B)| = |\Phi(A) + \Phi(B)| \le |\Phi(A)| + |\Phi(B)| = \|A\| + \|B\|.$$

最后证明矩阵乘积的范数不等式. 对任意$l, m, n \in \mathbb{N}$ 和任意$A = (a_{ij})_{l\times m} \in \mathbb{R}^{l\times m}, B = (b_{jk})_{m\times n} \in \mathbb{R}^{m\times n}$, 由矩阵乘法有

$$AB = \Big(\sum_{j=1}^m a_{ij}b_{jk}\Big)_{1\le i\le l, 1\le k\le n} \in \mathbb{R}^{l\times n}.$$

由此和Cauchy 不等式得到

$$\|AB\|^2 = \sum_{i=1}^l\sum_{k=1}^n \Big|\sum_{j=1}^m a_{ij}b_{jk}\Big|^2 \le \sum_{i=1}^l\sum_{k=1}^n \Big(\sum_{j=1}^m |a_{ij}|^2\Big)\Big(\sum_{j=1}^m |b_{jk}|^2\Big)$$
$$= \left(\sum_{i=1}^l\sum_{j=1}^m |a_{ij}|^2\right)\left(\sum_{j=1}^m\sum_{k=1}^n |b_{jk}|^2\right) = \|A\|^2\|B\|^2.$$

所以$\|AB\| \le \|A\|\|B\|$. $\square$

♠ **矩阵的算子范数：** 把矩阵A 作为线性变换(线性算子) $x \mapsto Ax$ 而导出的范数称为算子范数, 也即对任意$m, n \in \mathbb{N}$ 定义

$$\|A\| := \sup_{0 \ne x \in \mathbb{R}^n} \frac{|Ax|}{|x|}, \qquad A \in \mathbb{R}^{m\times n}$$

其中$|x|, |Ax|$ 是向量的欧几里德范数.

我们来证明$\|\cdot\|$确为矩阵范数且$\|\cdot\| \le \|\cdot\|_2$, 后者是矩阵的2-范数.

【证】对任意$m, n \in \mathbb{N}$ 和任意$A \in \mathbb{R}^{m\times n}$, 由Cauchy 不等式有

$$\begin{aligned}
&|Ax|^2 = \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij}x_j\Big)^2 \le \sum_{i=1}^m \Big(\sum_{j=1}^n |a_{ij}|^2\Big)\big(\sum_{j=1}^n |x_j|^2\big) \\
&= \Big(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2\Big)|x|^2 = \|A\|_2^2 |x|^2 \\
&\Longrightarrow \quad |Ax| \le \|A\|_2 |x| \qquad \text{for all} \quad x \in \mathbb{R}^n \\
&\Longrightarrow \quad \|A\| = \sup_{0\ne x\in\mathbb{R}^n} \frac{|Ax|}{|x|} \le \|A\|_2.
\end{aligned}$$

这同时证明了$\|A\| < +\infty$ for all $A \in \mathbb{R}^{m\times n}$.

其次证明$\|\cdot\|$ 在每个$\mathbb{R}^{m\times n}$上是范数.

$\|\cdot\|$ 的正定性: $\|A\| = 0 \Longrightarrow |Ax| = 0$ for all $0 \ne x \in \mathbb{R}^n \Longrightarrow Ax = 0$ for all $x \in \mathbb{R}^n$ $\Longrightarrow A = 0$.

$\|\cdot\|$ 的正齐次性: 这可由$\frac{|\alpha Ax|}{|x|} = |\alpha|\frac{|Ax|}{|x|}$推出.

$\|\cdot\|$ 满足三角不等式: 这可由

$$\frac{|(A+B)x|}{|x|} \le \frac{|Ax| + |Bx|}{|x|} = \frac{|Ax|}{|x|} + \frac{|Bx|}{|x|} \le \|A\| + \|B\|$$

推出.

最后证明证矩阵乘积的范数不等式. 对任意$A \in \mathbb{R}^{l\times m}$, $B \in \mathbb{R}^{m\times n}$, 据算子范数的定义有

$$|Ay| \le \|A\||y|, \quad y \in \mathbb{R}^m; \qquad |Bx| \le \|B\||x|, \quad x \in \mathbb{R}^n.$$

两者合成: 取$y = Bx$, 即得

$$|ABx| \le \|A\||Bx| \le \|A\|\|B\||x|, \quad x \in \mathbb{R}^n.$$

两边除以$|x|(>0)$ 再取上确界即得$\|AB\|\le\|A\|\|B\|$. $\square$

注：最后这个不等式的另一个“证明”如下：

$$\forall\,x\in\mathbb{R}^n\setminus\{0\}\implies\frac{|ABx|}{|x|}=\frac{|Ay|}{|y|}\Big|_{y=Bx}\frac{|Bx|}{|x|}\le\|A\|\|B\|.$$

取上确界即得$\|AB\|\le\|A\|\|B\|$. 但其中有一个疏忽: $x\neq 0 \not\Longleftrightarrow Bx\neq 0$. 而若$Bx=0$则$|Bx|$当然不能做分母. 虽然这个证明不严谨, 但它能引导到正确的结果. 先通过各种手段猜出结果, 然后再对结果给出严格证明, 这是研究过程中常用的方法.

【顺便说明:】如果限制在$m\times 1$ 或$1\times n$矩阵空间上, 则易见矩阵的2-范数和矩阵的算子范数都等于欧几里德范数. 事实上只需验证算子范数的情形:

对任意$A=(a_1,a_2,...,a_m)^\tau\in\mathbb{R}^{m\times 1}$, 易见对任意$x\in\mathbb{R}^1=\mathbb{R}$ 有$|Ax|=|xA|=|x||A|$. 所以$\|A\|=|A|$.

对任意$A=(a_1,a_2,...,a_n)\in\mathbb{R}^{1\times n}$, 由矩阵乘法和Cauchy不等式有

$$\frac{|Ax|}{|x|}=\frac{|Ax|}{|x|}=\frac{|\sum_{j=1}^n a_jx_j|}{|x|}\le\Big(\sum_{j=1}^n|a_j|^2\Big)^{1/2}=|A|$$

因此$\|A\|\le|A|$. 当$A\neq 0$ 时, 取$x=A^\tau$ 有

$$\|A\|\ge\frac{|AA^\tau|}{|A^\tau|}=\frac{|AA^\tau|}{|A|}=\frac{|A|^2}{|A|}=|A|.$$

所以$\|A\|=|A|$. $\square$

【约定】由于同一矩阵空间上的范数之间彼此等价, 因此使用特殊范数即可. 本讲义下面内容中只使用上述两种范数$\|\cdot\|$, 即$\|\cdot\|$是**矩阵的2-范数**或**矩阵的算子范数**. 根据这两种范数的定义易见有

$$\|Ax\|\le\|A\||x|\qquad\forall\,A\in\mathbb{R}^{m\times n},\ \forall\,x\in\mathbb{R}^{n\times 1}=\mathbb{R}^n.\qquad\square$$

【线性变换的Lipschitz 连续性】设$L\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$. 据线性变换的矩阵表示有$L(x)=Ax$, 其中$A\in\mathbb{R}^{m\times n}$. 设$\|\cdot\|$为一矩阵范数. 则由$L(x)-L(y)=Ax-Ay=A(x-y)$ 得到Lipschitz连续性估计:

$$|L(x)-L(y)|\le\|A\||x-y|,\quad\forall\,x,y\in\mathbb{R}^n.$$

【矩阵逆运算的连续性】

【命题8.4.】设$\|\cdot\|$是一矩阵范数. 则矩阵逆运算

$$\mathbb{R}^{n\times n}\ni A\mapsto A^{-1}$$

按范数$\|\cdot\|$ 是连续的, 即$\lim\limits_{B\to A}B^{-1}=A^{-1}$.

具体来说, 设$A\in\mathbb{R}^{n\times n}$ 可逆. 则当$B\in\mathbb{R}^{n\times n}$ 满足

$$\|B-A\|\le\frac{1}{2\|A^{-1}\|}$$

时, B 也可逆且

$$\|B^{-1}-A^{-1}\|\le 2\|A^{-1}\|^2\|B-A\|.$$

【证】设A 可逆, B 满足$\|B-A\|\le\frac{1}{2\|A^{-1}\|}$. 令$\alpha=\frac{1}{\|A^{-1}\|},\beta=\|B-A\|$. 则$\beta<\alpha$. $\forall\, x\in\mathbb{R}^n\equiv\mathbb{R}^{n\times 1}$ 来估计

$$\begin{aligned}&\alpha|x|=\alpha|A^{-1}Ax|\le\alpha\|A^{-1}\||Ax|=|Ax|\le|(A-B)x|+|Bx|\\&\le\|A-B\||x|+|Bx|=\beta|x|+|Bx|\end{aligned}$$

$$\Longrightarrow\ (\alpha-\beta)|x|\le|Bx|\qquad\forall\, x\in\mathbb{R}^n.$$

因$\alpha-\beta>0$ 故上式蕴含方程$Bx=0$ 只有零解. 因此B 可逆.

下证连续性估计：容易证明下列关系式:

$$B^{-1}-A^{-1}=B^{-1}(A-B)A^{-1}.\tag{$*$}$$

如将A^{-1} 移至右边则还有

$$B^{-1}=A^{-1}+B^{-1}(A-B)A^{-1}.$$

从而由矩阵乘积的范数不等式和$\|A-B\|\|A^{-1}\|\le\frac{1}{2}$得到

$$\|B^{-1}\|\le\|A^{-1}\|+\|B^{-1}\|\|A-B\|\|A^{-1}\|\le\|A^{-1}\|+\frac{1}{2}\|B^{-1}\|.$$

这给出B^{-1} 的范数估计:

$$\|B^{-1}\|\le 2\|A^{-1}\|.$$

于是由$(*)$即得

$$\|B^{-1}-A^{-1}\|\le\|B^{-1}\|\|A-B\|\|A^{-1}\|\le 2\|A^{-1}\|^2\|B-A\|.\qquad\square$$

【方阵的无穷级数】考虑$m=n$的情形. 此时对任意$A,B\in\mathbb{R}^{n\times n}$, 乘积$AB, BA, A^k, A^kB^l$等等仍属于$\mathbb{R}^{n\times n}$.

【定义(收敛和绝对收敛)】设$A_k\in\mathbb{R}^{n\times n}, k=0,1,2,...,$. 我们称矩阵级数$\sum\limits_{k=0}^{\infty}A_k$收敛如果存在$S\in\mathbb{R}^{n\times n}$ 使得

$$\lim_{N\to\infty}\Big\|S-\sum_{k=0}^{N}A_k\Big\|=0.$$

此时也称$\sum\limits_{k=0}^{\infty}A_k$ 收敛于S 并记

$$S=\lim_{N\to\infty}\sum_{k=0}^{N}A_k=\sum_{k=0}^{\infty}A_k.$$

进一步, 若

$$\sum_{k=0}^{\infty}\|A_k\|<+\infty$$

则称矩阵级数$\sum\limits_{k=0}^{\infty}A_k$绝对收敛. $\Box$

易见绝对收敛蕴含收敛: 我们有

$$\Big\|\sum_{k=0}^{p}A_k-\sum_{k=0}^{q}A_k\Big\|=\Big\|\sum_{k=q+1}^{p}A_k\Big\|\le\sum_{k=q+1}^{p}\|A_k\|\to 0\quad\text{as}\quad p>q\to\infty.$$

这说明$\{\sum_{k=0}^{p}A_k\}_{p=1}^{\infty}$是Caushy 列. 据$(\mathbb{R}^{n\times n},\|\cdot-\cdot\|)$的完备性知级数$\sum\limits_{k=0}^{\infty}A_k$收敛. 注意:当绝对收敛时还有

$$\Big\|\sum_{k=0}^{\infty}A_k\Big\|\le\sum_{k=0}^{\infty}\|A_k\|$$

它是不等式$\|\sum_{k=0}^{N}A_k\|\le\sum_{k=0}^{N}\|A_k\|$ 和$N\to\infty$的结果.

本讲义对矩阵级数的应用主要涉及幂级数. 即$A_k=a_kA^k$, 其中$a_k\in\mathbb{R}, A\in\mathbb{R}^{n\times n}$ 并且主要考虑绝对收敛的情形.

【命题8.5(矩阵级数的乘法).】设$a_k,b_k\in\mathbb{R}, A,B\in\mathbb{R}^{n\times n}$ 且A,B 可交换: $AB=BA$. 若$\sum\limits_{k=0}^{\infty}a_kA^k$ 和$\sum\limits_{k=0}^{\infty}b_kB^k$都绝对收敛, 则有

$$\Big(\sum_{k=0}^{\infty}a_kA^k\Big)\Big(\sum_{k=0}^{\infty}b_kB^k\Big)=\sum_{k=0}^{\infty}\Big(\sum_{j=0}^{k}a_{k-j}b_jA^{k-j}B^j\Big).$$

【证】由于A,B 可交换, 故下面的计算方式就与数值级数的情形完全相同. 对任意$N\in\mathbb{N}$, 计算

$$
\begin{aligned}
&\Big(\sum_{k=0}^{N}a_kA^k\Big)\Big(\sum_{k=0}^{N}b_kB^k\Big)=\sum_{k=0}^{N}\sum_{j=0}^{N}a_kA^kb_jB^j\\
&=\sum_{k\ge0,j\ge0,\,k+j\le N}^{N}a_kA^kb_jB^j+\sum_{0\le k\le N,0\le j\le N,\,k+j\ge N+1}^{N}a_kA^kb_jB^j\\
&:=\sum_{\le N}+\sum_{\ge N+1},
\end{aligned}
$$

$$
\sum_{\le N}=\sum_{k=0}^{N}\sum_{i+j=k}a_ib_jA^iB^j=\sum_{k=0}^{N}\Big(\sum_{j=0}^{k}a_{k-j}b_jA^{k-j}B^j\Big),
$$

$$
\begin{aligned}
&\Big\|\sum_{\ge N+1}\Big\|\le\sum_{0\le k\le N,0\le j\le N,\,k+j\ge N+1}^{N}|a_k|\|A^k\||b_j|\|B^j\|\\
&\le\sum_{0\le j\le N,\,N\ge k\ge (N+1)/2}^{N}|a_k|\|A^k\||b_j|\|B^j\|+\sum_{0\le k\le N,\,N\ge j\ge N/2}^{N}|a_k|\|A^k\||b_j|\|B^j\|\\
&\le\Big(\sum_{k=[(N+1)/2]}^{N}|a_k|\|A^k\|\Big)\Big(\sum_{j=0}^{N}|b_j|\|B^j\|\Big)+\Big(\sum_{k=0}^{N}|a_k|\|A^k\|\Big)\Big(\sum_{j=[(N+1)/2]}^{N}|b_j|\|B^j\|\Big)\\
&\to0\quad\text{as}\quad N\to\infty.
\end{aligned}
$$

于是得到

$$
\Big\|\Big(\sum_{k=0}^{N}a_kA^k\Big)\Big(\sum_{k=0}^{N}b_kB^k\Big)-\sum_{k=0}^{N}\Big(\sum_{j=0}^{k}a_{k-j}b_jA^{k-j}B^j\Big)\Big\|\le\Big\|\sum_{\ge N+1}\Big\|\longrightarrow0\quad(N\to\infty).
$$

因此

$$
\lim_{N\to\infty}\sum_{k=0}^{N}\Big(\sum_{j=0}^{k}a_{k-j}b_jA^{k-j}B^j\Big)=\lim_{N\to\infty}\Big(\sum_{k=0}^{N}a_kA^k\Big)\Big(\sum_{k=0}^{N}b_kB^k\Big)=\Big(\sum_{k=0}^{\infty}a_kA^k\Big)\Big(\sum_{k=0}^{\infty}b_kB^k\Big).
$$

这证明了矩阵级数$\sum_{k=0}^{\infty}(\sum_{j=0}^{k}a_{k-j}b_jA^{k-j}B^j)$收敛且所证等式成立. □

易见对任意$A\in\mathbb{R}^{n\times n}$ 和任意$k\in\mathbb{N}$ 有

$$
\|A^k\|\le\|A\|^k.
$$

[细: $\|A^k\|=\|A^{k-1}A\|\le\|A^{k-1}\|\|A\|\le\|A^{k-2}\|\|A\|^2\cdots\le\|A\|^k$.]

【**例** (矩阵空间指数映射)】对任意$A\in\mathbb{R}^{n\times n}$ 有绝对收敛:
$$\sum_{k=0}^{\infty}\frac{\|A^k\|}{k!}\le\sum_{k=0}^{\infty}\frac{\|A\|^k}{k!}=e^{\|A\|}<+\infty.$$
因此可以定义指数映射$\exp:\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}$ 如下
$$\exp(A)=\sum_{k=0}^{\infty}\frac{A^k}{k!}=\mathrm{I}+A+\frac{A^2}{2!}+\cdots,\quad A\in\mathbb{R}^{n\times n}.$$
如果$A,B\in\mathbb{R}^{n\times n}$可交换, 则有
$$\exp(A+B)=\exp(A)\exp(B).$$
事实上由上面命题和绝对收敛以及A,B 可交换有
$$\begin{aligned}\exp(A)\exp(B)&=\sum_{k=0}^{\infty}\Big(\sum_{j=0}^{k}\frac{1}{(k-j)!}\frac{1}{j!}A^{k-j}B^j\Big)\\&=\sum_{k=0}^{\infty}\frac{1}{k!}\Big(\sum_{j=0}^{k}\frac{k!}{(k-j)!j!}A^{k-j}B^j\Big)\\&=\sum_{k=0}^{\infty}\frac{1}{k!}(A+B)^k=\exp(A+B).\end{aligned}$$
□

作业题

1. 设$\mathrm{I}\in\mathbb{R}^{n\times n}$是单位矩阵. $\|\cdot\|,\|\cdot\|_2$ 分别是矩阵的算子范数和2-范数. 证明 $\|I\|=1$, $\|I\|_2=\sqrt{n}$. 一般地, 设$A=(a_{ij})_{n\times n}\in\mathbb{R}^{n\times n}$. 证明
$$\|A\|_2=\sqrt{\mathrm{trace}(A^\tau A)}$$
因此矩阵的2-范数也叫**迹范数**. 此外证明算子范数和2-范数(迹范数) 有具体的互相控制关系(见陈书习题8.1.4, 第二册page 107):
$$\|A\|\le\|A\|_2\le\sqrt{n}\|A\|.$$

2. 陈书习题8.1.2(i),(ii) (第二册page 106).

3. 设$\|\cdot\|$ 是矩阵的算子范数或矩阵的2-范数. 设$A\in\mathbb{R}^{n\times n}$ 满足$\|A\|<1$. 证明$\mathrm{I}-A$ 可逆并且
$$\|(\mathrm{I}-A)^{-1}\|\le\frac{\|\mathrm{I}\|}{1-\|A\|}.$$

一般地, 证明: 若存在$k \in \mathbb{N}$ 使得$\|A^k\| < 1$, 则$\mathrm{I} - A$ 可逆并且

$$\|(\mathrm{I} - A)^{-1}\| \le \frac{\|\mathrm{I}\| + \|A\| + \cdots + \|A^{k-1}\|}{1 - \|A^k\|}.$$

这里I 是单位矩阵.

4. $\|\cdot\|$ 是矩阵的算子范数, $A \in \mathbb{R}^{n\times n}$. 证明:

A 是正交阵 $\iff$ A 可逆且$\|A\| = \|A^{-1}\| = 1$.

[回忆: 矩阵$A \in \mathbb{R}^{n\times n}$称为正交阵如果它满足$A^\tau A = AA^\tau = \mathrm{I}$, 或等价地, A可逆且$A^{-1} = A^\tau$, 或等价地, 线性变换$x \mapsto Ax$ 是保距的(或保内积的), 见下面习题.]

5(正交变换的特征). 设映射$f: \mathbb{R}^n \to \mathbb{R}^n$ 满足$f(0) = 0$ (不要求f是线性的). 则以下(a),(b),(c) 等价:

(a) f 是一个正交线性变换, 即存在正交矩阵$A \in \mathbb{R}^n$使得$f(x) \equiv Ax$.

(b) f 保距, 即

$$|f(x) - f(y)| = |x - y| \qquad \forall\, x, y \in \mathbb{R}^n.$$

(c) f 保内积, 即

$$\langle f(x), f(y)\rangle = \langle x, y\rangle \qquad \forall\, x, y \in \mathbb{R}^n.$$

6. 设$A \in \mathbb{R}^{n\times n}$, 用$A^\tau$ 表示A的转置.

(1) 证明

$$\exp(A)\exp(-A) = \mathrm{I} \qquad \text{因此}\exp(A)\ \text{总是可逆阵且}\ (\exp(A))^{-1} = \exp(-A).$$

(2) 证明

$$(\exp(A))^\tau = \exp(A^\tau).$$

(3) 假设A 是反对称阵, 即$A^\tau + A = 0$. 证明$\exp(A)$是正交阵.

7. 陈书习题8.1.5(ii),(iii). [第二册pp. 107-108]

8. 设$2 \le n \in \mathbb{N}$. 证明: 当且仅当n 为偶数时, 存在$\mathcal{A} \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$ 使得

$$\mathcal{A} \circ \mathcal{A}(x) \equiv -x\,, \qquad x \in \mathbb{R}^n.$$

[化成矩阵来看.]

§8.2 可微性,微分和偏导数

我们先回顾一元(即单变量)向量值函数的可微性, 以$\mathbb{R}^3$中质点的运动轨迹为例: 设一质点的运动轨迹(即参数曲线)为

$$S=\{\mathbf{r}(t)\mid \mathbf{r}(t)=(x(t),y(t),z(t))^{\tau}\in\mathbb{R}^3, t\in[a,b]\}.$$

我们称参数曲线$\mathbf{r}$在$t\in[a,b]$ 处可微, 如果存在向量$\mathbf{r}'(t)\in\mathbb{R}^3$, 使得

$$\mathbf{r}(t+h)=\mathbf{r}(t)+\mathbf{r}'(t)h+o(h)\qquad (t+h\in[a,b],\ h\to 0).$$

它也等价于差商极限

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}(t)=\mathbf{r}'(t):=\lim_{t+h\in[a,b],\,h\to 0}\frac{\mathbf{r}(t+h)-\mathbf{r}(t)}{h}\quad \text{存在且属于}\mathbb{R}^3.$$

$\mathbf{r}'(t)$叫做向量值函数$\mathbf{r}$ 在时刻t的导数或微分。在物理上, 导数$\mathbf{r}'(t)$也称为质点在t时刻的速度, 记作$\mathbf{r}'(t)=\mathbf{v}(t)$.

因欧空间中, 向量有极限等价于向量的每个坐标函数都有极限, 故$\mathbf{r}(\cdot)$在t可微等价于每个坐标函数$x(\cdot),y(\cdot),z(\cdot)$在$t$可微. 此时有

$$\mathbf{r}'(t)=\begin{pmatrix}x'(t)\\ y'(t)\\ z'(t)\end{pmatrix}.$$

$x'(t),y'(t),z'(t)$ 分别是质点沿x-轴方向, y-轴方向和z-轴方向运动的速度.

由此例看出, 一元向量值函数(即**曲线**)的可微性和表示与一元实值函数的情形没有本质区别, 都是一维行为.

【切线的观点】在曲线S上取一点$\mathbf{p}$, 即$\mathbf{p}=\mathbf{r}(t_0), t_0\in[a,b]$. 考察曲线在点$\mathbf{p}$附近的行为, 主要是看$S$在$\mathbf{p}$处的切线的样子. 为分析方便, 我们假设$t_0\in(a,b)$. 此外假设这曲线$S$是简单的可微曲线且$t_0$ 不是参数奇点, 即$t\mapsto\mathbf{r}(t)$ 在$[a,b]$上是单射且可微并且$\mathbf{r}'(t_0)\neq 0$.

因有界闭区间$[a,b]$ 是紧集, 故逆映射$\mathbf{r}^{-1}:S\to[a,b]$ 也连续.[1]

设在同一曲线S上另一个质点的运动轨迹为$s\mapsto\gamma(s)\in S, s\in(-\delta,\delta)$ $(\delta>0)$. 我们假定γ 在$(-\delta,\delta)$上可微且经过点$\mathbf{p}$, 即$\gamma(0)=\mathbf{p}$. 导数$\gamma'(0)$ 是这个新质点在$\mathbf{p}$处的速度. 我们要找到两个速度$\gamma'(0)$与$\mathbf{r}'(t_0)$的关系.

[1] **因可微蕴含连续, 故$\mathbf{r}$在紧集$[a,b]$上连续, 从而逆映射$\mathbf{r}^{-1}:S\to[a,b]$连续.**

注意到$S=\mathbf{r}([a,b])$, 即γ 位于$\mathbf{r}$的值域, 故可令

$$t(s)=\mathbf{r}^{-1}(\gamma(s)),\quad s\in(-\delta,\delta).$$

由复合函数连续性知$t(s)$在$s\in(-\delta,\delta)$内连续, 且有

$$\lim_{s\to 0}t(s)=t(0)=\mathbf{r}^{-1}(\gamma(0))=\mathbf{r}^{-1}(\mathbf{p})=t_0,\quad \gamma(s)=\mathbf{r}(t(s)),\quad s\in(-\delta,\delta).$$

如果$\gamma'(0)=\mathbf{0}$ (零向量), 则有$\gamma'(0)=\mathbf{0}=0\mathbf{r}'(t_0)=\mathbf{r}'(t_0)0$.

设$\gamma'(0)\neq\mathbf{0}$. 则当$0<|s|<<\delta$ 时有$\gamma(s)\neq\gamma(0)$ 从而有$t(s)\neq t(0)$. 据此我们计算

$$\frac{\gamma(s)-\gamma(0)}{s-0}=\frac{\mathbf{r}(t(s))-\mathbf{r}(t(0))}{s-0}=\frac{\mathbf{r}(t(s))-\mathbf{r}(t(0))}{t(s)-t(0)}\cdot\frac{t(s)-t(0)}{s-0},$$

$$\Big\langle\frac{\mathbf{r}(t(s))-\mathbf{r}(t(0))}{t(s)-t(0)},\ \frac{\gamma(s)-\gamma(0)}{s-0}\Big\rangle=\Big|\frac{\mathbf{r}(t(s))-\mathbf{r}(t(0))}{t(s)-t(0)}\Big|^2\frac{t(s)-t(0)}{s-0}$$

$$\frac{\Big\langle\frac{\mathbf{r}(t(s))-\mathbf{r}(t(0))}{t(s)-t(0)},\ \frac{\gamma(s)-\gamma(0)}{s-0}\Big\rangle}{\Big|\frac{\mathbf{r}(t(s))-\mathbf{r}(t(0))}{t(s)-t(0)}\Big|^2}=\frac{t(s)-t(0)}{s-0}.$$

令$s\to 0$ 并注意$0\neq t(s)-t_0=t(s)-t(0)\to 0$ 得到

$$t'(0)=\lim_{s\to 0}\frac{t(s)-t(0)}{s-0}=\frac{\langle\mathbf{r}'(t_0),\,\gamma'(0)\rangle}{|\mathbf{r}'(t_0)|^2}.$$

这证明了$s\mapsto t(s)$在$s=0$ 可微. 于是再由$\gamma(s)=\mathbf{r}(t(s))$ 和复合函数求导法则知

$$\gamma'(0)=\mathbf{r}'(t(0))t'(0)=\mathbf{r}'(t_0)h.$$

因$h=t'(0)$是实数, 这表明$\gamma'(0)$ 与$\mathbf{r}'(t_0)$ 共线.

反之, 对任意$h\in\mathbb{R}$, 令

$$\delta=\frac{1}{1+|h|}\min\{t_0-a,b-t_0\},\quad \gamma(s)=\mathbf{r}(t_0+hs),\quad s\in(-\delta,\delta).$$

则由复合映射性质知$\gamma:(-\delta,\delta)\to S$ 可微且$\gamma(0)=\mathbf{r}(0)=\mathbf{p},\gamma'(0)=\mathbf{r}'(t_0)h$.

这样一来我们可以定义S在点$\mathbf{p}\in S$ 处的**切空间** $TS_{\mathbf{p}}$ 如下:

$$TS_{\mathbf{p}}=\{\gamma'(0)\mid \text{存在}\delta>0\text{ 和可微映射 }\gamma:(-\delta,\delta)\to S\ \text{ 使得 }\gamma(0)=\mathbf{p}\}$$

并由上面分析的结果知

$$TS_{\mathbf{p}}=\{\mathbf{r}'(t_0)h\mid h\in\mathbb{R}\}$$

它是一个一维线性空间. 这与切线本质上一致：二者是平行的直线(即只差一个平移):

$$S\text{在点}\mathbf{p}=\mathbf{r}(t_0)\text{的切线}=\{\mathbf{p}+\mathbf{r}'(t_0)h \mid h\in\mathbb{R}\}=\mathbf{p}+TS_{\mathbf{p}}.$$

【切平面的观点】为了把一元函数的微分的概念推广到多元函数上, 一个自然的做法是, 先从几何出发把切线推广到切平面, 然后再看应如何定义多元函数的可微性。我们以$\mathbb{R}^3$中的显示曲面为例, 设$D\subset\mathbb{R}^2$为一开集, $f:D\to\mathbb{R}$ 连续. 称

$$S=\{(x,y,z)^\tau \mid z=f(x,y),\,(x,y)\in D\}$$

为一个显示曲面. 为分析方便, 我们取

$$D=\mathbb{R}^2,\quad f(x,y)=x^2+y^2$$

即S 是一个旋转抛物面:

$$S=\{(x,y,z)^\tau \mid z=x^2+y^2,\,(x,y)\in\mathbb{R}^2\}.$$

任取一点$\mathbf{p}=(x_0,y_0,z_0)\in S$. 设想在$S$上有很多蚂蚁, 它们以不同速度都经过点$\mathbf{p}$. 每个蚂蚁的轨迹是一条位于曲面$S$上的曲线。我们把所有蚂蚁在$\mathbf{p}$处的速度的集合$TS_{\mathbf{p}}$叫做曲面$S$在点$\mathbf{p}$的**切空间**[2], 即(与上面完全一样)

$$TS_{\mathbf{p}}=\{\gamma'(0)\mid \text{存在}\delta>0\ \text{和可微映射}\ \gamma:(-\delta,\delta)\to S\ \text{使得}\ \gamma(0)=\mathbf{p}\,\}.$$

把括号$\{\cdots\}$中的可微映射$\gamma:(-\delta,\delta)\to S$ 写作$\gamma(t)=(x(t),y(t),z(t))^\tau$, 则由$S$的定义有$z(t)=x(t)^2+y(t)^2$. 因此

$$\gamma(t)=\begin{pmatrix} x(t)\\ y(t)\\ x(t)^2+y(t)^2\end{pmatrix},\quad \gamma'(t)=\begin{pmatrix} x'(t)\\ y'(t)\\ 2x(t)x'(t)+2y(t)y'(t)\end{pmatrix}$$

$$\gamma'(0)=\begin{pmatrix} x'(0)\\ y'(0)\\ 2x(0)x'(0)+2y(0)y'(0)\end{pmatrix}=\begin{pmatrix} 1 & 0\\ 0 & 1\\ 2x_0 & 2y_0\end{pmatrix}\begin{pmatrix} x'(0)\\ y'(0)\end{pmatrix}.$$

如令

$$A=\begin{pmatrix} 1 & 0\\ 0 & 1\\ 2x_0 & 2y_0\end{pmatrix},\qquad h=\begin{pmatrix} x'(0)\\ y'(0)\end{pmatrix}$$

[2] **顾名思义：切空间$TS_{\mathbf{p}}$就是由曲面S在点$\mathbf{p}\in S$ 处的所有切向量构成的空间.**

则上式写为

$$\gamma'(0) = Ah.$$

反之对任意$h = (h_1, h_2)^\tau \in \mathbb{R}^2$, 取

$$\gamma(t) = \begin{pmatrix} x_0 + h_1 t \\ y_0 + h_2 t \\ (x_0 + h_1 t)^2 + (y_0 + h_2 t)^2 \end{pmatrix}$$

则$\gamma : (-1, 1) \to S$ 且$\gamma'(0) = Ah$. 于是曲面S在点$\mathbf{p}$的切空间$TS_{\mathbf{p}}$ 等于

$$TS_{\mathbf{p}} = \{Ah \mid h \in \mathbb{R}^2\}.$$

把切空间$TS_{\mathbf{p}}$做平移使得$TS_{\mathbf{p}}$的原点平移到点$\mathbf{p}$ 就得到切平面:

$$\text{曲面}S\text{在点}\mathbf{p}\text{的切平面为} \quad \mathbf{p} + TS_{\mathbf{p}} = \{\mathbf{p} + Ah \,|\, h \in \mathbb{R}^2\}.$$

注意到$\mathbf{p} + Ah = (x_0 + h_1, y_0 + h_2, x_0^2 + y_0^2 + 2x_0h_1 + 2y_0h_2)^\tau$, 也即这向量的前两个分量无约束, 是自由变量, 只有第三个分量是前两个分量的线性函数. 因此这个切平面集合也可用第三个分量的函数关系来描述, 即用一个三元一次方程表示:

$$\text{曲面}S\text{在点}\mathbf{p}\text{的切平面方程为} \quad z = x_0^2 + y_0^2 + 2x_0h_1 + 2y_0h_2, \quad (h_1, h_2) \in \mathbb{R}^2.$$

作为对照, 我们把函数$f(x, y) = x^2 + y^2$ 在(x_0, y_0) 点作展开, 则有

$$(x_0 + h_1)^2 + (y_0 + h_2)^2 = x_0^2 + y_0^2 + 2x_0h_1 + 2y_0h_2 + h_1^2 + h_2^2, \quad (h_1, h_2) \in \mathbb{R}^2.$$

由此看出(结合图像), 在点$\mathbf{p} = (x_0, y_0, z_0) \in S$ 附近, S上的点与S在$\mathbf{p}$处的切平面上的点的距离是自由变量的增量$h = (h_1, h_2)$的绝对值$|h|$的高阶无穷小:

$$\frac{|(x_0 + h_1)^2 + (y_0 + h_2)^2 - (x_0^2 + y_0^2) - (2x_0h_1 + 2y_0h_2)|}{|h|} = |h| \to 0 \quad (h \to 0). \qquad (*)$$

如令

$$F(x, y) = \begin{pmatrix} x \\ y \\ f(x, y) \end{pmatrix} = \begin{pmatrix} x \\ y \\ x^2 + y^2 \end{pmatrix},$$

则$(*)$ 也可用映射$(x, y) \mapsto F(x, y)$写成

$$\frac{|F(x_0 + h_1, x_0 + h_2) - F(x_0, y_0) - Ah|}{|h|} = |h| \to 0 \quad (h \to 0).$$

这个“曲面上点**p** 附近的点与该曲面在**p**处的切平面上的点的距离是自由变量的增量的高阶无穷小”的性质就是“切”的概念, 也即“可微”的概念。

下面我们将给出可微的定义. 为了能包括函数在定义域的边界上某些点处的可微性(类似于一元函数在区间端点的单侧导数), 需要引进

【集合的锥点】设$E \subset \mathbb{R}^n$. 称$x \in E$是E的一个锥点(a cone point of E), 如果存在可逆矩阵$C \in \mathbb{R}^{n\times n}$和$\delta > 0$ 使得$x + Ct \subset E$ for all $t \in [0, \delta]^n$. □

易见E**的内点都是**E**的锥点**, 因为此时可以取$C = \mathrm{I}$(单位矩阵). 闭区间$\prod_{i=1}^n [a_i, b_i]$ $(a_i < b_i)$ 的边界点都是该区间的锥点.

【定义(可微与微分)】设$E \subset \mathbb{R}^n$, $f : E \to \mathbb{R}^m$, x 是E的一个锥点(例如x是E的内点). 我们称映射f在点x 可微, 如果存在一个线性变换(看作矩阵) $A = A_x \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ (看作$\mathbb{R}^{m\times n}$) 使得

$$\lim_{x+h\in E,\, h\to 0} \frac{|f(x + h) - f(x) - Ah|}{|h|} = 0$$

或等阶地

$$\lim_{E\ni y\to x} \frac{|f(y) - f(x) - A(y - x)|}{|y - x|} = 0.$$

这里$h, y - x$ 是列向量。

若x是E的内点, 即存在$\delta_x > 0$ 使得$B(x, \delta_x) \subset E$, 则以上极限可以写成展开式的形式:

$$f(x + h) = f(x) + Ah + \alpha(x; h), \qquad |h| < \delta_x$$

这里$\alpha(x, h)$ 是h 的**高阶无穷小**, 即

$$\alpha(x; h) = o(h) \qquad \text{i.e.} \qquad \lim_{h\to 0} \frac{|\alpha(x; h)|}{|h|} = 0.$$

当f 在锥点x 可微时, 称x对应的线性变换A_x 为f在点x的**微分**, 或**切映射**, 或**导映射**, 记为

$$A_x = \mathrm{d}f_x = \mathrm{d}f(x) = Df(x) = f'(x).$$

在这些记号下, 例如对于记号$A_x = f'(x)$, 上面关于可微性的关系式便写成

$$\lim_{x+h\in E,\, h\to 0} \frac{|f(x + h) - f(x) - f'(x)h|}{|h|} = 0$$

或等阶地

$$\lim_{E\ni y\to x} \frac{|f(y) - f(x) - f'(x)(y - x)|}{|y - x|} = 0.$$

以及

$$f(x+h)=f(x)+f'(x)h+\alpha(x;h)$$

其中$\alpha(x;h)$ 是$h\to 0$ 的高阶无穷小: $\lim\limits_{x+h\in E,\,h\to 0}|\alpha(x;h)|/|h|=0.$

如果E的每一点都是锥点且f在E上处处可微, 则称f 在E上可微. □

【**注1**】关于微分的记号和意义: 因为我们把一个线性变换等同于对它的表示形式(例如矩阵形式), 故这些微分记号$\mathrm{d}f_x, \mathrm{d}f(x), Df(x), f'(x)$就有两个身份:

一个是**表面身份**, 即对$\mathrm{d}f_x, \mathrm{d}f(x), Df(x), f'(x)$所采用的表示形式, 例如常用的表示形式是矩阵, 称之为Jacobi 矩阵(详见下面).

一个是**内在身份**, 即$\mathrm{d}f_x, \mathrm{d}f(x), Df(x), f'(x)$是从$f$的自变量$x$的增量$h$所在的线性空间$\mathbb{R}^n$ 到$\mathbb{R}^m$的**线性映射**.

无论这些映射$\mathrm{d}f_x, \mathrm{d}f(x), Df(x), f'(x)$用什么表示, 它们的映射值是相同的, 即都是$\mathbb{R}^m$是的向量, 即当这些映射作用到$h$时恒有

$$\mathrm{d}f_x(h)=\mathrm{d}f(x)(h)=Df(x)h=f'(x)h\in\mathbb{R}^m\qquad\forall\, h\in\mathbb{R}^n. \tag{$\clubsuit$}$$

如果只写映射关系而不写自变量h, 则上式被抽象为

$$\mathrm{d}f_x(\cdot)=\mathrm{d}f(x)(\cdot)=Df(x)\cdot=f'(x)\cdot\,. \tag{$\spadesuit$}$$

其中一点"$\cdot$"表示自变量h被填入的位置. 但是, 就像把一个具体的矩阵A看成是一个线性变换$h\mapsto A(h)=Ah$ 一样, $A(\cdot)$中的"$(\cdot)$"经常被省略, 即简写成A. 记住这点, 我们就可以省略"$(\cdot)$"和"$\cdot$"而写成

$$\mathrm{d}f_x=\mathrm{d}f(x)=Df(x)=f'(x).$$

【**注2**】很多教科书中只考虑在E的内点处定义微分, 因此直接假定E是开集. 我们考虑锥点的目的是要在E的边界上某些点处定义微分: 若$x\in\partial E$是E的锥点, 则可以考察f在x处的可微性. 这与一元函数在闭区间的端点处是否具有单侧导数一样.

但是需要证明上述微分(切映射、或导映射)作为x的映射是良好定义的(well defined), 也即需要证明在E的任一锥点x处,

微分$A_x=f'(x)$由f和x唯一决定.

(注: 如果x不是锥点而只是聚点, 则以上定义的微分不必唯一, 从而产生歧义. 因此我们只对锥点定义微分.)

事实上由锥点的定义, 存在可逆矩阵$C \in \mathbb{R}^{n\times n}$ 和$\delta > 0$使得$x + Ct \in E$ for all $t = (t_1, t_2, ..., t_n)^\tau \in [0, \delta]^n$. 如写$C = (\mathbf{c}_1\ \mathbf{c}_2\ ...\ \mathbf{c}_n)$ 其中$\mathbf{c}_j \in \mathbb{R}^{n\times 1}$为列向量, 则$\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n$ 线性无关且$Ct = \sum_{j=1}^{n} t_j \mathbf{c}_j$.

假设另有$B_x \in \mathbb{R}^{m\times n}$也满足微分的定义. 令$H = B_x - A_x$, 则有

$$Hh = o(h)$$

意即: 对任意$\varepsilon > 0$ 存在$\eta > 0$ 使得

$$\text{对所有满足}x + h \in E \text{ 和}|h| < \eta\text{的}h\text{都有} \quad |Hh| \le \varepsilon|h|.$$

取$t = (t_1, 0, ..., 0)^\tau$ 则$Ct = t_1\mathbf{c}_1$. 当$0 < t_1 < \min\{\delta, \frac{\eta}{|\mathbf{c}_1|}\}$时有$x + Ct \in E$ 且$|Ct| = t_1|\mathbf{c}_1| < \eta$ 从而有

$$t_1|H\mathbf{c}_1| = |Ht_1\mathbf{c}_1| = |HCt| \le \varepsilon|Ct| = \varepsilon t_1|\mathbf{c}_1|.$$

消去$t_1 > 0$ 得到$|H\mathbf{c}_1| \le \varepsilon|\mathbf{c}_1|$. 因$\varepsilon > 0$ 可以任意小, 令$\varepsilon \to 0^+$ 推出$|H\mathbf{c}_1| = 0$. 同理对$(0, t_2, 0, ..., 0)^\tau, ..., (0, 0, ..., 0, t_n)^\tau$ 应用上述论证可得$H\mathbf{c}_i = 0, i = 1, 2, ..., n$. 这就给出$HC = (H\mathbf{c}_1\ H\mathbf{c}_2\ ...\ H\mathbf{c}_m) = 0$. 但$C$可逆, 故得$H = 0$, 即$B_x = A_x$.

【注3】关于高阶无穷小的记号: 在数学表达和推导中, 为了突出主要因素, 经常把h的高阶无穷小$\alpha(x; h)$写成$o(h)$, 意即$o(h)$ 是满足下面极限关系的小向量:

$$\lim_{h\to 0} \frac{|o(h)|}{|h|} = 0$$

这里极限"$h \to 0$"在具体情况中可能是被约束的极限: 例如要求h 在满足$x+h \in E$的条件下趋于零. 当然若x 是E的内点, 则无约束, 只需$|h| << 1$.

按这个约定易见: 若A, B 是常数或常向量或常矩阵, $\varphi(h)$ 是函数或向量值函数或矩阵函数且$\varphi(h)$在$|h| << 1$ 内有界, 它们使得加法、乘法$Ao(h) + Bo(h), \varphi(h)o(h)$ 有意义, 则有

$$Ao(h) + Bo(h) = o(h), \quad \varphi(h)o(h) = o(h).$$

这是因为, 例如设$A, \varphi(h)$ 是向量或矩阵, 则有

$$|Ao(h)| \le \|A\||o(h)|, \quad \|\varphi(h)o(h)\| \le \|\varphi(h)\||o(h)| \le C|o(h)|.$$

【注4】今后凡说"映射f在点$x \in E$ 可微"时, 这点x都是E的锥点. 为了突出微分的实质性质, 我们后面将只考虑映射在其定义域的**内点处的可微性**, 但其中很多基本性质对锥点也成立.

此外我们再次强调: 若$A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, 则乘积Ax中的x自动转换为列向量.

【注5】微分的定义很容易推广到赋范线性空间上的映射, 例如设$(X, \|\cdot\|), (Y, \|\cdot\|)$为两个赋范线性空间, $f: X \to Y, x \in X$. 则在上面的定义中把欧氏范数$|\cdot|$ 换成$\|\cdot\|$就得到了f在点x可微的定义和微分的记号. 例如对$X = Y = \mathbb{R}^{n\times n}$ 的情形, 我们后面给出可微映射和计算微分的例子.

在本讲义中我们令$\mathbf{e}_1 = (1, 0, ..., 0)^\tau$, $\mathbf{e}_2 = (0, 1, 0, ..., 0)^\tau$, ..., $\mathbf{e}_n = (0, ..., 0, 1)^\tau$ 是$\mathbb{R}^n$中的通常使用的标准正交基.

【对坐标的微分和记号】考虑$m = 1$, 映射$f = f_j$为数值函数：f_j是自变量x的第j 个坐标, 即

$$f_j(x) = x_j, \quad x = (x_1, x_2, ..., x_n)^\tau \in \mathbb{R}^n.$$

此时显然有

$$f_j(x + h) = f_j(x) + h_j \qquad \forall\, h \in \mathbb{R}^n.$$

由可微的定义和微分的唯一性知f处处可微且

$$\mathrm{d}f_{j_x}(h) = \mathrm{d}f_j(x)(h) = f_j'(x)h = h_j \qquad \forall\, h \in \mathbb{R}^n.$$

现在看微分的三种表示方式:

第一种表示是用矩阵, 即

$$\mathrm{d}f_{j_x} = \mathrm{d}f_j(x) = f_j'(x) = \mathbf{e}_j^\tau = (0, ..., 0, 1, 0, ..., 0)$$

此时有

$$\mathrm{d}f_{j_x}(h) = \mathrm{d}f_j(x)(h) = f_j'(x)h = \mathbf{e}_j^\tau h = h_j, \qquad h = (h_1, h_2, ..., h_n)^\tau \in \mathbb{R}^n.$$

第二种表示是用内积(从而当做实值线性函数), 即

$$\mathrm{d}f_{j_x} = \mathrm{d}f_j(x) = f_j'(x) = \langle \mathbf{e}_j, \cdot \rangle$$

此时有

$$\mathrm{d}f_{j_x}(h)=\mathrm{d}f_j(x)(h)=f_j'(x)h=\langle \mathbf{e}_j,h\rangle=h_j,\qquad h=(h_1,h_2,...,h_n)^\tau\in\mathbb{R}^n.$$

第三种表示与第二种差不多但是采用了直接记号: 直接将f_j 写成x_j 即直接定义$\mathrm{d}x_j=\mathrm{d}f_j(x)=\mathrm{d}f_{j_x}=f_j'(x):\mathbb{R}^n\to\mathbb{R}$ 如下:

$$\mathrm{d}x_j(h)=h_j,\qquad h=(h_1,h_2,...,h_n)^\tau\in\mathbb{R}^n.$$

需要注意: 这个我们熟悉的微分记号$\mathrm{d}x_j=\mathrm{d}x_j(\cdot)$的一个身份是实值线性映射, 其中"$x_j$"不是具体数而只是提示读者: $\mathrm{d}x_j$是从$\mathbb{R}^n$向其第j个坐标轴的投影映射. 因此写法"d0.3"就属于概念错误.

对于通常的写法Δx_j 即第j个坐标上的增量, 可以理解为

$$\mathrm{d}x_j(\Delta x)=\Delta x_j\quad\text{其中}\quad \Delta x=(\Delta x_1,\Delta x_1,...,\Delta x_n)^\tau\in\mathbb{R}^n.$$

如果直接写"$\mathrm{d}x_j=\Delta x_j$"就造成小混乱: 等式左边$\mathrm{d}x_j$是映射而右边Δx_j是具体数值. □

【例(仿线性映射的微分)】设$A\in\mathbb{R}^{m\times n},b\in\mathbb{R}^n$, $f(x)=b+Ax,x\in\mathbb{R}^n$. 则有

$$f(x+h)=b+Ax+Ah=f(x)+Ah,\qquad h\in\mathbb{R}^n.$$

因常向量$\mathbf{0}$是h的高阶无穷小, 故由微分的唯一性知$f'(x)=A$. 这也可写成(为了熟悉微分的记号)

$$\mathrm{d}f_x(h)=\mathrm{d}f(x)(h)=Df(x)h=f'(x)h=Ah,\qquad x,h\in\mathbb{R}^n.\qquad\square$$

【例(二次型的微分)】设$A\in\mathbb{R}^{n\times n}$为对称阵, $Q(x)=x^\tau Ax,x\in\mathbb{R}^n$. 则

$$\begin{aligned}Q(x+h)&=(x^\tau+h^\tau)A(x+h)=x^\tau A(x+h)+h^\tau A(x+h)\\&=x^\tau Ax+x^\tau Ah+h^\tau Ax+h^\tau Ah=Q(x)+2x^\tau Ah+h^\tau Ah,\quad h\in\mathbb{R}^n.\end{aligned}$$

因$h\mapsto 2x^\tau Ah$ 是线性映射, 且按矩阵乘积的范数不等式有(例如使用矩阵的2−范数)

$$|h^\tau Ah|\le\|h^\tau A\|\|h\|\le\|h^\tau\|\|A\|\|h\|=\|A\|\|h\|^2=\|A\||h|^2$$

它显然是h的高阶无穷小, 故得知$Q'(x)=2x^\tau A$. 这也可写成(为了熟悉微分的记号)

$$\mathrm{d}Q_x(h)=\mathrm{d}Q(x)(h)=DQ(x)h=Q'(x)h=2x^\tau Ah,\quad x,h\in\mathbb{R}^n.\qquad\square$$

【例(矩阵的指数映射的微分)】

研究矩阵空间$\mathbb{R}^{n\times n}$上的指数映射$A\mapsto \exp(A)=\sum\limits_{k=0}^{\infty}\frac{A^k}{k!}$ 的微分. 设$A\in\mathbb{R}^{n\times n}$. 来证明

$$\mathrm{d}\exp_A(H)=\exp'(A)(H)=\sum_{k=1}^{\infty}\frac{1}{k!}\sum_{j=1}^{k}A^{k-j}HA^{j-1},\qquad H\in\mathbb{R}^{n\times n}.$$

由这个微分的表示式可知当H与A可交换时有

$$\mathrm{d}\exp_A(H)=\exp'(A)(H)=\exp(A)H.$$

【证】先证明对任意$A,H\in\mathbb{R}^{n\times n}$ 和$k\in\mathbb{N}$ 有

$$(A+H)^k=A^k+\sum_{j=1}^{k}A^{k-j}HA^{j-1}+\alpha_k(A;H)$$

其中

$$\alpha_1(A;H)=0,\quad \|\alpha_k(A;H)\|\le\frac{(k-1)k}{2}(\|A\|+\|H\|)^{k-2}\|H\|^2,\ \ k\ge 2.$$

$k=1,2$时, 是显然的. 设在某个$k\ge 2$时成立, 则在$k+1$ 时, 计算

$$\begin{aligned}
&(A+H)^{k+1}=\Big(A^k+\sum_{j=1}^{k}A^{k-j}HA^{j-1}+\alpha_k(A;H)\Big)(A+H)\\
&=A^{k+1}+\sum_{j=1}^{k}A^{k-j}HA^{j}+A^kH+\sum_{j=1}^{k}A^{k-j}HA^{j-1}H+\alpha_k(A;H)(A+H)\\
&=A^{k+1}+\sum_{j=0}^{k}A^{k-j}HA^{j}+\alpha_{k+1}(A;H)\\
&=A^{k+1}+\sum_{j=1}^{k+1}A^{k+1-j}HA^{j-1}+\alpha_{k+1}(A;H)
\end{aligned}$$

其中

$$\alpha_{k+1}(A;H)=\sum_{j=1}^{k}A^{k-j}HA^{j-1}H+\alpha_k(A;H)(A+H).$$

作估计:

$$\|\alpha_{k+1}(A;H)\|\le\sum_{j=1}^{k}\|A^{k-j}HA^{j-1}H\|+\|\alpha_k(A;H)(A+H)\|$$

$$\le\sum_{j=1}^{k}\|A\|^{k-j}\|\|H\|\|A\|^{j-1}\|H\|+\|\alpha_k(A;H)\|(\|A\|+\|H\|)$$

$$
\begin{aligned}
&\le k\|A\|^{k-1}\|H\|^2+\frac{k(k-1)}{2}\|(\|A\|+\|H\|)^{k-1}\|H\|^2\\
&\le (k+\frac{(k-1)k}{2})\|(\|A\|+\|H\|)^{k-1}\|H\|^2\\
&=\frac{k(k+1)}{2}\|(\|A\|+\|H\|)^{k-1}\|H\|^2.
\end{aligned}
$$

据归纳法原理, 所证关系式成立.

根据这一结果我们有

$$
\begin{aligned}
\exp(A+H)&=\mathrm{I}+\sum_{k=1}^{\infty}\frac{(A+H)^k}{k!}\\
&=\mathrm{I}+\sum_{k=1}^{\infty}\frac{A^k}{k!}+\sum_{k=1}^{\infty}\frac{1}{k!}\sum_{j=1}^{k}A^{k-j}HA^{j-1}+\sum_{k=1}^{\infty}\frac{1}{k!}\alpha_k(A;H)\\
&=\exp(A)+\sum_{k=1}^{\infty}\frac{1}{k!}\sum_{j=1}^{k}A^{k-j}HA^{j-1}+\sum_{k=2}^{\infty}\frac{1}{k!}\alpha_k(A;H).
\end{aligned}
$$

因

$$
\begin{aligned}
&\Big\|\sum_{k=2}^{\infty}\frac{1}{k!}\alpha_k(A;H)\Big\|\le\sum_{k=2}^{\infty}\frac{1}{k!}\|\alpha_k(A;H)\|\le\sum_{k=2}^{\infty}\frac{1}{k!}\frac{(k-1)k}{2}(\|A\|+\|H\|)^{k-2}\|H\|^2\\
&=\frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{k!}(\|A\|+\|H\|)^k\|H\|^2=\frac{1}{2}e^{\|A\|+\|H\|}\|H\|^2.
\end{aligned}
$$

故

$$
\sum_{k=2}^{\infty}\frac{1}{k!}\alpha_k(A;H)=o(H)\quad (H\to 0).
$$

又易见$\sum\limits_{k=1}^{\infty}\frac{1}{k!}\sum\limits_{j=1}^{k}A^{k-j}HA^{j-1}$ 绝对收敛且$H\mapsto\sum\limits_{k=1}^{\infty}\frac{1}{k!}\sum\limits_{j=1}^{k}A^{k-j}HA^{j-1}$ 是线性映射, 故由可微性和微分的定义知$\exp(\cdot)$ 在点A可微且$\exp'(A)(H)=\sum\limits_{k=1}^{\infty}\frac{1}{k!}\sum\limits_{j=1}^{k}A^{k-j}HA^{j-1}$.

上面的推导还给出了具体估计:

$$
\Big\|\exp(A+H)-\exp(A)-\sum_{k=1}^{\infty}\frac{1}{k!}\sum_{j=1}^{k}A^{k-j}HA^{j-1}\Big\|\le\frac{1}{2}e^{\|A\|+\|H\|}\|H\|^2.
$$

□

下面一些命题给出多元微分及其运算的简单性质.

【命题8.6(可微蕴含连续).】 设$E\subset\mathbb{R}^n$, 映射$f:E\to\mathbb{R}^m$ 在点$x\in E^\circ$可微. 则f在点x处连续 并且存在$\delta=\delta_x>0$ 和$0<L=L_x<+\infty$使得$B(x,\delta)\subset E$ 且

$$
|f(y)-f(x)|\le L|y-x|\qquad \forall\, y\in B(x,\delta).
$$

【证】由可微的定义知存在$\delta=\delta_x>0$ 充分小使得$B(x,\delta)\subset E$ 且

$$|f(y)-f(x)-f'(x)(y-x)|=|\alpha(x;y-x)|\le|y-x|\qquad\forall\, y\in B(x,\delta).$$

取$L=\|f'(x)\|+1$ 则有

$$\begin{aligned}&|f(y)-f(x)|\le|f'(x)(y-x)|+|y-x|\\&\le\|f'(x)\||y-x|+|y-x|=L|y-x|\qquad\forall\, y\in B(x,\delta).\end{aligned}$$

这蕴含f在点x连续. $\square$

【偏导数】

我们知道一个函数的微分$f'(x)$可表示成一个矩阵, 但进一步的表示是什么? 为此我们需要学习函数的偏导数, 偏导数本身也是研究问题的重要工具. 偏导数就是函数对自变量的某个分量(即坐标)的导数, 其中在求导数时其他分量保持不动, 因此是一元函数行为. 回忆: 对一元函数来说, 可微和可导是一回事。

【定义(偏导数)】设$\Omega\subset\mathbb{R}^n$为开集, $f:\Omega\to\mathbb{R}^m$, $x=(x_1,x_2,...,x_n)^\tau\in\Omega$, $j\in\{1,2,...,n\}$. 若差商极限

$$\frac{\partial f}{\partial x_j}(x):=\lim_{t\to0}\frac{f(x+t\mathbf{e}_j)-f(x)}{t}\in\mathbb{R}^m$$

存在, 则称$\frac{\partial f}{\partial x_j}(x)$为$f$ 在点x处对x的第j个坐标x_j 的偏导数.

偏导数$\frac{\partial f}{\partial x_j}(x)$ 也常记作$D_jf(x)$ 并被等价地定义成差商极限

$$\begin{aligned}&\frac{\partial f}{\partial x_j}(x)=D_jf(x)\\&=\lim_{y_j\to x_j}\frac{f(x_1,..,x_{j-1},y_j,x_{j+1},...,x_n)-f(x_1,...,x_{j-1},x_j,x_{j+1},...,x_n)}{y_j-x_j}.\end{aligned}$$

$\square$

偏导数的其他记号：例如二元函数或映射$f(x,y)$ 在定点(x_0,y_0)的偏导数可写为

$$f'_x(x_0,y_0)=\partial_xf(x_0,y_0)=\frac{\partial f}{\partial x}(x_0,y_0)=\lim_{x\to x_0}\frac{f(x,y_0)-f(x_0,y_0)}{x-x_0},$$

$$f'_y(x_0,y_0)=\partial_yf(x_0,y_0)=\frac{\partial f}{\partial y}(x_0,y_0)=\lim_{y\to y_0}\frac{f(x_0,y)-f(x_0,y_0)}{y-y_0}.$$

下面命题给出当可微时微分与偏导数的关系. 首先引进

【**定义(函数的梯度)**】设开集$\Omega \subset \mathbb{R}^n, x = (x_1, x_2, ..., x_n)^\tau \in \Omega$, 数值函数$f : \Omega \to \mathbb{R}$. 若$\forall\, j \in \{1, 2, ..., n\}$, 偏导数$\frac{\partial f}{\partial x_j}(x)$存在, 则称向量

$$\nabla f(x) = \mathrm{grad} f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} = \Big(\frac{\partial f}{\partial x_1}(x),\, \frac{\partial f}{\partial x_2}(x),\, ...,\, \frac{\partial f}{\partial x_n}(x)\Big)^\tau$$

为f在点x处的梯度. □

【**命题8.7(数值函数的可微性和微分的表示).**】设开集$\Omega \subset \mathbb{R}^n, x = (x_1, x_2, ..., x_n)^\tau \in \Omega$, 数值函数$f : E \to \mathbb{R}$. 则有:

f在x可微 $\Longleftrightarrow \forall\, j \in \{1, 2, ..., n\}$ 偏导数$\frac{\partial f}{\partial x_j}(x)$存在并有

$$f(x+h) = f(x) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) h_j + \alpha(x; h), \qquad |h| << 1$$

其中$h = (h_1, h_2, ..., h_n)^\tau$, $\alpha(x; h)$ 是h 的高阶无穷小.

因此当f在x 可微时, 微分$f'(x)$ 可表示成梯度$\nabla f(x)$的转置(行向量)

$$f'(x) = (\nabla f(x))^\tau = \Big(\frac{\partial f}{\partial x_1}(x),\, \frac{\partial f}{\partial x_2}(x),\, ...,\, \frac{\partial f}{\partial x_n}(x)\Big)$$

并有

$$f(x+h) = f(x) + f'(x)h + o(h) = f(x) + \langle \nabla f(x), h \rangle + o(h), \qquad |h| << 1.$$

因此

$$\mathrm{d} f_x(h) = \mathrm{d} f(x)(h) = f'(x)h = \langle \nabla f(x), h \rangle \qquad \forall\, h \in \mathbb{R}^n$$

即

$$\mathrm{d} f_x(h) = \mathrm{d} f(x)(h) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) h_j = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) \mathrm{d} x_j(h) \qquad \forall\, h \in \mathbb{R}^n$$

即, 作为映射, 微分等于

$$\mathrm{d} f_x = \mathrm{d} f(x) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) \mathrm{d} x_j.$$

【证】"$\Longrightarrow$": 设f在点x可微. 则由微分的定义知, 微分$f'(x) \in \mathbb{R}^{1\times n}$ (可看成行向量) 作为线性映射$h \mapsto f'(x)h$ 是$\mathbb{R}^n = \mathbb{R}^{n\times 1}$ 上的线性函数, 即存在唯一一组系

数$a_1(x), a_2(x), ..., a_n(x)$ 使得

$$f'(x)h = \sum_{j=1}^{n} a_j(x)h_j, \qquad h = (h_1, h_2, ..., h_n)^\tau \in \mathbb{R}^n.$$

因此有

$$f(x+h) = f(x) + \sum_{j=1}^{n} a_j(x)h_j + \alpha(x;h), \qquad |h| << 1.$$

其中$\alpha(x;h) = o(h)$ 是h 的高阶无穷小.

对每个$j \in \{1, 2, ..., n\}$, 取$h = t\mathbf{e}_j, t \in \mathbb{R}, 0 < |t| << 1$, 则有$h_j = t; h_i = 0, i \neq j$, 从而有

$$\frac{f(x+t\mathbf{e}_j) - f(x)}{t} = a_j(x) + \frac{\alpha(x;t\mathbf{e}_j)}{t}.$$

因

$$\frac{|\alpha(x;t\mathbf{e}_j)|}{|t|} = \frac{|\alpha(x;t\mathbf{e}_j)|}{|t\mathbf{e}_j|} \to 0 \quad (t \to 0)$$

故得知差商极限

$$\frac{\partial f}{\partial x_j}(x) := \lim_{t\to 0} \frac{f(x+t\mathbf{e}_j) - f(x)}{t} = a_j(x) \in \mathbb{R} \quad 存在, \quad j = 1, 2, ..., n.$$

"$\Longleftarrow$": 设f具有上述性质, 则由可微的定义(注意$\alpha(x;h)$ 是h 的高阶无穷小!) 和微分的唯一性即知$f'(x)h = \sum\limits_{j=1}^{n} \frac{\partial f}{\partial x_j}(x)h_j\ (\forall\, h \in \mathbb{R}^n)$, 因此$f'(x) = (\frac{\partial f}{\partial x_1}(x),\ \frac{\partial f}{\partial x_2}(x),\ ...,\ \frac{\partial f}{\partial x_n}(x))$. □

【顺便再看一下微分的几何意义】设开集$\Omega \subset \mathbb{R}^n$, 数值函数$f : \Omega \to \mathbb{R}$. 令

$$S = \{(x, y) \mid y = f(x),\ x \in \Omega\}$$

为f的图象, 即它是由$y = f(x)$ 确定的显示曲面 $S:\ y = f(x),\ x \in \Omega$.

设$x_0 \in \Omega$ 并设f 在点x_0处可微. 则由上面命题知在x_0 附近f近似等于仿线性函数:

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + o(x - x_0).$$

令$y_0 = f(x_0)$. 我们把平面

$$P: \quad y = y_0 + \langle \nabla f(x_0), x - x_0 \rangle, \quad x \in \mathbb{R}^n$$

也即把集合

$$P = \{(x, y) \mid y = y_0 + \langle \nabla f(x_0), x - x_0 \rangle,\ x \in \mathbb{R}^n\}$$

叫做曲面$S:\ y=f(x)$在点$(x_0,y_0)\in S$处的**切平面**.

易见这个切平面P的单位法向量$\mathbf{n}$为

$$\mathbf{n}=\frac{(\nabla f(x_0),-1)}{\sqrt{|\nabla f(x_0)|^2+1}}\qquad \text{或}\qquad \mathbf{n}=-\frac{(\nabla f(x_0),-1)}{\sqrt{|\nabla f(x_0)|^2+1}}.$$

利用法向量, 这个经过点(x_0,y_0)的平面P 也可唯一地表示成

$$P=\{(x,y)\in\mathbb{R}^n\times\mathbb{R}\mid (x,y)-(x_0,y_0)\perp\mathbf{n}\}.$$

正是由于这个垂直关系, $\mathbf{n}$ 才被称为平面P的法向量.

【命题8.8(一般映射的可微性和微分的表示).】设$\Omega\subset\mathbb{R}^n$ 为开集, $x=(x_1,x_2,...,x_n)^\tau\in\Omega$, 映射$f=(f_1,f_2,...,f_m)^\tau:\Omega\to\mathbb{R}^m$. 则有:

f在点x处可微 $\iff$ 每个坐标函数$f_i:\Omega\to\mathbb{R}$ 在点x处可微 $(i=1,2,...,m)$.

这时(即当f在点x处可微时) 按矩阵表示有

$$f'(x)=\begin{pmatrix} f_1'(x)\\ f_2'(x)\\ \vdots\\ f_m'(x)\end{pmatrix}=\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & ... & \frac{\partial f_1}{\partial x_n}(x)\\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & ... & \frac{\partial f_2}{\partial x_n}(x)\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & ... & \frac{\partial f_m}{\partial x_n}(x)\end{pmatrix}=\Big(\frac{\partial f_i}{\partial x_j}(x)\Big)_{1\le i\le m,1\le j\le n}$$

右边的矩阵称为f在点x处的**Jacobi 矩阵**, 记作$J_f(x)$, 即

$$J_f(x)=\Big(\frac{\partial f_i}{\partial x_j}(x)\Big)_{m\times n}.$$

【证】"$\Longrightarrow$": 设f在x可微. 使用矩阵表示将$f'(x)$写成

$$f'(x)=\begin{pmatrix} a_{11}(x) & a_{12}(x) & ... & a_{1n}(x)\\ a_{21}(x) & a_{22}(x) & ... & a_{2n}(x)\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1}(x) & a_{m2}(x) & ... & a_{mn}(x)\end{pmatrix}.$$

则

$$f(x+h)-f(x)=f'(x)h+\alpha(x,h)=\begin{pmatrix} \sum_{j=1}^n a_{1j}(x)h_j\\ \sum_{j=1}^n a_{2j}(x)h_j\\ \vdots\\ \sum_{j=1}^n a_{mj}(x)h_j\end{pmatrix}+\begin{pmatrix}\alpha_1(x,h)\\ \alpha_2(x,h)\\ \vdots\\ \alpha_m(x,h)\end{pmatrix}$$

即

$$f_i(x+h)-f_i(x)=\sum_{j=1}^{n}a_{ij}(x)h_j+\alpha_i(x,h),\quad i=1,2,...,m.$$

因$\alpha_i(x,h)$ 是$\alpha(x,h)$的分量, 故

$$\frac{|\alpha_i(x,h)|}{|h|}\le\frac{|\alpha(x,h)|}{|h|}\to 0\quad(h\to 0)$$

即$\alpha_i(x,h)=o(h)$. 因此f_i在x可微. 由可微的定义和**命题8.7(数值函数的可微性和微分的表示)**知

$$(a_{i1}(x),a_{i2}(x),...,a_{in}(x))=f_i'(x)=\left(\frac{\partial f_i}{\partial x_1}(x),\,\frac{\partial f_i}{\partial x_2}(x),\,...,\,\frac{\partial f_i}{\partial x_n}(x)\right),\quad i=1,2,...,m.$$

于是有

$$f'(x)=(a_{ij}(x))_{m\times n}=\left(\frac{\partial f_i}{\partial x_j}(x)\right)_{1\le i\le m,1\le j\le n}=J_f(x).$$

"$\Longleftarrow$": 设f的每个坐标函数f_j 在x 可微. 则由**命题8.7(数值函数的可微性和微分的表示)** 知

$$f_i'(x)=\left(\frac{\partial f_i}{\partial x_1}(x),\,\frac{\partial f_i}{\partial x_2}(x),\,...,\,\frac{\partial f_i}{\partial x_n}(x)\right),\quad i=1,2,...,m.$$

由此和可微的定义(注意高阶无穷小)有

$$f(x+h)-f(x)=J_f(x)h+\alpha(x;h),\qquad |h|<<1$$

其中

$$\alpha(x;h)=\begin{pmatrix}\alpha_1(x;h)\\ \alpha_2(x;h)\\ \vdots\\ \alpha_m(x;h)\end{pmatrix},\quad \alpha_i(x;h)=o(h),\quad i=1,2,...,m.$$

易见

$$|\alpha(x;h)|=\Big(\sum_{i=1}^{m}\alpha_i(x;h)^2\Big)^{1/2}=o(|h|)\qquad\text{即}\qquad \alpha(x;h)=o(h).$$

因此由可微的定义和微分的唯一性知f在点x可微且$f'(x)=J_f(x)$. $\square$

【**注**】如上一命题, 在这个命题中若将坐标的微分表示$\mathrm{d}x_j(h)=h_j$ 代入便有

$$h=\sum_{j=1}^{n}h_j\mathbf{e}_j=\sum_{j=1}^{n}\mathbf{e}_jh_j=\sum_{j=1}^{n}\mathbf{e}_j\mathrm{d}x_j(h)$$

从而有

$$\mathrm{d}f_x(h) = \mathrm{d}f(x)(h) = f'(x)h = \sum_{j=1}^{n} f'(x)\mathbf{e}_j \mathrm{d}x_j(h) \qquad \forall\, h \in \mathbb{R}^n$$

即

$$\mathrm{d}f_x = \mathrm{d}f(x) = \sum_{j=1}^{n} f'(x)\mathbf{e}_j \mathrm{d}x_j.$$

而

$$f'(x)\mathbf{e}_j = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(x) \\ \frac{\partial f_2}{\partial x_j}(x) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x) \end{pmatrix} =: \frac{\partial f}{\partial x_j}(x), \quad j = 1, 2, ..., n$$

故上述微分可以写成

$$\mathrm{d}f_x = \mathrm{d}f(x) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) \mathrm{d}x_j.$$

【Jacobi矩阵与微分的关系】

上面我们已证明如果一个映射$f = (f_1, f_2, ..., f_m)^\tau$ 在点x可微, 则(作为线性映射)微分$f'(x)$的矩阵表示就是Jacobi矩阵$J_f(x) = (\frac{\partial f_i}{\partial x_j}(x))_{m\times n}$, 即在表示的意义下有$f'(x) = J_f(x)$. 但是Jacobi矩阵$J_f(x)$本身的存在性条件则比较宽, 它只需矩阵的每个元素存在, 即只需f的每个坐标函数f_i的每个偏导数$\frac{\partial f_i}{\partial x_j}(x)$都在点$x$存在. 从本节后面的例子我们也将看到, 与一元情形不同, 当$n \geq 2$时, Jacobi矩阵$J_f(x)$的存在不一定蕴含f在点x可微, 甚至不蕴含f在x连续！不过在本讲义中和未来研究中的绝大多数情况下, 我们遇到和使用的映射, 它们的偏导数都是连续的从而都是可微的(见下面命题), 因此写法$f'(x) = J_f(x)$就是有意义的.

【命题8.9.】设$\Omega \subset \mathbb{R}^n$ 为开集, 映射$f = (f_1, f_2, ..., f_m)^\tau : \Omega \to \mathbb{R}^m$, $x_0 \in \Omega$.

(a) 若每个坐标函数f_i在x_0的某个邻域内有各一阶偏导数且这些偏导数都在此邻域内有界, 则f 在x_0连续.

(b) 若每个坐标函数f_i在x_0的某个邻域内有各一阶偏导数且这些偏导数都在x_0连续, 则f 在x_0可微, 从而(在表示的意义下)有$f'(x_0) = J_f(x_0)$.

【证】因$f = (f_1, f_2, ..., f_m)^\tau$的有界性、连续性、可微性分别等价于$f$的每个坐标函数$f_i$的有界性、连续性、可微性, 故我们只需对$m = 1$, 即$f$是数值函数的情形, 进行证明即可. 为书写方便, 以下我们将x, y 等写成行向量.

因Ω 是开集且$x_0 \in \Omega$, 故存在$r > 0$ 使得闭球$\overline{B}(x_0, r) \subset \Omega$. 写$x_0 = (x_{01}, x_{02}, ..., x_{0n})$. 令$\delta = r/\sqrt{n}$, $I(x_0) = \prod_{j=1}^{n}[x_{0j} - \delta, x_{0j} + \delta]$. 则$B(x_0, \delta) \subset I(x_0) \subset \overline{B}(x_0, r)$.

(a): 由假设知对于充分小的$r > 0$, 存在$0 < M < +\infty$ 使得

$$\left|\frac{\partial f}{\partial x_j}(x)\right| \le M \qquad \forall x \in \overline{B}(x_0, r), \quad j = 1, 2, ..., n.$$

来证明f在$I(x_0)$ 上满足Lipschitz 条件

$$|f(y) - f(x)| \le L|y - x|, \quad x, y \in I(x_0)$$

其中$L = M\sqrt{n}$. 这蕴含f在$I(x_0)$上连续, 特别蕴含f在点x_0 连续.

任取$x, y \in I(x_0)$, 按平行于坐标轴的方向加减相关项, 有

$$\begin{aligned}
&f(y) - f(x) = f(y_1, y_2, ..., y_n) - f(x_1, x_2, ..., x_n)\\
&= f(y_1, y_2, ..., y_n) - f(x_1, y_2, ..., y_n)\\
&+\sum_{j=2}^{n-1}\Big(f(x_1, ..., x_{j-1}, y_j, y_{j+1}, ..., y_n) - f(x_1, ..., x_{j-1}, x_j, y_{j+1}, ..., y_n)\Big)\\
&+f(x_1, ..., x_{n-1}, y_n) - f(x_1, ..., x_{n-1}, x_n).
\end{aligned}$$

这里当$n \le 2$ 时, “$\sum_{j=2}^{n-1}$” 不出现.

看上式中一般项的估计: 对一元函数

$$t \mapsto g_j(t) := f(x_1, ..., x_{j-1}, t, y_{j+1}, ..., y_n), \quad t \in [x_{0j} - \delta, x_{0j} + \delta]$$

应用中值定理, 存在ξ_j, 它介于x_j, y_j 之间(从而$\xi_j \in [x_{0j} - \delta, x_{0j} + \delta]$), 使得

$$\begin{aligned}
&f(x_1, ..., x_{j-1}, y_j, y_{j+1}, ..., y_n) - f(x_1, ..., x_{j-1}, x_j, y_{j+1}, ..., y_n)\\
&= g_j(y_j) - g_j(x_j) = g_j'(\xi_j)(y_j - x_j)\\
&= \frac{\partial f}{\partial x_j}(x_1, ..., x_{j-1}, \xi_j, y_{j+1}, ..., y_n)(y_j - x_j).
\end{aligned}$$

对第一项和第三项也有相同的关系式. 因此

$$\begin{aligned}
f(y) - f(x) = &\frac{\partial f}{\partial x_1}(\xi_1, y_2, ..., y_n)(y_1 - x_1)\\
&+\sum_{j=2}^{n-1}\frac{\partial f}{\partial x_j}(x_1, ..., x_{j-1}, \xi_j, y_{j+1}, ..., y_n)(y_j - x_j)\\
&+\frac{\partial f}{\partial x_n}(x_1, ..., x_{n-1}, \xi_n)(y_n - x_n)
\end{aligned}$$

从而有

$$|f(y)-f(x)| \le M\sum_{j=1}^{n}|y_j-x_j| \le M\sqrt{n}|y-x|.$$

(b): 在(a)中得到关系式中取$x=x_0$, $y=x_0+h$, $h\in[-\delta,\delta]^n$ 则有

$$\begin{aligned}
&\alpha(x_0,h) := f(x_0+h)-f(x_0)-\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(x_0)h_j\\
&=\left(\frac{\partial f}{\partial x_1}(\xi_1,x_{02}+h_2,...,x_{0n}+h_n)-\frac{\partial f}{\partial x_1}(x_0)\right)h_1\\
&+\sum_{j=2}^{n-1}\left(\frac{\partial f}{\partial x_j}(x_{01},...,x_{0,j-1},\xi_j,x_{0,j+1}+h_{j+1},...,x_{0n}+h_n)-\frac{\partial f}{\partial x_j}(x_0)\right)h_i\\
&+\left(\frac{\partial f}{\partial x_n}(x_{01},...,x_{0,n-1},\xi_n)-\frac{\partial f}{\partial x_n}(x_0)\right)h_n.
\end{aligned}$$

如令

$$\Delta_j(\varepsilon)=\sup_{|x-x_0|\le\varepsilon}\left|\frac{\partial f}{\partial x_j}(x)-\frac{\partial f}{\partial x_j}(x_0)\right|,\quad 0\le\varepsilon\le\delta$$

则当$|h|\le\delta$ 时

$$|\alpha(x_0,h)|\le\sum_{j=1}^{n}\Delta_j(|h|)|h_j|\le\left(\sum_{j=1}^{n}\Delta_j(|h|)^2\right)^{1/2}|h|.$$

由假设, 偏导数$\frac{\partial f}{\partial x_j}$ 在x_0 连续, 知

$$\lim_{\varepsilon\to0^+}\Delta_j(\varepsilon)=0,\quad j=1,2,...,n.$$

因此

$$\left(\sum_{j=1}^{n}\Delta_j(|h|)^2\right)^{1/2}\to 0\quad \text{as}\quad |h|\to0$$

从而有

$$\frac{|\alpha(x_0,h)|}{|h|}\to0\quad\text{as}\quad |h|\to0.$$

这证明了

$$f(x_0+h)=f(x_0)+\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(x_0)h_j+o(h).$$

所以f在x_0可微. □

使用Jacobi矩阵和矩阵范数, 上述**命题8.9之(b)**可以叙述成:

【命题8.10.】 设$\Omega\subset\mathbb{R}^n$为开集, 映射$f:\Omega\to\mathbb{R}^m$ 的Jacobi矩阵$J_f(x)=\left(\frac{\partial f_i}{\partial x_j}(x)\right)_{m\times n}$在邻域$B(x_0,\delta)\subset\Omega$ 内存在且在点x_0 连续, 即(按矩阵范数)

$$\lim_{x\to x_0}\|J_f(x)-J_f(x_0)\|=0$$

则f在点x_0可微. $\square$

与此相应, 我们引进

【定义(C^1映射类)】 设$\Omega\subset\mathbb{R}^n$, 映射$f=(f_1,f_2,...,f_m)^\tau:\Omega\to\mathbb{R}^m$ 的每个坐标函数f_i在Ω 内有各一阶偏导数$\frac{\partial f_i}{\partial x_j}$ 且所有这些偏导数都在Ω 内连续. 则称f 属于C^1类, 记作$f\in C^1(\Omega,\mathbb{R}^m)$,即

$$C^1(\Omega,\mathbb{R}^m)=\Big\{f:\Omega\to\mathbb{R}^m\ \Big|\ \frac{\partial f_i}{\partial x_j}\in C(\Omega,\mathbb{R}),\ i=1,2,...,m;j=1,2,...,n\Big\}.$$

当$m=1$ 时记$C^1(\Omega)=C^1(\Omega,\mathbb{R})$ 并称$C^1(\Omega)$为C^1 函数类. $\square$

使用Jacobi矩阵表达并用矩阵范数, 易见

$$C^1(\Omega,\mathbb{R}^m)=\{f:\Omega\to\mathbb{R}^m\ |\ J_f\in C(\Omega,\mathbb{R}^{m\times n})\}.$$

由**命题8.9** 或**命题8.10**知若$f\in C^1(\Omega,\mathbb{R}^m)$, 则$f$在$\Omega$内可微.

下面给出微分的算数运算法则。在运算中我们只要清楚矩阵运算法则, 就不会出错. 例如复习一下: 设$\mathbf{a}=(a_1,a_2,...,a_m)^\tau$为列向量, $\mathbf{b}=(b_1,b_2,...,b_n)$ 为行向量. 则

$$\mathbf{ab}=\begin{pmatrix}a_1\\ a_2\\ \vdots\\ a_m\end{pmatrix}(b_1,b_2,...,b_n)=\begin{pmatrix}a_1b_1 & a_1b_2 & \cdots & a_1b_n\\ a_2b_1 & a_2b_2 & \cdots & a_2b_n\\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots\\ a_mb_1 & a_mb_2 & \cdots & a_mb_n\end{pmatrix}=(a_ib_j)_{m\times n}.$$

【命题8.11(微分的算数运算).】 设$E\subset\mathbb{R}^n,x\in E^\circ$.

(a) 设映射$f,g:E\to\mathbb{R}^m$ 都在x可微. 则$f+g,cf$ (其中c为常数) 也在x可微且

$$(f+g)'(x)=f'(x)+g'(x),\quad (cf)'(x)=cf'(x)$$

即

$$\mathrm{d}(f+g)_x=\mathrm{d}f_x+\mathrm{d}g_x,\quad \mathrm{d}(cf)_x=c\mathrm{d}f_x.$$

以上也可等价地写成

$$(f+g)'(x)h = f'(x)h + g'(x)h,\quad (cf)'(x)h = cf'(x)h \qquad \forall\, h\in\mathbb{R}^n$$

即

$$\mathrm{d}(f+g)_x(h) = \mathrm{d}f_x(h) + \mathrm{d}g_x(h),\quad \mathrm{d}(cf)_x(h) = c\mathrm{d}f_x(h) \qquad \forall\, h\in\mathbb{R}^n.$$

因此对于固定的x, 微分运算$f\mapsto \mathrm{d}f_x = f'(x)$是$f$的线性运算.

(b) 设映射$f,g:E\to\mathbb{R}^m$ 都在x可微. 则内积$\langle f,g\rangle(x)=\langle f(x),g(x)\rangle$ 也在x可微且

$$\langle f,g\rangle'(x) = g(x)^{\tau}f'(x) + f(x)^{\tau}g'(x).$$

特别对于$g=f$的情形有

$$(|f|^2)'(x) = 2f(x)^{\tau}f'(x).$$

(c) 设映射$f:E\to\mathbb{R}^m$ 和函数$g:E\to\mathbb{R}$ 都在x可微. 则乘积gf也在x可微且

$$(gf)'(x) = f(x)g'(x) + g(x)f'(x).$$

进一步设$g(x)\neq 0$. 则除法$\frac{f}{g}$ 也在x可微且

$$\Big(\frac{f}{g}\Big)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}.$$

【证】(a): 由假设有

$$f(x+h) = f(x) + f'(x)h + \alpha(x;h),\quad g(x+h) = g(x) + g'(x)h + \beta(x;h)$$

其中$\alpha(x;h),\beta(x;h)$ 都是h的高阶无穷小,即$\alpha(x;h)=o(h),\beta(x;h)=o(h)$ $(h\to 0)$.

计算

$$\begin{aligned}
&(f+g)(x+h)\\
&= f(x+h)+g(x+h) = f(x)+g(x)+f'(x)h+g'(x)h+\alpha(x;h)+\beta(x;h)\\
&= (f+g)(x) + f'(x)h + g'(x)h + \alpha(x;h) + \beta(x;h),
\end{aligned}$$

$$\begin{aligned}
&(cf)(x+h) = cf(x+h) = cf(x) + cf'(x)h + c\alpha(x;h)\\
&= (cf)(x) + cf'(x)h + c\alpha(x;h).
\end{aligned}$$

因$\alpha(x;h)+\beta(x;h), c\alpha(x;h)$也都是$h$的高阶无穷小, 故由可微的定义和微分的唯一性即知所证成立.

(b): 为计算方便, 以下同一记号$o(h)$表示h的不同的高阶无穷小. 计算

$$
\begin{aligned}
&\langle f(x+h), g(x+h)\rangle = \langle f(x)+f'(x)h+o(h), g(x+h)\rangle \\
&= \langle f(x), g(x+h)\rangle + \langle f'(x)h, g(x+h)\rangle + o(h) \\
&= \langle f(x), g(x)+g'(x)h+o(h)\rangle + \langle f'(x)h, g(x)+g'(x)h+o(h)\rangle + o(h) \\
&= \langle f(x), g(x)\rangle + \langle f(x), g'(x)h\rangle + o(h) + \langle f'(x)h, g(x)\rangle + o(h) \\
&= \langle f(x), g(x)\rangle + f(x)^\tau g'(x)h + g(x)^\tau f'(x)h + o(h) \\
&= \langle f(x), g(x)\rangle + (\, f(x)^\tau g'(x) + g(x)^\tau f'(x)\,)h + o(h)
\end{aligned}
$$

这里(以及下面)我们用到了可微必连续, 因此例如$\langle o(h), g(x+h)\rangle = o(h)$ $(|h|\to 0)$. 所以由可微的定义和微分的唯一性知$\langle f, g\rangle$ 在x可微且$\langle f, g\rangle'(x) = g(x)^\tau f'(x) + f(x)^\tau g'(x)$.

(c): 计算

$$
\begin{aligned}
&(gf)(x+h) = g(x+h)f(x+h) = g(x+h)(f(x)+f'(x)h+o(h)) \\
&= g(x+h)f(x) + g(x+h)f'(x)h + o(h) \\
&= (g(x)+g'(x)h+o(h))f(x) + (g(x)+g'(x)h+o(h))f'(x)h + o(h) \\
&= g(x)f(x) + g'(x)hf(x) + o(h) + g(x)f'(x)h + o(h) \\
&= g(x)f(x) + f(x)g'(x)h + g(x)f'(x)h + o(h) \\
&= g(x)f(x) + (\, f(x)g'(x) + g(x)f'(x)\,)h + o(h)
\end{aligned}
$$

其中我们用到事实:

$$g'(x)h\,\text{是实数, 因此}\ \ g'(x)hf(x) = f(x)g'(x)h.$$

同上, 由可微的定义和微分的唯一性知gf 在x 可微且$(gf)'(x) = f(x)g'(x) + g(x)f'(x)$.

最后进一步假设$g(x)\neq 0$. 则由可微必连续知当$|h| << 1$时$|g(x+h)| \ge \frac{1}{2}|g(x)|\,(>0)$ 因此$h\mapsto \frac{1}{g(x+h)}$ 在$|h| << 1$内有界.

计算

$$\frac{1}{g(x+h)} - \frac{1}{g(x)} = \frac{g(x)-g(x+h)}{g(x+h)g(x)}$$

$$
\begin{aligned}
&= \frac{g(x)-g(x+h)}{g(x)}\Big(\frac{1}{g(x+h)}-\frac{1}{g(x)}\Big)+\frac{g(x)-g(x+h)}{g(x)^2} \\
&= \frac{(g(x)-g(x+h))^2}{g(x)^2 g(x+h)}-\frac{g(x+h)-g(x)}{g(x)^2} \\
&= \frac{(g'(x)h+o(h))^2}{g(x)^2 g(x+h)}-\frac{g'(x)h+o(h)}{g(x)^2}=-\frac{g'(x)}{g(x)^2}h+o(h)
\end{aligned}
$$

即

$$\frac{1}{g(x+h)}=\frac{1}{g(x)}-\frac{g'(x)}{g(x)^2}h+o(h).$$

所以$\frac{1}{g}$ 在x 可微且

$$(\frac{1}{g})'(x)=-\frac{g'(x)}{g(x)^2}.$$

于是$\frac{f}{g}=\frac{1}{g}f$ 在x 可微且

$$
\begin{aligned}
&\Big(\frac{f}{g}\Big)'(x)=f(x)(\frac{1}{g})'(x)+\frac{1}{g(x)}f'(x)=-f(x)\frac{g'(x)}{g(x)^2}+\frac{1}{g(x)}f'(x) \\
&=\frac{g(x)f'(x)-f(x)g'(x)}{g(x)^2}. \qquad\qquad \Box
\end{aligned}
$$

【复合映射的微分法】

复合映射及其微分一直都是最重要的因为它们复杂—— 复杂的结构能描述复杂现象, 其中复合映射的求导法则是主要功能。

【定理8.12(复合映射的可微性和微分的锁链法则).】设$U\subset\mathbb{R}^n, V\subset\mathbb{R}^m$ 分别是$\mathbb{R}^n, \mathbb{R}^m$的开集, 映射$f:U\to V$ 和$g:V\to\mathbb{R}^l$ 分别在点$x\in U$处和点$y=f(x)\in V$处可微. 则复合映射$g\circ f:U\to\mathbb{R}^l$ 在点x处可微且

$$(g\circ f)'(x)=g'(f(x))f'(x) \qquad 即 \qquad \mathrm{d}(g\circ f)_x=\mathrm{d}g_{f(x)}\circ\mathrm{d}f_x.$$

也可等价地写成

$$(g\circ f)'(x)h=g'(f(x))f'(x)h \quad 即 \quad \mathrm{d}(g\circ f)_x(h)=\mathrm{d}g_{f(x)}(\mathrm{d}f_x(h)) \quad \forall\, h\in\mathbb{R}^n.$$

【证】由f,g 分别在点x和$y=f(x)$处可微有

$$f(x+h)-f(x)=f'(x)h+\alpha(h),$$

$$g(y+k)-g(y)=g'(y)k+\beta(k)$$

其中$\alpha(h) = \alpha(x;h) = o(h), \beta(k) = \beta(y;k) = o(k)$分别是$h$ 和k的高阶无穷小. 注意可微蕴含连续. 因此当$|h| << 1$ 时$f(x+h)$位于$y = f(x)$的邻域内. 于是取$k = f(x+h) - f(x)$ 便有$f(x+h) = f(x) + k = y + k$,

$$\begin{aligned}
&g \circ f(x+h) - g \circ f(x) = g(f(x+h)) - g(f(x)) \\
&= g(y+k) - g(y) = g'(y)k + \beta(k) \qquad (\text{注意 } k = f'(x)h + \alpha(h)\ ) \\
&= g'(y)f'(x)h + g'(y)\alpha(h) + \beta(\,f'(x)h + \alpha(h)\,).
\end{aligned}$$

令

$$\gamma(h) = g'(y)\alpha(h) + \beta(\,f'(x)h + \alpha(h)\,), \qquad |h| << 1.$$

则上式写成

$$g \circ f(x+h) = g \circ f(x) + g'(y)f'(x)h + \gamma(h), \qquad |h| << 1.$$

下证$\gamma(h) = o(h)$ 即$\gamma(h)$ 是h 的高阶无穷小. 若此事成立, 则据可微的定义和微分的唯一性即知复合映射$g \circ f$ 在x 可微且$(g \circ f)'(x) = g'(y)f'(x) = g'(f(x))f'(x)$.

易见$g'(y)\alpha(h) = o(h)$. 因此为证$\gamma(h) = o(h)$, 只需证明$\beta(\,f'(x)h + \alpha(h)\,) = o(h)$.

根据高阶无穷小$\beta(k) = o(k)$的定义知, 对任意$\varepsilon > 0$ 存在$\eta > 0$ 使得

$$|\beta(k)| \le \frac{\varepsilon}{\|f'(x)\| + 1}|k| \qquad \forall\, k \in \mathbb{R}^n \quad \text{s.t.} \quad |k| < \eta.$$

而对此$\eta > 0$, 由$|f'(x)h + \alpha(h)| \le \|f'(x)\||h| + |\alpha(h)|$ 和$|\alpha(h)| = o(|h|)$ 知存在$\delta > 0$ 使得只要$|h| < \delta$ 就有$|f'(x)h + \alpha(h)| \le (\|f'(x)\| + 1)|h| < \eta$. 于是只要$|h| < \delta$ 就有

$$|\beta(f'(x)h + \alpha(h))| \le \frac{\varepsilon}{\|f'(x)\| + 1}|f'(x)h + \alpha(h)| \le \varepsilon|h|$$

因此

$$\text{当}\, 0 < |h| < \delta \text{ 时有} \quad \frac{|\beta(\,f'(x)h + \alpha(h)\,)|}{|h|} \le \varepsilon.$$

所以

$$\lim_{|h| \to 0} \frac{|\beta(\,f'(x)h + \alpha(h)\,)|}{|h|} = 0.$$

这就是$\beta(\,f'(x)h + \alpha(h)\,) = o(h)$. $\square$

以上的复合映射微分的锁链法则都是在内层函数的定义域为开集的情况下被证明的——因为开集中每一点都是内点. 其实在很多情况下这个法则对锥点也成立. 但我们不去探讨细节了而只对一种容易的也是常使用的情形给出证明.

【命题8.13.】 设$\Omega\subset\mathbb{R}^n$ 为开集, 映射$f:\Omega\to\mathbb{R}^m$ 在Ω上可微, 设$I\subset\mathbb{R}$为一区间, 它可以是$[a,b],[a,b),(a,b],(a,b)$ 中的任何一个, 设映射$\varphi:I\to\Omega$在I上可微。则复合映射$t\mapsto f\circ\varphi(t)$ 在I 上可微且

$$(f\circ\varphi)'(t)=f'(x)\varphi'(t)\Big|_{x=\varphi(t)}=f'(\varphi(t))\varphi'(t),\quad t\in I.$$

【证】只需对$I=[a,b]$是闭区间的情形进行证明即可(其他情形类似). 为应用上面已证的结果, 我们需要把映射$\varphi(t)$ 保持可微地延拓到一个大一点的开区间$(a-\delta,b+\delta)$上且使得延拓后的φ 仍位于Ω内.

因$\varphi([a,b])$ 是位于Ω内的紧集, 故不难证明存在$\varepsilon>0$ 使得

$$\{x\in\mathbb{R}^n\mid \mathrm{dist}(x,\varphi([a,b]))<\varepsilon\}\subset\Omega.$$

(见第七章作业题, 这里$\mathrm{dist}=\rho$ 是欧氏度量). 取$\delta>0$ 满足

$$|\varphi'(a)|\delta<\varepsilon,\quad |\varphi'(b)|\delta<\varepsilon.$$

然后按切线做延拓:

$$\varphi(t):=\varphi(a)+\varphi'(a)(t-a)\quad \text{当 } a-\delta<t\le a \text{ 时};$$

$$\varphi(t):=\varphi(b)+\varphi'(b)(t-b)\quad \text{当 } b\le t<b+\delta \text{ 时}.$$

易见延拓后的φ 在$(a-\delta,b+\delta)$ 内处处可微. 此外当$t\in(a-\delta,a]$时

$$\mathrm{dist}(\varphi(t),\varphi([a,b]))\le|\varphi(t)-\varphi(a)|\le|\varphi'(a)||t-a|\le|\varphi'(a)|\delta<\varepsilon.$$

所以$\varphi(t)\in\Omega$. 同理有$\varphi(t)\in\Omega$ for all $t\in[b,b+\delta)$. 这证明了$\varphi(t)\in\Omega$ for all $t\in(a-\delta,\,b+\delta)$.

做了这样处理后,我们就使用上面已证的复合映射微分法了: 因$(a-\delta,\,b+\delta)$是开区间, Ω 是开集且$\varphi(t)\in\Omega$ for all $t\in(a-\delta,\,b+\delta)$, 故由复合映射微分的锁链法则知复合映射$t\mapsto f\circ\varphi(t)$ 在$(a-\delta,\,b+\delta)$内处处可微且

$$(f\circ\varphi)'(t)=f'(x)\varphi'(t)\Big|_{x=\varphi(t)}=f'(\varphi(t))\varphi'(t)\qquad\forall\, t\in(a-\delta,b+\delta).$$

最后再由$[a,b]\subset(a-\delta,b+\delta)$ 即知命题成立. $\square$

【锁链法则的具体公式】

在微分验算中, 最复杂者当属复合映射的微分的计算, 其中将频繁使用偏导数. 下面我们给出详细的计算公式.

设$U \subset \mathbb{R}^n$ 是开集, $V \subset \mathbb{R}^m$是开集, 映射$f : U \to V$ 和$g : V \to \mathbb{R}^l$ 分别在点$x \in U$处和点$y = f(x) \in V$ 处可微. 写

$$g(y) = \begin{pmatrix} g_1(y) \\ g_2(y)) \\ \vdots \\ g_l(y) \end{pmatrix}, \qquad y = f(x) = \begin{pmatrix} f_1(x) \\ f_2(x)) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

则

$$(g \circ f)'(x) = \begin{pmatrix} (g_1 \circ f)'(x) \\ (g_2 \circ f)'(x) \\ \vdots \\ (g_l \circ f)'(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1 \circ f}{\partial x_1}(x) & \frac{\partial g_1 \circ f}{\partial x_2}(x) & ... & \frac{\partial g_1 \circ f}{\partial x_n}(x) \\ \frac{\partial g_2 \circ f}{\partial x_1}(x) & \frac{\partial g_2 \circ f}{\partial x_2}(x) & ... & \frac{\partial g_2 \circ f}{\partial x_n}(x) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_l \circ f}{\partial x_1}(x) & \frac{\partial g_l \circ f}{\partial x_2}(x) & ... & \frac{\partial g_l \circ f}{\partial x_n}(x) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(y) & \frac{\partial g_1}{\partial y_2}(y) & ... & \frac{\partial g_1}{\partial y_m}(y) \\ \frac{\partial g_2}{\partial y_1}(y) & \frac{\partial g_2}{\partial y_2}(y) & ... & \frac{\partial g_2}{\partial y_m}(y) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_l}{\partial y_1}(y) & \frac{\partial g_l}{\partial y_2}(y) & ... & \frac{\partial g_l}{\partial y_m}(y) \end{pmatrix}_{y=f(x)} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & ... & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & ... & \frac{\partial f_2}{\partial x_n}(x) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & ... & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^m \frac{\partial g_1}{\partial y_i}(y)\frac{\partial f_i}{\partial x_1}(x) & \sum_{i=1}^m \frac{\partial g_1}{\partial y_i}(y)\frac{\partial f_i}{\partial x_2}(x) & ...... & \sum_{i=1}^m \frac{\partial g_1}{\partial y_i}(y)\frac{\partial f_i}{\partial x_n}(x) \\ \sum_{i=1}^m \frac{\partial g_2}{\partial y_i}(y)\frac{\partial f_i}{\partial x_1}(x) & \sum_{i=1}^m \frac{\partial g_2}{\partial y_i}(y)\frac{\partial f_i}{\partial x_2}(x) & ...... & \sum_{i=1}^m \frac{\partial g_2}{\partial y_i}(y)\frac{\partial f_i}{\partial x_n}(x) \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^m \frac{\partial g_l}{\partial y_i}(y)\frac{\partial f_i}{\partial x_1}(x) & \sum_{i=1}^m \frac{\partial g_l}{\partial y_i}(y)\frac{\partial f_i}{\partial x_2}(x) & ...... & \sum_{i=1}^m \frac{\partial g_l}{\partial y_i}(y)\frac{\partial f_i}{\partial x_n}(x) \end{pmatrix}_{y=f(x)}.$$

即

$$\frac{\partial g_k \circ f}{\partial x_j}(x) = \sum_{i=1}^m \frac{\partial g_k}{\partial y_i}(y)\frac{\partial f_i}{\partial x_j}(x)\Big|_{y=f(x)}, \quad k = 1, 2, ..., l; \quad j = 1, 2, ..., n.$$

或写成

$$\frac{\partial g_k \circ f}{\partial x_j}(x) = \sum_{i=1}^m \Big(\frac{\partial g_k}{\partial y_i}\Big)(f(x))\frac{\partial f_i}{\partial x_j}(x), \quad k = 1, 2, ..., l; \quad j = 1, 2, ..., n.$$

以上包含了**外层映射g为数值函数的复合函数求导公式即$l = 1$的情形：**

$$\frac{\partial}{\partial x_j}\Big(g(f(x)\Big) = \Big(\frac{\partial g}{\partial y_1}\Big)(f(x))\frac{\partial f_1}{\partial x_j}(x) + \Big(\frac{\partial g}{\partial y_2}\Big)(f(x))\frac{\partial f_2}{\partial x_j}(x) + \cdots + \Big(\frac{\partial g}{\partial y_m}\Big)(f(x))\frac{\partial f_m}{\partial x_j}(x)$$

$j = 1, 2, ..., n.$

【例】$l = 1, m = 3, n = 2.$

设函数$\alpha(u,v), \beta(u,v), \gamma(u,v)$在点$(u,v) \in \mathbb{R}^2$ 可微, 而函数$g(x,y,z)$在对应点$(x,y,z) = (\alpha(u,v), \beta(u,v), \gamma(u,v))$可微. 则复合函数$(u,v) \mapsto g(\alpha(u,v), \beta(u,v), \gamma(u,v))$ 在(u,v)可微且

$$\begin{aligned}\frac{\partial}{\partial u}\Big(g(\alpha(u,v), \beta(u,v), \gamma(u,v))\Big) &= \Big(\frac{\partial g}{\partial x}\Big)(\alpha(u,v), \beta(u,v), \gamma(u,v))\frac{\partial \alpha}{\partial u}(u,v)\\ &+\Big(\frac{\partial g}{\partial y}\Big)(\alpha(u,v), \beta(u,v), \gamma(u,v))\frac{\partial \beta}{\partial u}(u,v)\\ &+\Big(\frac{\partial g}{\partial z}\Big)(\alpha(u,v), \beta(u,v), \gamma(u,v))\frac{\partial \gamma}{\partial u}(u,v),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial v}\Big(g(\alpha(u,v), \beta(u,v), \gamma(u,v))\Big) &= \Big(\frac{\partial g}{\partial x}\Big)(\alpha(u,v), \beta(u,v), \gamma(u,v))\frac{\partial \alpha}{\partial v}(u,v)\\ &+\Big(\frac{\partial g}{\partial y}\Big)(\alpha(u,v), \beta(u,v), \gamma(u,v))\frac{\partial \beta}{\partial v}(u,v)\\ &+\Big(\frac{\partial g}{\partial z}\Big)(\alpha(u,v), \beta(u,v), \gamma(u,v))\frac{\partial \gamma}{\partial v}(u,v).\end{aligned}$$

□

对Chain Rule 的解释和记忆方法...

【例】设$f(x,y,z) = \log\Big(e^{x+y} + z^2\cos(xz)\Big)$. 在其定义域内计算$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$:

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y,z) &= \frac{1}{e^{x+y} + z^2\cos(xz)}\frac{\partial}{\partial x}\Big(e^{x+y} + z^2\cos(xz)\Big) = \frac{e^{x+y} - z^3\sin(xz)}{e^{x+y} + z^2\cos(xz)},\\ \frac{\partial f}{\partial y}(x,y,z) &= \frac{1}{e^{x+y} + z^2\cos(xz)}\frac{\partial}{\partial y}\Big(e^{x+y} + z^2\cos(xz)\Big) = \frac{e^{x+y}}{e^{x+y} + z^2\cos(xz)},\\ \frac{\partial f}{\partial z}(x,y,z) &= \frac{1}{e^{x+y} + z^2\cos(xz)}\frac{\partial}{\partial z}\Big(e^{x+y} + z^2\cos(xz)\Big) = \frac{2z\cos(xz) - z^2x\sin(xz)}{e^{x+y} + z^2\cos(xz)}.\end{aligned}$$

□

【例】设$x = (x_1, x_2, ..., x_n)^\tau \in \mathbb{R}^n$ (列向量). 计算$x \mapsto |x|^\alpha$ 在$x \in \mathbb{R}^n \setminus \{0\}$的微分$(|x|^\alpha)'$, 这里$\alpha \in \mathbb{R}$是常数.

【解】计算

$$\begin{aligned}(|x|^\alpha)' &= \Big((|x|^2)^{\frac{\alpha}{2}}\Big)' = \frac{\alpha}{2}\Big((|x|^2)^{\frac{\alpha}{2}-1}\Big)(|x|^2)' = \frac{\alpha}{2}\Big((|x|^2)^{\frac{\alpha}{2}-1}\Big)2x^\tau\\ &= \alpha\Big((|x|^2)^{\frac{\alpha-2}{2}}\Big)x^\tau = \alpha|x|^{\alpha-2}x^\tau = \alpha\frac{x^\tau}{|x|^{2-\alpha}}, \quad x \in \mathbb{R}^n \setminus \{0\}.\end{aligned}$$

$\square$

【方向导数,梯度和等高集】

顾名思义, 方向导数就是函数沿着某个方向$\mathbf{v}$的导数. 要点是: 这种导数(特别是计算这些导数的过程)都是一元函数行为.

【定义】 设$\Omega \subset \mathbb{R}^n$为开集, $f : \Omega \to \mathbb{R}^m$, $x \in \Omega$, 设$\mathbf{v} \in \mathbb{R}^n$为一个单位向量即$|\mathbf{v}| = 1$. 若差商极限

$$D_{\mathbf{v}} f(x) := \lim_{t \to 0} \frac{f(x + t\mathbf{v}) - f(x)}{t} \in \mathbb{R}^m$$

存在, 则称$D_{\mathbf{v}} f(x)$为f 在点x处沿着方向$\mathbf{v}$的方向导数. $\square$

如果对于坐标轴的方向$\mathbf{v} = \mathbf{e}_j = (0, ..., 0, 1, 0, ..., 0)^\tau$, 方向导数$D_{\mathbf{e}_j} f(x) \in \mathbb{R}^m$存在, 则$D_{\mathbf{e}_j} f(x)$ 是f在点x处对于自变量的第j个坐标x_j的偏导数, 即此时有

$$D_{\mathbf{e}_j} f(x) = D_j f(x) = \frac{\partial f}{\partial x_j}(x).$$

一个自然的问题是: 如果映射$f : \Omega \to \mathbb{R}$ 在Ω内某点x_0处的所有方向导数都存在, 能否推出f在x_0可微？或连续？事实说明, 对于**多元函数**来说, 即便在f某点x_0处的所有方向导数都存在, 也不一定保证f在点x_0连续, 更不保证f在点x_0可微.

我们以二元函数为例说明这点.

【例】

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{if} \quad x^2 + y^2 > 0, \\ 0 & \text{if} \quad x = y = 0. \end{cases}$$

不难看出这函数f在$\mathbb{R}^2 \setminus \{(0, 0)\}$ 内处处可微.

看f在$(0, 0)$的方向导数: 任取单位向量$\mathbf{v} = (\cos\theta, \sin\theta)^\tau$, 计算

$$\frac{f(t\cos\theta, t\sin\theta) - f(0, 0)}{t} = \frac{1}{t} \cdot \frac{t^3 \cos^2\theta \sin\theta}{t^4 \cos^4\theta + t^2 \sin^2\theta} = \frac{\cos^2\theta \sin\theta}{t^2 \cos^4\theta + \sin^2\theta}, \quad t \neq 0.$$

通过考察$\sin\theta \neq 0$ 和$\sin\theta = 0$ 可知方向导数总存在:

$$D_{\mathbf{v}} f(0, 0) = \lim_{t \to 0} \frac{\cos^2\theta \sin\theta}{t^2 \cos^4\theta + \sin^2\theta} = \begin{cases} \frac{\cos^2\theta}{\sin\theta} & \text{if} \quad \sin\theta \neq 0, \\ 0 & \text{if} \quad \sin\theta = 0. \end{cases}$$

但是当(x, y) 沿着弯曲的曲线, 例如沿着抛物线$y = x^2$, 趋于原点$(0, 0)$ 时却有

$$\lim_{x \to 0} f(x, x^2) = \lim_{x \to 0} \frac{x^2 x^2}{x^4 + x^4} = \frac{1}{2} \neq f(0, 0) = 0.$$

这说明f在$(0,0)$ 不连续, 从而更不可微. $\square$

这个例子说明, “可偏导”与“可微”有较大区别. 可微性是通过**整体高阶无穷小**来定义的, 因此可微蕴含连续是显然的. 但偏导数和任意方向导数都是按一维方向定义的单变量的导数, 是一维行为, 它无法排除各一维方向的变化程度很可能不一致的情形.

这个例子同时表明: 对于多元函数来说, 连续性其实是比较强的性质(我们已在上一章已体验了这一点).

刚才说到“各一维方向的变化程度很不一致”导致函数可能不连续. 如果是“一致的”, 例如函数在某点附近函数的各偏导数有界, 或者所有偏导数都在该点连续, 则可得到肯定的结果, 这就是上面**命题8.9** 的结论.

下面的命题给出了 当可微时微分与方向导数的关系:

【命题8.14.】设$\Omega\subset\mathbb{R}^n$为开集, $f:\Omega\to\mathbb{R}^m$ 在点$x\in\Omega$可微. 则f在点x处沿任何方向$\mathbf{v}$的方向导数$D_{\mathbf{v}}f(x)$都存且

$$D_{\mathbf{v}}f(x)=\mathrm{d}f_x(\mathbf{v})=f'(x)\mathbf{v}.$$

特别f在点x处的每个偏导数都存在:

$$D_jf(x)=f'(x)\mathbf{e}_j=\frac{\partial f}{\partial x_j}(x),\quad j=1,2,...,n.$$

【证】因$x\in\Omega$而Ω是开集, 故存在$\delta>0$ 使得$B(x,\delta)\subset\Omega$. 令$\varphi(t)=x+t\mathbf{v},t\in(-\delta,\delta)$, 则$\varphi:(-\delta,\delta)\to B(x,\delta)\subset\Omega$ 可微且$\varphi'(t)=\mathbf{v}$. 由假设条件和复合映射可微性和锁链法则知$t\mapsto f\circ\varphi(t)=f(\varphi(t))$ $t=0$ 可微并由方向导数的定义有

$$\begin{aligned}D_{\mathbf{v}}f(x)&=\lim_{t\to0}\frac{f(x+t\mathbf{v})-f(x)}{t}=\lim_{t\to0}\frac{f(\varphi(t))-f(\varphi(0))}{t}\\&=(f\circ\varphi)'(0)=f'(\varphi(0))\varphi'(0)=f'(x)\mathbf{v}=\mathrm{d}f_x(\mathbf{v}).\end{aligned}$$

也可直接用微分的定义证明: 由f在点x处可微有

$$\begin{aligned}&\frac{f(x+t\mathbf{v})-f(x)}{t}=\frac{f'(x)t\mathbf{v}+o(t\mathbf{v})}{t}=f'(x)\mathbf{v}+\frac{o(t\mathbf{v})}{t}\\&\to f'(x)\mathbf{v}\quad\text{as}\quad t\to0.\end{aligned}$$

所以$D_{\mathbf{v}}f(x)$存在且等于$f'(x)\mathbf{v}$. $\square$

作为这一命题的推论我们有

【命题8.15(数值函数的情形).】 设$\Omega\subset\mathbb{R}^n$为开集, $f:\Omega\to\mathbb{R}$ 在点$x\in\Omega$可微. 则f在点x处沿任何方向$\mathbf{v}$的方向导数$D_{\mathbf{v}}f(x)$都存且它与梯度$\nabla f(x)$有如下关系:

$$D_{\mathbf{v}}f(x)=\langle\nabla f(x),\mathbf{v}\rangle. \qquad \square$$

【梯度的意义】 让我们从方向导数与梯度的上述关系出发做进一步分析. 假设$f:U(x_0)\to\mathbb{R}$在x_0 可微. 我们希望找到一个单位向量$\mathbf{e}$ 使得f沿着方向$\mathbf{e}$的变化律最大(在绝对值的意义下).

首先由Cauchy不等式有

$$|D_{\mathbf{v}}f(x_0)|=|\langle\nabla f(x_0),\,\mathbf{v}\rangle|\le|\nabla f(x_0)| \qquad \forall\,|\mathbf{v}|=1.$$

显然若$\nabla f(x_0)=0$ 则有

$$D_{\mathbf{v}}f(x_0)=0 \qquad \forall\,|\mathbf{v}|=1.$$

这种满足梯度$\nabla f(x)=0$ 的点叫做f的驻点或稳定点或临界点. 我们将在§**8.5** 节中对这种点做系统研究. 设$\nabla f(x_0)\neq 0$. 此时我们取

$$\mathbf{e}=\pm\frac{\nabla f(x_0)}{|\nabla f(x_0)|}$$

则有

$$D_{\mathbf{e}}f(x_0)=\langle\nabla f(x_0),\,\mathbf{e}\rangle=\pm|\nabla f(x_0)|.$$

于是得到

$$\max_{|\mathbf{v}|=1}|D_{\mathbf{v}}f(x_0)|=|D_{\mathbf{e}}f(x_0)|=|\nabla f(x_0)|.$$

因此梯度方向$\pm\frac{\nabla f(x_0)}{|\nabla f(x_0)|}$是$f$在点$x_0$变化律最大的方向. 进一步还有:

正梯度方向$\frac{\nabla f(x_0)}{|\nabla f(x_0)|}$是$f$在点$x_0$上升最快的方向;

负梯度方向$-\frac{\nabla f(x_0)}{|\nabla f(x_0)|}$是$f$在点$x_0$下降最快的方向.

物理学和数学模型中常出现梯度$\nabla f,\nabla u$, 等等. 这是因为自然界中一般来说若一物体不受复杂因素干扰, 则它在任意点的运动方向等于使得它容易达到最大变化的方向.

例如一个大石头从山上滚下来, 它基本上是沿着下降最快的方向, 即沿着山的梯度方向滚落的, 而不会沿着盘山公路滚落.[3]

让我们再看一个进一步的问题：考虑一块连续可微的山坡型曲面$y = f(x) \in \mathbb{R}$ $(x \in \Omega)$, "山坡型"指$\nabla f(x) \neq 0$ for all $x \in \Omega$. 我们要设计一条道路$t \mapsto x(t)$ 使得$f(x)$从x_0出发沿此道路下降最快. 这等于说$t \mapsto x(t)$的变化方向$x'(t)$ 时刻等于f在点$x(t)$处的负梯度方向, 也即$x(t)$应满足微分方程:

$$x'(t) = -\rho(x(t))\frac{\nabla f(x(t))}{|\nabla f(x(t))|}, \quad t \in [0, T]; \quad x(0) = x_0.$$

这里$\rho(x) > 0$是根据实际情况或根据需要而设计的连续函数. 此方程等价于积分方程:

$$x(t) = x_0 - \int_0^t \rho(x(s))\frac{\nabla f(x(s))}{|\nabla f(x(s))|}\mathrm{d}s, \quad t \in [0, T].$$

【等高集】

现在看函数沿着与梯度垂直的方向的行为. 设$\Omega \subset \mathbb{R}^n$为开集, 函数$f : \Omega \to \mathbb{R}$. 给定$c \in \mathbb{R}$. 我们称集合

$$H_c(\Omega) := \{x \in \Omega \mid f(x) = c\}$$

为f的高度等于c的等高集. 当$\Omega \subset \mathbb{R}^2$时, 也称等高集为等高线.

可以假定Ω 是连通开集(否则在它的每个连通分支上考虑问题). 设f在Ω内可微.

如果$\nabla f(x) \equiv 0$, 即$f'(x) \equiv 0$, 则应用下一节中的**定理8.17(一般映射的微分中值不等式)**的**推论2** 知f 为常值函数: $f(x) \equiv C$, $x \in \Omega$. 此时, 当$c \neq C$时等高集$H_c(\Omega) =$空集; 当$c = C$时$H_c(\Omega) = \Omega$.

假设$\nabla f(x) \not\equiv 0$. 为了看清事物, 我们进一步假定$x \mapsto \nabla f(x)$在Ω内连续. 后面我们将利用隐函数定理证明, 若$\nabla f(x_0) \neq 0$, 则在x_0的一个邻域$U = U(x_0)$内有$\nabla f(x) \neq 0$ 并且等高集$H_c(U)$为一条可微曲线(当$n = 2$) 或可微曲面(当$n \geq 3$), 这曲线/曲面称为等

[3]**顺便说几句：大自然中物体的变化, 在逻辑上有多种可能的情形, 其中最大可能的情形是耗能最小的情形, 即物体朝着使其耗能最小的方向进行自身的改变. 例如石头从山上滑落(受引力的作用), 大体上是沿着下降最快的路线进行, 因为这样最"省事"(耗能最小). 又例如, 光的传播总是沿着耗时最短(等价于耗能最小)的路线进行. 当然在实际情况中由于某些随机因素的作用, 物体的实际行为与"确定性的理论"的预测有一定偏差. 现代物理学实际上采取几率的观点：物体之所以能沿着耗能最小的方向发展, 是因为这种情况发生的几率最大. 至于为什么这种情况几率最大？没有答案, 可认为是自然的本性. 另一个观点是演化、动态的观点, 即随着时间的发展, 事物将产生变化. 如果把几率的观点(例如量子力学)、时间观念和宏观确定性理论相结合, 那么我们对"自然规律"便可以达到这样的认知程度：它既不以人的意志为转移, 也不完全是初态决定论的.**

高线/等高面. 等高面中的曲线当然也称为等高线. 眼下我们暂且承认这个结论, 来看发生什么. 不失一般性, 我们就假定$\nabla f(x) \neq 0$ 于Ω. 我们主要考察等高线的行为.

设$x = x(t)$ 是Ω 内的一条可微曲线, $t \in (\alpha, \beta)$. 由复合函数的微分公式我们有以下等价关系:

$$\begin{aligned}
& f(x(t)) = \text{const.}, \qquad t \in (\alpha, \beta) \\
& \Longleftrightarrow \\
& 0 = \frac{\mathrm{d}}{\mathrm{d}t} f(x(t)) = \langle \nabla f(x(t)),\, x'(t) \rangle \qquad \forall\, t \in (\alpha, \beta) \\
& \Longleftrightarrow x'(t) \perp \nabla f(x(t)) \qquad \forall\, t \in (\alpha, \beta).
\end{aligned}$$

这就是说: 曲线$x = x(t)$ 是f的一条等高线当且仅当该曲线在t 时刻的切线垂直于f在点$x(t)$处的梯度.

在$n = 2$ 即在$\Omega \subset \mathbb{R}^2$的实际问题中, 等高线有局部显示表示, 即$y = y(x)$ $(x \in (a, b))$ 的形式, 或$x = x(y)\,(y \in (a, b))$ 的形式. 以前者为例. 设$x_0 \in (a, b), y_0 = y(x_0)$. 则由等高性$f(x, y(x)) \equiv c$ 于(a, b) 并使用梯度记号

$$\nabla f(x, y) = \Big(\frac{\partial f}{\partial x}(x, y),\, \frac{\partial f}{\partial y}(x, y) \Big)$$

我们有

$$0 = \frac{\mathrm{d}}{\mathrm{d}x} \Big(f(x, y(x)) \Big) = \Big(\frac{\partial f}{\partial x} \Big)(x, y(x)) + \Big(\frac{\partial f}{\partial y} \Big)(x, y(x)) y'(x) = \langle \nabla f(x, y(x)),\, (1, y'(x)) \rangle$$

for all $x \in (a, b)$. 也即

$$(1, y'(x)) \perp \nabla f(x, y(x)) \qquad \forall\, x \in (a, b).$$

特别在考察点$(x_0, y_0) = (x_0, y(x_0))$ 处有

$$(1, y'(x_0)) \perp \nabla f(x_0, y_0) \qquad \text{即} \qquad \frac{(1, y'(x_0))}{\sqrt{1 + y'(x_0)^2}} \perp \frac{\nabla f(x_0, y_0)}{|\nabla f(x_0, y_0)|}.$$

因$\frac{(1, y'(x_0))}{\sqrt{1 + y'(x_0)^2}}$ 是等高线$y = y(x)$ 在点(x_0, y_0)处的切线

$$y = y_0 + y'(x_0)(x - x_0)$$

的(单位)切向量, 而此切线的法向量

$$\mathbf{n} = \frac{(y'(x_0), -1)}{\sqrt{y'(x_0)^2 + 1}} \perp \frac{(1, y'(x_0))}{\sqrt{1 + y'(x_0)^2}}$$

故(因$\dim(\mathbb{R}^2)=2$) 必有

$$\mathbf{n}=\pm\frac{\nabla f(x_0,y_0)}{|\nabla f(x_0,y_0)|}.$$

这就是说**等高线$y=y(x)$ 在点(x_0,y_0)处的切线垂直于梯度$\nabla f(x_0,y_0)$.** 利用这个原理人们能够画出山区的等高线地图(图示). 在图中, 等高线之间距离较近的地区比较陡, 距离较远的地区比较缓.

【作业题】

1. 设$A\in\mathbb{R}^{l\times m}$, $\Omega\subset\mathbb{R}^n$为开集, $f:\Omega\to\mathbb{R}^m=\mathbb{R}^{m\times 1}$ 在Ω上可微, $g(x)=Af(x), x\in\Omega$. 用微分的定义证明g也在Ω上可微且$g'(x)=Af'(x)$. (注意考察高阶无穷小.)

2. (1) 设

$$f(x,y)=\begin{pmatrix} e^x\cos y \\ e^x\sin y\end{pmatrix},\quad (x,y)\in\mathbb{R}^2.$$

证明f 处处可微并求f的微分(Jacobi矩阵) $f'(x,y)$ 和行列式$\det f'(x,y)$.

(2) 设

$$f(r,\theta,\phi)=\begin{pmatrix} r\sin\theta\cos\phi \\ r\sin\theta\sin\phi \\ r\cos\theta\end{pmatrix},\quad (r,\theta,\phi)\in\mathbb{R}^3.$$

证明f 处处可微并求f的微分(Jacobi矩阵) $f'(r,\theta,\phi)$ 和行列式$\det J_f(r,\theta,\phi)$.

(3) 说明: (1)中给出的映射f的微分(Jacobi矩阵) 处处可逆, 但f不是单射; (2)中的映射f在在其微分可逆的区域内也不是单射。

3.

(1) 设$u(x,y,z)=x^{\frac{y}{z}}, (x,y,z)\in\mathbb{R}^3_{>0}$. 求$\frac{\partial u}{\partial x},\frac{\partial u}{\partial y},\frac{\partial u}{\partial z}$.

(2) 设$u(x,y)=\frac{x-y}{x+y}\log\left(\frac{y}{x}\right)$, $(x,y)\in\mathbb{R}^2_{>0}$. 验证$u$是偏微分方程

$$x\frac{\partial u}{\partial x}+y\frac{\partial u}{\partial y}=0$$

的一个解.

(3) 设$u(x,y)=\arctan\left(\frac{x+y}{1-xy}\right)$, $xy\neq 1$. 验证u是偏微分方程

$$(1+x^2)\frac{\partial u}{\partial x}=(1+y^2)\frac{\partial u}{\partial y}$$

的一个解.

4. 设开集$\Omega\subset\mathbb{R}^2$, $f:\Omega\to\mathbb{R}$ 在Ω 上有偏导数, 即$\frac{\partial f}{\partial x}(x,y),\frac{\partial f}{\partial y}(x,y)$ 处处存在. 假设$\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}$中有一个在点$(x_0,y_0)\in\Omega$ 连续. 证明f 在(x_0,y_0) 可微.

5. 设$\Omega\subset\mathbb{R}^m$ 为开集, $f\in C^1(\Omega,\mathbb{R}^m)$. 设$x_0\in\Omega$ 且$f'(x_0)$可逆. 证明存在$\delta>0$ 使得$B(x_0,\delta)\subset\Omega$ 且

$$\|f'(x)-f'(x_0)\|\le\frac{1}{2\|[f'(x_0)]^{-1}\|}\qquad\forall x\in B(x_0,\delta)$$

进而证明当$x\in B(x_0,\delta)$ 时, $f'(x)$ 可逆且

$$\|[f'(x)]^{-1}-[f'(x_0)]^{-1}\|\le 2\|[f'(x_0)]^{-1}\|^2\|f'(x)-f'(x_0)\|\qquad \forall\, x\in B(x_0,\delta).$$

这表明逆矩阵$x\mapsto [f'(x)]^{-1}$在x_0 附近存在且在x_0 连续. (参见**命题8.4.**)

6. (1) 设$y\in\mathbb{R}^m$为列向量. 证明映射$y\mapsto y/|y|$ 在$\mathbb{R}^m\setminus\{0\}$上可微且

$$\left(\frac{y}{|y|}\right)'=\frac{1}{|y|}\left(\mathrm{I}-\frac{yy^\tau}{|y|^2}\right),\quad y\in\mathbb{R}^m\setminus\{0\}.$$

(2) 设$m\ge n, A\in\mathbb{R}^{m\times n}$ 为满秩矩阵. 证明

$$\left(\frac{Ax}{|Ax|}\right)'=\frac{1}{|Ax|}\left(\mathrm{I}-\frac{Ax(Ax)^\tau}{|Ax|^2}\right)A,\quad x\in\mathbb{R}^n\setminus\{0\}.$$

一般地, 设开集$\Omega\subset\mathbb{R}^n$, 映射$\varphi:\Omega\to\mathbb{R}^m$ 在点$x\in\Omega$可微且$\varphi(x)\ne 0$. 证明

$$\left(\frac{\varphi(x)}{|\varphi(x)|}\right)'=\frac{1}{|\varphi(x)|}\left(\mathrm{I}-\frac{\varphi(x)(\varphi(x))^\tau}{|\varphi(x)|^2}\right)\varphi'(x).$$

以上$\mathrm{I}\in\mathbb{R}^{m\times m}$为单位阵.

7. 陈书习题**8.3.1, 8.3.3, 8.3.4.** (第二册pp.115-117.)

§8.3 微分中值定理

我们知道在研究一元实函数的整体性质时, Lagrange 中值定理起了核心作用, 它建立了局部(导数$f'(\xi)$) 和整体(任意增量$f(x)-f(y)$)的关系. 这种中值定理对多元实函数也成立. 对于向量值函数即映射, 这个中值定理也基本被保持了, 只是中值等式将被减弱成中值不等式.

首先我们引进**直线段**的定义

【**定义**】若$x,y\in\mathbb{R}^n$, 则称

$$[x,y]=\{(1-t)x+ty\mid 0\le t\le 1\}=[y,x]$$

为连接x,y的**闭直线段**, 而称

$$(x,y)=\{(1-t)x+ty\mid 0<t<1\}=(y,x)$$

为是以x,y为端点的**开直线段**. [注: 当$x=y$时, $[x,x]=(x,x)=\{x\}$ 是单点集.] $\Box$

按照这个记号我们知道：$\mathbb{R}^n$中的一个集合E 是凸集当且仅当对任意$x,y\in E$ 都有$[x,y]\subset E$.

【**定理8.16(实值函数的微分中值等式).**】设$E\subset\mathbb{R}^n$, 函数$f:E\to\mathbb{R}$ 在E上连续, 在E°内可微. 则对任意$x,y\in E$ 满足闭直线段$[x,y]\subset E$ 且开直线段$(x,y)\subset E^\circ$ 者, 都存在$\xi\in(x,y)$ 使得

$$f(x)-f(y)=f'(\xi)(x-y).$$

这也可用梯度∇f写成

$$f(x)-f(y)=\langle\nabla f(\xi),x-y\rangle.$$

也常写成

$$\exists\, 0<\theta<1\quad\text{s.t.}\quad f(x)=f(y)+\langle\nabla f((1-\theta)y+\theta x),x-y\rangle.\tag{$\spadesuit$}$$

【证】设$[x,y]\subset E$ 满足$(x,y)\subset E^\circ$. 令

$$\varphi(t):=(1-t)y+tx=y+t(x-y)\in E,\qquad t\in[0,1].$$

则显然$\varphi'(t)=x-y$ for all $t\in(0,1)$. 令$F(t)=f(\varphi(t)),t\in[0,1]$, 则由假设和复合函数连续性知$F$在$[0,1]$ 上连续并且由**定理8.12(复合映射的可微性和微分的锁链法则)**

知F在开区间$(0,1)$内可微且

$$F'(t)=f'(\varphi(t))\varphi'(t)=f'(\varphi(t))(x-y)=\langle\nabla f(\varphi(t)),\,x-y\rangle,\quad t\in(0,1).$$

在$[0,1]$上对$F(t)$ 应用Lagrange 中值定理并注意$F(1)=f(\varphi(1))=f(x)$,
$F(0)=f(\varphi(0))=f(y)$ 知存在$\theta\in(0,1)$ 使得

$$f(x)-f(y)=F(1)-F(0)=F'(\theta)=f'(\xi)(x-y)=\langle\nabla f(\xi),\,x-y\rangle$$

其中$\xi=\varphi(\theta)\in[x,y]$. $\quad\Box$.

【定理8.17(一般映射的微分中值不等式).】设$E\subset\mathbb{R}^n$, 函数$f:E\to\mathbb{R}^m$ 在E上连续, 在E°内可微. 则对任意$x,y\in E$ 满足闭直线段$[x,y]\subset E$ 且开直线段$(x,y)\subset E^\circ$ 者, 都存在$\xi\in(x,y)$ 使得

$$|f(x)-f(y)|\le\|f'(\xi)\||x-y|.$$

其中$\|\cdot\|$是矩阵范数(例如是矩阵2-范数或算子范数).

【证】写$f=(f_1,f_2,...,f_m)^\tau$. 任取$\mathbf{a}=(a_1,a_2,...,a_m)^\tau\in\mathbb{R}^m$ 令

$$g(x)=\langle\mathbf{a},f(x)\rangle=\sum_{i=1}^m a_if_i(x),\quad x\in\Omega.$$

则由f 可微知每个f_j 可微从而g在E° 内可微且由“线性组合的微分等于微分的线性组合”有

$$g'(x)=\sum_{i=1}^m a_if_i'(x)=(a_1,a_2,...,a_m)\begin{pmatrix}f_1'(x)\\ f_2'(x)\\ \vdots\\ f_m'(x)\end{pmatrix}=\mathbf{a}^\tau f'(x),\quad x\in E^\circ.$$

应用上一定理知任意$x,y\in E$ 满足$[x,y]\subset E$ 且$(x,y)\subset E^\circ$, 都存在$\xi\in(x,y)$ 使得

$$g(x)-g(y)=g'(\xi)(x-y)=\mathbf{a}^\tau f'(\xi)(x-y).$$

由g的定义易见上式可写成

$$\langle\mathbf{a},f(x)-f(y)\rangle=\mathbf{a}^\tau f'(\xi)(x-y).\qquad(\P)$$

运用矩阵乘积的范数不等式有

$$
\begin{aligned}
|\langle \mathbf{a}, f(x)-f(y)\rangle| &= |\mathbf{a}^\tau f'(\xi)(x-y)| \le \|\mathbf{a}^\tau f'(\xi)\||x-y)| \\
&\le \|\mathbf{a}^\tau\|\|f'(\xi)\||x-y| = |\mathbf{a}|\|f'(\xi)\||x-y|.
\end{aligned}
$$

若$f(x)-f(y)\neq 0$, 则事先取$\mathbf{a}=\frac{f(x)-f(y)}{|f(x)-f(y)|}$ 得到

$$|f(x)-f(y)|\le \|f'(\xi)\||x-y|.$$

若$f(x)-f(y)=0$, 则这个不等式自动成立. □

【推论1(Lipschitz连续性)】设$\Omega\subset\mathbb{R}^n$ 为开凸集, 映射$f:\overline{\Omega}\to\mathbb{R}^m$ 在$\overline{\Omega}$上连续、在Ω内可微且微分$f'(x)$在Ω内有界, 即$L:=\sup\limits_{x\in\Omega}\|f'(x)\|<+\infty$. 则$f$在$\overline{\Omega}$上是Lipschitz连续的且Lipschitz 常数可取为L, 即

$$|f(x)-f(y)|\le L|x-y| \qquad \forall\, x,y\in\overline{\Omega}.$$

【证】因Ω 是开凸集, 故$\Omega=\Omega^\circ$ 且对任意$x,y\in\Omega$ 都有$[x,y]\subset\Omega$. 于是应用一般映射的微分中值不等式知存在$\xi\in(x,y)$ 使得

$$|f(x)-f(y)|\le\|f'(\xi)\||x-y|\le L|x-y|.$$

现在对任意$x,y\in\overline{\Omega}$, 由闭包的序列刻画知在Ω内存在点列$\{x_k\}_{k=1}^\infty,\{y_k\}_{k=1}^\infty$ 使得$x_k\to x$, $y_k\to y\,(k\to\infty)$. 于是由f在$\overline{\Omega}$上连续便有

$$|f(x)-f(y)|=\lim_{k\to\infty}|f(x_k)-f(y_k)|\le\lim_{k\to\infty}L|x_k-y_k|=L|x-y|.$$

所以f在$\overline{\Omega}$上Lipschitz 连续且Lipschitz 常数可取为L. □

【推论2(常值函数的判定)】设$\Omega\subset\mathbb{R}^n$ 为$\mathbb{R}^n$中的连通开集. 则有:

(a) 若函数$f:\Omega\to\mathbb{R}$ 在Ω内可微且$\nabla f(x)=0$ for all $x\in\Omega$, 则$f(x)\equiv$ 常值, $x\in\Omega$.

一般地, 若映射$f:\Omega\to\mathbb{R}^m$ 在Ω内可微且$f'(x)=0$ for all $x\in\Omega$, 则$f(x)\equiv$ 常值, $x\in\Omega$.

(b) 若函数$f:\overline{\Omega}\to\mathbb{R}$ 在$\overline{\Omega}$上连续在Ω内可微且$\nabla f(x)=0$ for all $x\in\Omega$, 则$f(x)\equiv$ 常值, $x\in\overline{\Omega}$.

一般地, 若映射$f:\overline{\Omega}\to\mathbb{R}^m$ 在$\overline{\Omega}$上连续在Ω内可微且$f'(x)=0$ for all $x\in\Omega$, 则$f(x)\equiv$ 常值, $x\in\overline{\Omega}$.

【证】直接证一般情形.

(a): f在Ω内是常值映射是指对于某个$\mathbf{c} \in \mathbb{R}^m$有$f(x) \equiv \mathbf{c}$ for all $x \in \Omega$. [也即对于某点$x_0 \in \Omega$ 有$f(x) \equiv f(x_0)$ for all $x \in \Omega$.] 因Ω 连通, 故由连续映射的性质(见本讲义第七章**命题7.37**) 知, 为证明f在Ω内是常值映射, 只需证明f在Ω内局部是常值映射.

任取$x \in \Omega$, 因Ω 是开集, 故存在$\delta = \delta_x > 0$ 使得$B(x, \delta) \subset \Omega$. 由于开球$B(x, \delta)$ 是开凸集, 故在$B(x, \delta)$上对f应用映射的微分中值不等式并注意$f'(\cdot) = 0$ 于Ω内, 得到

$$|f(y) - f(x)| \le \|f'(\xi)\| |y - x| = 0 \qquad \forall\, y \in B(x, \delta).$$

所以$f(y) \equiv f(x)$ for all $y \in B(x, \delta)$ 即f 在$B(x, \delta)$ 内是常值映射. 再由$x \in \Omega$的任意性知f局部是常值映射.

(b): 由假设知f已满足(a)中条件. 因此对于某个$\mathbf{c} \in \mathbb{R}^m$有$f(x) \equiv \mathbf{c}$ for all $x \in \Omega$. 对任意$x \in \overline{\Omega}$, 由闭包的序列刻画知存在$\{x_k\}_{k=1}^{\infty} \subset \Omega$使得$x_k \to x\,(k \to \infty)$. 于是由$f$在$\overline{\Omega}$上连续和$f(x_k) = \mathbf{c}$ 便有$f(x) = \lim\limits_{k\to\infty} f(x_k) = \mathbf{c}$. 所以$f(x) \equiv \mathbf{c}$, $x \in \overline{\Omega}$. $\Box$

【作业题】

1. 设$\Omega \subset \mathbb{R}^n$ 为开凸集, $\{a_1, a_2, ..., a_N\} \subset \Omega$, $f : \Omega \to \mathbb{R}^m$ 连续且在$\Omega \setminus \{a_1, a_2, ..., a_N\}$内处处可微. 证明对任意$x, y \in \Omega$ 存在$\xi \in (x, y) \setminus \{a_1, a_2, ..., a_N\}$ 使得

$$|f(x) - f(y)| \le \|f'(\xi)\| |x - y|.$$

这里(x, y) 是以x, y为端点的开直线段.

2. 设$\Omega \subset \mathbb{R}^n$ 为开集, 满足Ω中的任何两点x, y都能用一条位于Ω内的折线连接, 并且这折线上的直线段的个数≤ 3 且每个直线段的长度$\le 2|x - y|$. 设$f : \overline{\Omega} \to \overline{\Omega}$ 连续且f在Ω 内处处可微且

$$\|f'(x)\| \le \frac{1}{7} \qquad \forall\, x \in \Omega.$$

证明f 是$\overline{\Omega}$上的压缩映射从而有唯一不动点.

3. 设$\Omega \subset \mathbb{R}^n$ 为开集, $g : \Omega \to \mathbb{R}^n$在$\Omega$ 内处处可微. 证明: 对任意$x, y \in \Omega$满足$[x, y] \subset \Omega$者, 存在$\xi = y + \theta(x - y), \theta \in (0, 1)$, 使得

$$|g(x) - g(y) - g'(y)(x - y)| \le \|g'(\xi) - g'(y)\| |x - y|.$$

其中$\|\cdot\|$是矩阵范数(例如可取为矩阵的2-范数或矩阵的算子范数).

[提示: 固定$y \in \Omega$. 考虑x的映射

$$f(x) = g(x) - g(y) - g'(y)(x - y), \quad x \in \Omega.$$

计算: $f'(x) = g'(x) - g'(y)$, $f(y) = 0$.]

§8.4. 高阶偏导数和Taylor公式

若一个函数的某个一阶偏导数处处存在, 则这一阶偏导数就是一个函数, 称为偏导函数. 对这个新的函数同样可以看它的一阶偏导数是否存在, 即考虑原来函数的二阶偏导数的存在性, 等等.

【定义】设开集$\Omega \subset \mathbb{R}^n$, 函数$f : \Omega \to \mathbb{R}$ 在Ω 内处处存在一阶偏导数$D_j f(x) = \frac{\partial f}{\partial x_j}(x)$. 若偏导函数$x \mapsto \frac{\partial f}{\partial x_j}(x)$ 在点$x_0 \in \Omega$ 有一阶偏导数$\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j})(x_0)$, 则记

$$D_i D_j f(x_0) = D_i(D_j f)(x_0) = \frac{\partial}{\partial x_i}\Big(\frac{\partial f}{\partial x_j}\Big)(x_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$$

并称之为f 在x_0的一个二阶偏导数.

若二阶偏导数$\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$在$\Omega$ 内处处存在, 则称之为f在Ω内的一个二阶偏导函数.

一般地, 若f的一个r阶偏导数

$$D_{j_r} D_{j_{r-1}} \cdots D_{j_1} f(x) := \frac{\partial^r f}{\partial x_{j_r} \partial x_{j_{r-1}} \cdots \partial x_{j_1}}(x)$$

在Ω内处处存在, 则称之为f在Ω内的一个n 阶偏导函数. 若函数$x \mapsto \frac{\partial^r f}{\partial x_{j_r} \partial x_{j_{r-1}} \cdots \partial x_{j_1}}(x)$ 在点$x_0 \in \Omega$ 有一阶偏导数$\frac{\partial}{\partial x_{j_{r+1}}}\left(\frac{\partial^r f}{\partial x_{j_r} \partial x_{j_{r-1}} \cdots \partial x_{j_1}}\right)(x_0)$, 则记

$$\begin{aligned} D_{j_{r+1}} D_{j_r} D_{j_{r-1}} \cdots D_{j_1} f(x_0) &:= \frac{\partial}{\partial x_{j_{r+1}}}\left(\frac{\partial^r f}{\partial x_{j_r} \partial x_{j_{r-1}} \cdots \partial x_{j_1}}\right)(x_0) \\ &= \frac{\partial^{r+1} f}{\partial x_{j_{r+1}} \partial x_{j_r} \partial x_{j_{r-1}} \cdots \partial x_{j_1}}(x_0) \end{aligned}$$

并称之为f 在x_0的一个$r+1$阶偏导数. □

偏导数的其它记号:

$$\frac{\partial^r f}{\partial x_{j_r} \partial x_{j_{r-1}} \cdots \partial x_{j_1}}(x) = \Big(\frac{\partial}{\partial x_{j_r}}\Big)\Big(\frac{\partial}{\partial x_{j_{r-1}}}\Big) \cdots \Big(\frac{\partial}{\partial x_{j_1}}\Big) f(x) = \partial_{x_{j_r}} \partial_{x_{j_{r-1}}} \cdots \partial_{x_{j_1}} f(x).$$

二元函数的情形: 自变量为$(x, y) \in \mathbb{R}^2$.

$$\begin{aligned} &\Big(\frac{\partial}{\partial y}\Big)\Big(\frac{\partial}{\partial x}\Big) f(x,y) = \frac{\partial^2 f}{\partial y \partial x}(x,y), \\ &\Big(\frac{\partial}{\partial x}\Big)^2 f(x,y) = \frac{\partial^2 f}{\partial x \partial x}(x,y) = \frac{\partial^2 f}{\partial x^2}(x,y), \\ &\frac{\partial^3 f}{\partial x \partial y \partial x}(x,y), \quad \frac{\partial^3 f}{\partial y \partial y \partial x}(x,y) := \frac{\partial^3 f}{\partial y^2 \partial x}(x,y), \end{aligned}$$

$$\frac{\partial^n}{\partial x \partial x \cdots \partial x \partial y \partial y \cdots \partial y} = \Big(\frac{\partial}{\partial x}\Big)^k \Big(\frac{\partial}{\partial y}\Big)^{n-k} = \frac{\partial^n}{\partial x^k \partial y^{n-k}}.$$

问题: 偏导算子$\frac{\partial}{\partial x}$ 与$\frac{\partial}{\partial y}$ 是否可交换? 也即是否有

$$\Big(\frac{\partial}{\partial x}\Big)\Big(\frac{\partial}{\partial y}\Big)=\Big(\frac{\partial}{\partial y}\Big)\Big(\frac{\partial}{\partial x}\Big)\ ?\qquad \text{i.e.}\qquad \frac{\partial^2}{\partial x\partial y}=\frac{\partial^2}{\partial y\partial x}\ ?$$

这问题依赖于偏导算子作用的对象：作用在较好的函数上, 它们可交换; 作用在较差的函数上, 它们不可交换. 何谓较好的函数? 例如偏导数连续的函数就属于较好的函数.

【命题8.18(偏导数的换序).】设$n\ge 2,\Omega\subset\mathbb{R}^n$ 为开集, 函数$f:\Omega\to\mathbb{R}$ 在Ω内的两个偏导函数$\frac{\partial^2 f}{\partial x_j\partial x_i},\frac{\partial^2 f}{\partial x_j\partial x_i}$存在. 若这两个偏导函数都在点$x_0\in\Omega$ 连续, 则

$$\frac{\partial^2 f}{\partial x_j\partial x_i}(x_0)=\frac{\partial^2 f}{\partial x_i\partial x_j}(x_0).$$

因此若两个偏导函数在Ω内处处连续，则

$$\frac{\partial^2 f}{\partial x_j\partial x_i}(x)=\frac{\partial^2 f}{\partial x_i\partial x_j}(x)\qquad \forall\, x\in\Omega.$$

【证】可设$i\ne j$. 由偏导数的定义易知, 对于定点x_0, 偏导数$\frac{\partial^2 f}{\partial x_j\partial x_i}(x_0)$, $\frac{\partial^2 f}{\partial x_i\partial x_j}(x_0)$仅由自变量分量中的第$i$个和第$j$ 个分量决定, 与其它分量无关. 因此如令(不妨$i<j$)

$$\widetilde{f}(x_i,x_j)=f(x_{01},...,x_{0,i-1},x_i,x_{0,i+1},...,x_{0,j-1},x_j,x_{0,j+1},...,x_{0n})$$

则$\widetilde{f}$的两个偏导函数$\frac{\partial^2 \widetilde{f}}{\partial x_j\partial x_i}(x_i,x_j)$, $\frac{\partial^2 \widetilde{f}}{\partial x_i\partial x_j}(x_i,x_j)$ 在(x_{0i},x_{0j})的一个邻域内存在且

$$\frac{\partial^2 \widetilde{f}}{\partial x_j\partial x_i}(x_{0i},x_{0j})=\frac{\partial^2 f}{\partial x_j\partial x_i}(x_0),\qquad \frac{\partial^2 \widetilde{f}}{\partial x_i\partial x_j}(x_{0i},x_{0j})=\frac{\partial^2 f}{\partial x_i\partial x_j}(x_0).$$

同时由对f的假设知F的偏导函数$\frac{\partial^2 \widetilde{f}}{\partial x_j\partial x_i}(x_i,x_j)$, $\frac{\partial^2 \widetilde{f}}{\partial x_i\partial x_j}(x_i,x_j)$在点$(x_{0i},x_{0j})$ 连续.

于是问题归结为$n=2$的情形, 也即我们只需证明: 设$U\subset\mathbb{R}^2$ 为开集, 函数$f:U\to\mathbb{R}$ 在U内的两个偏导函数$\frac{\partial^2 f}{\partial x\partial y},\frac{\partial^2 f}{\partial y\partial x}$存在. 若这两个偏导函数都在点$(x,y)\in U$ 连续, 则

$$\frac{\partial^2 f}{\partial x\partial y}(x,y)=\frac{\partial^2 f}{\partial y\partial x}(x,y).$$

取$\delta>0$充分小使得$[x-\delta,x+\delta]\times[y-\delta,y+\delta]\subset U$. 考察函数

$$F(h,k)=f(x+h,y+k)-f(x+h,y)-f(x,y+k)+f(x,y),\quad |h|,|k|\le\delta.$$

令

$$\varphi(t)=f(x+th,y+k)-f(x+th,y),\quad t\in[0,1].$$

则应用一元函数中值定理有

$$
\begin{aligned}
&F(h,k)=\varphi(1)-\varphi(0)=\varphi'(\theta_1)=\frac{\partial f}{\partial x}(x+\theta_1 h,y+k)h-\frac{\partial f}{\partial x}(x+\theta_1 h,y)h\\
&=\left[\frac{\partial f}{\partial x}(x+\theta_1 h,y+k)-\frac{\partial f}{\partial x}(x+\theta_1 h,y)\right]h\\
&=\frac{\partial^2 f}{\partial y\partial x}(x+\theta_1 h,y+\theta_2 k)kh,\qquad 0<\theta_1,\theta_2<1.
\end{aligned}
$$

再令

$$\psi(t)=f(x+h,y+tk)-f(x,y+tk),\quad t\in[0,1].$$

则如上有

$$
\begin{aligned}
&F(h,k)=\psi(1)-\psi(0)=\psi'(\theta_3)=\frac{\partial f}{\partial y}(x+h,y+\theta_3 k)k-\frac{\partial f}{\partial y}(x,y+\theta_3 k)k\\
&=\left[\frac{\partial f}{\partial y}(x+h,y+\theta_3 k)k-\frac{\partial f}{\partial y}(x,y+\theta_3 k)\right]k\\
&=\frac{\partial^2 f}{\partial x\partial y}(x+\theta_4 h,y+\theta_3 k)hk,\qquad 0<\theta_3,\theta_4<1.
\end{aligned}
$$

因此

$$\frac{\partial^2 f}{\partial y\partial x}(x+\theta_1 h,y+\theta_2 k)kh=\frac{\partial^2 f}{\partial x\partial y}(x+\theta_4 h,y+\theta_3 k)hk.$$

在此等式中取$0<|h|,|k|<\delta$ 然后消去$hk\neq 0$ 得到

$$\frac{\partial^2 f}{\partial y\partial x}(x+\theta_1 h,y+\theta_2 k)=\frac{\partial^2 f}{\partial x\partial y}(x+\theta_4 h,y+\theta_3 k).$$

至此我们并没有用到偏导数的连续性. 现在让$h\to 0,k\to 0$. 则由两个二阶偏导数在(x,y) 连续即得

$$\frac{\partial^2 f}{\partial y\partial x}(x,y)=\frac{\partial^2 f}{\partial x\partial y}(x,y).\qquad\square$$

下面介绍常用的光滑函数类.

【定义(C^r 函数类)】设开集$\Omega\subset\mathbb{R}^n$, r 正整数. 我们用记号

$$C^r(\Omega)=C^r(\Omega,\mathbb{R})$$

表示定义在Ω上、所有阶数$\le r$ 的偏导数都在Ω内存在且都在Ω内连续的实值函数的全体. 如果$f\in C^r(\Omega)$ 则称f 属于C^r 函数类, 也称f **在Ω上是r阶光滑的**. 为统一起见我们定义$C^0(\Omega)=C(\Omega,\mathbb{R})$. 易见有

$$C^0(\Omega)\supset C^1(\Omega)\supset C^2(\Omega)\supset C^3(\Omega)\supset\cdots.$$

C^∞ 函数类$C^\infty(\Omega)$ 是Ω上任意阶光滑的函数的全体, 即

$$C^\infty(\Omega) = \bigcap_{r=0}^{\infty} C^r(\Omega). \qquad \Box$$

由于C^1 函数都是可微函数, 故若$f \in C^2(\Omega)$ 则f的各个一阶偏导函数$D_i f$ 都在Ω内可微. 此时称f在Ω内二次可微, 等等. 于是若$f \in C^\infty(\Omega)$, 则f的所有偏导函数都在Ω内可微. 由于可微蕴含连续, 故我们也称$C^\infty(\Omega)$为在Ω内无限次可微的函数类.

• **光滑函数类上的偏导数算子.** 回忆偏导数算子:

$$D_j = \frac{\partial}{\partial x_j} = \partial_{x_j}, \quad j = 1, 2, ..., n.$$

令

$$D = (D_1, D_2, ..., D_n).$$

对于非负整数$\alpha_1, \alpha_2, ..., \alpha_n$, 我们称$\alpha = (\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Z}^n_{\geq 0}$为一个**多重指标** 并称

$$|\alpha| = \alpha_1 + \alpha_2 + ... + \alpha_n$$

为多重指标α的阶. 定义

$$D^\alpha = (D_1)^{\alpha_1}(D_2)^{\alpha_2} \cdots (D_n)^{\alpha_n} = D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n}$$

称它是一个$|\alpha|$阶的偏导数算子. 按此记号, 当$f \in C^r(\Omega)$ 且$|\alpha| \leq r$ 时有

$$D^\alpha f(x) = D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f(x) = \Big(\frac{\partial}{\partial x_1}\Big)^{\alpha_1} \Big(\frac{\partial}{\partial x_2}\Big)^{\alpha_2} \cdots \Big(\frac{\partial}{\partial x_n}\Big)^{\alpha_n} f(x)$$

其中对某个$\alpha_i = 0$的情形, 规定

$$\Big(\frac{\partial}{\partial x_i}\Big)^0 f = D_i^0 f \equiv f.$$

作为**命题8.18(偏导数的换序)**的推论我们有

【命题8.19.】设$n \geq 2, \Omega \subset \mathbb{R}^n$ 为开集, 函数$f \in C^r(\Omega)\,(2 \leq r \leq \infty)$. 则对任意正整数$2 \leq k \leq r$, f的k 阶偏导数$\frac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(x)$ 的值不依赖于偏导数的次序, 也即任意排列指标$j_1, j_2, ..., j_k$都不会改变偏导数的值. 因此可将f的所有k阶偏导数写成如下标准形式:

$$\frac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(x) = \Big(\frac{\partial}{\partial x_1}\Big)^{\alpha_1} \Big(\frac{\partial}{\partial x_2}\Big)^{\alpha_2} \cdots \Big(\frac{\partial}{\partial x_n}\Big)^{\alpha_n} f(x)$$
$$= D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f(x) = D^\alpha f(x)$$

其中$\alpha=(\alpha_1,\alpha_2,...,\alpha_n)$为$k$阶多重指标.

【证】我们对偏导数的阶数k 用归纳法. 当$k=2$时这是上一命题的结论. 假定当偏导数阶数$=k\,(\ge 2)$时所证成立, 看$k+1(\le r)$时. 任取$j_1,j_2,...,j_k,j_{k+1}\in\{1,2,...,m\}$. 记$i=j_{k+1}$. 由归纳假设有

$$D_{j_{k+1}}D_{j_k}D_{j_{k-1}}\cdots D_{j_1}f=D_i(D_{j_k}D_{j_{k-1}}\cdots D_{j_1}f)=D_i\Big(D_1^{\alpha_1}D_2^{\alpha_2}\cdots D_n^{\alpha_n}f\Big)$$

其中$\alpha_1+\alpha_2+\cdots+\alpha_n=k$. 若$i=1$, 则上式右端已是标准形式. 设$2\le i\le n$. 则上式右端写为

$$D_i\Big(D_1^{\alpha_1}D_2^{\alpha_2}\cdots D_n^{\alpha_n}f\Big)=D_iD_1^{\alpha_1}D_2^{\alpha_2}\cdots D_{i-1}^{\alpha_{i-1}}D_i^{\alpha_i}\cdots D_n^{\alpha_n}f.$$

令

$$g=D_2^{\alpha_2}\cdots D_n^{\alpha_n}f.$$

则$g\in C^{1+\alpha_1}(\Omega)$. 若$\alpha_1=0$. 则有$D_iD_1^0=D_i=D_1^0D_i$. 当$\alpha_1\ge 1$时, 反复利用$k=2$时的结果有

$$D_iD_1^{\alpha_1}g=D_iD_1D_1^{\alpha_1-1}g=D_1D_iD_1^{\alpha_1-1}g=\cdots=D_1^{\alpha_1}D_ig$$

即

$$D_iD_1^{\alpha_1}D_2^{\alpha_2}\cdots D_{i-1}^{\alpha_{i-1}}D_i^{\alpha_i}\cdots D_n^{\alpha_n}f=D_1^{\alpha_1}D_iD_2^{\alpha_2}\cdots D_{i-1}^{\alpha_{i-1}}D_i^{\alpha_i}\cdots D_n^{\alpha_n}f.$$

对这种相邻换序的次数运用归纳法, 经有限步后得到

$$D_iD_1^{\alpha_1}D_2^{\alpha_2}\cdots D_{i-1}^{\alpha_{i-1}}D_i^{\alpha_i}\cdots D_n^{\alpha_n}f=D_1^{\alpha_1}D_2^{\alpha_2}\cdots D_{i-1}^{\alpha_{i-1}}D_i^{1+\alpha_i}\cdots D_n^{\alpha_n}f=D^{\widetilde{\alpha}}f$$

其中$\widetilde{\alpha}=(\alpha_1,...,1+\alpha_i,...,\alpha_n)$ 是一个$k+1$ 阶多重指标. 据归纳法原理, 命题得证. □

【例】看二元函数. 设$\Omega\subset\mathbb{R}^2$ 为开集, $f\in C^6(\Omega)$. 则

$$\frac{\partial^6 f}{\partial x\partial y\partial x^3\partial y}(x,y)=\frac{\partial^6 f}{\partial x^4\partial y^2}(x,y)=\Big(\frac{\partial}{\partial x}\Big)^4\Big(\frac{\partial}{\partial y}\Big)^2 f(x,y).$$

□

看一个稍复杂但常用的性质：复合函数的光滑性和高阶偏导数.

【例】设$I\subset\mathbb{R}$ 为开区间, $\Omega\subset\mathbb{R}^n$ 为开集, $p,q\in\mathbb{N}$, $g\in C^p(I)$, $f\in C^q(\Omega)$ 且$f(\Omega)\subset I$. 则复合函数$g\circ f\in C^r(I)$ 其中$r=\min\{p,q\}$. 此外对任意$1\le m\le r$ 和任意$\alpha\in\mathbb{Z}_{\ge 0}^n$ 满

足$|\alpha| = m$ 有

$$\begin{aligned}
&D^\alpha (g\circ f)(x)\\
&=\sum_{k=1}^m g^{(k)}(f(x)) \sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m} C_{\alpha,\beta_1,\beta_2,...,\beta_k} D^{\beta_1} f(x) D^{\beta_2} f(x)\cdots D^{\beta_k} f(x)
\end{aligned}$$

其中$D=(D_1,D_2,...,D_n), D_j=\frac{\partial}{\partial x_j}, 0\neq \beta_i\in\mathbb{Z}^n_{\geq 0}$, $C_{\alpha,\beta_1,\beta_2,...,\beta_k}$ 是常数.

【证】不妨设$r\geq 3$. 由复合函数微分法有

$$\begin{aligned}
&D_j(g\circ f)(x)=g'(f(x))D_jf(x),\\
&D_iD_j(g\circ f)(x)=g''(f(x))D_if(x)D_jf(x)+g'(f(x))D_iD_jf(x),\\
&D_kD_iD_j(g\circ f)(x)\\
&=g'''(f(x))D_kf(x)D_if(x)D_jf(x)\\
&+g''(f(x))\Big((D_kD_if(x))D_jf(x)+D_if(x)(D_kD_jf(x))+D_kf(x)D_iD_jf(x)\Big)\\
&+g'(f(x))D_kD_iD_jf(x),\quad i,j,k=1,2,...,n.
\end{aligned}$$

由假设条件易见$g\circ f\in C^3(I)$ 同时由

$$(D_kD_if(x))D_jf(x)=D^{\mathbf{e}_k+\mathbf{e}_i}f(x)D^{\mathbf{e}_j}f(x)\quad 等等$$

可知所证等式对于$m=1,2,3$ 成立.

假设对于$1\leq m\leq r-1$ 有$g\circ f\in C^m(I)$ 且所证等式对于任意$1\leq|\alpha|\leq m$ 成立. 则对任意$\alpha\in\mathbb{Z}^n_{\geq 0}$ 满足$|\alpha|=m$, 由$D^\alpha(g\circ f)(x)$的上述表示和$m\leq r-1$可知$D^\alpha(g\circ f)(x)$仍有一阶偏导数且任意$j\in\{1,2,...,n\}$ 有

$$\begin{aligned}
&D_jD^\alpha(g\circ f)(x)\\
&=\sum_{k=1}^m D_j\Big\{g^{(k)}(f(x))\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m} C_{\alpha,\beta_1,\beta_2,...,\beta_k} D^{\beta_1} f(x) D^{\beta_2} f(x)\cdots D^{\beta_k} f(x)\Big\}\\
&=\sum_{k=1}^m \Big\{g^{(k+1)}(f(x))D_jf(x)\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m} C_{\alpha,\beta_1,\beta_2,...,\beta_k} D^{\beta_1} f(x) D^{\beta_2} f(x)\cdots D^{\beta_k} f(x)\\
&+g^{(k)}(f(x))\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m} C_{\alpha,\beta_1,\beta_2,...,\beta_k} D_j\Big(D^{\beta_1} f(x) D^{\beta_2} f(x)\cdots D^{\beta_k} f(x)\Big)\Big\}.
\end{aligned}$$

因为f的偏导数可换序, 故对于新的一组常数$A_{j,\alpha,\beta_1,\beta_2,...,\beta_{k+1}}$ (其中有些常数为0) 可写

$$\begin{aligned}
&D_jf(x)\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m} C_{\alpha,\beta_1,\beta_2,...,\beta_k} D^{\beta_1} f(x) D^{\beta_2} f(x)\cdots D^{\beta_k} f(x)\\
&=\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_{k+1}|=m+1} A_{j,\alpha,\beta_1,\beta_2,...,\beta_{k+1}} D^{\beta_1} f(x) D^{\beta_2} f(x)\cdots D^{\beta_{k+1}} f(x).
\end{aligned}$$

其次我们计算

$$
\begin{aligned}
&D_j\Big(D^{\beta_1}f(x)D^{\beta_2}f(x)\cdots D^{\beta_k}f(x)\Big)\\
&=(D_jD^{\beta_1}f(x))D^{\beta_2}f(x)\cdots D^{\beta_k}f(x)+D^{\beta_1}f(x)(D_jD^{\beta_2}f(x))\cdots D^{\beta_k}f(x)\\
&+\cdots+D^{\beta_1}f(x)\cdots D^{\beta_{k-1}}f(x)(D_jD^{\beta_k}f(x))\\
&=D^{\beta_1+\mathbf{e}_j}f(x)D^{\beta_2}f(x)\cdots D^{\beta_k}f(x)+D^{\beta_1}f(x)D^{\beta_2+\mathbf{e}_j}f(x)\cdots D^{\beta_k}f(x)\\
&+\cdots+D^{\beta_1}f(x))\cdots D^{\beta_{k-1}}f(x)D^{\beta_k+\mathbf{e}_j}f(x).
\end{aligned}
$$

因此(再由f的偏导数可换序) 对于新的一组常数$B_{j,\alpha,\beta_1,\beta_2,...,\beta_k}$ (其中有些常数为0) 有

$$
\begin{aligned}
&\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m}C_{\alpha,\beta_1,\beta_2,...,\beta_k}D_j\Big(D^{\beta_1}f(x)D^{\beta_2}f(x)\cdots D^{\beta_k}f(x)\Big)\\
&=\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m+1}B_{j,\alpha,\beta_1,\beta_2,...,\beta_k}D^{\beta_1}f(x)D^{\beta_2}f(x)\cdots D^{\beta_k}f(x)
\end{aligned}
$$

于是得到

$$
\begin{aligned}
&D_jD^{\alpha}(g\circ f)(x)\\
&=g^{(m+1)}(f(x))\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_{m+1}|=m+1}A_{j,\alpha,\beta_1,\beta_2,...,\beta_{m+1}}D^{\beta_1}f(x)D^{\beta_2}f(x)\cdots D^{\beta_{m+1}}f(x)\\
&+\sum_{k=1}^{m-1}g^{(k+1)}(f(x))\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_{k+1}|=m+1}A_{j,\alpha,\beta_1,\beta_2,...,\beta_{k+1}}D^{\beta_1}f(x)D^{\beta_2}f(x)\cdots D^{\beta_{k+1}}f(x)\\
&+\sum_{k=2}^{m}g^{(k)}(f(x))\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m+1}B_{j,\alpha,\beta_1,\beta_2,...,\beta_k}D^{\beta_1}f(x)D^{\beta_2}f(x)\cdots D^{\beta_k}f(x)\\
&+g^{(1)}(f(x))\sum_{|\beta_1|=m+1}B_{j,\alpha,\beta_1}D^{\beta_1}f(x)\\
&=\sum_{k=1}^{m+1}g^{(k)}(f(x))\sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m+1}C_{j,\alpha,\beta_1,\beta_2,...,\beta_k}D^{\beta_1}f(x)D^{\beta_2}f(x)\cdots D^{\beta_k}f(x)
\end{aligned}
$$

其中

$$
\begin{aligned}
&C_{j,\alpha,\beta_1,\beta_2,...,\beta_{m+1}}=A_{j,\alpha,\beta_1,\beta_2,...,\beta_{m+1}},\\
&C_{j,\alpha,\beta_1,\beta_2,...,\beta_k}=A_{j,\alpha,\beta_1,\beta_2,...,\beta_k}+B_{j,\alpha,\beta_1,\beta_2,...,\beta_k},\quad 2\le k\le m\\
&C_{j,\alpha,\beta_1}=B_{j,\alpha,\beta_1}.
\end{aligned}
$$

由g,f各自的光滑性知$D_jD^{\alpha}(g\circ f)$ 连续($\forall\,|\alpha|=m,j\in\{1,2,...,n\}$), 所以$g\circ f\in C^{m+1}(I)$. 现在任取多重指标$\alpha=(\alpha_1,\alpha_2,...,\alpha_n)$ 满足$|\alpha|=m+1$. 令$j\in\{1,2,...,n\}$为

最小下标使得$a_j \ge 1$. 令$\widetilde{\alpha} = \alpha - \mathbf{e}_j$. 则$\widetilde{\alpha} \in \mathbb{Z}^n_{\ge 0}$ 且$|\widetilde{\alpha}| = m, \alpha = \widetilde{\alpha} + \mathbf{e}_j$. 因$C^{m+1}(I)$中的函数偏导数可换序, 故有

$$D^{\alpha}(g \circ f)(x) = D^{\widetilde{\alpha}+\mathbf{e}_j}(g \circ f)(x) = D^{\mathbf{e}_j} D^{\widetilde{\alpha}}(g \circ f)(x) = D_j D^{\widetilde{\alpha}}(g \circ f)(x).$$

于是由刚才得到的结果有

$$\begin{aligned} &D^{\alpha}(g \circ f)(x) \\ &= \sum_{k=1}^{m+1} g^{(k)}(f(x)) \sum_{|\beta_1|+|\beta_2|+\cdots+|\beta_k|=m+1} C_{\alpha,\beta_1,\beta_2,\dots,\beta_k} D^{\beta_1} f(x) D^{\beta_2} f(x) \cdots D^{\beta_k} f(x) \end{aligned}$$

其中$C_{\alpha,\beta_1,\beta_2,\dots,\beta_k} = C_{j,\widetilde{\alpha},\beta_1,\beta_2,\dots,\beta_k}$. 这证明了所证等式对于$|\alpha| = m+1$ 也成立. 据归纳法原理我们证明了$g \circ f \in C^r(I)$ (包括$r = \infty$的情形) 且所证等式成立. $\Box$

【例】设

$$g(t) = e^{-1/t} 1_{\{t>0\}} = \begin{cases} e^{-1/t} & \text{if} \quad t > 0 \\ 0 & \text{if} \quad t \le 0 \end{cases}$$

$$\psi(x) = g\Big(1 - \frac{|x-a|^2}{\delta^2}\Big), \quad x \in \mathbb{R}^n$$

其中$a \in \mathbb{R}^n, \delta > 0$为常数. 则$\psi \in C^{\infty}(\mathbb{R}^n)$, $\psi \ge 0$, 且支集$\mathrm{supp}\psi = \overline{B}(a,\delta)$.

这是因为周知$g \in C^{\infty}(\mathbb{R})$, 而二次多项式$f(x) = 1 - \frac{|x-a|^2}{\delta^2}$ 显然属于$C^{\infty}(\mathbb{R}^n)$. 所以由上一例题知$\psi = g \circ f \in C^{\infty}(\mathbb{R}^n)$. 此外由$g(t)$的定义知: $\psi(x) \ne 0$ 当且仅当$1 - \frac{|x-a|^2}{\delta^2} > 0$, 即当且仅当$x \in B(a,\delta)$. 所以$\{x \in \mathbb{R}^n \mid \psi(x) \ne 0\} = B(a,\delta)$. 因此$\mathrm{supp}\psi = \overline{B}(a,\delta)$. $\psi(x) \ge 0$ 于$\mathbb{R}^n$是显然的. $\Box$

下面我们将导出一些与偏导数有关的展开式, 为建立多元函数的Taylor 公式作准备. 先看二元函数. 设开集$\Omega \subset \mathbb{R}^2$, $f \in C^r(\Omega)$. 则f 的每个r 阶偏导数的标准形式为

$$\frac{\partial^r f}{\partial x^i \partial y^{r-i}}(x,y) = \Big(\frac{\partial}{\partial x}\Big)^i \Big(\frac{\partial}{\partial y}\Big)^{r-i} f(x,y), \quad i \in \{1,2,...,r\}.$$

由于在$C^r(\Omega)$ $(r \ge 2)$ 上$\frac{\partial}{\partial x}$ 与$\frac{\partial}{\partial y}$可交换, 我们可以定义线性组合$h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}$的乘方:

$$\Big(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\Big)^r = \sum_{i=0}^{r} \left(\begin{array}{c} r \\ i \end{array}\right) \Big(h\frac{\partial}{\partial x}\Big)^i \Big(k\frac{\partial}{\partial y}\Big)^{r-i} = \sum_{i=0}^{r} \left(\begin{array}{c} r \\ i \end{array}\right) h^i k^{r-i} \Big(\frac{\partial}{\partial x}\Big)^i \Big(\frac{\partial}{\partial y}\Big)^{r-i}.$$

也即

$$\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^r f(x,y)=\sum_{i=0}^{r}\begin{pmatrix} r \\ i \end{pmatrix} h^i k^{r-i}\Big(\frac{\partial}{\partial x}\Big)^i\Big(\frac{\partial}{\partial y}\Big)^{r-i} f(x,y)$$

或写成哑元形式(略去自变量, 只写函数关系)

$$\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^r f=\sum_{i=0}^{r}\begin{pmatrix} r \\ i \end{pmatrix} h^i k^{r-i}\Big(\frac{\partial}{\partial x}\Big)^i\Big(\frac{\partial}{\partial y}\Big)^{r-i} f.$$

这里

$$\begin{pmatrix} r \\ i \end{pmatrix}=\frac{r!}{i!(r-i)!},\quad i\in\{1,2,...,r\}.$$

易见

$$\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^r (c_1f+c_2g)=c_1\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^r f+c_2\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^r g.$$

□

再看二元函数:

【例】设$\Omega\subset\mathbb{R}^2$ 为开集, $m\in\mathbb{N}$, $f\in C^m(\Omega)$, 闭直线段$[(x_0,y_0),(x_0+h,y_0+k)]\subset\Omega$. 则函数$t\mapsto f(x_0+th,y_0+tk)$ 属于$C^m([0,1])$ 且

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\Big(f(x_0+th,y_0+tk)\Big)=\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m f(x_0+th,y_0+tk),\quad t\in[0,1].$$

其中右边表示**先对独立变量**(x,y)**求导, 得到导数的表达式之后再代入**$(x,y)=(x_0+th,y_0+tk)$**等具体值:**

$$\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m f(x,y)\Big|_{(x,y)=(x_0+th,y_0+tk)}.$$

【证】当$m=1$时, 设$f\in C^1(\Omega)$. 由复合函数求导法有

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(f(x_0+th,y_0+tk)\Big)=\Big(\frac{\partial}{\partial x}f\Big)(x_0+th,y+tk)h+\Big(\frac{\partial}{\partial y}f\Big)(x_0+th,y+tk)k$$
$$=\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)f(x_0+th,y_0+tk).$$

因$f\in C^1(\Omega)$故上式右边是t的连续函数, 因此$t\mapsto f(x_0+th,y_0+tk)$ 属于$C^1([0,1])$.

设$m\in\mathbb{N}$ 并设所证性质对$C^m(\Omega)$中的函数成立. 则任取$f\in C^{m+1}(\Omega)$. 此时分别对$C^m(\Omega)$中的函数

$$\frac{\partial}{\partial x}f,\quad \frac{\partial}{\partial y}f$$

应用归纳假设有

$$\begin{aligned}
&\frac{\mathrm{d}^{m+1}}{\mathrm{d}t^{m+1}}\Big(f(x_0+th,y_0+tk)\Big)=\frac{\mathrm{d}^m}{\mathrm{d}t^m}\left(\frac{\mathrm{d}}{\mathrm{d}t}\Big(f(x_0+th,y_0+tk)\Big)\right)\\
&=\frac{\mathrm{d}^m}{\mathrm{d}t^m}\left(h\Big(\frac{\partial}{\partial x}f\Big)(x_0+th,y+tk)\Big)+k\Big(\frac{\partial}{\partial y}f\Big)(x_0+th,y_0+tk)\Big)\right)\\
&=h\frac{\mathrm{d}^m}{\mathrm{d}t^m}\left(\Big(\frac{\partial}{\partial x}f\Big)(x_0+th,y+tk)\Big)\right)+k\frac{\mathrm{d}^m}{\mathrm{d}t^m}\left(\Big(\frac{\partial}{\partial y}f\Big)(x_0+th,y_0+tk)\Big)\right)\\
&=h\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m\Big(\frac{\partial}{\partial x}f\Big)(x_0+th,y+tk)+k\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m\Big(\frac{\partial}{\partial x}f\Big)(x_0+th,y+tk)\\
&=\left(h\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m\frac{\partial}{\partial x}f+k\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m\frac{\partial}{\partial y}f\right)(x,y)\Big|_{(x,y)=(x_0+th,y+tk)}\\
&=\left(\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m\left(h\frac{\partial}{\partial x}f+k\frac{\partial}{\partial y}f\right)\right)(x,y)\Big|_{(x,y)=(x_0+th,y+tk)}\\
&=\left(\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^{m+1}f\right)(x,y)\Big|_{(x,y)=(x_0+th,y+tk)}\\
&=\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^{m+1}f(x_0+th,y_0+tk).
\end{aligned}$$

同时由$f\in C^{m+1}(\Omega)$知上式右边是t的连续函数, 因此$t\mapsto f(x_0+th,y_0+tk)$ 属于$C^{m+1}([0,1])$. 据归纳法原理知所证性质成立.

【推广到n 元光滑函数】为此我们引入数学上常用记号:

对于为多重指标$\alpha=(\alpha_1,\alpha_2,...,\alpha_n)\in\mathbb{Z}^n_{\geq 0}$, 向量$h=(h_1,h_2,...,h_n)\in\mathbb{R}^n$ 和$D=(D_1,D_2,...,D_n), D_j=\frac{\partial}{\partial x_j}$, 定义

$$\begin{aligned}
&\alpha!=\alpha_1!\alpha_2!\cdots\alpha_n!,\qquad |\alpha|=\alpha_1+\alpha_2+\cdots+\alpha_n,\\
&h^\alpha=h_1^{\alpha_1}h_2^{\alpha_2}\cdots h_n^{\alpha_n},\\
&D^\alpha f(x)=D_1^{\alpha_1}D_2^{\alpha_2}\cdots D_n^{\alpha_n}f(x),\\
&D_j^0 f(x)\equiv f(x).
\end{aligned}$$

【命题8.20.】设$\Omega\subset\mathbb{R}^n$ 为开集, $m\in\mathbb{N}$, $f\in C^m(\Omega)$. 则

(a) 对任意$h=(h_1,h_2,...,h_n)\in\mathbb{R}^n$ 有

$$\Big(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\Big)^m f(x)=\Big(\sum_{j=1}^n h_jD_j\Big)^m f(x)=\sum_{|\alpha|=m}\frac{m!}{\alpha!}h^\alpha D^\alpha f(x),\quad x\in\Omega.$$

(b) 设闭直线段$[x_0,x_0+h]\in\Omega$. 则函数$t\mapsto f(x_0+th)$ 属于$C^m([0,1])$ 且对任

意$t\in[0,1]$有

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\Big(f(x_0+th)\Big)=\left(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\right)^m f(x_0+th)=\sum_{|\alpha|=m}\frac{m!}{\alpha!}h^\alpha D^\alpha f(x_0+th).$$

【证】(a): 任取定$h=(h_1,h_2,...,h_n)\in\mathbb{R}^n$. 我们对光滑阶数$m$ 用归纳法. 当$m=1$ 时, 易见$\alpha\in\mathbb{Z}_{\ge0}^n$ 满足$|\alpha|=1$ 当且仅当存在$j\in\{1,2,...,n\}$ 使得$\alpha=\mathbf{e}_j$. 同时有$h_j=h^{\mathbf{e}_j}$, 并由$D_i^0g(x)\equiv g(x)$ 知有$D_jf(x)=D^{\mathbf{e}_j}f(x)$. 因此

$$\begin{aligned}&\Big(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\Big)f(x)=\Big(\sum_{j=1}^n h_jD_j\Big)f(x)=\sum_{j=1}^n h_jD_jf(x)=\sum_{j=1}^n h^{\mathbf{e}_j}D^{\mathbf{e}_j}f(x)\\&=\sum_{|\alpha|=1}h^\alpha D^\alpha f(x)=\sum_{|\alpha|=1}\frac{1}{\alpha!}h^\alpha D^\alpha f(x),\quad x\in\Omega.\end{aligned}$$

假设当光滑阶数为$m\in\mathbb{N}$时所证等式成立, 看$m+1$时. 设$f\in C^{m+1}(\Omega)$. 则$D_if\in C^m(\Omega), i=1,2,...,n$. 因此由归纳假设和偏导数可换序有

$$\begin{aligned}&\Big(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\Big)^{m+1}f(x)=\Big(\sum_{j=1}^n h_jD_j\Big)^{m+1}f(x)\\&=\Big(\sum_{j=1}^n h_jD_j\Big)^m\Big(\Big(\sum_{i=1}^n h_iD_i\Big)f\Big)(x)=\Big(\sum_{j=1}^n h_jD_j\Big)^m\Big(\sum_{i=1}^n h_iD_if\Big)(x)\\&=\sum_{i=1}^n h_i\Big(\sum_{j=1}^n h_jD_j\Big)^m D_if(x)=\sum_{i=1}^n h_i\sum_{|\beta|=m}\frac{m!}{\beta!}h^\beta D^\beta D_if(x)\\&=\sum_{i=1}^n\sum_{|\beta|=m}\frac{m!}{\beta!}h^{\beta+\mathbf{e}_i}D^{\beta+\mathbf{e}_i}f(x)\quad(\text{内层求和中作平移 }\beta+\mathbf{e}_i=\alpha)\\&=\sum_{i=1}^n\sum_{|\alpha|=m+1}\alpha_i1_{\alpha_i\ge1}\frac{m!}{\alpha!}h^\alpha D^\alpha f(x)=\sum_{|\alpha|=m+1}\Big(\sum_{i=1}^n\alpha_i1_{\alpha_i\ge1}\Big)\frac{m!}{\alpha!}h^\alpha D^\alpha f(x)\\&=\sum_{|\alpha|=m+1}|\alpha|\frac{m!}{\alpha!}h^\alpha D^\alpha f(x)=\sum_{|\alpha|=m+1}\frac{(m+1)!}{\alpha!}h^\alpha D^\alpha f(x).\end{aligned}$$

据归纳法原理知所证等式成立.

(b): 由(a) 知为证(b) 只需证明(b)中的第一个等号, 即只需证明

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\Big(f(x_0+th)\Big)=\Big(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\Big)^m f(x_0+th),\quad t\in[0,1].$$

还是对光滑阶数m 用归纳法. 当$m=1$ 时, 设$f\in C^1(\Omega)$, 则由复合函数的微分法或**命题8.13*** 有

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(f(x_0+th)\Big)=f'(x_0+th)h=\langle\nabla f(x_0+th),h\rangle$$

$$= \sum_{j=1}^{n} \Big(\frac{\partial}{\partial x_j} f\Big)(x_0 + th) h_j = \Big(\sum_{j=1}^{n} h_j \frac{\partial}{\partial x_j}\Big) f(x_0 + th).$$

同时由$f \in C^1(\Omega)$知$t \mapsto f(x_0 + th)$ 属于$C^1([0,1])$

假设当光滑阶数为m时所证等式成立, 看$m+1$时. 设$f \in C^{m+1}(\Omega)$. 分别对$C^m(\Omega)$中的函数

$$\frac{\partial}{\partial x_1} f, \quad \frac{\partial}{\partial x_2} f, \quad \ldots, \quad \frac{\partial}{\partial x_n} f$$

应用归纳假设有

$$\begin{aligned}
&\frac{\mathrm{d}^{m+1}}{\mathrm{d}t^{m+1}}\Big(f(x_0+th)\Big) = \frac{\mathrm{d}^m}{\mathrm{d}t^m}\left(\frac{\mathrm{d}}{\mathrm{d}t} f(x_0+th)\right) = \frac{\mathrm{d}^m}{\mathrm{d}t^m}\Big(\sum_{i=1}^{n}\Big(\frac{\partial}{\partial x_i} f\Big)(x_0+th) h_i\Big) \\
&= \sum_{i=1}^{n} h_i \frac{\mathrm{d}^m}{\mathrm{d}t^m}\Big(\Big(\frac{\partial}{\partial x_i} f\Big)(x_0+th)\Big) = \sum_{i=1}^{n} h_i \Big(\sum_{j=1}^{n} h_j \frac{\partial}{\partial x_j}\Big)^m \Big(\frac{\partial}{\partial x_i} f\Big)(x_0+th) \\
&= \Big(\sum_{i=1}^{n} h_i \Big(\sum_{j=1}^{n} h_j \frac{\partial}{\partial x_j}\Big)^m \frac{\partial}{\partial x_i} f\Big)(x)\Big|_{x=x_0+th} \\
&= \Big(\Big(\sum_{j=1}^{n} h_j \frac{\partial}{\partial x_j}\Big)^m \sum_{i=1}^{n} h_i \frac{\partial}{\partial x_i} f\Big)(x)\Big|_{x=x_0+th} \\
&= \Big(\Big(\sum_{j=1}^{n} h_j \frac{\partial}{\partial x_j}\Big)^{m+1} f\Big)(x)\Big|_{x=x_0+th} = \Big(\sum_{j=1}^{n} h_j \frac{\partial}{\partial x_j}\Big)^{m+1} f(x_0+th).
\end{aligned}$$

同时由$f \in C^{m+1}(\Omega)$知$t \mapsto f(x_0 + th)$ 属于$C^{m+1}([0,1])$. 据归纳法原理知所证性质成立. □

为熟悉多重指标和偏导算子的运算, 我们来看

乘积型函数的偏导数计算公式. 设$f_1(t), f_2(t), ..., f_n(t)$ 为一元光滑函数. 则对于多重指标$\alpha = (\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Z}^n_{\ge 0}$ 和偏导算子$D^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_2}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}$ 有

$$D^\alpha\Big(f_1(x_1) f_2(x_2) \cdots f_n(x_n)\Big) = f_1^{(\alpha_1)}(x_1) f_2^{(\alpha_2)}(x_2) \ \cdots f_n^{(\alpha_n)}(x_n).$$

其中$f_j^{(k)}(t)$表示一元函数$f_j(t)$ 的k 阶导数.

【证】当$n=1$时自然成立. 假设在$n-1$ 时成立. 则在n时有

$$\begin{aligned}
&D^\alpha\Big(f_1(x_1) f_2(x_2) \cdots f_n(x_n)\Big) \\
&= D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_{n-1}^{\alpha_{n-1}} D_n^{\alpha_n}\Big(f_1(x_1) f_2(x_2) \cdots f_{n-1}(x_{n-1}) f_n(x_n)\Big) \\
&= D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_{n-1}^{\alpha_{n-1}}\Big(D_n^{\alpha_n}\Big(f_1(x_1) f_2(x_2) \cdots f_{n-1}(x_{n-1}) f_n(x_n)\Big)\Big)
\end{aligned}$$

$$
\begin{aligned}
&= D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_{n-1}^{\alpha_{n-1}} \Big(f_1(x_1) f_2(x_2) \cdots f_{n-1}(x_{n-1}) D_n^{\alpha_n} f_n(x_n) \Big) \\
&= \Big[D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_{n-1}^{\alpha_{n-1}} \Big(f_1(x_1) f_2(x_2) \cdots f_{n-1}(x_{n-1}) \Big) \Big] D_n^{\alpha_n} f_n(x_n) \\
&= \Big[D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_{n-1}^{\alpha_{n-1}} \Big(f_1(x_1) f_2(x_2) \cdots f_{n-1}(x_{n-1}) \Big) \Big] f_n^{(\alpha_n)}(x_n) \\
&= f_1^{(\alpha_1)}(x_1) f_2^{(\alpha_2)}(x_2) \ \cdots f_{n-1}^{(\alpha_{n-1})}(x_{n-1}) f_n^{(\alpha_n)}(x_n). \qquad \square
\end{aligned}
$$

• **单项式的偏导数计算公式**.

设$\alpha = (\alpha_1, \alpha_2, ..., \alpha_n), \beta = (\beta_1, \beta_2, ..., \beta_n) \in \mathbb{Z}_{\geq 0}^n$ (多重指标). 对单项式$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$的偏导数$D^\beta x^\alpha$ 的计算公式与一元函数形式上相同, 即

$$
D^\beta x^\alpha = \begin{cases} \dfrac{\alpha!}{(\alpha-\beta)!} x^{\alpha-\beta} & \text{if} \quad \alpha - \beta \in \mathbb{Z}_{\geq 0}^n, \\ \quad 0 & \text{if} \quad \alpha - \beta \notin \mathbb{Z}_{\geq 0}^n. \end{cases}
$$

【证】回忆: 对一元单项式t^k的导数, 有

$$
\begin{aligned}
&(t^k)^{(j)} = \frac{k!}{(k-j)!} t^{k-j} && \text{if} \quad k \geq j, \\
&(t^k)^{(j)} \equiv 0 && \text{if} \quad k < j.
\end{aligned}
$$

对一元函数$t^{\alpha_1}, t^{\alpha_2}, ..., t^{\alpha_m}$ 应用乘积型函数的偏导数计算公式有

$$
D^\beta x^\alpha = D_1^{\beta_1} D_2^{\beta_2} \cdots D_n^{\beta_n} \Big(x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \Big) = \Big(x_1^{\alpha_1} \Big)^{(\beta_1)} \Big(x_2^{\alpha_2} \Big)^{(\beta_2)} \cdots \Big(x_n^{\alpha_n} \Big)^{(\beta_n)}.
$$

设$\alpha - \beta \notin \mathbb{Z}_{\geq 0}^n$ 也即存在$j \in \{1, 2, ..., n\}$ 使得$\alpha_j < \beta_j$. 此时有

$$
(x_j^{\alpha_j})^{(\beta_j)} = 0 \qquad \forall\, x_j \in \mathbb{R}.
$$

因此这时有$D^\beta x^\alpha = 0$.

设$\alpha - \beta \in \mathbb{Z}_{\geq 0}^n$, 即$\alpha_j \geq \beta_j$, $j = 1, 2, ..., n$. 此时有

$$
\begin{aligned}
D^\beta x^\alpha &= \frac{\alpha_1!}{(\alpha_1 - \beta_1)!} x_1^{\alpha_1 - \beta_1} \frac{\alpha_2!}{(\alpha_2 - \beta_2)!} x_2^{\alpha_2 - \beta_2} \cdots \frac{\alpha_n!}{(\alpha_n - \beta_n)!} x_n^{\alpha_n - \beta_n} \\
&= \frac{\alpha!}{(\alpha-\beta)!} x^{\alpha-\beta}. \qquad \square
\end{aligned}
$$

有了以上准备, 我们可以进入

♠ **多变量函数的Taylor公式**

与单变量函数相同, 我们可以将一个光滑多元函数表示成一个多元多项式与一个“小项”之和, 以便进行数学分析.

【命题8.21(具有Lagrange型余项的Taylor公式).】

设$\Omega\subset\mathbb{R}^n$为开集, r为非负整数, 函数$f\in C^{r+1}(\Omega)$, $x_0\in\Omega$. 则对任意$x\in\Omega$, 只要闭直线段$[x_0,x]\subset\Omega$ 就有

$$f(x)=\sum_{k=0}^{r}\sum_{|\alpha|=k}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha+\sum_{|\alpha|=r+1}\frac{D^\alpha f(\xi)}{\alpha!}(x-x_0)^\alpha$$

也写成

$$f(x)=\sum_{|\alpha|\le r}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha+\sum_{|\alpha|=r+1}\frac{D^\alpha f(\xi)}{\alpha!}(x-x_0)^\alpha$$

其中$\xi=x_0+\theta(x-x_0)$, $0<\theta=\theta_{x_0,x}<1$.

上式右边第一项称为f在x_0 展开的r阶Taylor 多项式, 第二项称为Lagrange型余项.

【证】令$h=x-x_0$, $\varphi(t)=f(x_0+th),t\in[0,1]$. 据**命题8.20**之**(b)**的证明知$\varphi\in C^{r+1}([0,1])$ 且

$$\varphi^{(k)}(t)=\sum_{|\alpha|=k}\frac{k!}{\alpha!}h^\alpha D^\alpha f(x_0+th),\quad t\in[0,1],\quad k=0,1,2,...,r+1;$$

$$\varphi^{(k)}(0)=\sum_{|\alpha|=k}\frac{k!}{\alpha!}h^\alpha D^\alpha f(x_0),\quad k=0,1,2,...,r+1.$$

由一元函数的Taylor 公式(Lagrange 型余项), 存在$0<\theta<1$ 使得

$$\varphi(1)=\sum_{k=0}^{r}\frac{\varphi^{(k)}(0)}{k!}+\frac{\varphi(^{(r+1)}(\theta)}{(r+1)!}.$$

但$\varphi(1)=f(x_0+h)=f(x)$, 故得到

$$f(x_0+h)=\sum_{k=0}^{r}\sum_{|\alpha|=k}\frac{D^\alpha f(x_0)}{\alpha!}h^\alpha+\sum_{|\alpha|=r+1}\frac{D^\alpha f(x_0+\theta h)}{\alpha!}h^\alpha.$$

把$h=x-x_0$ 带入上式即得证. □

实际上由于f 是C^{r+1}类光滑函数, 上述Taylor公式的余项可以取为积分型的: 沿用上面记号我们有

$$\varphi(1)=\sum_{k=0}^{r}\frac{\varphi^{(k)}(0)}{k!}+\frac{1}{r!}\int_0^1(1-t)^r\varphi^{(r+1)}(t)\mathrm{d}t.$$

这就给出

【**命题8.22(具有积分型余项的Taylor公式).**】

设$\Omega\subset\mathbb{R}^n$为开集, r 为非负整数, 函数$f\in C^{r+1}(\Omega)$, $x_0\in\Omega$. 则对任意$x\in\Omega$, 只要闭直线段$[x_0,x]\subset\Omega$ 就有

$$f(x)=\sum_{|\alpha|\le r}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha+(r+1)\sum_{|\alpha|=r+1}\frac{(x-x_0)^\alpha}{\alpha!}\int_0^1(1-t)^r D^\alpha f(x_0+t(x-x_0))\mathrm{d}t.$$

特别对$r=0$ 即$f\in C^1(\Omega)$ 有

$$f(x)=f(x_0)+\sum_{j=1}^n(x_j-x_{0j})\int_0^1 D_jf(x_0+t(x-x_0))\mathrm{d}t.$$

□

注: 积分型余项有两个优点:

(1) 有显示结构, 便于各种估算, 以后研究函数空间, Sobolev 空间、数值分析等等, 会用到它.

(2) 可用于复值函数, 即

$$f(x)=f_1(x)+\mathrm{i}f_2(x),\qquad \mathrm{i}=\sqrt{-1}$$

其中$f_1,f_2\in C^{n+1}(\Omega,\mathbb{R})$. 复值函数的偏导数的定义为

$$D^\alpha(f_1+\mathrm{i}f_2)(x)=D^\alpha f_1(x)+\mathrm{i}D^\alpha f_2(x).$$

□

如果把C^{r+1} 类减弱成C^r 类但仍保持同阶Taylor 多项式, 则有下面的小欧余项 $o(|x-x_0|^r)$ 的Taylor 公式:

【**命题8.23(余项为$o(|x-x_0|^r)$的Taylor公式).**】设$\Omega\subset\mathbb{R}^n$为开集, r 为正整数, 函数$f\in C^r(\Omega)$, $B(x_0,\delta)\subset\Omega$. 则对任意$x\in B(x_0,\delta)$ 有

$$f(x)=\sum_{|\alpha|\le r}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha+o(|x-x_0|^r)$$

其中$o(|x-x_0|^r)$ 为$|x-x_0|^r$ 的高阶无穷小, 即$\frac{|o(|x-x_0|^r)|}{|x-x_0|^r}\to 0$ as $x\to x_0$.

【证】如前, 为记号简便, 令$h=x-x_0$. 应用**具有Lagrange型余项的Taylor公式**于r 阶光滑函数$f\in C^r(\Omega)$ 有

$$f(x_0+h)=\sum_{|\alpha|\le r-1}\frac{D^\alpha f(x_0)}{\alpha!}h^\alpha+\sum_{|\alpha|=r}\frac{D^\alpha f(x_0+\theta h)}{\alpha!}h^\alpha$$

$$= \sum_{|\alpha|\le r} \frac{D^\alpha f(x_0)}{\alpha!} h^\alpha + \sum_{|\alpha|=r} \frac{D^\alpha f(x_0+\theta h) - D^\alpha f(x_0)}{\alpha!} h^\alpha, \quad 0<\theta<1.$$

为证本命题, 只需证明

$$R(x_0,h) := \sum_{|\alpha|=r} \frac{D^\alpha f(x_0+\theta h) - D^\alpha f(x_0)}{\alpha!} h^\alpha = o(|h|^r).$$

事实上我们有

$$|R(x_0,h)| \le \sum_{|\alpha|=r} \frac{|D^\alpha f(x_0+\theta h) - D^\alpha f(x_0)|}{\alpha!} |h^\alpha|,$$

$$|h^\alpha| = |h_1|^{\alpha_1}|h_2|^{\alpha_2}\cdots|h_n|^{\alpha_n} \le |h|^{\alpha_1+\alpha_2+\cdots+\alpha_n} = |h|^{|\alpha|} = |h|^r$$

从而有

$$|R(x_0,h)| \le \alpha(x_0,h)|h|^r$$

其中

$$\alpha(x_0,h) = \sum_{|\alpha|=r} \frac{1}{\alpha!} \sup_{0\le\theta\le1} |D^\alpha f(x_0+\theta h) - D^\alpha f(x_0)|.$$

由$x \mapsto D^\alpha f(x)$ 的连续性有

$$\lim_{h\to0} \sup_{0\le\theta\le1} |D^\alpha f(x_0+\theta h) - D^\alpha f(x_0)| = 0 \qquad \forall\, |\alpha| = r.$$

因有限多个无穷小量之和还是无穷小量, 故得

$$\lim_{h\to0} \alpha(x_0,h) = 0.$$

因此

$$\frac{|R(x_0,h)|}{|h|^r} \le \alpha(x_0,h) \to 0 \quad (h\to0).$$

所以$R(x_0,h) = o(|h|^r)$. $\square$

• **Taylor 多项式的唯一性**.

【命题8.24(Taylor 多项式的唯一性).】设$\Omega \subset \mathbb{R}^n$为开集, r 为正整数, 函数$f \in C^r(\Omega)$, $B(x_0,\delta) \subset \Omega$. 假设存在常数组$\{c_\alpha\}_{|\alpha|\le n}$ 使得

$$f(x) = \sum_{|\alpha|\le r} c_\alpha (x-x_0)^\alpha + o(|x-x_0|^r) \qquad \forall\, x \in B(x_0;\delta).$$

则必有

$$c_\alpha = \frac{D^\alpha f(x_0)}{\alpha!} \qquad \forall\, |\alpha| \le r.$$

【证】由假设和Taylor公式有

$$0 = \sum_{|\alpha|\le r} \Big(c_\alpha - \frac{D^\alpha f(x_0)}{\alpha!}\Big)(x-x_0)^\alpha + o(|x-x_0|^r) \qquad \forall\, x \in B(x_0;\delta)$$

即

$$\sum_{|\alpha|\le r} \Big(c_\alpha - \frac{D^\alpha f(x_0)}{\alpha!}\Big)(x-x_0)^\alpha = o(|x-x_0|^r) \qquad \forall\, x \in B(x_0;\delta).$$

于是只需证明下列性质: 若存在常数组$\{C_\alpha\}_{|\alpha|\le r}$ 使得

$$\sum_{|\alpha|\le r} C_\alpha h^\alpha = o(|h|^r) \quad (|h| \to 0)$$

则必有

$$C_\alpha = 0 \qquad \forall\, |\alpha| \le r.$$

回忆一元函数: 在一元微分学中我们已证明了若常数$a_0, a_1, ..., a_r$ 满足

$$a_0 + a_1 t + a_2 t^2 + \cdots + a_r t^r = o(t^r) \quad (t \to 0)$$

则$a_0 = a_1 = \cdots = a_r = 0$. [直接证明也不难.]

固定任意$x \in \mathbb{R}^n$, 取$h = tx$, $|t| < \delta/(1+|x|)$. 则有$h^\alpha = t^{|\alpha|}x^\alpha$,

$$\sum_{k=0}^{r} t^k \left(\sum_{|\alpha|=k} C_\alpha x^\alpha\right) = \sum_{|\alpha|\le r} C_\alpha (tx)^\alpha = o(|t|^r|x|^r) = o(t^r) \qquad \forall\, |t| < \delta/(1+|x|).$$

据一元函数结果和x的任意性知

$$\sum_{|\alpha|=k} C_\alpha x^\alpha = 0 \qquad \forall\, x \in \mathbb{R}^n, \quad k = 0, 1, 2, ..., r.$$

也即

$$\sum_{|\alpha|\le r} C_\alpha x^\alpha \equiv 0.$$

对此恒等式两边取偏导数：应用单项式的偏导数计算公式, 对任意多重指标β 满足$|\beta| \le r$, 有

$$\begin{aligned}0 \equiv D^\beta \sum_{|\alpha|\le r} C_\alpha x^\alpha &= \sum_{|\alpha|\le n} C_\alpha D^\beta x^\alpha = \sum_{|\alpha|\le r,\, \alpha-\beta\in\mathbb{Z}^n_{\ge 0}} C_\alpha D^\beta x^\alpha \\ &= \sum_{|\alpha|\le r,\, \alpha-\beta\in\mathbb{Z}^n_{\ge 0}} C_\alpha \frac{\alpha!}{(\alpha-\beta)!} x^{\alpha-\beta} = C_\beta \beta! + \sum_{|\alpha|\le n,\, 0\ne\alpha-\beta\in\mathbb{Z}^n_{\ge 0}} C_\alpha \frac{\alpha!}{(\alpha-\beta)!} x^{\alpha-\beta}.\end{aligned}$$

取$x=0=(0,0,...,0)$, 则有

$$0^{\alpha-\beta}=0 \quad \text{if} \quad 0\neq \alpha-\beta\in\mathbb{Z}^n_{\geq 0}.$$

因此

$$0=C_\beta\beta!+\sum_{|\alpha|\leq r, 0\neq\alpha-\beta\in\mathbb{Z}^n_{\geq 0}} C_\alpha\frac{\alpha!}{(\alpha-\beta)!}0^{\alpha-\beta}=C_\beta\beta!$$

即$C_\beta=0$. 据β的任意性即得$C_\beta=0$ for all $|\beta|\leq r$. $\square$

【例：二元函数的Taylor公式】

Lagrange 型余项:

$$\begin{aligned}&f(x,y)=\sum_{i+j\leq r}\frac{1}{i!j!}\cdot\frac{\partial^{i+j}f}{\partial x^i\partial y^j}(x_0,y_0)(x-x_0)^i(y-y_0)^j\\&+\sum_{i+j=r+1}\frac{1}{i!j!}\cdot\frac{\partial^{r+1}f}{\partial x^i\partial y^j}(x_0+\theta(x-x_0),y_0+\theta(y-y_0))(x-x_0)^i(y-y_0)^j.\end{aligned}$$

积分型余项:

$$\begin{aligned}&f(x,y)=\sum_{i+j\leq r}\frac{1}{i!j!}\cdot\frac{\partial^{i+j}f}{\partial x^i\partial y^j}(x_0,y_0)(x-x_0)^i(y-y_0)^j\\&+(r+1)\sum_{i+j=r+1}\frac{1}{i!j!}(x-x_0)^i(y-y_0)^j\int_0^1(1-\theta)^r\frac{\partial^{r+1}f}{\partial x^i\partial y^j}(x_0+\theta(x-x_0),y_0+\theta(y-y_0))\mathrm{d}\theta.\end{aligned}$$

小欧余项:

$$f(x,y)=\sum_{i+j\leq r}\frac{1}{i!j!}\cdot\frac{\partial^{i+j}f}{\partial x^i\partial y^j}(x_0,y_0)(x-x_0)^i(y-y_0)^j+o([(x-x_0)^2+(y-y_0)^2]^{r/2}).$$

【例】求二元函数$f(x,y)=\cos(x^2+y)$在原点$(0,0)$ 展开的具有小欧余项的三阶Taylor公式.

方法一：用定义(永远可行).

$$\begin{aligned}&\frac{\partial f}{\partial x}(x,y)=-\sin(x^2+y)2x, \quad \frac{\partial f}{\partial y}(x,y)=-\sin(x^2+y)\\&\frac{\partial^2 f}{\partial x^2}(x,y)=-2\sin(x^2+y)-2x\cos(x^2+y)2x=-2\sin(x^2+y)-4x^2\cos(x^2+y),\\&\frac{\partial^2 f}{\partial x\partial y}(x,y)=-2x\cos(x^2+y), \quad \frac{\partial^2 f}{\partial y^2}(x,y)=-\cos(x^2+y),\end{aligned}$$

$$
\begin{aligned}
&\frac{\partial^3 f}{\partial x^3}(x,y) = -2\cos(x^2+y)2x - 8x\cos(x^2+y) + 4x^2\sin(x^2+y)2x,\\
&\frac{\partial^3 f}{\partial x^2\partial y}(x,y) = -2\cos(x^2+y) + 4x^2\sin(x^2+y),\\
&\frac{\partial^3 f}{\partial x\partial y^2}(x,y) = 2x\sin(x^2+y),\\
&\frac{\partial^3 f}{\partial y^3}(x,y) = \sin(x^2+y).
\end{aligned}
$$

$\Longrightarrow$

$$
\begin{aligned}
&\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0,\\
&\frac{\partial^2 f}{\partial x^2}(0,0) = 0, \quad \frac{\partial^2 f}{\partial x\partial y}(0,0) = 0, \quad \frac{\partial^2 f}{\partial y^2}(0,0) = -1,\\
&\frac{\partial^3 f}{\partial x^3}(0,0) = 0, \quad \frac{\partial^3 f}{\partial x^2\partial y}(0,0) = -2, \quad \frac{\partial^3 f}{\partial x\partial y^2}(0,0) = 0, \quad \frac{\partial^3 f}{\partial y^3}(0,0) = 0.
\end{aligned}
$$

$\Longrightarrow$

$$f(x,y) = \cos(x^2+y) = 1 - \frac{y^2}{2} - x^2y + o((x^2+y^2)^{3/2}).$$

方法二：利用Taylor公式中Taylor多项式的唯一性. 由$\cos u$ 的Taylor展开有

$$\cos u = \sum_{k=0}^{\infty}(-1)^k\frac{u^{2k}}{(2k)!} = 1 - \frac{u^2}{2} + O(u^4).$$

取$u = x^2 + y$得到

$$
\begin{aligned}
&\cos(x^2+y) = 1 - \frac{(x^2+y)^2}{2} + O((x^2+y)^4) = 1 - \frac{x^4+2x^2y+y^2}{2} + O((x^2+y)^4)\\
&= 1 - \frac{y^2}{2} - x^2y - \frac{x^4}{2} + O((x^2+y)^4) = 1 - \frac{y^2}{2} - x^2y + O((x^2+y^2)^2).
\end{aligned}
$$

最后注意显然有

$$O((x^2+y^2)^2) = o((x^2+y^2)^{3/2}) \qquad \left(\sqrt{x^2+y^2}\to 0\right).$$

以上用到: 当$x^2+y^2\le 1$ 时

$$\frac{x^4}{2} + |O((x^2+y)^4)| \le \frac{x^4}{2} + C(|x|+|y|)^4 \le C_1(x^2+y^2)^2.$$

其中C, C_1 为正常数. $\square$

【作业题】

1. 对任意$r \in \mathbb{N}$, 求函数$f(x) = e^{|x|^2}, x \in \mathbb{R}^n$,在$x_0 = 0$展开的具有Lagrange型余项的Taylor公式, 其中Taylor多项式的阶数(次数)为$2r$.

2. 设$\Omega \subset \mathbb{R}^n$ 是一个开凸集且$0 \in \Omega$. 设$f \in C^1(\Omega)$. 证明存在$g_j \in C(\Omega), i = 1, 2, ..., n$使得

$$f(x) = \sum_{j=1}^{n} x_j g_j(x), \quad x \in \Omega$$

且

$$g_j(0) = D_j f(0) = \frac{\partial f}{\partial x_j}(0), \quad j = 1, 2, ..., n.$$

[注意: 集合$\{ty \mid t \in [0,1], |y - x| \le \delta\}$ 是紧集.]

3. 陈书习题**8.4.3.**

4. 陈书习题**8.4.6.**

5. 陈书习题**8.4.8.**

6. 设$\Omega \subset \mathbb{R}^n$ 为开集, $f \in C^2(\Omega)$. 设Δ 是关于变量x的Laplace 算子, 即

$$\Delta f(x) = \frac{\partial^2 f}{\partial x_1^2}(x) + \frac{\partial^2 f}{\partial x_2^2}(x) + \cdots + \frac{\partial^2 f}{\partial x_n^2}(x), \quad x \in \Omega.$$

(1) 证明Laplace 算子在正交变换下保持形式不变, 即对任意正交矩阵$B = (b_{ij})_{n\times n}$, 令$g(y) = f(By)$, 则有

$$\Delta f(x) = \Delta g(y) \quad \text{其中} \quad x = By$$

$y \in B^{-1}(\Omega) = \{B^{-1}x \,|\, x \in \Omega\}$. 上式右端是关于$y$ 的Laplace 算子.

(2) 一般地, 设$A = (a_{ij})_{n\times n}$ 是一个对称正定矩阵. 则可确定一个可逆矩阵$B = (b_{ij})_{n\times n}$使得

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \Delta g(y) \qquad \text{其中} \quad x = By$$

$g(y) = f(By), y \in B^{-1}(\Omega) = \{B^{-1}x \,|\, x \in \Omega\}$, 而$\Delta$ 是关于y的Laplace 算子, 即

$$\Delta g(y) = \frac{\partial^2 g}{\partial y_1^2}(y) + \frac{\partial^2 g}{\partial y_2^2}(y) + \cdots + \frac{\partial^2 g}{\partial y_n^2}(y).$$

§8.5. 用偏导数研究极值和凹凸性

在一元函数微分学中我们已看到导数是研究函数的极值、单调性、凹凸性等方面的重要信息和有力工具. 本节表明在多元函数的情形更是如此. 主要工具是——

【梯度, Hesse 矩阵, 二阶Taylor公式(最常用)】

由**命题8.20之(a)** 有

$$\sum_{|\alpha|=m}\frac{D^\alpha f(x)}{\alpha!}h^\alpha=\frac{1}{m!}\Big(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\Big)^m f(x),\quad t\in[0,1].$$

特别对$m=1,2$ 分别有

$$\sum_{|\alpha|=1}\frac{D^\alpha f(x)}{\alpha!}h^\alpha=\Big(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\Big)f(x)=\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}f(x)=\langle\nabla f(x),h\rangle,$$

$$\sum_{|\alpha|=2}\frac{D^\alpha f(x)}{\alpha!}h^\alpha=\frac{1}{2!}\Big(\sum_{j=1}^n h_j\frac{\partial}{\partial x_j}\Big)^2 f(x)$$
$$=\frac12\sum_{i=1}^n\sum_{j=1}^n h_ih_j\frac{\partial^2}{\partial x_i\partial x_j}f(x)=\frac12\sum_{i=1}^n\sum_{j=1}^n\frac{\partial^2 f}{\partial x_i\partial x_j}(x)h_ih_j,\quad x\in\Omega.$$

假设$f\in C^2(\Omega)$. 则由对称性

$$\frac{\partial^2 f}{\partial x_i\partial x_j}(x)=\frac{\partial^2 f}{\partial x_j\partial x_i}(x)$$

知

$$h\mapsto\sum_{i=1}^n\sum_{j=1}^n\frac{\partial^2 f}{\partial x_i\partial x_j}(x)h_ih_j=h^\tau H_f(x)h,\qquad h\in\mathbb{R}^n$$

是h的一个二次型. 它对应的对称矩阵

$$H_f(x)=\Big(\frac{\partial^2 f}{\partial x_i\partial x_j}(x)\Big)_{n\times n}$$

叫做f 在点x的Hesse 矩阵, 即

$$H_f(x)=\begin{pmatrix}\frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1\partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_n}(x)\\ \frac{\partial^2 f}{\partial x_2\partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2\partial x_n}(x)\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ \frac{\partial^2 f}{\partial x_n\partial x_1}(x) & \frac{\partial^2 f}{\partial x_n\partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x)\end{pmatrix}.$$

使用Hesse 矩阵, Taylor公式中的二阶项可写为

$$\sum_{|\alpha|=2}\frac{D^\alpha f(x)}{\alpha!}h^\alpha=\frac12 h^\tau H_f(x)h.$$

这样一来, 对于C^2类函数, 它的Taylor公式可以用梯度和Hesse 矩阵表示:

【命题8.25(二阶Taylor公式).】设$\Omega \subset \mathbb{R}^n$为开集, 函数$f \in C^2(\Omega)$, $x_0 \in \Omega$. 则对任意$x \in \Omega$, 只要闭直线段$[x_0, x] \subset \Omega$ 就有下列三种二阶Taylor公式成立:

(a) Lagrange型:

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2}(x - x_0)^\tau H_f(\xi)(x - x_0)$$

其中$\xi = x_0 + \theta(x - x_0), 0 < \theta < 1$.

(b) 积分型:

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \int_0^1 (1 - t)(x - x_0)^\tau H_f(\xi(t))(x - x_0)\mathrm{d}t$$

其中$\xi(t) = x_0 + t(x - x_0), t \in [0, 1]$.

特别当Ω 是开凸集时, 以上(a),(b) 两型对任何$x_0, x \in \Omega$ 成立.

(c) 高阶无穷小型:

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2}(x - x_0)^\tau H_f(x_0)(x - x_0) + o(|x - x_0|^2). \qquad \square$$

上述二阶Taylor公式是最常用的, 因为它们是由两个最重要的力学量和几何量—— 梯度和Hesse 矩阵——组成的.

【函数的极值】

【定义(局部极值)】设实函数$f : E \subset \mathbb{R}^n \to \mathbb{R}$, x_0 是E的一个**内点**.

若存在x_0的一个邻域$B(x_0, \delta) \subset E$ 使得

$$f(x) \geq f(x_0) \qquad (\text{resp.} \quad f(x) \leq f(x_0) \ ) \qquad \forall\, x \in B(x_0, \delta)$$

则称x_0是f的一个局部极小(大)值点同时称$f(x_0)$ 为f的一个局部极小(大)值. f的局部极小和极大值点统称为f的局部极值点; 局部极小和极大值统称为f的局部极值.

而若存在x_0的一个邻域$B(x_0, \delta) \subset E$ 使得

$$f(x) > f(x_0) \qquad (\text{resp.} \quad f(x) < f(x_0) \ ) \qquad \forall\, x \in B(x_0, \delta) \setminus \{x_0\}.$$

则称x_0是f的一个局部严格极小(大)值点同时称$f(x_0)$ 为f的一个局部严格极小(大)值. f的局部严格极小和严格极大值点统称为f的局部严格极值点; 局部严格极小和严格极大值统称为f的局部严格极值. $\square$

【**强调**】函数的**局部极值点**首先是函数定义域的**内点**, 这点很重要因为这时我们可以使用微分技术对其进行研究.

【**定义(整体最值)**】设实函数$f : E \subset \mathbb{R}^n \to \mathbb{R}$,

若存在$x_0 \in E$使得

$$f(x) \geq f(x_0) \qquad (\text{resp.} \quad f(x) \leq f(x_0) \ ) \qquad \forall\, x \in E$$

即

$$f(x_0) = \min_{x \in E} f(x) \qquad (\ \text{resp.}\ \ f(x_0) = \max_{x \in E} f(x)\ )$$

则称x_0是f的一个整体最小(大)值点, 同时称$f(x_0)$ 为f的整体最小(大)值. f的整体最小值点和整体最大值点统称为f的整体最值点; 整体最小值和整体最大值统称为f的整体最值.

而若

$$f(x) > f(x_0) \qquad (\text{resp.} \quad f(x) < f(x_0) \ ) \qquad \forall\, x \in E \setminus \{x_0\}.$$

则称x_0是f的一个整体严格最小(大)值点, 同时称$f(x_0)$ 为f的整体严格最小(大)值. f的整体严格最小值点和整体严格最大值点统称为f的整体严格最值点; 整体严格最小值和整体严格最大值统称为f的整体严格最值. □

【**定义(临界点)**】设映射$f : E \subset \mathbb{R}^n \to \mathbb{R}^m$在$E$的内点$x_0$ 处可微且微分$f'(x_0) = J_f(x_0)$ (Jacobi矩阵) 不是满秩的, 即$\text{rank}(f'(x_0)) < \min\{m, n\}$, 则称$x_0$ 为f的一个临界点(critical point) 同时称$f(x_0)$ 为f的一个临界值(critical value).

特别当$m = 1$ 即f 是实值函数时, 若梯度$\nabla f(x_0) = 0$, 即

$$\frac{\partial f}{\partial x_1}(x_0) = \frac{\partial f}{\partial x_2}(x_0) = \cdots = \frac{\partial f}{\partial x_n}(x_0) = 0$$

则称x_0 是函数f的一个临界点. □

【**命题8.26(取得极值的必要条件——Fermat极值原理).**】设实函数f在其定义域的内点x_0处可微. 则x_0 是f的局部极值点的必要条件是:

x_0 是f的临界点, 即$f'(x_0) = 0$ 也即$\nabla f(x_0) = 0$.

换言之, 若x_0不是f的临界点, 即$\nabla f(x_0) \neq 0$, 则x_0 不是f的局部极值点.

【证】取$\delta > 0$满足$B(x_0, \delta) \subset f$的定义域. 不妨设$f(x_0)$ 为局部极小值. 则对任意单位向量$\mathbf{v} \in \mathbb{R}^m$ 有

$$\frac{f(x_0 + t\mathbf{v}) - f(x_0)}{t} \geq 0 \qquad \forall\, t \in (-\delta, \delta) \setminus \{0\}.$$

令$t \to 0$, 则由方向导数与梯度的关系有

$$0 \leq \lim_{t \to 0} \frac{f(x_0 + t\mathbf{v}) - f(x_0)}{t} = D_{\mathbf{v}} f(x_0) = \langle \nabla f(x_0),\, \mathbf{v} \rangle.$$

以$-\mathbf{v}$ 代替$\mathbf{v}$ 得到$-\langle \nabla f(x_0),\, \mathbf{v} \rangle = \langle \nabla f(x_0),\, -\mathbf{v} \rangle \geq 0$ 即$\langle \nabla f(x_0),\, \mathbf{v} \rangle \leq 0$. 所以

$$\langle \nabla f(x_0),\, \mathbf{v} \rangle = 0 \qquad \forall\, |\mathbf{v}| = 1.$$

这蕴含$\nabla f(x_0) = 0$.

另法证明: 假设$\nabla f(x_0) \neq 0$. 记

$$\mathbf{e} = \frac{\nabla f(x_0)}{|\nabla f(x_0)|}.$$

因f在x_0 可微, 故

$$\lim_{t \to 0} \frac{f(x_0 + t\mathbf{e}) - f(x_0)}{t} = D_{\mathbf{e}} f(x_0) = \langle \nabla f(x_0), \mathbf{e} \rangle = |\nabla f(x_0)| > 0.$$

因此存在$\eta > 0\, (\eta < \delta)$ 使得

$$\frac{f(x_0 + t\mathbf{e}) - f(x_0)}{t} > 0 \qquad \forall\, t \in (-\eta, \eta) \setminus \{0\}.$$

从而有

$$f(x_0 + t\mathbf{e}) > f(x_0) \quad \forall\, t \in (0, \eta); \qquad f(x_0 + t\mathbf{e}) < f(x_0) \quad \forall\, t \in (-\eta, 0).$$

所以x_0 不是f的局部极值点. $\square$

Fermat极值原理的一个推论——

【**Rolle 定理**】设$\Omega \subset \mathbb{R}^m$为有界开集, $f : \overline{\Omega} \to \mathbb{R}$ 连续且在Ω内可微. 若f在边界$\partial\Omega$上为常数, 则存在$x_0 \in \Omega$ 使得$\nabla f(x_0) = 0$.

【证】由假设知存在常数c 使得$f(x) = c$ for all $x \in \partial\Omega$. 因$\overline{\Omega}$是紧集, 故存在$x_0, y_0 \in \overline{\Omega}$ 使得

$$f(x_0) = \min_{x \in \overline{\Omega}} f(x), \quad f(y_0) = \max_{x \in \overline{\Omega}} f(x).$$

若$x_0, y_0 \in \partial\Omega$, 则$f(x_0) = f(y_0) = c$ 因而$f(x) \equiv c, x \in \overline{\Omega}$. 这蕴含$\nabla f(x) = 0$ for all $x \in \Omega$.

若x_0, y_0 不都属于$\partial\Omega$, 例如设$x_0 \notin \partial\Omega$, 则$x_0 \in \Omega$. 这蕴含$x_0$ 也是f 在Ω内的一个局部极小值点. 据Fermat极值原理即知$\nabla f(x_0) = 0$. $\Box$

【整体最值点的搜索范围】设$\Omega \subset \mathbb{R}^n$为有界开集, 函数$f : \overline{\Omega} \to \mathbb{R}$ 连续且在Ω 内可微. 则

$$\max_{x\in\overline{\Omega}} f(x) = \max\{f(x) \mid x \in \Omega \ \text{且}\ \nabla f(x) = 0;\ \ \text{或者}\ \ x \in \partial\Omega\},$$
$$\min_{x\in\overline{\Omega}} f(x) = \min\{f(x) \mid x \in \Omega \ \text{且}\ \nabla f(x) = 0;\ \ \text{或者}\ \ x \in \partial\Omega\}. \qquad \Box$$

【用Hesse 矩阵判断函数取得局部极值的充分条件和必要条件】

【命题8.27.】设开集$\Omega \subset \mathbb{R}^m$, $f \in C^2(\Omega, \mathbb{R})$. 设$x_0 \in \Omega$ 是f的一个临界点, 即$\nabla f(x_0) = 0$. 则

(a) 若$H_f(x_0)$ 正定, 则x_0 是f的一个局部严格极小值点.

若$H_f(x_0)$ 负定, 则x_0 是f的一个局部严格极大值点.

(b) 若x_0 是f的一个局部极小值点, 则$H_f(x_0)$半正定;

若x_0 是f的一个局部极大值点, 则$H_f(x_0)$半负定.

(c) 若$H_f(x_0)$ 不定, 则x_0 不是f的极值点.

【注：使Hesse 矩阵$H_f(x_0)$为不定的点x_0 称为f的**鞍点**.】

【证】(a): 设Hesse 矩阵$A := H_f(x_0)$正定. 设λ 为A的最小特征值. 则

$\lambda > 0$ 且$h^\tau A h \ge \lambda |h|^2$ for all $h \in \mathbb{R}^m$. 由二阶Taylor公式和$\nabla f(x_0) = 0$ 有

$$f(x) - f(x_0) = \frac{1}{2}(x - x_0)^\tau A(x - x_0) + o(|x - x_0|^2) \ge \frac{\lambda}{2}|x - x_0|^2 + o(|x - x_0|^2).$$

取$\delta > 0$ 使得$B(x_0; \delta) \subset \Omega$ 且$|o(|x - x_0|^2)| \le \frac{\lambda}{4}|x - x_0|^2$ for all $x \in B(x_0; \delta)$. 则有

$$f(x) - f(x_0) \ge \frac{\lambda}{2}|x - x_0|^2 - \frac{\lambda}{4}|x - x_0|^2 = \frac{\lambda}{4}|x - x_0|^2 > 0 \qquad \forall x \in \breve{B}(x_0; \delta).$$

这表明x_0 是f的一个局部严格极小值点.

其次设$H_f(x_0)$ 负定. 由于$H_{-f}(x_0) = -H_f(x_0)$, 故$H_f(x_0)$ 负定等价于$-H_f(x_0)$正定, 即$H_{-f}(x_0)$正定. 又$\nabla(-f)(x_0) = -\nabla f(x_0) = 0$, 故$x_0$ 是$-f$的一个局部严格极小值点, 也即x_0 是f的一个局部严格极大值点.

(b): 设x_0 是f的一个局部极小值点. 要证明$H_f(x_0)$半正定, 即证明$h^\tau H_f(x_0)h \ge 0$ for all $h \in \mathbb{R}^m$.

任取定$h \in \mathbb{R}^m$. 由Lagrange型的二阶Taylor公式知当$0 < t << 1$ 时有

$$0 \le f(x_0 + th) - f(x_0) = \frac{1}{2}t^2 h^\tau H_f(x_0 + \theta th)h \qquad (0 < \theta < 1\,).$$

这给出

$$h^\tau H_f(x_0 + \theta th)h \ge 0 \qquad \forall\, 0 < t << 1.$$

而由$f \in C^2$知所有二阶偏导数$x \mapsto \frac{\partial^2 f}{\partial x^i \partial x^j}(x)$ 连续. 故

$$h^\tau H_f(x_0)h = \lim_{t \to 0^+} h^\tau H_f(x_0 + \theta th)h \ge 0.$$

所以$H_f(x_0)$半正定.

同理可证, 当x_0 是f的一个局部极大值点时, $H_f(x_0)$半负定.

(c): 设$H_f(x_0)$ 不定. 则$H_f(x_0)$ 既非半正定也非半负定. 由(b) 知x_0 既不是f的局部极小值点, 也不是f的局部极大值点. 因此x_0 不是f的极值点. □

【例】设$f(x, y) = (y - y_0)^2 - (x - x_0)^2$, $(x, y) \in \mathbb{R}^2$. 则f的梯度为(行向量)

$$\nabla f(x, y) = (-2(x - x_0), 2(y - y_0)).$$

由此看出f的临界点为(x_0, y_0) 即$\nabla f(x_0, y_0) = (0, 0)$. 但$f$的Hesse 矩阵

$$H_f(x, y) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$$

是不定的. 因此f 没有极值点. 特别,临界点(x_0, y_0)不是极值点. 这一点也可直接从定义推出. 图象上看出: 曲面$z = f(x, y) = (y - y_0)^2 - (x - x_0)^2$ 是一个马鞍面, 临界点(x_0, y_0)位于马鞍面的中心. “鞍点”一词由此而来.

【函数的凹凸性】

【定义】设$E \subset \mathbb{R}^n$ 为一凸集, $f : E \to \mathbb{R}$.

我们称f是凸函数(convex function) 如果

$$f((1-t)x+ty) \le (1-t)f(x)+tf(y) \qquad \forall\, x,y\in E, t\in[0,1].$$

我们称f是凹函数(concave function) 如果

$$f((1-t)x+ty) \ge (1-t)f(x)+tf(y) \qquad \forall\, x,y\in E, t\in[0,1].$$

$\square$

易见 $\quad f$ 是凹函数$\iff -f$是凸函数.

因此在基本性质推导中只需考虑凸函数.

【命题8.28(凸函数的微分刻画).】设$\Omega\subset\mathbb{R}^n$ 为开凸集, 函数$f:\Omega\to\mathbb{R}$ 在Ω上可微. 则

f是凸函数$\iff$ 曲面$y=f(x)$位于它的任意点处的切平面的上方, 即

$$f(x)\ge f(x_0)+\langle\nabla f(x_0), x-x_0\rangle \qquad \forall\, x,x_0\in\Omega.$$

【证】"$\Longrightarrow$": 设f 是凸函数. 对任意$x,x_0\in\Omega$ 有

$$f((1-t)x_0+tx)) \le (1-t)f(x_0)+tf(x) \qquad \forall\, 0\le t\le 1.$$

如写成

$$(1-t)x_0+tx = x_0+t(x-x_0), \quad (1-t)f(x_0)+tf(x)=f(x_0)+t(f(x)-f(x_0))$$

则得到

$$f(x)-f(x_0)\ge \frac{f(x_0+t(x-x_0))-f(x_0)}{t} \qquad \forall\, 0<t<1.$$

因f在x_0可微, 令$t\to 0^+$ 即得

$$f(x)-f(x_0)\ge \lim_{t\to0^+}\frac{f(x_0+t(x-x_0))-f(x_0)}{t} = \frac{\mathrm{d}}{\mathrm{d}t}\Big(f(x_0+t(x-x_0))\Big)\Big|_{t=0}$$
$$=\Big\langle\nabla f(x_0+t(x-x_0)), x-x_0\Big\rangle\Big|_{t=0}=\langle\nabla f(x_0), x-x_0\rangle.$$

"$\Longleftarrow$": 对任意$x,y\in\Omega, t\in[0,1]$, 令$z=(1-t)x+ty$. 则由假设有

$$f(x)\ge f(z)+\langle\nabla f(z), x-z\rangle, \quad f(y)\ge f(z)+\langle\nabla f(z), y-z\rangle.$$

将第一个不等式两边乘以$(1-t)$, 第二个乘以t, 然后二者相加并注意$(1-t)x+ty=z$, 便有

$$(1-t)f(x)+tf(y)\ge f(z)+\Big\langle \nabla f(z),(1-t)x+ty-z\Big\rangle = f(z)$$

也即$f((1-t)x+ty)\le(1-t)f(x)+tf(y)$. 所以$f$是凸函数. □

由上述命题我们立即得到关于凸函数的最小值的一个充分条件:

【命题8.29.】设$\Omega\subset\mathbb{R}^n$ 为开凸集, 函数$f:\Omega\to\mathbb{R}$为凸函数. 若$x_0\in\Omega$ 是f的一个临界点, 即$\nabla f(x_0)=0$, 则$f(x_0)$ 是整体最小值, 即$f(x_0)=\min\limits_{x\in\Omega}f(x)$. □

下面看Hesse 矩阵的作用.

【命题8.30.(凸函数的Hesse矩阵刻画)】设$\Omega\subset\mathbb{R}^n$ 为开凸集, 函数$f\in C^2(\Omega)$. 则:

f是凸函数$\Longleftrightarrow$ Hesse 矩阵$H_f(x)$处处半正定.

【证】"$\Longleftarrow$": 设Hesse 矩阵$H_f(x)$处处半正定. 对任意$x,x_0\in\Omega$, 由Ω 是凸集知闭直线段$[x_0,x]\subset\Omega$. 应用Lagrange型的二阶Taylor 公式和H_f处处半正定即得

$$f(x)-f(x_0)-\langle\nabla f(x_0),x-x_0\rangle=\frac{1}{2}(x-x_0)^\tau H_f(\xi)(x-x_0)\ge 0$$

即

$$f(x)\ge f(x_0)+\langle\nabla f(x_0),x-x_0\rangle$$

其中$\xi=x_0+\theta(x-x_0),0<\theta<1$. 据**凸函数的微分刻画**即知$f$ 是凸函数.

"$\Longrightarrow$": 设f是凸函数. 任取$x\in\Omega$. 因Ω是开集故存在$\delta>0$ 使得$B(x,\delta)\subset\Omega$. 对任意$h\in\mathbb{R}^n$, 当$0<t<\frac{\delta}{1+|h|}$ 时有$x+th\in B(x,\delta)$. 应用**Lagrange型的二阶Taylor 公式** 和**凸函数的微分刻画** 有

$$\frac{1}{2}t^2h^\tau H_f(x+\theta th)h=f(x+th)-f(x)-\langle\nabla f(x),th\rangle\ge 0$$

从而有

$$h^\tau H_f(x+\theta th)h\ge 0\qquad \forall\, 0<t<<\frac{\delta}{1+|h|}.$$

再由$f\in C^2$ 即得

$$h^\tau H_f(x)h=\lim_{t\to 0^+}h^\tau H_f(x+\theta th)h\ge 0.$$

据$h\in\mathbb{R}^n$的任意性可知$H_f(x)$半正定. □

【作业题】

1. 陈书习题**8.1.3.**

2. 设$f \in C^1(\mathbb{R}^n)$ 满足

$$\langle \nabla f(x), x\rangle \geq 0 \qquad \forall\, x \in \mathbb{R}^n.$$

证明$f(0) = \min\limits_{x\in\mathbb{R}^n} f(x),\ \nabla f(0) = 0.$

[学习**定理8.16(实值函数的微分中值等式)** 和那里的等式(♠).]

3. 陈书习题**8.4.1.**

4. 令

$$f(x,y) = (x-y)\log(\frac{x}{y}), \quad (x,y) \in \mathbb{R}^2_{>0}.$$

证明f 是非负的凸函数.

5. 设$\Omega \subset \mathbb{R}^n$ 为开集, $f \in C^2(\Omega)$. 证明f 的梯度的微分等于f的Hesse 矩阵, 即

$$(\nabla f)'(x) = H_f(x), \quad x \in \Omega.$$

6. 设$f \in C^2(\mathbb{R}^n)$.

(1) 证明对任意$x, y \in \mathbb{R}^n$, 存在$\xi \in (x,y)$ (开线段) 使得

$$\langle x-y, \nabla f(x) - \nabla f(y)\rangle = (x-y)^\tau H_f(\xi)(x-y).$$

[参见**定理8.17(一般映射的微分中值不等式)**证明中的内积等式(¶)].

(2) 假设f 的Hesse 矩阵$H_f(x)$ 处处正定. 证明f 是严格凸函数. 此外证明梯度映射$x \mapsto \nabla f(x)$ 是严格保序的, 即

$$\langle \nabla f(x) - \nabla f(x),\, x-y\rangle > 0 \qquad \forall\, x, y \in \mathbb{R}^n,\ \ x \neq y.$$

§8.6. 单位分解

单位分解(partition of unity) 是研究几何, 拓扑和大范围分析时使用的一种技术, 用于沟通局部与局部之间、局部与整体之间的关系. 本课程中, 在学习微分形式的积分时要用到这个技术.

首先引进函数类$C_c^r(\mathbb{R}^n)$ (其中$0 \le r \le \infty$) 其定义为

$$f \in C_c^r(\mathbb{R}^n) \iff f \in C^r(\mathbb{R}^n) \text{ 且支集 } \mathrm{supp} f \text{ 为紧集.}$$

【定理8.31(单位分解).】 设$\{U_\alpha\}_{\alpha \in A}$ 为$\mathbb{R}^n$中的一族开集, 则存在**非负**函数列$\{\phi_j\}_{j=1}^\infty$具有下列性质:

(i) 对每个$j \in \mathbb{N}$, $\phi_j \in C_c^\infty(\mathbb{R}^n)$ 且存在$\alpha_j \in A$ 使得$\mathrm{supp}\phi_j \subset U_{\alpha_j}$.

(ii) 函数列$\{\phi_j\}_{j=1}^\infty$具有局部有限性, 即对每个紧集$K \subset \bigcup_{\alpha \in A} U_\alpha$, 下标集 $J := \{j \in \mathbb{N} \mid \exists x \in K \text{ s.t. } \phi_j(x) \neq 0\}$ 是有限集. 因此

$$\text{对任意} j \in \mathbb{N} \setminus J \text{ 有} \quad \phi_j(x) \equiv 0, \quad x \in K.$$

(iii)

$$\sum_{j=1}^\infty \phi_j(x) \equiv 1, \qquad x \in \bigcup_{\alpha \in A} U_\alpha.$$

[具有性质(i)-(iii)的**非负**函数列$\{\phi_j\}_{j=1}^\infty$ 称为从属于开集族$\{U_\alpha\}_{\alpha \in A}$的一个光滑的单位分解.]

为证这一定理我们需要两个引理.

【引理8.32.】 对每个非空开集$U \subset \mathbb{R}^n$ 可以构造一列渐升的紧集$\{K_i\}_{i=1}^\infty$ 使得

$$K_i \subset K_{i+1}^\circ \subset U, \quad i = 1, 2, 3, ...; \quad U = \bigcup_{i=1}^\infty K_i.$$

【证】当$U = \mathbb{R}^n$ 时, 取$K_i = \overline{B}(0, i)(i = 1, 2, 3, ...)$ 即可. 设$U \neq \mathbb{R}^n$ 即$U^c \neq \emptyset$. 令

$$K_i = \{x \in \mathbb{R}^n \mid B(x, 1/i) \subset U, |x| \le i\}$$

$$= \{x \in \mathbb{R}^n \mid \rho(x, U^c) \ge 1/i\} \cap \overline{B}(0, i).$$

易见第二个等号确实成立. 因$x \mapsto \rho(x, U^c)$ 连续, 故这第二个表示说明K_i是有界闭集, 即是$\mathbb{R}^n$中的紧集.

对任意$x \in U$, 有$x \notin U^c$ 因而有$\rho(x,U^c) > 0$. 取自然数$i \geq \max\{|x|, \frac{1}{\rho(x,U^c)}\}$ 则有$\rho(x,U^c) \geq \frac{1}{i}$ 且$|x| \leq i$. 因此$x \in K_i$. 这证明了$U \subset \bigcup_{i=1}^{\infty} K_i$. 而显然有$K_i \subset U, i = 1,2,3,\dots$. 所以$U = \bigcup_{i=1}^{\infty} K_i$.

对任意$x \in K_i$, 令$\delta = \frac{1}{i} - \frac{1}{i+1}$. 则对任意$y \in B(x,\delta)$ 有

$$\rho(y,U^c) \geq \rho(x,U^c) - |x-y| > \frac{1}{i} - \delta = \frac{1}{i+1}, \quad |y| \leq |y-x| + |x| < 1 + i.$$

因此$y \in K_{i+1}$. 这表明$B(x,\delta) \subset K_{i+1}$. 所以$x \in K_{i+1}^\circ$. 所以$K_i \subset K_{i+1}^\circ\,(i = 1,2,3,...)$.
□

【引理8.33.】设$\{U_\alpha\}_{\alpha\in A}$ 为$\mathbb{R}^n$中的一族非空开集. 则存在一列开球$\{V_j\}_{j=1}^{\infty}$ 具有下列性质:

(i)

$$\bigcup_{\alpha\in A} U_\alpha = \bigcup_{j=1}^{\infty} V_j.$$

(ii) 对任意$j \in \mathbb{N}$ 存在$\alpha_j \in A$ 使得$\overline{V}_j \subset U_{\alpha_j}$.

(iii) 对任意紧集$K \subset \bigcup_{\alpha\in A} U_\alpha$, 只有有限多个$V_j$与$K$ 相交, 即$\{j \in \mathbb{N} \,|\, K \cap V_j \neq \emptyset\}$ 是有限集.

【证】令$U = \bigcup_{\alpha\in A} U_\alpha$ 并设$\{K_i\}_{i=1}^{\infty}$ 是上一引理中对于U构作的紧集列. 令

$$K_{-1} = K_0 = \emptyset, \quad L_i = K_{i+1} \setminus K_i^\circ, \quad i = 0,1,2,\dots.$$

由K_i 的性质有

$U = \bigcup_{i=1}^{\infty} K_i = \bigcup_{i=1}^{\infty}(K_i \setminus K_{i-1}) \subset \bigcup_{i=1}^{\infty}(K_i \setminus K_{i-1}^\circ) = \bigcup_{i=0}^{\infty}(K_{i+1} \setminus K_i^\circ) = \bigcup_{i=0}^{\infty} L_i \subset U$.
因此

$$U = \bigcup_{i=0}^{\infty} L_i \quad 且 \quad L_i \cap K_{i-1} \subset L_i \cap K_i^\circ = \emptyset, \quad i = 0,1,2,\dots.$$

下面来找开球V_j.

对任意$i \in \mathbb{N} \cup \{0\}$, 对任意$x \in L_i$ 存在$\alpha = \alpha_x \in A$ 使得$x \in U_\alpha$ 同时有$x \notin K_{i-1}$. 这表明$x \in U_\alpha \setminus K_{i-1}$. 因$U_\alpha \setminus K_{i-1}$ 是开集, 故存在$\delta_x > 0$ 使得闭球$\overline{B}(x,\delta_x) \subset U_\alpha \setminus K_{i-1}$. 这样我们得到

$$L_i \subset \bigcup_{x\in L_i} B(x,\delta_x).$$

但L_i 是紧集, 故在开球族$\{B(x,\delta_x)\}_{x\in L_i}$ 中存在有限多个开球, 记作$B_{i,1},B_{i,2},...,B_{i,N_i}$, 使得$L_i\subset\bigcup_{j=1}^{N_i}B_{i,j}$ 且

$$\text{对每个}B_{i,j},\ \text{存在}\alpha\in A\ \text{使得}\ \overline{B}_{i,j}\subset U_\alpha\setminus K_{i-1}.$$

于是由$U=\bigcup_{i=0}^{\infty}L_i$ 和$B_{i,j}\subset\overline{B}_{i,j}\subset U$ 得到

$$U=\bigcup_{i=0}^{\infty}\bigcup_{j=1}^{N_i}B_{i,j}=\bigcup_{i=0}^{\infty}\bigcup_{j=1}^{N_i}\overline{B}_{i,j}.$$

将所有这些球做成一个集合

$$\mathcal{B}=\{B_{i,j}\mid j=1,2,...,N_i;\ i=0,1,2,...\}.$$

易见$\mathcal{B}$ 是可数集. 来证明$\mathcal{B}$ 是无限集. 假设$\mathcal{B}$ 是有限集, 则将其元素排成$\mathcal{B}=\{B_{i_1,j_1},\ B_{i_2,j_2},\ ...,\ B_{i_p,j_p}\}$ 其中$p\in\mathbb{N}$, 便有

$$U=\bigcup_{s=1}^{p}B_{i_s,j_s}=\bigcup_{s=1}^{p}\overline{B}_{i_s,j_s}.$$

因有限多个紧集的并还是紧集, 故上式表明U既是开集又是紧集, 但U非空, 这便是矛盾. [细：U既开又紧蕴含U既开又闭, 但U非空, 故必有$U=\mathbb{R}^n$. 但这又与U 是紧集矛盾.] 这矛盾证明了$\mathcal{B}$ 是可数无限集. 将$\mathcal{B}$ 排列成$\mathcal{B}=\{V_j\}_{j=1}^{\infty}$, 则易见$\{V_j\}_{j=1}^{\infty}$ 满足引理中的(i),(ii).

对于(iii), 任取紧集$K\subset U$. 由K_i的性质有$U=\bigcup_{i=1}^{\infty}K_i=\bigcup_{i=1}^{\infty}K_i^\circ$ 故$\{K_i^\circ\}_{i=1}^{\infty}$ 是K的一族开覆盖，因此由K紧知存在$i_1,i_2,...,i_N\in\mathbb{N}$ 使得$K\subset\bigcup_{j=1}^{N}K_{i_j}^\circ$. 令$m=\max\{i_1,i_2,...,i_N\}$ 则由K_i°单调增加有$\bigcup_{j=1}^{N}K_{i_j}^\circ=K_m^\circ$. 于是有$K\subset K_m^\circ\subset K_m$. 现在对任意$i\ge m+1$ 和任意$j\in\{1,2,...,N_i\}$ 有

$$K\cap B_{i,j}\subset K_m\cap B_{i,j}=K_{i-1}\cap B_{i,j}=\emptyset.$$

因此若$K\cap B_{i,j}\ne\emptyset$, 则必有$1\le i\le m, 1\le j\le N_i$. 因此若$K\cap V_j\ne\emptyset$, 则必有$V_j\in\{B_{i,j}\mid j=1,2,...,N_i\,;\ i=0,1,...,m\}$ 从而这样的V_j 只有有限多个, 即$\{j\in\mathbb{N}\,|\,K\cap V_j\ne\emptyset\}$ 是有限集. $\square$

【注】有同学可能问: 既然以有理点为中心、正有理数为半径的球只有可数多个且在任何开集内稠密, 为何不考虑它们? 主要难点在于如何选取这些开球使他它们具有引理中的性质(iii).

【单位分解定理的证明:】 令$U=\bigcup_{\alpha\in A}U_\alpha$ 并设$\{V_j\}_{j=1}^\infty$ 是上面引理中的一列开球. 将每个开球V_j写成$V_j=B(a_j,\delta_j)$ 其中a_j且球心, $\delta_j>0$为半径. 令$\psi_j(x)=g(1-\frac{|x-a_j|^2}{\delta_j^2})$ 其中$g(t)=e^{-1/t}1_{\{t>0\}}$. 则我们在前面例子中已证明了$0\le\psi_j\in C_c^\infty(\mathbb{R}^n),\psi_j(x)>0$ for all $x\in V_j$, 以及$\mathrm{supp}\psi_j=\overline{V}_j$. 令

$$\psi(x)=\sum_{j=1}^\infty\psi_j(x),\quad x\in U.$$

[注意, 对每个$x\in U$ 这是一个正项级数, 因此其和有意义.] 对每个紧集$K\subset U$, 由$\{V_j\}_{j=1}^\infty$的性质知$J:=\{j\in\mathbb{N}\mid K\cap V_j\neq\emptyset\}$ 是有限集. 而易见$K\cap V_j\neq\emptyset\iff$ 存在$x\in K$ 使得$\psi_j(x)\neq0$. 因此

$$J=\{j\in\mathbb{N}\mid\exists\, x\in K\ \text{s.t.}\ \phi_j(x)\neq0\}=\{j\in\mathbb{N}\mid K\cap V_j\neq\emptyset\}\ \ \text{是有限集.}$$

特别, 对每个$x_0\in U$, 取$\delta>0$ 充分小使得闭球$\overline{B}(x_0,\delta)\subset U$. 则$J:=\{j\in\mathbb{N}\mid\exists\, x\in\overline{B}(x_0,\delta)$ s.t. $\phi_j(x)\neq0\}$ 是有限集. 因此

$$\psi(x)=\sum_{j\in J}\psi_j(x)<+\infty\qquad\forall\, x\in\overline{B}(x_0,\delta).$$

由$x_0\in U$的任意性知$\psi(x)$在U上处处有定义且有限. 又由$U=\bigcup_{j=1}^\infty V_j$ 可知$\psi(x)>0$ for all $x\in U$. 进一步, 因在$\overline{B}(x_0,\delta)$ 上ψ 等于有限多个C^∞函数之和, 故至少在开球$B(x_0,\delta)$ 有$\psi|_{B(x_0,\delta)}\in C^\infty(B(x_0,\delta))$. 再由$x_0\in U$的任意性即知$\psi\in C^\infty(U)$. 于是再由$\psi(x)>0$ for all $x\in U$ 可知$1/\psi\in C^\infty(U)$. 对每个$j\in\mathbb{N}$, 令

$$\phi_j(x)=\begin{cases}\psi_j(x)/\psi(x) & \text{if}\quad x\in U\\ 0 & \text{if}\quad x\in\mathbb{R}^n\setminus U\end{cases}$$

来证明$\phi_j\in C_c^\infty(\mathbb{R}^n)$ 且$\mathrm{supp}\phi_j=\mathrm{supp}\psi_j=\overline{V}_j$. 这个支集关系显然成立且闭球$\overline{V}_j$是紧集, 因此只需证明$\phi_j\in C^\infty(\mathbb{R}^n)$. 任取$x_0\in\mathbb{R}^n$. 若$x_0\in\mathrm{supp}\phi_j$, 则由$\mathrm{supp}\phi_j=\overline{V}_j\subset U$ 知存在$\delta>0$ 使得$B(x_0,\delta)\subset U$. 于是在$B(x_0,\delta)$ 内有$\phi_j=\psi_j/\psi\in C^\infty(B(x_0,\delta))$. 若$x_0\not\in\mathrm{supp}\phi_j$, 即$x_0\in(\mathrm{supp}\phi_j)^c$, 则由于后者是开集, 故存在$\delta>0$ 使得$B(x_0,\delta)\subset(\mathrm{supp}\phi_j)^c$. 由此和支集的定义可知在$B(x_0,\delta)$ 内有$\phi_j\equiv0\in C^\infty(B(x_0,\delta))$. 这就证明了$\phi_j$ 在$\mathbb{R}^n$中每一点的一个邻域内都是无限次可微的. 所以$\phi_j\in C^\infty(\mathbb{R}^n)$.

最后由ϕ_j 与ψ_j的关系和V_j的性质易见$\{\phi_j\}_{j=1}^\infty$ 满足单位分解定理中的(i),(ii),(iii). □

【推论】 设$K\subset U\subset\mathbb{R}^n$ 其中K为紧集, U为开集. 则存在$\phi\in C_c^\infty(\mathbb{R}^n)$ 满足

$$0\le\phi\le1\ \text{于}\mathbb{R}^n,\ \mathrm{supp}\phi\subset U,\ \text{且}\ \phi(x)=1\quad\forall\, x\in K.$$

【证】由假设知U^c为闭集且$K\cap U^c=\emptyset$, 因此$\mathrm{dist}(K,U^c)>0$. [注: 当$U^c=\emptyset$ 即当$U=\mathbb{R}^n$时, 定义$\mathrm{dist}(K,\emptyset)=+\infty$.] 取$0<\varepsilon<\mathrm{dist}(K,U^c)$. 则易见

$$K\subset U_\varepsilon:=\{x\in\mathbb{R}^n\mid \mathrm{dist}(x,K)<\varepsilon\}\subset\overline{U}_\varepsilon\subset\{x\in\mathbb{R}^n\mid \mathrm{dist}(x,K)\le\varepsilon\}\subset U.$$

对这一个开集U_ε, 由单位分解定理, 存在**非负**函数列$\{\phi_j\}_{j=1}^\infty$ 具有下列性质:

(i) 对每个$j\in\mathbb{N}$, $\phi_j\in C_c^\infty(\mathbb{R}^n)$ 且$\mathrm{supp}\phi_j\subset U_\varepsilon$.

(ii) $J:=\{j\in\mathbb{N}\mid \exists x\in K \text{ s.t. } \phi_j(x)\ne 0\}$ 是有限集. 因此

$$\text{对任意}j\in\mathbb{N}\setminus J\ \text{有}\quad \phi_j(x)\equiv 0,\qquad x\in K.$$

(iii) $\sum\limits_{j=1}^\infty\phi_j(x)\equiv 1,\quad x\in U_\varepsilon.$

令

$$\phi(x)=\sum_{j\in J}\phi_j(x),\qquad x\in\mathbb{R}^n.$$

则ϕ是有限多个$C^\infty(\mathbb{R}^n)$函数之和因此也属于$C^\infty(\mathbb{R}^n)$. 又当$x\in U_\varepsilon$ 时$0\le\phi(x)\le\sum\limits_{j=1}^\infty\phi_j(x)=1$, 当$x\notin U_\varepsilon$ 时一切$\phi_j(x)=0$ 从而$\phi(x)=0$. 因此$0\le\phi(x)\le 1$ for all $x\in\mathbb{R}^n$. 同时还得到$\{x\in\mathbb{R}^n\mid\phi(x)\ne 0\}\subset U_\varepsilon$ 从而有

$$\mathrm{supp}\phi\subset\overline{U}_\varepsilon\subset U.$$

注意$\overline{U}_\varepsilon$是紧集($\mathbb{R}^n$中的有界闭集), 因此$\mathrm{supp}\phi$是紧集. 所以$\phi\in C_c^\infty(\mathbb{R}^n)$. 最后对任意$x\in K$ 有$\phi(x)=\sum\limits_{j\in J}\phi_j(x)=\sum\limits_{j=1}^\infty\phi_j(x)=1$. 所以这个$\phi$是一个满足要求的函数. □

利用单位分解, 我们能将光滑函数的概念推广到一般点集上的函数上.

【定义】设$E\subset\mathbb{R}^n$, 映射$f:E\to\mathbb{R}^m$ 称为C^r类的($1\le r\le\infty$) 如果每一点$a\in E$ 都有一个邻域$U_a\subset\mathbb{R}^n$ 和一个C^r 类的映射$F_a:U_a\to\mathbb{R}^m$ 使得当$x\in E\cap U_a$ 时$F_a(x)=f(x)$. □

【命题8.34.】设$E\subset\mathbb{R}^n$, 映射$f:E\to\mathbb{R}^m$, $1\le r\le\infty$. 则f在E上是C^r类的当且仅当存在开集$U\supset E$ 和C^r类的映射$F:U\to\mathbb{R}^m$ 使得当$x\in E$ 时$F(x)=f(x)$.

[因此这个命题可以当做C^r类映射的另一个(等价的)定义.]

【证】显然若存在开集$U\supset E$ 和C^r类的映射$F:U\to\mathbb{R}^m$ 使得当$x\in E$ 时$F(x)=f(x)$, 则f在E上是C^r类的.

反之, 设f在E上是C^r类的. 则每一点$a\in E$ 都有一个邻域$U_a\subset\mathbb{R}^n$ 和一个C^r 类的映射$F_a: U_a\to\mathbb{R}^m$ 使得当$x\in E\cap U_a$ 时$F_a(x)=f(x)$. 令$U=\bigcup_{a\in E}U_a$. 设$\{\phi_j\}_{j=1}^{\infty}$是从属于开集族$\{U_a\}_{a\in E}$的一个光滑的单位分解. 则对每个$j\in\mathbb{N}$ 存在$a_j\in E$ 使得$\mathrm{supp}\phi_j\subset U_{a_j}$. 作零延拓: $\phi_j(x)F_{a_j}(x)=0$ 当$x\in\mathbb{R}^n\setminus U_{a_j}$ 时. 则易见延拓后的$\phi_jF_{a_j}$属于$C_c^r(\mathbb{R}^n)$ 且$\mathrm{supp}(\phi_jF_{a_j})\subset\mathrm{supp}\phi_j$. 令

$$F(x)=\sum_{j=1}^{\infty}\phi_j(x)F_{a_j}(x),\quad x\in U.$$

则$F\in C^r(U,\mathbb{R}^m)$. 事实上对任意$x_0\in U$, 存在$\delta>0$ 使得闭球$\overline{B}(x_0,\delta)\subset U$. 由单位分解的性质知$J:=\{j\in\mathbb{N}\mid\exists x\in\overline{B}(x_0,\delta)\text{ s.t. }\phi_j(x)\neq 0\}$ 是有限集. 因此对任意$j\in\mathbb{N}\setminus J$ 有$\phi_j(x)\equiv 0$, $x\in\overline{B}(x_0,\delta)$ 从而有

$$F(x)=\sum_{j\in J}\phi_j(x)F_{a_j}(x),\quad x\in\overline{B}(x_0,\delta).$$

这等式右端是有限多个C^r类函数的和, 因此这蕴含F在开球$B(x_0,\delta)$内属于C^r类. 由$x_0\in U$的任意性知$F\in C^r(U,\mathbb{R}^m)$. 最后对任意$x\in E$, 令$J_x=\{j\in\mathbb{N}\,|\,U_{a_j}\ni x\}$. 则当$j\in\mathbb{N}\setminus J_x$ 时$x\notin U_{a_j}$ 因此$\phi_j(x)=0$. 而当$j\in J_x$时有$x\in U_{a_j}$从而有$F_{a_j}(x)=f(x)$. 于是有

$$\begin{aligned}F(x)&=\sum_{j=1}^{\infty}\phi_j(x)F_{a_j}(x)=\sum_{j\in J_x}^{\infty}\phi_j(x)F_{a_j}(x)=\sum_{j\in J_x}^{\infty}\phi_j(x)f(x)=\Big(\sum_{j\in J_x}\phi_j(x)\Big)f(x)\\&=\Big(\sum_{j=1}^{\infty}\phi_j(x)\Big)f(x)=f(x).\end{aligned}\qquad\square$$

【作业题】

待定.

§8.7 反函数定理与隐函数定理

反函数定理与隐函数定理、不动点定理、Sard 定理[4] 是数学分析中三类最重要的定理(重要工具!).

反函数定理也叫逆映射定理, 主要用于研究同胚或局部同胚映射的存在性和光滑性. 隐函数定理用于研究由方程确定的函数的性质如连续和光滑性质.

这两类定理有很大共性因为从其中一个可以推出另一个. 二者证明的公共技术是压缩映射的不动点定理—— 即不动函数定理。

研究思路如下: 设我们要求解隐函数方程$F(x,y)=0$, 就是说要把y 解出来作为x的函数$y=f(x)$ 使满足恒等式$F(x,f(x))\equiv 0, x\in U$. 例如若取$F(x,y)=g(y)-x$, 则求解方程$F(x,y)=0$就等于求$g$的反函数.

为使此方程有解, 事先需要确定(x_0,y_0) 使得$F(x_0,y_0)=0$. 然后设法求方程在(x_0,y_0)附近的解$y=f(x)$, 它满足$y_0=f(x_0)$. 通常的做法是选取一个可逆矩阵A使得对x_0的邻域$U(x_0)$中的每个x, 映射

$$y\mapsto \Phi(x,y)=y-A^{-1}F(x,y)$$

关于y 在y_0的一个闭邻域$\overline{V}(y_0)$上是压缩的自映射. 如果此事办成, 则由不动点定理, 对每个$x\in U(x_0)$ 存在唯一的y, 记作$y=f(x)$使得$\Phi(x,f(x))=f(x)$, 也即$F(x,f(x))=0$. 由于已有$\Phi(x_0,y_0)=y_0$, 故由唯一性便有$f(x_0)=y_0$. 数学分析表明, 如果在点(x_0,y_0), 映射$F(x,y)$ 关于y的微分也即关于y的Jacobi矩阵$F'_y(x_0,y_0):=D_yF(x_0,y_0)$ 可逆, 则可取$A=F'_y(x_0,y_0)$.

上面指出的不动点, 即不动函数, 可以概括成如下命题, 它是Banach压缩映射不动点定理的一个重要应用.

首先我们定义C^r 类矩阵值的映射.

【定义】 $m,n\in\mathbb{N}, r\in\mathbb{N}\cup\{\infty\}$. 设$\Omega\subset\mathbb{R}^n$ 为开集. 我们称矩阵映射

$$x\mapsto A(x)=(a_{ij}(x))_{m\times n}\in\mathbb{R}^{m\times n},\quad x\in\Omega$$

属于$C^r(\Omega,\mathbb{R}^{m\times n})$类, 如果$A(\cdot)$ 的每个元素$a_{ij}(\cdot)\in C^r(\Omega)$. 注意, 取$m=1$ 或$n=1$ 则以上包含了C^r类向量值映射的定义. □

[4]Sard 定理是: 光滑映射的临界值的集合是一个Lebesgue 零测集, 即若$\Omega\subset\mathbb{R}^n$ 为开集, 映射$f:\Omega\to\mathbb{R}^m$ 属于C^1类, 则f 的临界值的集合是$\mathbb{R}^m$中的Lebesgue零测集.

【**命题8.35(不动函数).**】设$U\subset\mathbb{R}^n$, $V\subset\mathbb{R}^m$分别是$\mathbb{R}^n$和$\mathbb{R}^m$的开集, 映射$\Phi=(\Phi_1,\Phi_2,...,\Phi_m)^\tau:U\times\overline{V}\to\mathbb{R}^m$ 连续且满足

(i) $\Phi:U\times\overline{V}\to V$, 即$\Phi$把$U\times\overline{V}$ 映入V内.

(ii) 映射$y\mapsto\Phi(x,y)$在$\overline{V}$上是一致压缩的, 即存在常数$0\le q<1$ 使得

$$|\Phi(x,y)-\Phi(x,z)|\le q|y-z|\qquad\forall\, y,z\in\overline{V},\quad\forall\, x\in U.$$

则有

(a) 存在唯一的映射$f:U\to V$ 满足$\Phi(x,f(x))=f(x)\ \forall\, x\in U$. 此外有$f\in C(U,V)$.

(b) 若$\Phi\in C^r(U\times V,V)$ $(r\ge 1)$, 则对所有$(x,y)\in U\times V$矩阵$\mathrm{I}-\Phi'_y(x,y)$可逆, 并且(a)中得到的映射f 满足$f\in C^r(U,V)$和

$$f'(x)=[\mathrm{I}-\Phi'_y(x,y)]^{-1}\Phi'_x(x,y)\Big|_{y=f(x)}=[\mathrm{I}-\Phi'_y(x,f(x))]^{-1}\Phi'_x(x,f(x)),\quad x\in U.\ (*)$$

这里$\mathrm{I}\in\mathbb{R}^{m\times m}$是单位矩阵, $\Phi'_x(x,y),\Phi'_y(x,y)$ 分别是$\Phi(x,y)$关于x 和y的微分(Jacobi矩阵), 即按矩阵分块对应有

$$\begin{pmatrix}\Phi'_x(x,y) & \Phi'_y(x,y)\end{pmatrix}=D\Phi(x,y)=\begin{pmatrix}\frac{\partial\Phi_1}{\partial x_1} & \frac{\partial\Phi_1}{\partial x_2} & \cdots & \frac{\partial\Phi_1}{\partial x_n} & \frac{\partial\Phi_1}{\partial y_1} & \frac{\partial\Phi_1}{\partial y_2} & \cdots & \frac{\partial\Phi_1}{\partial y_m}\\ \frac{\partial\Phi_2}{\partial x_1} & \frac{\partial\Phi_2}{\partial x_2} & \cdots & \frac{\partial\Phi_2}{\partial x_n} & \frac{\partial\Phi_2}{\partial y_1} & \frac{\partial\Phi_2}{\partial y_2} & \cdots & \frac{\partial\Phi_2}{\partial y_m}\\ \cdots\cdots & & & & \cdots\cdots & & & \\ \frac{\partial\Phi_m}{\partial x_1} & \frac{\partial\Phi_m}{\partial x_2} & \cdots & \frac{\partial\Phi_m}{\partial x_n} & \frac{\partial\Phi_m}{\partial y_1} & \frac{\partial\Phi_m}{\partial y_2} & \cdots & \frac{\partial\Phi_m}{\partial y_m}\end{pmatrix}.$$

【证】(a): 对任意$x\in U$, 映射$y\mapsto\Phi(x,y)$ 是闭集$\overline{V}$上的压缩自映射, 据压缩映射不动点定理, 在$\overline{V}$中存在唯一的一点y, 记作$y=f(x)$, 使得$\Phi(x,f(x))=f(x)$. 又因当$y\in\overline{V}$时$\Phi(x,y)\in V$, 故$f(x)\in V$. 下面的$f:U\to V$表示这样得到的唯一映射.

来证明$f\in C(U,V)$. 对任意$x_0\in U$, 存在$\delta>0$使得$B(x_0,\delta)\subset U$ (因U 是开集). 由$\Phi(\cdot,f(\cdot))=f(\cdot)$ 有: 当$|h|<\delta$时

$$|f(x_0+h)-f(x_0)|=|\Phi(x_0+h,f(x_0+h))-\Phi(x_0,f(x_0))|$$
$$\le|\Phi(x_0+h,f(x_0+h))-\Phi(x_0+h,f(x_0))|+|\Phi(x_0+h,f(x_0))-\Phi(x_0,f(x_0))|$$
$$\le q|f(x_0+h)-f(x_0)|+|\Phi(x_0+h,f(x_0))-\Phi(x_0,f(x_0))|$$

$\Longrightarrow$

$$|f(x_0+h)-f(x_0)| \le \frac{1}{1-q}|\Phi(x_0+h, f(x_0)) - \Phi(x_0, f(x_0)|, \quad |h|<\delta. \qquad (**)$$

据Φ 的连续性即知f在任一点$x_0 \in U$ 连续.

(b): 设$\Phi \in C^r(U\times V, V)$ $(r\ge 1)$. 因C^r类映射是可微的, 故Φ在$U\times V$上处处可微. 先证明方阵$\mathrm{I}-\Phi_y'(x,y)$在$U\times V$上处处可逆. 取定任意$(x,y)\in U\times V$. 因V 是开集, 故存在$\delta>0$ 使得$B(y,\delta)\subset V$. 对任意$k\in\mathbb{R}^m\setminus\{0\}$, 当$0<|t|<<1$ 时有$y+tk\in B(y,\delta)$ 从而由Φ 可微有

$$\Phi(x,y+tk)-\Phi(x,y) = \Phi_y'(x,y)tk + o(|tk|).$$

另一方面由Φ 关于y 的压缩性有

$$|\Phi(x,y+tk)-\Phi(x,y)| \le q|t||k|.$$

因此

$$|\Phi_y'(x,y)tk| \le q|t||k| + o(|t||k|)$$

从而有

$$|\Phi_y'(x,y)k| \le q|k| + \frac{o(|t||k|)}{|t|} \to q|k| \quad \text{as} \quad t\to 0.$$

这证明了

$$|\Phi_y'(x,y)k| \le q|k| \qquad \forall\, k\in\mathbb{R}^m\setminus\{0\}.$$

使用矩阵的算子范数我们得到

$$\|\Phi_y'(x,y)\| = \sup_{k\in\mathbb{R}^m\setminus\{0\}} \frac{|\Phi_y'(x,y)k|}{|k|} \le q < 1.$$

由此即知$\mathrm{I}-\Phi_y'(x,y)$可逆(也见§7.1的作业题).

其次证明f处处可微且关系式$(*)$成立. 任取$x_0\in U$. 取$\delta>0$ 充分小使得闭球$\overline{B}(x_0,\delta)\subset U$. 设$h\in\mathbb{R}^n$ 且$|h|<\delta$. 记

$$y_0 = f(x_0), \quad k = f(x_0+h)-f(x_0) = f(x_0+h)-y_0.$$

则$f(x_0+h) = y_0+k$. 由Φ 可微有

$$k = f(x_0+h)-f(x_0) = \Phi(x_0+h, y_0+k) - \Phi(x_0,y_0)$$

$$
\begin{aligned}
&= (D\Phi)(x_0, y_0)\begin{pmatrix} h \\ k \end{pmatrix} + o(\sqrt{|h|^2+|k|^2}) \\
&= \Big(\Phi'_x(x_0,y_0) \quad \Phi'_y(x_0,y_0)\Big)\begin{pmatrix} h \\ k \end{pmatrix} + o(\sqrt{|h|^2+|k|^2}) \\
&= \Phi'_x(x_0,y_0)h + \Phi'_y(x_0,y_0)k + o(\sqrt{|h|^2+|k|^2})
\end{aligned}
$$

$\Longrightarrow$

$$\Big(\mathrm{I} - \Phi'_y(x_0,y_0)\Big)k = \Phi'_x(x_0,y_0)h + o(\sqrt{|h|^2+|k|^2}).$$

因$k = f(x_0+h) - f(x_0)$, 故这给出

$$f(x_0+h) - f(x_0) = [\mathrm{I} - \Phi'_y(x_0,y_0)]^{-1}\Phi'_x(x_0,y_0)h + o(\sqrt{|h|^2+|k|^2}).$$

这里用到

$$[\mathrm{I} - \Phi'_y(x_0,y_0)]^{-1}o(\sqrt{|h|^2+|k|^2}) = o(\sqrt{|h|^2+|k|^2})$$

这是因为$[\mathrm{I} - \Phi'_y(x_0,y_0)]^{-1}$ 是一个固定的矩阵且当$h \to 0$ 时$k = f(x_0+h) - f(x_0) \to 0$. 为证明$f$在$x_0$ 可微, 我们必须证明

$$o(\sqrt{|h|^2+|k|^2}) = o(|h|) \qquad (|h| \to 0).$$

为此只需证明

$$|k| \le C|h|$$

其中C 是常数. 由$\Phi \in C^r \subset C^1$ 可知$(x,y) \mapsto \Phi'_x(x,y)$ 按矩阵范数连续. 因此由闭球$\overline{B}(x_0,\delta) \subset U$ 知

$$M := \max_{x\in\overline{B}(x_0;\delta)} \|\Phi'_x(x,y_0)\| \quad \text{存在(因而有限)}.$$

这里$\|\cdot\|$ 是矩阵范数. (注意同一矩阵空间上的矩阵范数都是互相等价的, 所以例如你可以取$\|\cdot\|$为欧几里德范数. 此外, 一个矩阵映射$x \mapsto A(x) = (a_{ij}(x))_{m\times n}$按矩阵范数$\|\cdot\|$ 连续当且仅当该矩阵的每个元素$x \mapsto a_{ij}(x)$都连续. etc.)

对映射$x \mapsto \Phi(x,y_0)$ 在开凸集$B(x_0,\delta)$ 上应用微分中值不等式有

$$|\Phi(x_0+h,y_0) - \Phi(x_0,y_0| \le \|\Phi'_x(x_0+\theta h, y_0\||h| \le M|h|, \quad |h| < \delta.$$

然后应用(a)中的估计式$(**)$ (注意$y_0 = f(x_0)$) 即知当$|h| < \delta$时

$$|k| = |f(x_0+h) - f(x_0)| \le \frac{1}{1-q}|\Phi(x_0+h,y_0) - \Phi(x_0,y_0)| \le \frac{M}{1-q}|h| =: C|h|.$$

这就证明了

$$f(x_0+h)-f(x_0)=[\mathrm{I}-\Phi'_y(x_0,y_0)]^{-1}\Phi'_x(x_0,y_0)h+o(|h|).$$

所以f在点x_0 可微且

$$f'(x_0)=[\mathrm{I}-\Phi'_y(x_0,y_0)]^{-1}\Phi'_x(x_0,y_0)\Big|_{y_0=f(x_0)}.$$

再由$x_0\in U$ 的任意性即知f在U上处处可微且公式$(*)$ 成立, 即

$$f'(x)=[\mathrm{I}-\Phi'_y(x,f(x))]^{-1}\Phi'_x(x,f(x)),\quad x\in U. \tag{$*$}$$

最后证明$f\in C^r(U,V)$. 为此我们令

$$A(x,y)=\mathrm{I}-\Phi'_y(x,y),\quad B(x,y)=\Phi'_x(x,y).$$

回忆逆矩阵的表达式:

$$[A(x,y)]^{-1}=\frac{1}{\det A(x,y)}A^*(x,y)=\Big(\frac{A_{ij}(x,y)}{\det A(x,y)}\Big)_{m\times m}$$

其中$A^*(x,y)=(A_{ij}(x,y))_{m\times m}$ 是$A(x,y)$的伴随矩阵. 因$\det A(x,y)$ 和$A_{ij}(x,y)$都是$A(x,y)$ 的元素的有限次乘积的线性组合, 故它们与$A(x,y)$ 的元素有相同的光滑性即都属于C^{r-1} 类(当$r\in\mathbb{N}$ 时) 或属于C^∞类(当$r=\infty$时). 又因在$U\times V$ 上$\det A(x,y)\neq 0$, 故每个元素$\frac{A_{ij}(x,y)}{\det A(x,y)}$ 也都属于C^{r-1}类或属于C^∞类. 所以在$U\times V$ 上$(x,y)\mapsto[A(x,y)]^{-1}$ 属于C^{r-1}类或属于C^∞类. 因此在$U\times V$ 上$(x,y)\mapsto[A(x,y)]^{-1}B(x,y)$ 属于C^{r-1}类或属于C^∞类.

现在由

$$f'(x)=[A(x,f(x))]^{-1}B(x,f(x)),\quad x\in U \tag{$*$}$$

和$f(x)$在U 上连续可知$f'(x)$在U 上连续, 也即f 的每个分量都属于$C^1(U,\mathbb{R})$. 于是再由关系式$(*)$ 可知$f'(x)$的每个元素都属于$C^1(U,\mathbb{R})$ 也即f 的每个分量都属于$C^2(U,\mathbb{R})$. 如此操作, 归纳地可导出f的每个坐标函数都属于$C^r(U,\mathbb{R})$, 也即$f\in C^r(U,\mathbb{R}^m)$. 但$f(U)\subset V$, 故$f\in C^r(U,V)$. $\Box$

公式的记忆—只需知道复合映射微分法: 对恒等式$f(x)\equiv\Phi(x,f(x))$ 两边取微分, 应用复合映射的微分法有

$$f'(x)=\Phi'_x(x,y)\Big|_{y=f(x)}+\Phi'_y(x,y)\Big|_{y=f(x)}f'(x)=\Phi'_x(x,f(x))+\Phi'_y(x,f(x))f'(x)$$

因此
$$\Big(\mathrm{I}-\Phi_y'(x,f(x))\Big)f'(x)=\Phi_x'(x,f(x)),\quad x\in U.$$
假如还知道$\mathrm{I}-\Phi_y'(x,f(x))$ 可逆, 便得到
$$f'(x)=[\mathrm{I}-\Phi_y'(x,f(x))]^{-1}\Phi_x'(x,f(x)),\quad x\in U.$$

【反函数定理】 在数学上, “反函数”和“逆映射”是一个意思.

先介绍微分同胚.

【定义(C^r 类微分同胚)】设U,V 均为$\mathbb{R}^n$的开集(非空), $r\ge 1$, 映射$f:U\to V$是一个双射并且$f\in C^r(U,V)$ 且$f^{-1}\in C^r(V,U)$, 则称$f:U\to V$ 是一个C^r类的微分同胚. 若两个开集U,V 之间存在一个C^r 类微分同胚映射, 则称U,V是C^r 类微分同胚的或称U与V为C^r 微分同胚. □

【注】这个定义中只考虑了U,V 是同维数欧空间$\mathbb{R}^n$中的开集, 这是因为以后学习拓扑学会知道: 若U,V 分别是欧空间$\mathbb{R}^m$ 和$\mathbb{R}^n$ 中的开集且$m\ne n$, 则U,V 一定不同胚, 更谈不上微分同胚. 也见下面**§8.9.映射的秩与函数相关** 中的**例**, 那里2014级王志涵同学对于“不微分同胚”给了一很自然的证明. □

在进入反函数定理之前我们顺便指出C^r 类微分同胚的一个基本性质：**$\mathbb{R}^n$ 中开集之间的C^r 类微分同胚是$\mathbb{R}^n$ 的开集类上的一个等价关系.**

事实上, 对任意开集$U\subset\mathbb{R}^n$, 显然恒等映射$x\mapsto x$ 是U 与自身的C^r 类微分同胚.

设$f:U\to V$ 为C^r 类微分同胚, 则由定义知$f^{-1}:V\to U$ 也是C^r 类微分同胚.

设U 与V为C^r微分同胚, 且V与W为C^r微分同胚. 则由定义, 存在C^r类微分同胚$f:U\to V, g:V\to W$. 据复合映射微分法可知$g\circ f:U\to W$是C^r类微分同胚. 因此U与W 为C^r类微分同胚.

比同胚映射弱一点但也很重要的一种映射是开映射：

【定义(开映射)】设$\Omega\subset\mathbb{R}^n$ 为开集, 映射$f:\Omega\to\mathbb{R}^n$ 把Ω 中的开集映$\mathbb{R}^n$中的为开集, 则称f 是一个开映射.

我们将在第十二章中证明一个非常重要且好用的定理(它不要求可微）——

【<u>**Brouwer区域不变定理**</u>】

若$\Omega \subset \mathbb{R}^n$ 为开集, $f : \Omega \to \mathbb{R}^n$ 连续且局部一对一, 则f 是开映射.

[局部一对一是指: 对任意$x \in \Omega$ 存在$\delta > 0$ 使得$B(x, \delta) \subset \Omega$ 且f 在$B(x, \delta)$ 上是单射.]

$\Box$

眼下我们可以证明一个关于可微映射的开映射定理, 它在证明反函数定理时被用到.

【**命题8.36(可微映射的开映射定理).**】设$\Omega \subset \mathbb{R}^n$ 为开集, $f : \Omega \to \mathbb{R}^n$ 可微且$f'(x)$ 在Ω上处处可逆, 即$\det f'(x) \neq 0$ for all $x \in \Omega$. 则f 是开映射, 即对任意开集$U \subset \Omega$, $f(U)$ 是$\mathbb{R}^n$中的开集.

注意: 因定义域Ω 和值域$f(\Omega)$都含于$\mathbb{R}^n$, 故Jacobi矩阵$f'(x)$ 是方阵!

【证】设开集$U \subset \Omega$, 来证明$f(U)$ 是开集. 任取$y_0 \in f(U)$, 须证明存在$\varepsilon > 0$ 使得$B(y_0, \varepsilon) \subset f(U)$. 写$y_0 = f(x_0), x_0 \in U$. 令$A = f'(x_0)$.

由f 可微知$f(x) = f(x_0) + A(x - x_0) + o(|x - x_0|)$ 即

$$\lim_{U \ni x \to x_0} \frac{|f(x) - f(x_0) - A(x - x_0)|}{|x - x_0|} = 0.$$

据此有: 存在$\delta > 0$ 使得闭球$\overline{B}(x_0, \delta) \subset U$ 且

$$\frac{|f(x) - f(x_0) - A(x - x_0)|}{|x - x_0|} < \frac{1}{2\|A^{-1}\|} \qquad \forall\, x \in \overline{B}(x_0, \delta) \setminus \{x_0\}.$$

即

$$|f(x) - f(x_0) - A(x - x_0)| < \frac{|x - x_0|}{2\|A^{-1}\|} \qquad \forall\, x \in \overline{B}(x_0, \delta) \setminus \{x_0\}.$$

这里$\|\cdot\|$为矩阵范数. 由此有

$$|f(x) - f(x_0)| \geq |A(x - x_0)| - |f(x) - f(x_0) - A(x - x_0)|$$

$$> |A(x - x_0)| - \frac{|x - x_0|}{2\|A^{-1}\|} \geq \frac{|x - x_0|}{2\|A^{-1}\|} = \frac{\delta}{2\|A^{-1}\|} \quad \forall\, x \in \partial B(x_0, \delta) \qquad (*)$$

其中用到不等式

$$|x - x_0| = |A^{-1}A(x - x_0)| \leq \|A^{-1}\||A(x - x_0)|.$$

令

$$\varepsilon = \frac{1}{2} \min_{x \in \partial B(x_0, \delta)} |f(x) - f(x_0)|.$$

则由$(*)$ 知

$$\varepsilon \geq \frac{\delta}{4\|A^{-1}\|} > 0.$$

下证$B(f(x_0),\varepsilon) \subset f(B(x_0,\delta))$. 任取$y \in B(f(x_0),\varepsilon)$. 我们有: 对任意$x \in \partial B(x_0,\delta)$

$$|f(x)-y| \geq |f(x)-f(x_0)| - |f(x_0)-y| \geq 2\varepsilon - |f(x_0)-y| > |f(x_0)-y|. \qquad (**)$$

设$x_* \in \overline{B}(x_0,\delta)$ 是函数$x \mapsto |f(x)-y|^2$ 在闭球$\overline{B}(x_0,\delta)$上的最小值点, 即

$$|f(x_*)-y|^2 = \min_{x\in\overline{B}(x_0,\delta)} |f(x)-y|^2$$

则由严格不等式$(**)$知$x_* \notin \partial B(x_0,\delta)$. 因此$x_* \in B(x_0,\delta)$. 因$B(x_0,\delta)$ 是开集, 故存在$\eta > 0$ 使得$B(x_*,\eta) \subset B(x_0,\delta)$. 于是有$|f(x)-y|^2 \geq |f(x_*)-y|^2$ for all $x \in B(x_*,\eta)$. 这表明点$x_* \in B(x_0,\delta)$ 是函数$x \mapsto |f(x)-y|^2$ 在开球$B(x_0,\delta)$ 内的一个极小值点. 据极值原理知

$$\nabla(|f(x)-y|^2)\Big|_{x=x_*} = 0.$$

在梯度与Jacobian的关系式

$$\nabla(|g(x)|^2) = 2g(x)^{\tau} g'(x)$$

中取$g(x) = f(x) - y$ 得到

$$0 = \nabla(|f(x)-y|^2)\Big|_{x=x_*} = 2(f(x_*)-y)^{\tau} f'(x_*).$$

而由假设知$f'(x_*)$ 可逆, 故得$2(f(x_*)-y)^{\tau} = 0$ 即$f(x_*) - y = 0$. 所以$y = f(x_*) \in f(B(x_0,\delta))$. 这证明了$B(y_0,\varepsilon) = B(f(x_0),\varepsilon) \subset f(B(x_0,\delta)) \subset f(U)$. 由$y_0 \in f(U)$的任意性即知$f(U)$ 是开集. $\square$

区域: 数学教科书和文献中常出现“区域(region, domain)”. 区域是指连通开集. 例如“$G \subset \mathbb{R}^n$ 为一个区域”等于“$G \subset \mathbb{R}^n$ 为$\mathbb{R}^n$中的一个连通开集.” “闭区域”是指“区域的闭包”.

一点的邻域 今后我们总假定: 一点x_0的一个邻域U就是包含点x_0的一个连通开集.

[因为开集可以被分解成互不相交的连通开集的并, 故这个假定不失一般性.]

【定理8.37(局部反函数定理).】 设$\Omega \subset \mathbb{R}^n$ 为开集, 映射$f : \Omega \to \mathbb{R}^n$ 属于C^r类, $r \geq 1$. 设$x_0 \in \Omega$ 满足$f'(x_0)$可逆, 即

$$\det f'(x_0) \neq 0.$$

则存在x_0的一个邻域U 和$y_0 = f(x_0)$的一个邻域V 使得$f : U \to V$ 是C^r 类微分同胚. 此外有

$$(f^{-1})'(y) = [f'(x)]^{-1}\Big|_{x=f^{-1}(y)}, \quad y \in V.$$

使用Jacobi矩阵, 上式也可写成

$$J_{f^{-1}}(y) = [J_f(x)]^{-1}\Big|_{x=f^{-1}(y)}, \quad y \in V.$$

【证】我们将使用**命题8.35(不动函数)**. 为此我们暂时将字母x, y 对调, 即考虑$\Omega \ni y \mapsto f(y)$, $x_0 = f(y_0)$, 和$\det f'(y_0) \ne 0$. 令

$$\begin{aligned} &A = f'(y_0), \\ &\Phi(x, y) = y - A^{-1}(f(y) - x), \quad (x, y) \in \mathbb{R}^n \times \Omega. \end{aligned}$$

则$\Phi \in C^r(\mathbb{R}^n \times \Omega, \mathbb{R}^n)$ 且

$$\Phi'_y(x, y) = \mathrm{I} - A^{-1}f'(y) = A^{-1}(f'(y_0) - f'(y)), \quad (x, y) \in \mathbb{R}^n \times \Omega.$$

因f 是光滑映射, 故存在$\eta > 0$ 使得$\overline{B}(y_0, \eta) \subset \Omega$ 且按矩阵范数有

$$\|\Phi'_y(x, y)\| \le \|A^{-1}\|\|f'(y_0) - f'(y)\| \le \frac{1}{2} \quad \forall y \in \overline{B}(y_0, \eta),\, x \in \mathbb{R}^n.$$

由$x_0 = f(y_0)$ 有

$$\Phi(x, y) - y_0 = -A^{-1}(\, f(y) - f(y_0) - A(y - y_0)\,) - A^{-1}(x - x_0),$$

$$|\Phi(x, y) - y_0| \le \|A^{-1}\||f(y) - f(y_0) - A(y - y_0)| + \|A^{-1}\||x - x_0|.$$

由可微性知$\|A^{-1}\||f(y)-f(y_0)-A(y-y_0)|$ 是$|y-y_0|$的高阶无穷小. 因此存在$0 < \varepsilon \le \eta$ 使得

$$\|A^{-1}\||f(y) - f(y_0) - A(y - y_0)| \le \frac{1}{3}|y - y_0| \le \frac{\varepsilon}{3} \quad \forall\, y \in \overline{B}(y_0, \varepsilon).$$

令

$$\delta = \frac{\varepsilon}{2\|A^{-1}\|}.$$

则当$(x, y) \in \overline{B}(x_0, \delta) \times \overline{B}(y_0, \varepsilon)$时

$$|\Phi(x, y) - y_0| \le \frac{\varepsilon}{3} + \|A^{-1}\|\delta = \frac{5}{6}\varepsilon < \varepsilon.$$

这表明$\Phi:\overline{B}(x_0,\delta)\times\overline{B}(y_0,\varepsilon)\to B(y_0,\varepsilon)$, 即$\Phi$ 把$\overline{B}(x_0,\delta)\times\overline{B}(y_0,\varepsilon)$ 映入$B(y_0,\varepsilon)$ 内. 又由微分中值不等式有 (注意闭球$\overline{B}(y_0,\varepsilon)$是含于开集$\Omega$内的凸集)

$$|\Phi(x,y)-\Phi(x,z)|\le\|\Phi'_y(x,\xi)\||y-z|\le\frac{1}{2}|y-z|\quad\forall\,y,z\in\overline{B}(y_0,\varepsilon),\ x\in\overline{B}(x_0,\delta)$$

因此映射$y\mapsto\Phi(x,y)$在$\overline{B}(y_0,\varepsilon)$上是一致压缩的. 于是这个$C^r$类映射$\Phi$满足**命题8.35(不动函数)** 的全部条件. 因此存在唯一的映射$g\in C^r(B(x_0,\delta),B(y_0,\varepsilon))$ 使得

$$\Phi(x,g(x))\equiv g(x),\quad x\in B(x_0,\delta);\qquad g(x_0)=y_0$$

即

$$f(g(x))\equiv x,\quad x\in B(x_0,\delta);\qquad g(x_0)=y_0.$$

令$V=B(x_0,\delta),U=g(B(x_0,\delta))=g(V)$. 则$f:U\to V$ 为双射, 其逆映射为

$$f^{-1}=g:V\to U,\quad f^{-1}(x)=g(x).$$

同时由复合映射微分法知

$$\mathrm{I}=f'(g(x))g'(x),\quad x\in V.$$

因此$f'(g(x)),g'(x)$ 都可逆且

$$g'(x)=[f'(y)]^{-1}\Big|_{y=g(x)}\qquad\forall\,x\in V.$$

据开映射定理(**命题8.36(可微映射的开映射定理)**)和连续映射把连通集映为连通集可知$U=g(V)$ 是连通开集. 显然$y_0\in U$. 最后再将字母x,y对调回来, 将x_0,y_0对调回来. 则有$V=B(y_0,\delta)$ 是y_0的邻域, U 是x_0的邻域. □

反函数(逆映射)微分法的记忆: 只需记住复合映射的微分法.

从恒等式

$$y\equiv f(f^{-1}(y)),\quad x\equiv f^{-1}(f(x))$$

出发: 由恒同映射$x\mapsto x$ 的微分等于单位矩阵I 以及复合映射的微分法可知: 若$f(x)$在点x 可微**并且**其逆映射$f^{-1}(y)$ 在点$y=f(x)$ 可微, 则有

$$\mathrm{I}=f'(x)(f^{-1})'(y)\Big|_{x=f^{-1}(y)},\quad \mathrm{I}=(f^{-1})'(y)f'(x)\Big|_{y=f(x)}.$$

再注意这些微分(Jacobi) 都是方阵, 故得知它们都可逆从而有

$$(f^{-1})'(y) = [f'(x)]^{-1}\Big|_{x=f^{-1}(y)}, \quad f'(x) = [(f^{-1})'(y)]^{-1}\Big|_{y=f(x)}.$$

整体反函数定理.

【命题8.38.】 设$\Omega \subset \mathbb{R}^n$ 为开集, 映射$f:\Omega \to \mathbb{R}^n$ 属于C^r类, $r \geq 1$. 假设f 满足

(i) $f:\Omega \to \mathbb{R}^n$ 是单射 (这是应用中主要验证部分).

(ii) $\det f'(x) \neq 0 \quad \forall\, x \in \Omega$.

则$f(\Omega)$是$\mathbb{R}^n$的开集且$f:\Omega \to f(\Omega)$ 是C^r类微分同胚.

【证】条件(i)说明逆映射$f^{-1}: f(\Omega) \to \Omega$ 存在, 条件(ii)和开映射定理保证$f(\Omega)$ 是开集！下面证明$f^{-1} \in C^r(f(\Omega), \Omega)$. 为此我们只需证明对任意$y_0 \in f(\Omega)$ 存在$\delta > 0$ 使得$B(y_0;\delta) \subset f(\Omega)$ 且f^{-1} 在$B(y_0,\delta)$内属于C^r类. 如果此事成立, 则说明f^{-1}的每个坐标函数在$B(y_0,\delta)$内有直到r 阶的连续偏导数. 再由y_0的任意性知f^{-1}的每个坐标函数在$f(\Omega)$内有直到r阶的连续偏导数 [强调: 偏导数是在每一点定义的, 偏导数的连续性也是在每一点定义的], 此即f^{-1}在$f(\Omega)$ 内属于C^r类.

任取$y_0 \in f(\Omega)$. 写$y_0 = f(x_0), x_0 \in \Omega$. 由局部反函数定理, 存在$x_0$的一个邻域$U(x_0) \subset \Omega$ 和y_0的一个邻域$V(y_0) = B(y_0,\delta) \subset f(\Omega)$ 使得$f: U(x_0) \to V(y_0)$ 是C^r类微分同胚, 即$(f|_{U(x_0)})^{-1}: V(y_0) \to U(x_0)$属于$C^r$类. **但$f:\Omega \to f(\Omega)$ 是单射, 也即整体逆映射$f^{-1}: f(\Omega) \to \Omega$ 已经存在**, 故必有

$$f^{-1}(y) = (f|_{U(x_0)})^{-1}(y) \quad \forall\, y \in V(y_0) \qquad \text{i.e.}\quad f^{-1}\Big|_{V(y_0)} = (f|_{U(x_0)})^{-1}.$$

所以f^{-1} 在$V(y_0) = B(y_0,\delta)$内属于C^r类. 如上分析, 这证明了f^{-1}在$f(\Omega)$ 内属于C^r类. □

【注】与一元函数不同, 当$n \geq 2$时, 上述定理中的第(ii)条“$\det f'(x) \neq 0 \quad \forall\, x \in \Omega$”一般不能保证“$f:\Omega \to \mathbb{R}^n$ 是单射.”也即定理中的第(i)条一般不能去掉.

例如对$n=2$的情形, 设$\Omega = \mathbb{R}^2$, $f(x,y) = (e^x \cos y, e^x \sin y)$. 则$f:\mathbb{R}^2 \to \mathbb{R}^2$ 是光滑映射且

$$\det f'(x,y) = e^x \neq 0 \qquad \forall\,(x,y) \in \mathbb{R}^2.$$

但显然f 不是单射: $f(x,y+2\pi)\equiv f(x,y)$.

当然, 若将f的定义域缩小为$\mathbb{R}\times(0,2\pi)$, 则f 在$\mathbb{R}\times(0,2\pi)$ 上是单射. 对这类例子的一般分析见下面.

【例 (极坐标和球坐标变换)】设$n\ge 2$, 映射$\Psi:\mathbb{R}^n\to\mathbb{R}^n$

$$(r,\theta_1,\theta_2,...,\theta_{n-1})\mapsto x=(x_1,x_2,...,x_n)=\Psi(r,\theta_1,\theta_2,...,\theta_{n-1})$$

由下式给出

$$\begin{aligned}
&x_1=r\cos\theta_1\,,\\
&x_2=r\sin\theta_1\cos\theta_2\,,\\
&x_3=r\sin\theta_1\sin\theta_2\cos\theta_3\,,\\
&\cdots\cdots\\
&x_{n-1}=r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\cos\theta_{n-1}\,,\\
&x_n\ =\ r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\sin\theta_{n-1}\,.
\end{aligned}$$

显然 $\Psi\in C^\infty(\mathbb{R}^n,\mathbb{R}^n)$. 注意以上前 $n-1$ 行的最后一个都因子是余弦函数, 第 n 行的最后一个因子是正弦函数. 当 $n=2$ 时, 只有 θ_1 出现. **球极坐标变换定义为映射 Ψ 在 $[0,+\infty)\times[0,\pi]^{n-2}\times[0,2\pi]$ 上的限制.**

常用情形是极坐标变换 $(n=2)$ 和球坐标变换 $(n=3)$:

$n=2$:

$$\begin{aligned}
&x=r\cos\phi,\\
&y=r\sin\phi,\quad r\ge 0,\ \ \phi\in[0,2\pi].
\end{aligned}$$

$n=3$:

$$\begin{aligned}
&x=r\cos\theta,\\
&y=r\sin\theta\cos\phi,\\
&z=r\sin\theta\sin\phi,\quad r\ge 0,\ \ \theta\in[0,\pi],\ \phi\in[0,2\pi]\,.
\end{aligned}$$

【命题8.39(球极坐标变换).】球极坐标变换 $\Psi:[0,+\infty)\times[0,\pi]^{n-2}\times[0,2\pi]\to\mathbb{R}^n$ 是满射, 而 Ψ 在开区间 $(0,+\infty)\times(0,\pi)^{n-2}\times(0,2\pi)$ 上是单射并且

$$\det\Psi'(r,\theta_1,\theta_2,...,\theta_{n-1})=r^{n-1}\sin^{n-2}\theta_1\sin^{n-3}\theta_2\cdots\sin\theta_{n-2}.$$

【证】记 $\mathbb{R}^n$ 中的单位球面为

$$\mathbb{S}^{n-1}=\{\omega\in\mathbb{R}^n \mid |\omega|=1\}.$$

任一 $x\in\mathbb{R}^n$ 可表示为 $x=r\omega$, $r=|x|, \omega\in\mathbb{S}^{n-1}$, 且当 $x\neq 0$ 时表示是唯一的：$\omega=\frac{x}{|x|}$. 令

$$\Phi(\theta_1,\theta_2,...,\theta_{n-1}):=\Psi(1,\theta_1,\theta_2,...,\theta_{n-1}).$$

则

$$\Psi(r,\theta_1,\theta_2,...,\theta_{n-1})=r\Phi(\theta_1,\theta_2,...,\theta_{n-1}).$$

因此为证 Ψ 的上述性质, 只需证明 $\Phi:[0,\pi]^{n-2}\times[0,2\pi]\to\mathbb{S}^{n-1}$ 且是满射并且 Φ 在开区间 $(0,\pi)^{n-2}\times(0,2\pi)$ 上是单射.

我们对维数 n 用归纳法. 记 $\Phi_{n-1}=\Phi$. 当 $n=2$ 时, 这是单位圆周 $\mathbb{S}^1$ 表示的已知事实. 假设对于 $n-1\,(\geq 2)$ 所证成立, 来证明对于 n 也成立. 对任意 $(\theta_1,\theta_2,...,\theta_{n-1})\in[0,\pi]^{n-2}\times[0,2\pi]$ 由归纳假设有 $\Phi_{n-2}(\theta_2,...,\theta_{n-1})\in\mathbb{S}^{n-2}$. 因此利用递归关系

$$\Phi_{n-1}(\theta_1,\theta_2,...,\theta_{n-1})=(\cos\theta_1\,,\sin\theta_1\Phi_{n-2}(\theta_2,...,\theta_{n-1}))$$

得到

$$|\Phi_{n-1}(\theta_1,\theta_2,...,\theta_{n-1})|=\sqrt{\cos^2\theta_1+\sin^2\theta_1|\Phi_{n-2}(\theta_2,...,\theta_{n-1})|^2}=1$$

所以 Φ_{n-1} 把 $[0,\pi]^{n-2}\times[0,2\pi]$ 映入 $\mathbb{S}^{n-1}$. 其次任取 $\omega=(\omega_1,\omega_2,...,\omega_n)\in\mathbb{S}^{n-1}$. 若 $|\omega_1|=1$, 则 $\omega_j=0, j=2,3,...,n$. 此时当 $\omega_1=1$ 时, $\omega=(1,0,...,0)=\Phi_{n-1}(0,0,....,0)$; 当 $\omega_1=-1$ 时, $\omega=(-1,0,...,0)=\Phi_{n-1}(\pi,\pi,....,\pi)$. 假设 $|\omega_1|<1$. 则存在 $\theta_1\in(0,\pi)$ 使得 $\omega_1=\cos\theta_1$, 同时由 $|\omega_2|^2+|\omega_3|^2+\cdots+|\omega_n|^2=1-|\omega_1|^2=\sin^2\theta_1$ 知

$$\left(\frac{\omega_2}{\sin\theta_1},\frac{\omega_3}{\sin\theta_1},...,\frac{\omega_n}{\sin\theta_1}\right)\in\mathbb{S}^{n-2}.$$

因此由归纳假设存在 $(\theta_2,...,\theta_{n-1})\in[0,\pi]^{n-3}\times[0,2\pi]$ 使得

$$\left(\frac{\omega_2}{\sin\theta_1},\frac{\omega_3}{\sin\theta_1},...,\frac{\omega_n}{\sin\theta_1}\right)=\Phi_{n-2}(\theta_2,...,\theta_{n-1})$$

于是

$$\omega=(\cos\theta_1,\,\sin\theta_1\Phi_{n-2}(\theta_2,...,\theta_{n-1}))=\Phi_{n-1}(\theta_1,\theta_2,...,\theta_{n-1}).$$

这证明了 Φ_{n-1} 是满射.

为证明 Φ_{n-1} 在 $(0,\pi)^{n-2}\times(0,2\pi)$ 上是单射, 设 $(\theta_1,\theta_2,...,\theta_{n-1}),(\tilde\theta_1,\tilde\theta_2,...,\tilde\theta_{n-1})\in(0,\pi)^{n-2}\times(0,2\pi)$ 使得

$$\Phi_{n-1}(\theta_1,\theta_2,...,\theta_{n-1})=\Phi_{n-1}(\tilde\theta_1,\tilde\theta_2,...,\tilde\theta_{n-1})\,.$$

由 $0<\theta_1,\tilde\theta_1<\pi$ 和 $\cos\theta_1=\cos\tilde\theta_1\Longrightarrow\theta_1=\tilde\theta_1$ 且 $\sin\theta_1>0$. 消去 $\sin\theta_1$ 得到

$$\Phi_{n-2}(\theta_2,...,\theta_{n-1})=\Phi_{n-2}(\tilde\theta_2,...,\tilde\theta_{n-1})$$

据归纳假设有 $(\theta_2,...,\theta_{n-1})=(\tilde\theta_2,...,\tilde\theta_{n-1})$. 所以 Φ_{n-1} 在 $(0,\pi)^{n-2}\times(0,2\pi)$ 上是单射.
□

【例】(1) $\mathbb{R}^n$中的任意n维开区间都与$\mathbb{R}^n$ 是C^∞同胚的.

(2) $\mathbb{R}^n$中的任意n维开球$B(x_0,r)$都与$\mathbb{R}^n$ 是C^∞同胚的. 因此$B(x_0,r)$ 与$\mathbb{R}^n$中的任意n维开区间C^∞同胚.

【证】(1) 设$I=\prod_{i=1}^n(a_i,b_i)$, $-\infty\le a_i<b_i\le+\infty$. 考虑映射

$$f(x)=(f_1(x_1),f_2(x_2),...,f_n(x_n)),\quad x=(x_1,x_2,...,x_n)\in I$$

其中

$$f_i(x_i)=\begin{cases}\tan\left(\pi\frac{x_i-c_i}{b_i-a_i}\right),\quad c_i=\frac{a_i+b_i}{2}, & \text{if}\quad -\infty<a_i<b_i<+\infty,\\ -\log(b_i-x_i) & \text{if}\quad -\infty=a_i<b_i<+\infty,\\ \log(x_i-a_i) & \text{if}\quad -\infty<a_i<b_i=+\infty,\\ x_i & \text{if}\quad -\infty=a_i<b_i=+\infty.\end{cases}$$

易见$f_i\in C^\infty((a_i,b_i))$ 且严格单调增加, 其反函数属于$C^\infty(\mathbb{R})$:

$$(f_i)^{-1}(y_i)=\begin{cases}c_i+\frac{b_i-a_i}{\pi}\arctan y_i & \text{if}\quad -\infty<a_i<b_i<+\infty,\\ b_i-e^{-y_i} & \text{if}\quad -\infty=a_i<b_i<+\infty,\\ a_i+e^{y_i} & \text{if}\quad -\infty<a_i<b_i=+\infty,\\ y_i & \text{if}\quad -\infty=a_i<b_i=+\infty.\end{cases}$$

因此$f_i:(a_i,b_i)\to\mathbb{R}$ 是C^∞ 同胚$(i=1,2,...,n)$ 从而$f:I\to\mathbb{R}^n$ 是C^∞ 同胚.

(2). 令

$$y=f(x)=\frac{x-x_0}{\sqrt{r^2-|x-x_0|^2}},\quad x\in B(x_0,r).$$

易见$f\in C^\infty(B(x_0,r),\mathbb{R}^n)$, 且逆映射为

$$f^{-1}(y)=x_0+\frac{r}{\sqrt{1+|y|^2}}y,\quad y\in\mathbb{R}^n$$

它也是C^∞的. 因此$f:B(x_0,r)\to\mathbb{R}^n$ 是C^∞ 同胚. □

下面这个引理和例子在后面理解和建立曲面维数概念时有重要作用!

【引理8.40(保持原点不动的C^∞同胚).】 设$a>0,b>0$. 令

$$s=h(t)=\frac{(a+b)t}{2ab-(a-b)t},\quad t\in[-a,b].$$

则h 严格单调增加, $h(-a)=-1,h(b)=1,h(0)=0$, $h:[-a,b]\to[-1,1]$是C^∞ 同胚, 从而$h:(-a,b)\to(-1,1)$是C^∞ 同胚. 此外有

$$\begin{aligned}
&2ab-(a-b)t\ge(a+b)\min\{a,b\}>0\qquad\forall\, t\in[-a,b],\\
&h'(t)=\frac{2ab(a+b)}{(2ab-(a-b)t)^2}>0\qquad\forall\, t\in[-a,b],\\
&t=h^{-1}(s)=\frac{2abs}{a+b+(a-b)s}\qquad\forall\, s\in[-1,1],\\
&b-a-(a+b)s\ge2\min\{a,b\}>0\qquad\forall\, s\in[-1,1].
\end{aligned}$$

【证】计算表明$[-a,b]$ 上的单调函数$t\mapsto 2ab-(a-b)t$ 在区间$[-a,b]$的两个端点处的值$\ge(a+b)\min\{a,b\}>0$. 因此$2ab-(a-b)t\ge(a+b)\min\{a,b\}>0$ for all $t\in[-a,b]$. 这同时说明$h(t)$在$[a,b]$ 上是良好定义的. 计算

$$h'(t)=\frac{2ab(a+b)}{(2ab-(a-b)t)^2}>0,\quad t\in[-a,b].$$

因此h在$[-a,b]$ 上严格单调增加. 再计算可知$h(-a)=-1,h(b)=1,h(0)=0$. 因此$h([-a,b])=[-1,1]$, $h((-a,b))=(-1,1)$. 此外容易验证反函数$h^{-1}(s)$的上述表达式成立. 同样对于$[-1,1]$ 上的单调函数$s\mapsto a+b+(a-b)s$ 有$a+b+(a-b)s\ge2\min\{a,b\}>0$ for all $s\in[-1,1]$. 最后由$h(t)$ 和$h^{-1}(s)$ 的表达式易见$h\in C^\infty([a,b]),h^{-1}\in C^\infty([-1,1])$. □

【命题8.41(光滑曲面的表示).】 设$n\ge2$, $k\in\{1,2,...,n-1\}$,

$\mathbf{x}_0=(x_0,y_0)\in\mathbb{R}^k\times\mathbb{R}^{n-k}=\mathbb{R}^n$, 和$C^r$类($r\ge1$)的映射$f:I(x_0)\to J(y_0)$, $f(x_0)=y_0$, 其中$x_0=(x_{01},x_{02},...,x_{0k})\in\mathbb{R}^k,y_0=(y_{01},y_{02},...,y_{0,n-k})\in\mathbb{R}^{n-k}$,

$$I(x_0)=\prod_{i=1}^{k}(x_{0i}-\delta,\,x_{0i}+\delta),\quad J(y_0)=\prod_{j=1}^{n-k}(y_{0j}-\eta,\,y_{0j}+\eta),\quad\delta>0,\eta>0.$$

令$U(\mathbf{x}_0)=I(x_0)\times J(y_0)$. 则对于$U(\mathbf{x}_0)$内包含$\mathbf{x}_0$的显式曲线或曲面

$$S=\{(x,f(x))\mid x\in I(x_0)\},\quad \mathbf{x}_0\in S$$

可以构造一个C^r类同胚$\psi:U(\mathbf{x}_0)\to(-1,1)^n$ 使得ψ 把S映成$(-1,1)^n$的k维子区间$(-1,1)^k\times\{0\}$ 且把$\mathbf{x}_0$ 映为原点, 即

$$\psi(S)=(-1,1)^k\times\{0\},\quad \psi(\mathbf{x}_0)=0.$$

于是令$\varphi(t)=\psi^{-1}(t,0)$, 则$\varphi:(-1,1)^k\to S$ 是同胚且$\varphi\in C^r((-1,1)^k,\mathbb{R}^n)$.

【证】先做两个约定:

(1) 对应于集合乘积$U(\mathbf{x}_0)=I(x_0)\times J(y_0)$, 我们作分解$\mathbb{R}^n=\mathbb{R}^k\times\mathbb{R}^{n-k}$, 因此对于$\mathbb{R}^n$中的点$\mathbf{x},\mathbf{y}$的分解$\mathbf{x}=(x,y),\mathbf{y}=(x,z)$ 中便有$x\in\mathbb{R}^k,y,z\in\mathbb{R}^{n-k}$, 等等.

(2) 为了看得更清楚, 我们有时将映射的值写成列向量. 但自变量还是写成行向量以使表达式紧凑些。

考虑映射

$$g:U(\mathbf{x}_0)\to\mathbb{R}^n,\quad g(\mathbf{x})=g(x,y)=\begin{pmatrix} x\\ y-f(x)\end{pmatrix}.$$

易见g是单射, 其逆映射为

$$g^{-1}:g(U(\mathbf{x}_0))\to U(\mathbf{x}_0),\quad g^{-1}(u,v)=\begin{pmatrix} u\\ v+f(u)\end{pmatrix}.$$

由假设知$g:U(\mathbf{x}_0)=I(x_0)\times J(y_0)\to\mathbb{R}^k\times\mathbb{R}^{n-k}=\mathbb{R}^n$是$C^r$类的. 因

$$g'(\mathbf{x})=g'(x,y)=\begin{pmatrix} \mathrm{I} & 0\\ -f'(x) & \mathrm{I}\end{pmatrix},\quad \det g'(\mathbf{x})\equiv 1\neq 0$$

故由开映射定理知g 是开映射, 特别可知$g(U(\mathbf{x}_0))$ 是包含$g(\mathbf{x}_0)$的连通开集. 再根据整体逆映射定理即知$g:U(\mathbf{x}_0)\to g(U(\mathbf{x}_0))$ 是C^r类同胚.

然后考虑映射

$$h:g(U(\mathbf{x}_0))\to(-1,1)^n,\quad h(\mathbf{y})=h(x,z)=\begin{pmatrix} \frac{1}{\delta}(x-x_0)\\ H(x,z)\end{pmatrix}$$

其中$\mathbf{y}=(x,z)\in g(U(\mathbf{x}_0))$

$$\begin{aligned}
&H(x,z)=(h_1(x,z_1),h_2(x,z_2),...,h_{n-k}(x,z_{n-k}))^\tau,\\
&h_j(x,z_j)=\frac{\eta z_j}{\eta^2-(f_j(x)-y_{0j})^2-(f_j(x)-y_{0j})z_j},\\
&j=1,2,...,n-k.
\end{aligned}$$

这里我们用到了上面**引理8.40**：如写$f(x)=(f_1(x),f_2(x),...,f_k(x))^\tau$, 则由假设$f(x)\in J(y_0)$ 知

$$\eta+f_j(x)-y_{0j}>0,\quad \eta-(f_j(x)-y_{0j})>0\qquad \forall\, x\in I(x_0),\quad j=1,2,...,k.$$

因此由**引理8.40** 知对每个$x\in I(x_0)$ 和$j\in\{1,2,...,n-k\}$, 函数

$$z_j\mapsto s_j:=h_j(x,z_j)=\frac{\eta z_j}{\eta^2-(f_j(x)-y_{0j})^2-(f_j(x)-y_{0j})z_j}$$

是从区间$(-\eta-(f_j(x)-y_{0j}),\eta-(f_j(x)-y_{0j}))$ 到$(-1,1)$的C^∞ 同胚, 并且

$$\eta^2-(f_j(x)-y_{0j})^2-(f_j(x)-y_{0j})z_j\ge\eta\min\{\eta+f_j(x)-y_{0j},\,\eta-(f_j(x)-y_{0j})\}>0$$

for all $z_j\in[-\eta-(f_j(x)-y_{0j}),\eta-(f_j(x)-y_{0j})],j=1,2,...,n-k$. 这说明确实有$h(g(U(\mathbf{x}_0)))\subset(-1,1)^n$, 并且分母的严格正性表明映射$(x,z)\mapsto h(x,z)$ 属于C^r类. 对任意$x\in I(x_0)$, 由**引理8.40** 可知映射

$$H(x,\cdot):\prod_{j=1}^{n-k}\Big(-\eta-(f_j(x)-y_{0j}),\,\eta-(f_j(x)-y_{0j})\Big)\to(-1,1)^{n-k},\quad z\mapsto H(x,z)$$

是单射且实际上是C^∞ 同胚. 于是由h的构造可知h是单射. 其次来证明$h(g(U(\mathbf{x}_0))=(-1,1)^n$. 已有$h(g(U(\mathbf{x}_0))\subset(-1,1)^n$. 而对任意$(t,s)\in(-1,1)^k\times(-1,1)^{n-k}=(-1,1)^n$, 取

$$\begin{aligned}
&x=x_0+\delta t,\\
&z=H(x,\cdot)^{-1}(s)\subset\prod_{j=1}^{n-k}\Big(-\eta-(f_j(x)-y_{0j}),\,\eta-(f_j(x)-y_{0j})\Big),\\
&y=z+f(x).
\end{aligned}$$

则$x\in I(x_0),t=\frac{1}{\delta}(x-x_0),s=H(x,z),y\in J(y_0),z=y-f(x)$ 从而有

$$(t,s)=(\frac{1}{\delta}(x-x_0),H(x,z))=h(x,z)=h(x,y-f(x))=h(g(x,y)))\in h(g(U(\mathbf{x}_0))).$$

这证明了$(-1,1)^n \subset h(g(U(\mathbf{x}_0)))$. 所以$h(g(U(\mathbf{x}_0))) = (-1,1)^n$.

计算Jacobi矩阵: 对任意$\mathbf{y} = (x,z) \in g(U(\mathbf{x}_0))$ 有

$$h'(\mathbf{y}) = h'(x,z) = \begin{pmatrix} \frac{1}{\delta}\mathrm{I} & 0 \\ H'_x(x,z) & H'_z(x,z) \end{pmatrix}$$

$$\Longrightarrow \quad \mathrm{det}h'(\mathbf{y}) = \frac{1}{\delta^k}\mathrm{det}H'_z(x,z) = \frac{1}{\delta^k}\prod_{j=1}^{n-k}\frac{\partial h_j}{\partial z_j}(x,z_j) > 0.$$

这儿“>0”用到上面**引理8.40**. 于是由整体反函数定理知$h: g(U(\mathbf{x}_0)) \to (-1,1)^n$是$C^r$ 类同胚.

令$\psi = h \circ g$. 则$\psi: U(\mathbf{x}_0) \to (-1,1)^n$是$C^r$ 类同胚. 最后来证明

$$\psi(\mathbf{x}_0) = 0, \quad \psi(S) = (-1,1)^k \times \{0\}.$$

由g 和h的定义以及$y_0 = f(x_0)$ 我们有$g(\mathbf{x}_0) = g(x_0,y_0) = (x_0,0)$, $h(x_0,0) = 0$ 从而有

$$\psi(\mathbf{x}_0) = \psi(x_0,y_0) = h(g(x_0,y_0)) = h(x_0,0) = 0.$$

同理对任意$\mathbf{y} \in \psi(S)$, 写$\mathbf{y} = \psi(\mathbf{x})$ 其中$\mathbf{x} \in S$, 即$\mathbf{x} = (x,f(x)), x \in I(x_0)$. 此时由$g$的定义有$g(x,f(x)) = (x,0)$ 从而由h的定义有

$$\psi(\mathbf{x}) = \psi(x,f(x)) = h(g(x,f(x))) = h(x,0)$$

再结合$H(x,0) = 0$ 即得

$$\mathbf{y} = \psi(\mathbf{x}) = h(x,0) = (\frac{1}{\delta}(x-x_0),0) \in (-1,1)^k \times \{0\}.$$

因此$\psi(S) \subset (-1,1)^k \times \{0\}$. 反之对任意$(t,0) \in (-1,1)^k \times \{0\}$ 取$x = u_0 + \delta t$ 则$x \in I(x_0), t = \frac{1}{\delta}(x-x_0)$. 此外有$f(x) \in J(y_0)$. 于是有$(x,f(x)) \in S$ 从而有

$$\psi(S) \ni \psi(x,f(x)) = h(g(x,f(x))) = h(x,0) = (\frac{1}{\delta}(x-x_0),0) = (t,0).$$

因此$(-1,1)^k \times \{0\} \subset \psi(S)$. 这就证明了$\psi(S) = (-1,1)^k \times \{0\}$. $\square$

【隐函数】

- **隐函数的概念, 由方程确定的函数**. 给定一个方程或方程组

$$F(x_1, x_2, ..., x_n, x_{n+1}, x_{n+2}, ..., x_{n+m}) = 0 \tag{1}$$

并假定某点(向量) $(x_{01}, x_{02}, ..., x_{0n}, x_{0,n+1}, x_{0,n+2}, ..., x_{0,n+m})$ 满足此方程, 即

$$F(x_{01}, x_{02}, ..., x_{0n}, x_{0,n+1}, x_{0,n+2}, ..., x_{0,n+m}) = 0. \tag{2}$$

我们希望至少在点$(x_{01}, x_{02}, ..., x_{0n}, x_{0,n+1}, x_{0,n+2}, ..., x_{0,n+m})$的某个邻域内能够把$(x_1, x_2, ..., x_n, x_{n+1}, x_{n+2}, ..., x_{n+m})$的某些分量, 例如前$n$个$x_1, x_2, ..., x_n$ 作为自由变量, 使得其余分量, 例如后m个$x_{n+1}, ..., x_{n+m}$, 都是$(x_1, x_2, ..., x_n)$的函数. 如果此事办成, 并记

$$x = (x_1, x_2, ..., x_n), \quad y = (x_{n+1}, x_{n+2}, ..., x_{n+m}) = f(x) = (f_1(x), f_2(x), ..., f_m(x))$$

$$x_0 = (x_{01}, x_{02}, ..., x_{0n}), \quad y_0 = (x_{0,n+1}, x_{0,n+2}, ..., x_{0,n+m}) = f(x_0)$$

则得到恒等式

$$F(x, f(x)) \equiv 0, \quad x \in U(x_0); \quad y_0 = f(x_0). \tag{3}$$

这里$U(x_0)$是x_0的一个邻域.

方程(1) 叫做由已知函数F决定的经过点(x_0, y_0)一个隐函数方程(也可以理解为是F决定的经过点(x_0, y_0)一个曲面方程), 它是关于$(x_1, x_2, ..., x_n, x_{n+1}, x_{n+2}, ..., x_{n+m})$的**方程**而**不是**关于$(x_1, x_2, ..., x_n, x_{n+1}, x_{n+2}, ..., x_{n+m})$的恒等式! 而(3)中的恒等式当然是关于$x$的恒等式. 这时我们说函数或映射$y = f(x)$ 是隐函数方程(1)-(2)的一个解. 集合$S : y = f(x)$也称为是一个由$F(x, y) = 0$决定的且经过点(x_0, y_0)的显示曲线/曲面.

来看具体例子. 考虑圆周方程

$$F(x, y) := x^2 + y^2 - 1 = 0.$$

它有两个解

$$y = f(x) = \sqrt{1 - x^2}, \quad x \in [-1, 1]$$

$$y = g(x) = -\sqrt{1 - x^2}, \quad x \in [-1, 1]$$

进一步如果要求曲线$(x, f(x))$ 或$(x, g(x))$ 经过圆周上给定的点(x_0, y_0), 则

当$y_0>0$时, 只有$y=f(x)=\sqrt{1-x^2}$ 满足要求.

当$y_0<0$时, 只有$y=g(x)=-\sqrt{1-x^2}$ 满足要求.

但是当$y_0=0$时, 这两个解都满足要求, 我们无法分辨(除非再加其它条件).

什么原因导致两个解? ———————— 由$F(x,y)=x^2+y^2-1$ 有

$$\frac{\partial F}{\partial y}(x,y)=2y$$

因此

$$y_0\neq 0\quad\Longleftrightarrow\quad\frac{\partial F}{\partial y}(x_0,y_0)\neq 0;\qquad y_0=0\quad\Longleftrightarrow\quad\frac{\partial F}{\partial z}(x_0,y_0)=0.$$

联系线性方程组问题:

$$Bx+Ay=b$$

其中A 是方阵. 假定已经过调整使得A具有最大秩. 周知若系数矩阵A可逆, 则方程有唯一解:

$$y=A^{-1}(b-Bx).$$

若A不可逆, 即行列式$\det A=0$, 则此方程**如果有解** 就一定有无穷多解.

以后大家会体会到, 这里矩阵A的地位就相当于圆周方程对应的偏导数$\frac{\partial F}{\partial y}(x_0,y_0)$.

为什么强调**"如果有解"**? 因为并非所考虑的方程都有解或有想要的解. 例如方程

$$G(x,y):=x^2+y^2+1=0$$

便没有实解—— 尽管$G(x,y)$的偏导数与$F(x,y)=x^2+y^2-1$的偏导数完全相同. 所以为了避免在空集上做文章, 理论和逻辑上需要至少存在一点(x_0,y_0) 它满足隐函数方程.

【定理8.42(局部隐函数定理).】设$\Omega\subset\mathbb{R}^n\times\mathbb{R}^m=\mathbb{R}^{n+m}$为开集, 映射$F:\Omega\to\mathbb{R}^m$属于$C^r$ 类$(r\geq 1)$, $(x_0,y_0)\in\Omega$ (其中$x_0\in\mathbb{R}^n, y_0\in\mathbb{R}^m$) 满足

$$F(x_0,y_0)=0,\quad \det F'_y(x_0,y_0)\neq 0.$$

则存在$\delta>0,\varepsilon>0$ 使得隐函数方程

$$F(x,y)=0,\quad (x,y)\in\Omega$$

在$B(x_0,\delta)\times B(y_0,\varepsilon)\subset\Omega$ 内存在唯一的经过点(x_0,y_0)的C^r 类解$f: B(x_0,\delta)\to B(y_0,\varepsilon)$, 即$F(x,f(x))\equiv 0, x\in B(x_0,\delta)$ 且$f(x_0)=y_0$. 此外有

$$f'(x)=-[F_y'(x,y)]^{-1}F_x'(x,y)\Big|_{y=f(x)}=-[F_y'(x,f(x))]^{-1}F_x'(x,f(x)),\quad x\in B(x_0,\delta).$$

【证】令

$$A=F_y'(x_0,y_0),$$
$$\Phi(x,y)=y-A^{-1}F(x,y),\quad (x,y)\in\Omega.$$

易见$\Phi\in C^r(\Omega,\mathbb{R}^m)$,

$$\Phi(x_0,y_0)=y_0,\quad \Phi_y'(x_0,y_0)=0.$$

$$\Phi(x,y)=y\quad\Longleftrightarrow\quad F(x,y)=0.$$

由$\Phi_y'(x_0,y_0)=0$ (零矩阵) 和Φ是光滑的可知存在$\eta>0,\varepsilon>0$ 使得$\overline{B}(x_0,\eta)\times\overline{B}(y_0,\varepsilon)\subset\Omega$ 且

$$\max_{(x,y)\overline{B}(x_0,\eta)\times\overline{B}(y_0,\varepsilon)}\|\Phi_y'(x,y)\|\le\frac{1}{2}.$$

令

$$M=\max_{(x,y)\overline{B}(x_0,\eta)\times\overline{B}(y_0,\varepsilon)}\|\Phi_x'(x,y)\|.$$

$$\delta=\min\{\frac{\varepsilon}{1+2M},\eta\},\quad U=B(x_0,\delta),\quad V=B(y_0,\varepsilon).$$

来证明这个C^r类的映射Φ 在$U\times\overline{V}$上满足**命题8.35(不动函数)** 中的所有条件.

(i) Φ把$U\times\overline{V}$ 映入V内: 由微分中值不等式有

$$\begin{aligned}\forall\,(x,y)\in U\times\overline{V}\implies\quad &|\Phi(x,y)-y_0|=|\Phi(x,y)-\Phi(x_0,y_0)|\\ &\le|\Phi(x,y)-\Phi(x_0,y)|+|\Phi(x_0,y)-\Phi(x_0,y_0)|\\ &\le\|\Phi_x'(\xi,y)\||x-x_0|+\|\Phi_y'(x_0,\tilde{\xi})\||y-y_0|\\ &\le M|x-x_0|+\frac{1}{2}|y-y_0|\le M\delta+\frac{1}{2}\varepsilon<\varepsilon\\ \implies\quad &\Phi(x,y)\in V\end{aligned}$$

其中$\xi=x_0+\theta(x-x_0)\in U,\ \tilde{\xi}=y_0+\tilde{\theta}(y-y_0)\in V,\ 0<\theta,\tilde{\theta}<1.$

(ii) 映射$y\mapsto\Phi(x,y)$在$\overline{V}$上是一致压缩的: $\forall\,x\in U,\ \forall\,y,z\in\overline{V}$ 再次应用微分中值不等式有

$$|\Phi(x,y)-\Phi(x,z)|\le\|\Phi_y'(x,\zeta)\||y-z|\le\frac{1}{2}|y-z|.$$

据**命题8.35(不动函数)** 知存在唯一的映射$f:U\to V$ 满足

$$\Phi(x,f(x))\equiv f(x),\quad x\in U;\quad f(x_0)=y_0.$$

上面已指出, 这等价于

$$F(x,f(x))\equiv 0,\quad x\in U;\qquad f(x_0)=y_0.$$

进一步, 由Φ 属于C^r 类和**命题8.35(不动函数)** 还知$\mathrm{I}-\Phi_y'(x,y)$ 可逆for all $(x,y)\in U\times V$, 并有$f\in C^r(U,V)$ 且

$$f'(x)=[\mathrm{I}-\Phi_y'(x,y)]^{-1}\Phi_x'(x,y)\Big|_{y=f(x)},\quad x\in U.$$

最后计算

$$\Phi_y'(x,y)=\mathrm{I}-A^{-1}F_y'(x,y),\quad \Phi_x'(x,y)=-A^{-1}F_x'(x,y),\quad (x,y)\in I\times J.$$

因此由$A^{-1}F_y'(x,y)=\mathrm{I}-\Phi_y'(x,y)$ 可逆知$F_y'(x,y)$ 可逆for all $(x,y)\in U\times V$, 并且

$$\begin{aligned}&[\mathrm{I}-\Phi_y'(x,y)]^{-1}\Phi_x'(x,y)=-[A^{-1}F_y'(x,y)]^{-1}A^{-1}F_x'(x,y)\\&=-F_y'(x,y)]^{-1}AA^{-1}F_x'(x,y)=-F_y'(x,y)]^{-1}F_x'(x,y),\quad (x,y)\in U\times V.\end{aligned}$$

所以

$$f'(x)=-[F_y'(x,y)]^{-1}F_x'(x,y)\Big|_{y=f(x)}=-[F_y'(x,f(x))]^{-1}F_x'(x,f(x)),\quad x\in U.\qquad\Box$$

【关于隐函数f的值域】在隐函数定理中我们不仅写明了由方程$F(x,y)=0$决定的隐函数f的定义域, 还刻意标明f的值域. 原因是: 不仅证明中需要这样的限制, 而且容易举出例子, 它表明: 在同一个定义域I上可以有多个不同的隐函数, 例如对于$y_0\in J,\widetilde{y}_0\in\widetilde{J}$, 二者都满足$F(x_0,y_0)=0,F(x_0,\widetilde{y}_0)=0$, 但$J\cap\widetilde{J}=\emptyset$, 所以产生两个不同的隐函数$f:I\to J$, $\widetilde{f}:I\to\widetilde{J}$. 因此需要用隐函数$f,\widetilde{f}$ 的值域$J,\widetilde{J}$来区分它们(图示). 从相关的习题的证明中, 大家也可看到对隐函数值域限制的必要性!

隐函数微分公式的记忆—只需知道复合映射微分法: 对恒等式$0\equiv F(x,f(x))$ 两边取微分, 应用复合映射的微分法有

$$0=F_x'(x,y)\Big|_{y=f(x)}+F_y'(x,y)\Big|_{y=f(x)}f'(x)=F_x'(x,f(x))+F_y'(x,f(x))f'(x)$$

$\Longrightarrow$

$$F_y'(x,f(x))f'(x) = -F_x'(x,f(x)), \qquad f'(x) = -[F_y'(x,f(x))]^{-1}F_x'(x,f(x)).$$

【例】$m=2, n=1$, 隐函数方程决定的曲面以及曲面的切平面.

设开集$\Omega\subset\mathbb{R}^3$, 函数$F:\Omega\to\mathbb{R}$ 属于C^1类. 设$(x_0,y_0,z_0)\in\Omega$ 满足

$$F(x_0,y_0,z_0)=0, \quad \nabla F(x_0,y_0,z_0)\neq(0,0,0).$$

来研究由方程

$$F(x,y,z)=0, \quad (x,y,z)\in\Omega$$

决定的局部隐函数和一块曲面.

不妨设$F_z'(x_0,y_0,z_0)\neq 0$. 据局部隐函数定理, 方程$F=0$ 决定了空间$\mathbb{R}^3$中的一块C^1显式曲面

$$S_{xy}: \quad z=z(x,y), \quad (x,y)\in U(x_0,y_0); \quad z(x_0,y_0)=z_0$$

其中$U(x_0,y_0)$ 是(x_0,y_0)的一个邻域, 例如是以(x_0,y_0) 为中心的圆盘$U:(x-x_0)^2+(y-y_0)^2<r^2$ 或正方形$U=(x_0-\delta,x_0+\delta)\times(y_0-\delta,y_0+\delta)$. 在$(x_0,y_0)$ 点做微分展开有

$$z(x,y)-z_0 = z_x'(x_0,y_0)(x-x_0)+z_y'(x_0,y_0)(y-y_0)+o\Big(\sqrt{(x-x_0)^2+(y-y_0)^2}\Big).$$

略去无穷小余项, 我们得到曲面S_{xy} 在点(x_0,y_0,z_0) 处的切平面方程

$$P: \quad z-z_0 = z_x'(x_0,y_0)(x-x_0)+z_y'(x_0,y_0)(y-y_0), \quad (x,y,z)\in\mathbb{R}^3.$$

由本章开头所讲, 这个切平面的一个单位法向量可取为

$$\mathbf{n} = \frac{(z_x'(x_0,y_0),\ z_y'(x_0,y_0),\ -1)}{\sqrt{(z_x'(x_0,y_0))^2+(z_y'(x_0,y_0))^2+1}}.$$

也就是说

$$(x,y,z)\in P \quad\Longleftrightarrow\quad (x,y,z)-(x_0,y_0,z_0)\perp\mathbf{n}.$$

另一方面由隐函数偏导数的计算公式有

$$z_x'(x,y) = -\frac{F_x'(x,y,z)}{F_z'(x,y,z)}\Big|_{z=z(x,y)} = -\frac{F_x'(x,y,z(x,y))}{F_z'(x,y,z(x,y))},$$

$$z'_y(x,y) = -\frac{F'_y(x,y,z)}{F'_z(x,y,z)}\Big|_{z=z(x,y)} = -\frac{F'_y(x,y,z(x,y))}{F'_z(x,y,z(x,y))},$$

或写成

$$\frac{\partial z(x,y)}{\partial x} = -\frac{\frac{\partial}{\partial x}F(x,y,z)}{\frac{\partial}{\partial z}F(x,y,z)}\Big|_{z=z(x,y)} = -\frac{\left(\frac{\partial}{\partial x}F\right)(x,y,z(x,y))}{\left(\frac{\partial}{\partial z}F\right)(x,y,z(x,y))},$$

$$\frac{\partial z(x,y)}{\partial y} = -\frac{\frac{\partial}{\partial y}F(x,y,z)}{\frac{\partial}{\partial z}F(x,y,z)}\Big|_{z=z(x,y)} = -\frac{\left(\frac{\partial}{\partial y}F\right)(x,y,z(x,y))}{\left(\frac{\partial}{\partial z}F\right)(x,y,z(x,y))}.$$

特别在$(x,y)=(x_0,y_0)$ 有 (注意$z(x_0,y_0)=z_0$)

$$z'_x(x_0,y_0) - \frac{F'_x(x_0,y_0,z_0)}{F'_z(x_0,y_0,z_0)},$$

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

将切平面方程两边乘以非零数$F'_z(x_0,y_0,z_0)$ 然后移项得到切平面P的对称表达式:

$$P:\quad F'_x(x_0,y_0,z_0)(x-x_0)+F'_y(x_0,y_0,z_0)(y-y_0)+F'_z(x_0,y_0,z_0)(z-z_0)=0,\quad (x,y,z)\in\mathbb{R}^3.$$

也即

$$P:\quad \Big\langle (x-x_0,y-y_0,z-z_0,\ \nabla F(x_0,y_0,z_0)\Big\rangle = 0,\quad (x,y,z)\in\mathbb{R}^3$$

或写成

$$P = \Big\{(x,y,z)\in\mathbb{R}^3 \,\Big|\, \Big\langle (x-x_0,y-y_0,z-z_0),\nabla F(x_0,y_0,z_0)\Big\rangle = 0\Big\}.$$

相应地, 切平面P的法向量$\mathbf{n}$ 也写成了对称形式:

$$\mathbf{n} = \frac{\nabla F(x_0,y_0,z_0)}{|\nabla F(x_0,y_0,z_0)|}.$$

这种对称表示的好处是: 它只依赖于总体信息(x_0,y_0,z_0) 和$\nabla F(x_0,y_0,z_0)$. 事实上如果还知道偏导数$F'_x(x_0,y_0,z_0)\neq 0$, 则用相同推导可得一块C^1显示曲面$S_{yz}: x=x(y,z)$满足$x(y_0,z_0)=x_0$. 它在$\mathbb{R}^3$中的几何形状与上面得到的曲面S_{xy}在(x_0,y_0,z_0)附近完全相同; 曲面块S_{xy} 和S_{yz}(实为同一曲面块)在点(x_0,y_0,z_0)处的切平面方程当然具有相同形式.

再次指出, 为得到隐函数的偏导数关系, 只需知道复合函数求导法则: 例如

$$\begin{aligned}
&0 \equiv F(x,y,z(x,y))\\
&\Longrightarrow\ 0 = \frac{\partial}{\partial x}\Big(F(x,y,z(x,y))\Big) = F'_x(x,y,z(x,y)) + F'_z(x,y,z(x,y))z'_x(x,y)\\
&\Longrightarrow\ z'_x(x,y) = -\frac{F'_x(x,y,z)}{F'_z(x,y,z)}\Big|_{z=z(x,y)} = -\frac{F'_x(x,y,z(x,y))}{F'_z(x,y,z(x,y))}
\end{aligned}$$

等等.

【平面曲线的情形】 $m=1, n=1$, **由隐函数方程决定的曲线以及该曲线的切线方程.**

设开集$\Omega\subset\mathbb{R}^2$, $f\in C^r(\Omega,\mathbb{R})\,(r\ge 1)$. 设$c\in\mathbb{R}$, $(x_0,y_0)\in\Omega$ 满足

$$f(x_0,y_0)=c,\quad \nabla f(x_0,y_0)\neq(0,0).$$

研究由方程

$$f(x,y)=c,\quad (x,y)\in\Omega$$

决定的曲线—— 称之为**等高线**.

不妨设$f'_y(x_0,y_0)\neq 0$. 根据隐函数定理, 方程$f(x,y)-c=0$ 决定了平面$\mathbb{R}^2$上的一段C^1显式曲线

$$S_x:\quad y=y(x),\quad x\in(x_0-\delta,x_0+\delta);\quad y(x_0)=y_0.$$

在x_0 点做微分展开有

$$y(x)-y_0=y'_x(x_0)(x-x_0)+o(x-x_0).$$

略去无穷小余项, 我们得到曲线S_x 在点(x_0,y_0) 处的切线方程

$$P:\quad y-y_0=y'_x(x_0)(x-x_0),\quad (x,y)\in\mathbb{R}^2.$$

这个切线的一个单位法向量可取为

$$\mathbf{n}=\frac{(y'_x(x_0),\,-1)}{\sqrt{(y'_x(x_0))^2+1}}.$$

也就是说

$$(x,y)\in P\quad\Longleftrightarrow\quad (x,y)-(x_0,y_0)\perp\mathbf{n}.$$

另一方面由隐函数偏导数的计算公式有

$$y'_x(x)=-\frac{f'_x(x,y)}{f'_y(x,y)}\Big|_{y=y(x)}=-\frac{f'_x(x,y(x))}{f'_y(x,y(x))},$$

或写成

$$\frac{\partial y(x)}{\partial x}=-\frac{\frac{\partial}{\partial x}f(x,y)}{\frac{\partial}{\partial y}f(x,y)}\Big|_{y=y(x)}=-\frac{\left(\frac{\partial}{\partial x}f\right)(x,y(x))}{\left(\frac{\partial}{\partial z}f\right)(x,y(x))}.$$

特别在$x = x_0$ 有 (注意$y(x_0) = y_0$)

$$y'_x(x_0) = -\frac{f'_x(x_0, y_0)}{f'_y(x_0, y_0)}.$$

将切线方程两边乘以非零数$f'_y(x_0, y_0)$ 然后移项得到切线P的对称表达式:

$$P: \quad f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0) = 0, \quad (x, y) \in \mathbb{R}^2.$$

也即

$$P: \quad \Big\langle (x - x_0, y - y_0), \nabla f(x_0, y_0) \Big\rangle = 0, \quad (x, y) \in \mathbb{R}^2$$

或写成

$$P = \Big\{ (x, y) \in \mathbb{R}^2 \;\Big|\; \Big\langle (x - x_0, y - y_0), \nabla f(x_0, y_0) \Big\rangle = 0 \Big\}.$$

相应地, P的法向量$\mathbf{n}$ 也写成了对称形式:

$$\mathbf{n} = \frac{\nabla f(x_0, y_0)}{|\nabla f(x_0, y_0)|}.$$

同理若偏导数$f'_x(x_0, y_0) \neq 0$, 则用相同推导可得一段C^r显示曲线$S_y : x = x(y), y \in (y_0 - \eta, y_0 + \eta)$, 满足$x(y_0) = y_0$. 如上, 曲线$S_y$在点$(x_0, y_0)$的切线方程的形式不变, 也即它只由$(x_0, y_0)$和$\nabla f(x_0, y_0)$决定.

平面曲线的情形比较简单、对称, 但仍包含了重要的本质概念. 我们用下面隐函数定理的推论来概括它的一个重要性质. 这个性质对于理解隐函数方程解的局部唯一性和分叉现象很有帮助. 首先我们介绍光滑曲线的奇点和正则点的概念.

【定义(奇点与正则点)】设$I \subset \mathbb{R}$为开区间, 曲线$t \mapsto \gamma(t) \in \mathbb{R}^m$ 在I 上属于C^1 类. 对于$t_0 \in I$, 若$\gamma'(t_0) = 0$, 则称t_0 是这曲线的一个参数奇点, 简称奇点; 若$\gamma'(t_0) \neq 0$, 则称t_0 是这曲线的一个正则点. 若这曲线在I 内是处处正则的(即无奇点), 则称它是一条正则曲线. □

例如$t = 0$ 是曲线$t \mapsto (t^3, t^6)$的唯一奇点. 但这曲线的另一表示$x \mapsto (x, x^2)$ 便无奇点, 即$x \mapsto (x, x^2)$是一条正则曲线. 这例子说明有些曲线的参数奇点经过参数变换是可以消除的. 在实际问题的研究中, 一般尽可能选择无奇点的(即正则)的参数表示.

【隐函数定理的推论1】设开集$\Omega \subset \mathbb{R}^2$, $f \in C^r(\Omega, \mathbb{R})$ $(r \geq 1)$, $c \in \mathbb{R}$, $(x_0, y_0) \in \Omega$ 满足

$$f(x_0, y_0) = c, \quad \nabla f(x_0, y_0) \neq 0.$$

则除去曲线的参数表示不同之外, 在(x_0,y_0)的一个邻域内$U((x_0,y_0))$存在唯一一条满足方程$f(x,y)=c$且经过点(x_0,y_0)的正则曲线.

【证】不妨设$f'_y(x_0,y_0)\neq 0$. 由$f\in C^r$ 和局部隐函数定理(在隐函数定理中取$F(x,y)=f(x,y)-c$) 知存在$\delta>0,\eta>0$ 使得$U((x_0,y_0)):=(x_0-\delta,x_0+\delta)\times(y_0-\eta,y_0+\eta)\subset\Omega$, $f'_y(x,y)\neq 0$ for all $(x,y)\in U((x_0,y_0))$, 并且对每个$x\in(x_0-\delta,x_0+\delta)$ 存在唯一的$y\in(y_0-\eta,y_0+\eta)$ 使得$f(x,y)=c$. 记这一对应关系为$y=\varphi(x)$. 则隐函数定理还告诉我们曲线$\kappa(x):=(x,\varphi(x))$ 是唯一一条位于$U((x_0,y_0))$内且经过点(x_0,y_0)的C^1 显式曲线, 满足$f(\kappa(x))\equiv c$, $x\in(x_0-\delta,x_0+\delta)$. 显然$x\mapsto\kappa(x)$ 是一条正则曲线.

设$I\subset\mathbb{R}$ 为开区间, $t\mapsto\gamma(t)=(\gamma_1(t),\gamma_2(t))\in U((x_0,y_0))$ $(t\in I)$ 为一条正则曲线满足$f(\gamma(t))\equiv c$ $(t\in I)$ 和$\gamma(t_0)=(x_0,y_0)$ for some $t_0\in I$. 来证明存在一个包含x_0的开区间J 和一个C^1双射$h:I\to J$ 使得

$$h(t_0)=x_0;\quad \gamma(t)\equiv\kappa(h(t)),\quad t\in I;\quad \kappa(x)\equiv\gamma(h^{-1}(x)),\quad x\in J.$$

如果此事成立, 则它表明, 过点(x_0,y_0)的两条曲线$t\mapsto\gamma(t),t\in I$ 与$x\mapsto\kappa(x),x\in J$ 是重合的, 也即它们是同一条曲线的不同参数表示.

首先由$f(\gamma_1(t),\gamma_2(t))=c, f(\gamma_1(t),\varphi(\gamma_1(t))=c$, 和方程$f(x,y)=c$ 在$U((x_0,y_0))$内的解的唯一性知$\gamma_2(t)=\varphi(\gamma_1(t))$ for all $t\in I$. 这证明了$\gamma(t)=\kappa(\gamma_1(t)),t\in I$.

其次由复合函数求导法则有

$$0\equiv\frac{\mathrm{d}}{\mathrm{d}t}(f(\gamma(t)))=f'_x(\gamma(t))\gamma'_1(t)+f'_y(\gamma(t))\gamma'_2(t).$$

因$f'_y(\gamma(t))\neq 0$ for all $t\in I$, 故得

$$\gamma'_2(t)=-\frac{f'_x(\gamma(t))}{f'_y(\gamma(t))}\gamma'_1(t),\quad t\in I_1.$$

又因$t\mapsto\gamma(t)$ 是正则的, 即$(\gamma'_1(t),\gamma'_2(t))\neq 0$, 故由上面关系式知$\gamma'_1(t)\neq 0$ for all $t\in I$. 这表明$\gamma_1(t)$ 在I内是严格单调的, 其反函数$\gamma_1^{-1}:\gamma_1(I)\to I$ 也是C^1类的.

由$x_0=\gamma_1(t_0)\in\gamma_1(I)$ 和$\gamma_1(t)$ 在I内严格单调且连续知$\gamma_1(I)$ 是包含x_0的开区间. 取$J=\gamma_1(I)$, $h(t)=\gamma_1(t)$. 则$h:I\to J$ 是C^1 双射且$h(t_0)=x_0$, $\gamma(t)\equiv\kappa(h(t)),t\in I$; $\kappa(x)\equiv\gamma(h^{-1}(x))$, $x\in J$. $\square$

曲线的交叉点.[5] 设D_1, D_2 为开区间, $t \mapsto \gamma(t) \in \mathbb{R}^n, t \in D_1$ 和$\tau \mapsto \beta(\tau) \in \mathbb{R}^n, \tau \in D_2$ 为两条曲线, $\mathbf{x}_0 \in \mathbb{R}^n$. 若存在包含$t_0$的开区间$I_1 \subset D_1$ 和包含τ_0的开区间$I_2 \subset D_2$ 使得

$$\gamma(t_0) = \beta(\tau_0) = \mathbf{x}_0; \quad \gamma(t) \neq \beta(\tau) \quad \forall\, t \in I_1 \setminus \{t_0\},\ \forall\, \tau \in I_2 \setminus \{\tau_0\}$$

则称这两条曲线在点$\mathbf{x}_0$交叉. $\square$

从曲线交叉的角度看, 隐函数定理的推论1的实质上等价的陈述为

【隐函数定理的推论2】设开集$\Omega \subset \mathbb{R}^2$, $f \in C^r(\Omega, \mathbb{R})$ $(r \geq 1)$, $c \in \mathbb{R}$, $(x_0, y_0) \in \Omega$ 满足$f(x_0, y_0) = c$. 若方程$f(x, y) = c$在(x_0, y_0) 附近的解包含了至少两条在点(x_0, y_0)交叉的正则曲线, 则必有$\nabla f(x_0, y_0) = 0$.

【证】设两条正则曲线$t \mapsto \gamma(t) \in \Omega, t \in D_1$ 和$\tau \mapsto \beta(\tau) \in \Omega, \tau \in D_2$ 满足$f(\gamma(t)) \equiv c, t \in D_1; f(\beta(\tau)) \equiv c, \tau \in D_2$ 且$\gamma(t_0) = \beta(\tau_0) = (x_0, y_0)$ for some $t_0 \in D_1, \tau_0 \in D_2$, 并设这两条曲线在点$(x_0, y_0)$交叉, 即存在包含$t_0$的开区间$I_1 \subset D_1$ 和包含τ_0的开区间$I_2 \subset D_2$, 使得$\gamma(t) \neq \beta(\tau)\ \forall\, t \in I_1 \setminus \{t_0\}, \forall\, \tau \in I_2 \setminus \{\tau_0\}$. 现在让我们把$\gamma, \beta$的定义域局限在$I_1, I_2$ 上.

假设$\nabla f(x_0, y_0) \neq 0$. 不妨设$f'_y(x_0, y_0) \neq 0$. 则从隐函数定理的推论1 的证明知存在$(x_0, y_0)$的一个邻域$U((x_0, y_0))$和唯一一条$C^r$ 显示曲线$x \mapsto \kappa(x) \in U((x_0, y_0)), x \in (x_0 - \delta, x_0 + \delta)\,(\delta > 0)$ 满足$\kappa(x_0) = (x_0, y_0)$ 和$f(\kappa(x)) \equiv c$, $x \in (x_0 - \delta, x_0 + \delta)$, 并且存在包含$x_0$的开区间$J_1$, J_2 和C^r 双射$h_1 : I_1 \to J_1, h_2 : I_2 \to J_2$ 满足

$$h_1(t_0) = h_2(\tau_0) = x_0; \quad \kappa(x) = \gamma(h_1^{-1}(x)), x \in J_1; \quad \kappa(x) = \beta(h_2^{-1}(x)), x \in J_2.$$

易见$J_1 \cap J_2$仍是包含x_0的开区间, 因此$h_1^{-1}(J_1 \cap J_2), h_2^{-1}(J_1 \cap J_2)$分别是包含$t_0, \tau_0$的开区间. 显然$h_1^{-1}(J_1 \cap J_2) \subset I_1$, $h_2^{-1}(J_1 \cap J_2) \subset I_2$. 现在任取$t \in h_1^{-1}(J_1 \cap J_2) \setminus \{t_0\}$, 令$x = h_1(t)$, 则$x \neq x_0$ 且$x \in J_1 \cap J_2 \subset J_2$ 从而有$h_2^{-1}(x) \in I_2$. 取$\tau = h_2^{-1}(x)$, 则由$x \neq x_0$, 也即由$h_2(\tau) \neq h_2(\tau_0)$ 可知$\tau \neq \tau_0$ 因而有$\tau \in I_2 \setminus \{\tau_0\}$. 但同时还有

$$\gamma(t) = \gamma(h_1^{-1}(x)) = \kappa(x) = \beta(h_2^{-1}(x)) = \beta(\tau).$$

这就与两曲线在点(x_0, y_0)交叉矛盾. 此矛盾说明$\nabla f(x_0, y_0) \neq 0$ 不成立. 因此只能有$\nabla f(x_0, y_0) = 0$. $\square$

[5]这里描述的交叉本质上是十字形交叉, 它是比较容易研究的一种分叉. 对于丁字形分叉, 其研究要稍困难些.

【例】设二元函数$f:\mathbb{R}^2\to\mathbb{R}$ 属于C^r类($r\ge 1$). 假设由方程$f(x,y)=1$ 确定的集合(即等高线) 是8字形C^r曲线. 证明梯度方程$\nabla f(x,y)=0$ 在$\mathbb{R}^2$上至少有三个不同的解.

【证】设Ω_1,Ω_2 分别为8字形曲线围成的两个有界开区域. 则由$f(x,y)=1$ for all $(x,y)\in\partial\Omega_i$ 和多元函数的Rolle 定理知存在$(x_i,y_i)\in\Omega_i$ 使得$\nabla f(x_i,y_i)=0$, $i=1,2$.

令(x_0,y_0) 为8字形曲线中的交叉点. 来证明$\nabla f(x_0,y_0)=0$. 由题意知存在$\delta>0,\eta>0$ 和两个C^r函数$y=y_1(x),y=y_2(x)$ 满足

$$f(x,y_i(x))\equiv 1,\quad x\in(x_0-\delta,x_0+\delta),\ \ y_i(x)\in(y_0-\eta,y_0+\eta);\quad i=1,2$$

$$y_1(x_0)=y_2(x_0)=y_0,\quad y_1(x)\ne y_2(x)\quad \forall\, x\in(x_0-\delta,x_0+\delta)\setminus\{x_0\}. \tag{$*$}$$

这两条曲线$y=y_1(x),y=y_2(x)$显然是正则的. 因($*$)表明这两条曲线在点(x_0,y_0)交叉, 故由隐函数定理的推论2 知必有$\nabla f(x_0,y_0)=0$.

最后因$\Omega_1,\Omega_2,\partial\Omega_1\cap\partial\Omega_2$ 三者互不相交且$(x_1,y_1)\in\Omega_1,(x_2,y_2)\in\Omega_2,(x_0,y_0)\in\partial\Omega_1\cap\partial\Omega_2$, 所以$(x_1,y_1)$, (x_2,y_2), (x_0,y_0) 互不相同. 这证明了梯度方程$\nabla f(x,y)=0$ 在$\mathbb{R}^2$上至少有三个不同的解. $\Box$

• 隐函数问题的其它形式：例如由下列方程组确定的隐函数问题

$$\begin{cases} u=f(x,y,z,t),\\ g_1(x,y,z,t)=g_2(x,y,z)\\ h(y,t)=0.\end{cases}$$

等价于右端等于零的形式:

$$\begin{cases} f^1(x,y,z,t,u)=f(x,y,z,t)-u=0,\\ f^2(x,y,z,t,u)=g_1(x,y,z,t)-g_2(x,y,z)=0\\ f^3(x,y,z,t,u)=h(y,t)=0.\end{cases}$$

若某点(x_0,y_0,z_0,t_0,u_0)满足上述方程并且映射$(f^1(x,y,z,t,u),f^1(x,y,z,t,u),f^1(x,y,z,t,u))$ 在点(x_0,y_0,z_0,t_0,u_0)处关于(z,t,u)的Jacobi矩阵可逆, 则上述方程在该点附近便确定出一个(x,y)的隐函数：$(z,t,u)=(z(x,y),t(x,y),u(x,y))$.

【作业题】

1. 设$U,V\subset\mathbb{R}^n$ 为开集, 映射$f:U\to V$ 为单满射. 假设$f:U\to V$ 和$f^{-1}:V\to U$ 分别在U,V内处处可微. 证明$\det f'(x)\neq 0$ for all $x\in U$.

2. 陈书习题**8.5.1.**

3. 设$\Omega\subset\mathbb{R}^n$ 为有界开集, $f\in C^1(\Omega,\mathbb{R}^n)$ 满足

$$\det f'(x)\neq 0\qquad \forall\, x\in\Omega$$

$$\forall\, x\in\partial\Omega\quad 有\quad \lim_{\Omega\ni y\to x}|f(x)|=+\infty.$$

证明$f(\Omega)=\mathbb{R}^n$, 即f是满射. [参见**命题8.36(可微映射的开映射定理)**的证明.]

4. 设$f\in C^1(\mathbb{R}^n,\mathbb{R}^n)$ 满足

$$\det f'(x)\neq 0\qquad \forall\, x\in\mathbb{R}^n$$

$$\lim_{|x|\to+\infty}|f(x)|=+\infty.$$

证明$f(\mathbb{R}^n)=\mathbb{R}^n$, 即$f$是满射. [与上题证法类似.]

5. 设$g\in C^1(\mathbb{R}^n,\mathbb{R}^n)$ 满足

$$\|g'(x)\|\le\frac{1}{2}\qquad \forall\, x\in\mathbb{R}^n.$$

令$f(x)=x+g(x)$. 证明$f:\mathbb{R}^n\to\mathbb{R}^n$ 是C^1同胚.

6. 设$f\in C^2(\mathbb{R}^n)$.

(1) 假设f的Hesse 矩阵$H_f(x)$处处正定并且

$$\lim_{|x|\to+\infty}|\nabla f(x)|=+\infty.$$

证明梯度映射$\nabla f:\mathbb{R}^n\to\mathbb{R}^n$ 是C^1同胚(即$\nabla f:\mathbb{R}^n\to\mathbb{R}^n$ 是单满射且逆映射也是C^1的).

(2) [(1)的一个具体情形] 假设f的Hesse 矩阵$H_f(x)$ 一致正定, 即存在常数$\lambda>0$ 使得

$$h^{\tau}H_f(x)h\ge\lambda|h|^2\qquad \forall\, x,h\in\mathbb{R}^n.$$

证明

$$\lim_{|x|\to+\infty}|\nabla f(x)|=+\infty$$

从而梯度映射$\nabla f:\mathbb{R}^n\to\mathbb{R}^n$ 是C^1同胚.

6. 设$\varphi:\mathbb{R}\to\mathbb{R}$ 可微, 满足$\varphi(0)=0$ 且存在$a>0$ 和$0<k<1$ 使得当$y\in(-a,a)$ 时$|\varphi'(y)|\le k$. 证明存在$0<\delta\le a$ 使得当$x\in(-\delta,\delta)$ 时方程

$$x=y+\varphi(y)$$

存在唯一的解$y=y(x)$, 并且$y(x)$ 在$(-\delta,\delta)$ 上可微. [注：为了应用上面已证明的命题, 你可以假定φ 是C^1类的. 但由于$\mathbb{R}$ 是有序的, 故可微条件就够了.]

7. 陈书习题**8.3.2.**

8.(吉米多维奇3420). 证明：由方程$u=y+x\varphi(u)$ 确定的函数$u=u(x,y)$ 满足微分方程

$$\frac{\partial u}{\partial x}=\varphi(u)\frac{\partial u}{\partial y},\qquad \frac{\partial^2 u}{\partial x^2}=\frac{\partial}{\partial y}\Big([\varphi(u)]^2\frac{\partial u}{\partial y}\Big).$$

一般地对任意自然数n 成立Lagrange 公式:

$$\frac{\partial^n u}{\partial x^n}=\frac{\partial^{n-1}}{\partial y^{n-1}}\Big([\varphi(u)]^n\frac{\partial u}{\partial y}\Big).$$

[对证明中所需要的条件, 可以自行补充, 例如需要φ有足够高的光滑性, 在$u(x,y)$的定义域内某些分式的分母不为零, 等等.]

9. (参见陈书习题8.5.4.)

设$\mathbf{a}_0=(a_{01},a_{02},...,a_{0n})\in\mathbb{R}^n$ 使n次多项式

$$P(\mathbf{a}_0,x)=x^n+a_{01}x^{n-1}+a_{02}x^{n-2}+\cdots+a_{0(n-1)}x+a_{0n},\quad x\in\mathbb{R}$$

有n 个互不相同的实根$x_1<x_2<\cdots<x_n$. 证明: 存在$\delta>0$ 使得当$\mathbf{a}\in B(\mathbf{a}_0,\delta)$ 时, x的多项式$P(\mathbf{a},x)$ 仍有n个互不相同的实根且这些根保持大小顺序：$x_1(\mathbf{a})<x_2(\mathbf{a})<\cdots<x_n(\mathbf{a})$. 因此, 对每个$k\in\{1,2,...,n\}$, 映射$\mathbf{a}\mapsto x_k(\mathbf{a})$ 在$B(\mathbf{a}_0,\delta)$ 上是良好定义的. 进一步证明对每个$k\in\{1,2,...,n\}$, 映射$\mathbf{a}\mapsto x_k(\mathbf{a})$ 在在$B(\mathbf{a}_0,\delta)$ 上属于C^∞类. [本题说明, 在这种情况下, 多项式的根是多项式系数的光滑函数. 这与Abel理论(五次和五次以上的多项式的根一般无公式解)并无矛盾.]

10. 陈书习题**8.5.10.** (**Legendre 变换**).

11.陈书习题**8.5.12.** (**Cauchy-Riemann 方程**).

§8.8 映射的秩与函数相关

本节是局部反函数定理的重要应用. 函数组或映射的秩与函数组的相关、无关性等是研究函数组或映射的行为的重要概念，其基本思想就是要把一组函数(或一个映射)中起决定作用的某个最小子组确定下来，使得这个子组以外的其它函数都可由这个子组的函数来表示，当然这里的“表示”不必是线性表示. 我们将看到，这与线性代数中向量组的极大无关组或矩阵的秩等在概念上是相同的.

回忆线性代数. 以下的分析对行向量和列向量是相同的. 我们以行向量为例. 设$a_1, a_2, ..., a_m$ 是$\mathbb{R}^n$中的m个行向量. 设A是由$a_1, a_2, ..., a_m$ 构成的$m \times n$阶矩阵即

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \multicolumn{4}{c}{\dotfill} \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

假设$\{a_1, a_2, ..., a_m\}$ 的秩是k, 即$a_1, a_2, ..., a_m$ 中的极大线性无关组的向量的个数为k, $1 \le k \le \min\{n, m\}$. 则通过先做行置换(即用置换矩阵$P \in \mathbb{R}^{m\times m}$左乘$A$)再做列置换(即用置换矩阵$Q^{-1} \in \mathbb{R}^{n\times n}$右乘$PA$) 得到

$$PAQ^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

其中$A_{11} \in \mathbb{R}^{k\times k}$ 可逆即$\det(A_{11}) \neq 0$. 因$\mathrm{rank}(PAQ^{-1}) = \mathrm{rank}(A) = k$, 故矩阵$PAQ^{-1}$中的后$m-k$个行向量中的每一个都可以用前$k$个行向量的线性组合表示, 因此存在矩阵$B \in \mathbb{R}^{(m-k)\times k}$使得

$$(A_{21} \quad A_{22}) = B(A_{11} \quad A_{12}) = (BA_{11} \quad BA_{12}).$$

由此得到$A_{21} = BA_{11}, A_{22} = BA_{12}$ 从而得到$B = A_{21}A_{11}^{-1}$

$$A_{22} = A_{21}A_{11}^{-1}A_{12}. \tag{¶}$$

这个关系式将在下面证明秩定理时用到.

令I_k表示k阶单位矩阵. 则有

$$\begin{pmatrix} \mathrm{I}_k & 0 \\ -A_{21}(A_{11})^{-1} & \mathrm{I}_{m-k} \end{pmatrix} PAQ^{-1} \begin{pmatrix} (A_{11})^{-1} & -(A_{11})^{-1}A_{12} \\ 0 & \mathrm{I}_{n-k} \end{pmatrix} = \begin{pmatrix} \mathrm{I}_k & 0 \\ 0 & 0 \end{pmatrix}.$$

因此我们得到了可逆矩阵$\Psi \in \mathbb{R}^{m\times m}, \Phi \in \mathbb{R}^{n\times n}$ 使得

$$\Psi A \Phi^{-1} = \begin{pmatrix} \mathrm{I}_k & 0 \\ 0 & 0 \end{pmatrix}.$$

把A, Φ, Ψ 看成线性变换即

$$A(\mathbf{x}) = A\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n, \qquad \text{etc.}$$

或将$\mathbf{x}$写成行向量而定义

$$A(\mathbf{x}) = (A\mathbf{x}^\tau)^\tau = \mathbf{x}A^\tau, \quad \mathbf{x} \in \mathbb{R}^n, \qquad \text{etc.}$$

则我们得到

$$\Psi \circ A \circ \Phi^{-1}(u, v) = (u, 0), \qquad u \in \mathbb{R}^k, \quad v \in \mathbb{R}^{n-k}.$$

这种表示形式就是我们将要学习的所谓秩定理. 从表达式看出, 秩定理是说一个秩为常数k的映射在经过自变量的可逆变换和函数值的可逆变换后便化成了一个投影$(u, v) \mapsto (u, 0)$, 从而说明使得(变换后的)映射真正其作用的只是n个自变量中的前k个自变量. 对于非线性映射来说, 通过局部线性化, 上述结果局部地依然成立, 也即此时我们可以得到局部形式的秩定理. 这个局部秩定理的证明要用到局部反函数定理.

为简化记号, 我们可以将映射的坐标函数和自变量的分量做位置调整. 对任一映射(按列向量表示)

$$f(\mathbf{x}) = \begin{pmatrix} f_1(x_1, x_2, ..., x_n) \\ f_2(x_1, x_2, ..., x_n) \\ \vdots \\ f_m(x_1, x_2, ..., x_n) \end{pmatrix}, \qquad \mathbf{x} = (x_1, x_2, ..., x_n)^\tau$$

以及f_i的下标$(1, 2, ..., m)$ 和x_j的下标$(1, 2, ..., n)$的任意置换

$$(\sigma(1), \sigma(2), ..., \sigma(m)), \quad (\tau(1), \tau(2), ..., \tau(n))$$

如上分析, 我们可以确定置换矩阵$P \in \mathbb{R}^{m\times m}, Q \in \mathbb{R}^{n\times m}$ 使得映射

$$\widetilde{f}(\mathbf{x}) := Pf(Q^{-1}\mathbf{x})$$

的坐标函数和自变量的位置是按照置换$\sigma(i),\tau(j)$ 排列的, 也即

$$\widetilde{f}(\mathbf{x})=\begin{pmatrix}\widetilde{f}_1(x_1,x_2,...,x_n)\\ \widetilde{f}_2(x_1,x_2,...,x_n)\\ \vdots\\ \widetilde{f}_m(x_1,x_2,...,x_n)\end{pmatrix}=Pf(Q^{-1}\mathbf{x})=\begin{pmatrix}f_{\sigma(1)}(x_{\tau(1)},x_{\tau(2)},...,\ x_{\tau(n)})\\ f_{\sigma(2)}(x_{\tau(1)},x_{\tau(2)},...,\ x_{\tau(n)})\\ \vdots\\ f_{\sigma(m)}(x_{\tau(1)},x_{\tau(2)},...,\ x_{\tau(n)})\end{pmatrix}.$$

应注意: 矩阵P,Q^{-1} 只与f_i和x_j的下标(i,j)的位置有关而与$f_i(x_1,x_2,...,x_n)$值无关!
从这种变换我们看出, f 与$\widetilde{f}$ 的地位是相同的, 即

$$\widetilde{f}(\mathbf{x})=Pf(Q^{-1}\mathbf{x}),\quad f(\mathbf{x})=P^{-1}\widetilde{f}(Q\mathbf{x})$$

也即

$$\widetilde{f}=P\circ f\circ Q^{-1},\quad f=P^{-1}\circ\widetilde{f}\circ Q.$$

因此f 与$\widetilde{f}$ 具有相同的主要性质.

【定理8.43(秩定理).】设$\mathbf{x}_0\in\mathbb{R}^n$, $U(\mathbf{x}_0)\subset\mathbb{R}^n$ 为$\mathbf{x}_0$的一个邻域. 设$r\geq 1$, C^r 类映射$f:U(\mathbf{x}_0)\to\mathbb{R}^m$在$U(\mathbf{x}_0)$上的秩为常数$k$, 即

$$\mathrm{rank}f(\mathbf{x}):=\mathrm{rank}f'(\mathbf{x})\equiv k,\qquad \mathbf{x}\in U(\mathbf{x}_0).$$

令$\mathbf{y}_0=f(\mathbf{x}_0)$. 则有:

(a) 若$k=n<m$, 则存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)\subset U(\mathbf{x}_0)$, n维开区间$I\subset\mathbb{R}^n$, C^r类同胚$\varphi:O(\mathbf{x}_0)\to I$, 以及$\mathbf{y}_0$的一个邻域$O(\mathbf{y}_0)\subset\mathbb{R}^m$ 和C^r类同胚$\psi:O(\mathbf{y}_0)\to\psi(O(\mathbf{y}_0))\,(\subset\mathbb{R}^m)$, 使得$f(O(\mathbf{x}_0))\subset O(\mathbf{y}_0)$ 且

$$\psi\circ f\circ\varphi^{-1}(u)\equiv(u,0),\qquad u\in I.$$

(b) 若$k=m<n$, 则存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)\subset U(\mathbf{x}_0)$, 以$\mathbf{y}_0$ 为中心的m 维开区间$I\subset\mathbb{R}^m$, $n-m$维开区间$J\subset\mathbb{R}^{n-m}$, 以及C^r类同胚$\varphi:O(\mathbf{x}_0)\to I\times J$, 使得

$$f\circ\varphi^{-1}(u,v)\equiv u,\quad (u,v)\in I\times J.$$

(c) 若$1\leq k<\min\{m,n\}$, 则存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)\subset U(\mathbf{x}_0)$, k维开区间$I\subset\mathbb{R}^k$, $n-k$ 维开区间$J\subset\mathbb{R}^{n-k}$, C^r类同胚$\varphi:O(\mathbf{x}_0)\to I\times J$, 以及$\mathbf{y}_0$的一个邻域$O(\mathbf{y}_0)\subset\mathbb{R}^m$ 和C^r类同胚$\psi:O(\mathbf{y}_0)\to\psi(O(\mathbf{y}_0))\,(\subset\mathbb{R}^m)$, 使得$f(O(\mathbf{x}_0))\subset O(\mathbf{y}_0)$ 且

$$\psi\circ f\circ\varphi^{-1}(u,v)\equiv(u,0),\qquad (u,v)\in I\times J.$$

(d) 若$k=0$, 则f在$U(\mathbf{x}_0)$上为常值映射.

【注】上述命题中没有涉及$k=n=m$的简单情形, 因为这种情形属于局部反函数定理: 存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)$ 使得象集$I=f(O(\mathbf{x}_0))$是以$\mathbf{y}_0$ 为中心的一个开区间并且$f:O(\mathbf{x}_0)\to I$ 为C^r 类同胚并有$f\circ f^{-1}(u)\equiv u,\, u\in I$.

【定理的证明】首先做记号简化：将映射$f(\mathbf{x})$的坐标函数和自变量$\mathbf{x}$的分量进行位置调整—— 为简化记号, 调整后的映射仍记作$f(\mathbf{x})$—— 使得调整后的映射$f(\mathbf{x})$的Jacobi 矩阵$f'(\mathbf{x}_0)$ 中的**左上角**的主对角$k\times k$ 方阵可逆.

在推导中由于要计算Jacobi矩阵, 我们有时将映射写成列向量.

(a): 设$k=n<m$. 此时我们有分解$f(\mathbf{x})=(\varphi(\mathbf{x}),g(\mathbf{x}))$, 或写成列向量

$$f(\mathbf{x})=\begin{pmatrix}\varphi(\mathbf{x})\\ g(\mathbf{x})\end{pmatrix},\quad \mathbf{x}\in U(\mathbf{x}_0)$$

其中

$$\varphi(\mathbf{x})=\begin{pmatrix}f_1(\mathbf{x})\\ f_2(\mathbf{x})\\ \vdots\\ f_n(\mathbf{x})\end{pmatrix},\quad g(\mathbf{x})=\begin{pmatrix}f_{n+1}(\mathbf{x})\\ f_{n+2}(\mathbf{x})\\ \vdots\\ f_m(\mathbf{x})\end{pmatrix}.$$

由假设有

$$f'(\mathbf{x})=\begin{pmatrix}\varphi'(\mathbf{x})\\ g'(\mathbf{x})\end{pmatrix},\quad \det\varphi'(\mathbf{x}_0)\neq 0.$$

由光滑性假设和局部反函数定理知存在$\mathbf{x}_0$的邻域$W\subset U(\mathbf{x}_0)$ 使得$\varphi(W)$ 是$\varphi(\mathbf{x}_0)$的一个邻域并且$\varphi:W\to\varphi(W)$ 是C^r 类微分同胚. 取以$\varphi(\mathbf{x}_0)$为中心的一个n维开区间$I\subset\mathbb{R}^n$ 满足$I\subset\varphi(W)$. 然后令$O(\mathbf{x}_0)=\varphi^{-1}(I)$. 则由$\varphi:W\to\varphi(W)$为$C^r$同胚，也即$\varphi^{-1}:\varphi(W)\to W$ 为C^r同胚，可知$O(\mathbf{x}_0)\,(\subset W)$是$\mathbf{x}_0$的一个邻域且$\varphi^{-1}:I\to O(\mathbf{x}_0)$仍为$C^r$类微分同胚. 也即$\varphi:O(\mathbf{x}_0)\to I$仍为$C^r$类微分同胚.

由反函数关系我们有

$$f(\varphi^{-1}(u))=(u,\,g(\varphi^{-1}(u))):=(u,h(u)),\quad u\in I$$

其中$h(u)=g(\varphi^{-1}(u))$. 由复合映射性质知$h:I\to\mathbb{R}^{m-n}$ 是C^r 类的. 如令

$$\psi(u,w)=(u,w-h(u)),\quad (u,w)\in I\times\mathbb{R}^{m-n}$$

则得到标准表示

$$\psi\circ f\circ\varphi^{-1}(u)=\psi(u,h(u))=(u,0),\quad u\in I.$$

不难看出ψ 是C^r类的单射且逆映射$\psi^{-1}(x,y)=(x,y+h(x))$ 也是C^r类的, 其中$(x,y)\in I\times\mathbb{R}^{m-n}$. 因此$\psi: I\times\mathbb{R}^{m-n}\to\psi(I\times\mathbb{R}^{m-n})$ 是C^r 类同胚. 而由$\mathbf{y}_0=f(\mathbf{x}_0)=(\varphi(\mathbf{x}_0),g(\mathbf{x}_0))\in I\times\mathbb{R}^{m-n}$ 可知$O(\mathbf{y}_0):=I\times\mathbb{R}^{m-n}$ 是$\mathbf{y}_0$的一个邻域. 最后由$u\in I\Longrightarrow f(\varphi^{-1}(u))=(u,g(u))\in I\times\mathbb{R}^{m-n}$ 有$f(O(\mathbf{x}_0))=f(\varphi^{-1}(I))\subset I\times\mathbb{R}^{m-n}=O(\mathbf{y_0})$. 这证明了(a)成立.

(b): 设$k=m<n$. 此时我们将自变量$\mathbf{x}$划分为两块:

$$\mathbf{x}=(x,y),\quad \mathbf{x}_0=(x_0,y_0)\quad \text{其中}\quad x,x_0\in\mathbb{R}^m,\ \ y,y_0\in\mathbb{R}^{n-m}.$$

考虑C^r类光滑映射$\varphi:U(\mathbf{x}_0)\to\mathbb{R}^n$,

$$(u,v)=\varphi(\mathbf{x})=(f(x,y),y),\quad \mathbf{x}=(x,y)\in U(\mathbf{x}_0).$$

或写成列向量

$$\begin{pmatrix}u\\ v\end{pmatrix}=\varphi(\mathbf{x})=\begin{pmatrix}f(x,y)\\ y\end{pmatrix},\quad \mathbf{x}=(x,y)\in U(\mathbf{x}_0).$$

由假设知在f的Jacobi 矩阵

$$f'(\mathbf{x})=f'(x,y)=(f'_x(x,y)\quad f'_y(x,y))$$

中$f'_x(x,y)$在$\mathbf{x}_0=(x_0,y_0)$可逆. 于是有

$$\varphi'(\mathbf{x})=\begin{pmatrix}f'_x(x,y) & f'_y(x,y)\\ 0 & \mathrm{I}_{n-m}\end{pmatrix},\quad \det\varphi'(\mathbf{x}_0)=\det f'_x(x_0,y_0)\neq 0.$$

由光滑性假设和局部反函数定理知存在$\mathbf{x}_0=(x_0,y_0)$的邻域$W\subset U(\mathbf{x}_0)$ 使得$\varphi(W)$是$(u_0,v_0):=\varphi(\mathbf{x}_0)$的一个邻域并且$\varphi:W\to\varphi(W)$ 是C^r 同胚. 取以$u_0=f(\mathbf{x}_0)=f(x_0,y_0)$为中心的一个$n$维开区间$I\subset\mathbb{R}^n$, 以$v_0=y_0$为中心的一个$m-n$维开区间$J\subset\mathbb{R}^{m-n}$, 满足$I\times J\subset\Phi(W)$. 然后令$O(\mathbf{x}_0)=\varphi^{-1}(I\times J)$. 则由$\varphi:W\to\varphi(W)$为$C^r$同胚, 也即$\varphi^{-1}:\varphi(W)\to W$ 为C^r同胚, 可知$O(\mathbf{x}_0)\,(\subset W)$是$\mathbf{x}_0=(x_0,y_0)$的邻域且$\varphi^{-1}:I\times J\to O(\mathbf{x}_0)$仍为$C^r$同胚. 也即$\varphi:O(\mathbf{x}_0)\to I\times J$仍为$C^r$同胚.

从恒等式

$$(u,v)=\varphi(\varphi^{-1}(u,v))=(f(\varphi^{-1}(u,v)),v),\quad (u,v)\in I\times J$$

出发我们得到关系式

$$f(\varphi^{-1}(u,v))=u,\quad (u,v)\in I\times J.$$

这证明了(b)成立.

(c): 设$1\le k<\min\{m,n\}$. 此时, 如上, 我们将自变量$\mathbf{x}$划分为两块:

$$\mathbf{x}=(x,y),\ \ \mathbf{x}_0=(x_0,y_0)\quad \text{其中}\quad x,x_0\in\mathbb{R}^k,\ \ y,y_0\in\mathbb{R}^{n-k}.$$

相应地做映射的分解

$$f(\mathbf{x})=(\Phi(\mathbf{x}),\ \Psi(\mathbf{x}))$$

或写成列向量

$$f(\mathbf{x})=\begin{pmatrix}\Phi(\mathbf{x})\\ \Psi(\mathbf{x})\end{pmatrix}=\begin{pmatrix}\Phi(x,y)\\ \Psi(x,y)\end{pmatrix},\quad \mathbf{x}=(x,y)\in U(\mathbf{x}_0)$$

其中

$$\Phi(x,y)=\begin{pmatrix}f_1(x,y)\\ f_2(x,y)\\ \vdots\\ f_k(x,y)\end{pmatrix},\quad \Psi(x,y)=\begin{pmatrix}f_{k+1}(x,y)\\ f_{k+2}(x,y)\\ \vdots\\ f_m(x,y)\end{pmatrix}.$$

计算Jacobi矩阵:

$$f'(\mathbf{x})=\begin{pmatrix}\Phi'_x(x,y) & \Phi'_y(x,y)\\ \Psi'_x(x,y) & \Psi'_y(x,y)\end{pmatrix},\quad \det\Phi'_x(x_0,y_0)\neq 0.$$

考虑C^r类光滑映射$\varphi:U(\mathbf{x}_0)\to\mathbb{R}^n$,

$$(u,v)=\varphi(\mathbf{x})=(\Phi(x,y),\ y),\quad x=(x,y)\in U(\mathbf{x}_0).$$

或写成列向量

$$\begin{pmatrix}u\\ v\end{pmatrix}=\varphi(\mathbf{x})=\begin{pmatrix}\Phi(x,y)\\ y\end{pmatrix},\quad \mathbf{x}=(x,y)\in U(\mathbf{x}_0).$$

我们有

$$\varphi'(x)=\begin{pmatrix}\Phi'_x(x,y) & \Phi'_y(x,y)\\ 0 & \mathrm{I}_{n-k}\end{pmatrix},\quad \det\varphi'(\mathbf{x}_0)=\det\Phi'_x(x_0,y_0)\neq 0.$$

由光滑性假设和局部反函数定理知存在$\mathbf{x}_0=(x_0,y_0)$的邻域$W\subset U(\mathbf{x}_0)$ 使得$\varphi(W)$ 是

$$(u_0,v_0):=\varphi(\mathbf{x}_0)=\varphi(x_0,y_0)=(\Phi(x_0,y_0),y_0)$$

的一个邻域并且$\varphi:W\to\varphi(W)$ 是C^r 同胚. 注意这蕴含Jacobi矩阵$\varphi'(\mathbf{x})$ 可逆, 也即

$$0\neq \det\varphi'(\mathbf{x})=\det\Phi'_x(x,y)\qquad \forall\, \mathbf{x}=(x,y)\in W.$$

另一方面由假设$\mathrm{rank}f'(\mathbf{x})\equiv k$ 可知Jacobi 矩阵$f'(\mathbf{x})=f'(x,y)$ 中的后$m-k$ 行中的每个行向量都可由前面k个行向量来表示, 从而得到如下重要关系式(见前面的分析和关系式(¶))

$$\Psi'_y(x,y)=\Psi'_x(x,y)[\Phi'_x(x,y)]^{-1}\Phi'_y(x,y),\quad (x,y)=\mathbf{x}\in W \tag{$*$}$$

现在我们取以$u_0=\Phi(\mathbf{x}_0)=\Phi(x_0,y_0)$为中心的一个$k$维开区间$I\subset\mathbb{R}^k$, 以$v_0=y_0$为中心的一个$n-k$维开区间$J\subset\mathbb{R}^{n-k}$, 满足$\overline{I}\times\overline{J}\subset\varphi(W)$. 这里考虑闭包的目的是为后面证明关于函数相关的命题时做准备. 然后令$O(\mathbf{x}_0)=\varphi^{-1}(I\times J)$. 则由$\varphi:W\to\varphi(W)$ 为C^r同胚, 也即$\varphi^{-1}:\varphi(W)\to W$ 为C^r同胚，可知φ^{-1} 在$\overline{I}\times\overline{J}$上连续, $O(\mathbf{x}_0)\,(\subset W)$是$\mathbf{x}_0=(x_0,y_0)$的邻域且$\varphi^{-1}:I\times J\to O(\mathbf{x}_0)$仍为$C^r$同胚. 也即$\varphi:O(\mathbf{x}_0)\to I\times J$仍为$C^r$同胚.

为导出本定理中的(c)中的表示, 考虑恒等式

$$\begin{pmatrix} u\\ v\end{pmatrix}=\varphi(\varphi^{-1}(u,v))=\begin{pmatrix}\Phi(\varphi^{-1}(u,v))\\ \varphi^{-1}(u,v)_2\end{pmatrix},\quad (u,v)\in\overline{I}\times\overline{J}$$

这里$\varphi^{-1}(u,v)_2\in\mathbb{R}^{n-k}$ 是由向量$\varphi^{-1}(u,v)$ 的后$n-k$个分量组成的向量. 这就给出关系式

$$\Phi(\varphi^{-1}(u,v))=u,\quad (u,v)\in\overline{I}\times\overline{J}$$

从而有

$$f(\varphi^{-1}(u,v))=(\Phi(\varphi^{-1}(u,v)),\Psi(\varphi^{-1}(u,v)))=(u,\Psi(\varphi^{-1}(u,v))),\quad (u,v)\in\overline{I}\times\overline{J}.$$

来证明映射$(u,v)\mapsto\Psi(\varphi^{-1}(u,v))$ 关于v 是常值, 也即$\Psi(\varphi^{-1}(u,v))\equiv\Psi(\varphi^{-1}(u,v_0))$, $(u,v)\in\overline{I}\times\overline{J}$.

为此我们先证明映射$\Psi(\varphi^{-1}(u,v))$ 关于v的微分恒等于零. 据复合映射的微分法和逆映射的微分计算公式我们计算对任意$(u,v)\in I\times J$

$$\Big(\Psi(\varphi^{-1}(u,v))\Big)'=\Psi'(x,y)\Big|_{(x,y)=\varphi^{-1}(u,v)}(\varphi^{-1})'(u,v)$$

$$= \Big(\Psi'_x(x,y),\ \ \Psi'_y(x,y)\Big)[\varphi'(x,y)]^{-1}\Big|_{(x,y)=\varphi^{-1}(u,v)}.$$

因

$$[\varphi'(x,y)]^{-1} = \begin{pmatrix} \Phi'_x(x,y) & \Phi'_y(x,y) \\ 0 & \mathrm{I}_{n-k} \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} [\Phi'_x(x,y)]^{-1} & -[\Phi'_x(x,y)]^{-1}\Phi'_y(x,y) \\ 0 & \mathrm{I}_{n-k} \end{pmatrix}$$

故

$$\begin{aligned}
&\Big(\Psi'_x(x,y),\ \ \Psi'_y(x,y)\Big)[\varphi'(x,y)]^{-1} \\
&= \Big(\Psi'_x(x,y),\ \ \Psi'_y(x,y)\Big)\begin{pmatrix} [\Phi'_x(x,y)]^{-1} & -[\Phi'_x(x,y)]^{-1}\Phi'_y(x,y) \\ 0 & \mathrm{I}_{n-k} \end{pmatrix} \\
&= \Big(\Psi'_x(x,y)[\Phi'_x(x,y)]^{-1},\ \ -\Psi'_x(x,y)[\Phi'_x(x,y)]^{-1}\Phi'_y(x,y)+\Psi'_y(x,y)\Big) \\
&= \Big(\Psi'_x(x,y)[\Phi'_x(x,y)]^{-1},\ \ 0\Big)
\end{aligned}$$

其中用到$(*)$ 即$\Psi'_y(x,y) = \Psi'_x(x,y)[\Phi'_x(x,y)]^{-1}\Phi'_y(x,y)$. 因此

$$\Big(\Psi(\varphi^{-1}(u,v))\Big)' = \Big(\Psi'_x(x,y)[\Phi'_x(x,y)]^{-1},\ \ 0\Big)\Big|_{(x,y)=\varphi^{-1}(u,v)}, \quad (u,v)\in I\times J.$$

比较Jacobi矩阵中分块对应的位置我们得到

$$\Big(\Psi(\varphi^{-1}(u,v))\Big)'_v = 0, \quad (u,v)\in I\times J.$$

因J 是开凸集, 对$v\in J$ 应用微分中值不等式即知$\Psi(\varphi^{-1}(u,v))$ 关于v 是常值, 也即$\Psi(\varphi^{-1}(u,v)) \equiv \Psi(\varphi^{-1}(u,v_0))$, $(u,v)\in I\times J$. 因φ^{-1}在闭包$\overline{I}\times\overline{J}\subset\varphi(W)$ 上连续, 故得到

$$\Psi(\varphi^{-1}(u,v)) \equiv \Psi(\varphi^{-1}(u,v_0)) =: g(u), \quad (u,v)\in\overline{I}\times\overline{J}.$$

[注意这个函数$g(u)=\Psi(\varphi^{-1}(u,v_0))$在$k$维闭区间$\overline{I}$上连续. 后面在证明**命题8.44(函数相关)** 时要用到这点.]

联合上面结果我们得到表示

$$f\circ\varphi^{-1}(u,v) = f(\varphi^{-1}(u,v)) = (u, g(u)), \quad (u,v)\in I\times J.$$

如令

$$\psi(u,w) = (u, w-g(u)), \quad (u,w)\in I\times\mathbb{R}^{m-k}$$

则得到标准表示

$$\psi\circ f\circ\varphi^{-1}(u,v)=\psi(u,g(u))=(u,0),\quad (u,v)\in I\times J.$$

最后易见ψ 是C^r类的单射且逆映射$\psi^{-1}(x)=\psi^{-1}(x,y)=(x,y+g(x))$ 也是C^r类的. 因此$\psi: I\times\mathbb{R}^{m-k}\to\psi(I\times\mathbb{R}^{m-k})$ 是C^r 类同胚. 再由$f(\mathbf{x}_0)=(\Phi(\mathbf{x}_0),\Psi(\mathbf{x}_0))=(u_0,\Psi(\mathbf{x}_0))\in I\times\mathbb{R}^{m-k}$ 知$O(\mathbf{y}_0):=I\times\mathbb{R}^{m-k}$ 是$\mathbf{y}_0=f(\mathbf{x}_0)$的一个邻域. 最后由$(u,v)\in I\times J\Longrightarrow f\circ\varphi^{-1}(u,v)=(u,g(u))\in I\times\mathbb{R}^{m-k}$ 易见有$f(O(\mathbf{x}_0))=f(\varphi^{-1}(I\times J))\subset I\times\mathbb{R}^{m-k}=O(\mathbf{y}_0)$. 这证明了(c)成立.

(d): 设$k=0$, 即$\mathrm{rank}(f'(\mathbf{x}))\equiv 0,\mathbf{x}\in U(\mathbf{x}_0)$. 则$f'(\mathbf{x})\equiv 0,\mathbf{x}\in U(\mathbf{x}_0)$. 因$U(\mathbf{x}_0)$是连通开集, 故$f(\mathbf{x})\equiv$ 常值, $\mathbf{x}\in U(\mathbf{x}_0)$. 至此命题证毕. $\square$

【关于秩的直观解释】秩$\mathrm{rank}f(\mathbf{x})\equiv k,\mathbf{x}\in U(\mathbf{x}_0)$, 等于说当$\mathbf{x}$ 取遍$U(\mathbf{x}_0)$中所有点时, $f(\mathbf{x})$充满了一块k维区间.

作为**秩定理**的一个重要应用, 我们有下面

【例】设$\Omega\subset\mathbb{R}^n$为开集, 映射$f:\Omega\to\mathbb{R}^m$ 属于C^1类. 若$n>m$, 则f一定不是单射.

【证(2014级王志涵)】考虑Ω的连通分支, 我们可以假设Ω还是连通的.

由矩阵秩的定义和$n>m$知集合

$$\{\mathrm{rank}(f'(\mathbf{x}))\mid \mathbf{x}\in\Omega\}\subset\{0,1,2,...,m\}.$$

据整数集的性质可知该集合有最大元, 即存在$\mathbf{x}_0\in\Omega$ 使得

$$k:=\mathrm{rank}(f'(\mathbf{x}_0))=\max_{\mathbf{x}\in\Omega}\mathrm{rank}(f'(\mathbf{x}))\qquad (\text{即等于} f \text{ 的最大秩}).$$

若$k=0$, 则$f'(\mathbf{x})=0$ for all $\mathbf{x}\in\Omega$ 从而f在Ω上为常值映射,它当然不是单射.

以下设$k\ge 1$. 由$\mathrm{rank}(f'(\mathbf{x}_0))=k$的定义知$f'(\mathbf{x}_0)$有一个$k\times k$ 阶方阵$A(\mathbf{x}_0)$ 使得$\det A(\mathbf{x}_0)\neq 0$. 因$\Omega$是开集且$f$ 属于C^1类, 故存在$\delta>0$ 使得开球$B(\mathbf{x}_0,\delta)\subset\Omega$且当$\mathbf{x}\in B(\mathbf{x}_0,\delta)$ 时$\det A(\mathbf{x})\neq 0$. 这蕴含$\mathrm{rank}(f'(\mathbf{x}))\ge k$ for all $\mathbf{x}\in B(\mathbf{x}_0,\delta)$. 但$k$是最大秩, 故又有$\mathrm{rank}(f'(\mathbf{x}))\le k$ for all $\mathbf{x}\in B(\mathbf{x}_0,\delta)$. 所以$\mathrm{rank}(f'(\mathbf{x}))=k$ for all $\mathbf{x}\in B(\mathbf{x}_0,\delta)$. 这证明了$f$在$B(\mathbf{x}_0,\delta)$上的秩为常数$k$.

注意$n>m$, 由**定理8.43(秩定理)**我们有:

若$k = m$, 则存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0) \subset B(\mathbf{x}_0, \delta)$、以$f(\mathbf{x}_0)$为中心的$m$维开区间$I \subset \mathbb{R}^m$、$n-m$ 维开区间$J \subset \mathbb{R}^{n-m}$、以及$C^1$类同胚$\varphi : O(\mathbf{x}_0) \to I \times J$, 使得

$$f(\varphi^{-1}(u,v)) \equiv u, \quad (u,v) \in I \times J.$$

因$\varphi^{-1} : I \times J \to O(\mathbf{x}_0)$是单射, 故上式蕴含$f$不是单射。

若$k < m$, 则存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0) \subset B(\mathbf{x}_0, \delta)$、$k$维开区间$I \subset \mathbb{R}^k$、$n-k$ 维开区间$J \subset \mathbb{R}^{n-k}$、$C^1$类同胚$\varphi : O(\mathbf{x}_0) \to I \times J$、以及$\mathbf{y}_0 = f(\mathbf{x}_0)$的一个邻域$O(\mathbf{y}_0)) \subset \mathbb{R}^m$和$C^1$类同胚$\psi : O(\mathbf{y}_0) \to \psi(O(\mathbf{y}_0))\,(\subset \mathbb{R}^m)$, 使得$f(O(\mathbf{x}_0)) \subset O(\mathbf{y}_0)$ 且

$$\psi(f(\varphi^{-1}(u,v))) \equiv (u,0), \qquad (u,v) \in I \times J.$$

即

$$f(\varphi^{-1}(u,v)) \equiv \psi^{-1}(u,0), \qquad (u,v) \in I \times J.$$

因$\varphi^{-1} : I \times J \to O(\mathbf{x}_0)$是单射, 故上式蕴含$f$不是单射. □

【函数相关性】

【定义】设$f_1, f_2, ..., f_m$ 是在开集$U \subset \mathbb{R}^n$ 上连续的一组实值函数. 令$f = (f_1, f_2, ..., f_m)$. 如果对每个满足

$$F(f(\mathbf{x}))) \equiv 0, \quad \mathbf{x} \in U$$

的连续函数$F : \mathbb{R}^m \to \mathbb{R}$, 都存在开集$V \supset f(U)$使得$F(\mathbf{y}) \equiv 0, \mathbf{y} \in V$, 则称函数组$f_1, f_2, ..., f_m$ 在U 内为函数无关的(或称为函数独立的).

若函数组$f_1, f_2, ..., f_m$ 在U 内不是函数无关的, 则称函数组$f_1, f_2, ..., f_m$ 在U 内是函数相关的. □

由此定义易见, 连续函数组$f_1, f_2, ..., f_m$ 在U内是函数相关的当且仅当映射$f = (f_1, f_2, ..., f_m) : U \to \mathbb{R}^m$ 满足: 对任意开集$V \supset f(U)$, 存在连续函数$F : \mathbb{R}^m \to \mathbb{R}$ 使得$F(f(\mathbf{x}))) \equiv 0$, $\mathbf{x} \in U$, 但$F(\mathbf{y}) \not\equiv 0$, $\mathbf{y} \in V$.

【命题8.44(函数相关性).】设$\mathbf{x}_0 \in \mathbb{R}^n$, $U(\mathbf{x}_0) \subset \mathbb{R}^n$ 为$\mathbf{x}_0$的一个邻域, 设C^r 类($r \geq 1$)的函数组$f_1, f_2, ..., f_m : U(\mathbf{x}_0) \to \mathbb{R}$ 在$U(\mathbf{x}_0)$上的秩为常数k, 即映射$f = (f_1, f_2, ..., f_m) : U(\mathbf{x}_0) \to \mathbb{R}^m$ 满足$\mathrm{rank} f(\mathbf{x}) := \mathrm{rank} f'(\mathbf{x}) \equiv k$, $\mathbf{x} \in U(\mathbf{x}_0)$. 则有:

(a) 若$k=m$, 则存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)\subset U(\mathbf{x}_0)$ 使得函数组$f_1,f_2,...,f_m$ 在$O(\mathbf{x}_0)$内是函数无关的.

(b) 若$k<m$, 则存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)\subset U(\mathbf{x}_0)$ 使得函数组$f_1,f_2,...,f_m$ 在$O(\mathbf{x}_0)$内是函数相关的. 进一步, 若$k\ge 1$, 则在$f_1,f_2,...,f_m$中有k个函数在$O(\mathbf{x}_0)$ 内为函数独立, 而其余$m-k$个函数都可表示成这k个独立函数的函数.

【证】(a): 先设$k=m=n$. 此时由局部反函数定理, 存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)$和$\mathbf{y}_0=f(\mathbf{x}_0)=(f_1(\mathbf{x}_0),f_2(\mathbf{x}_0),...,f_m(\mathbf{x}_0))$的一个邻域$O(\mathbf{y}_0)$ 使得$f=(f_1,f_2,...,f_m):O(\mathbf{x}_0)\to O(\mathbf{y}_0)$ 是C^r 同胚. 设$F:\mathbb{R}^m\to\mathbb{R}$ 为任一连续函数满足

$$F(f(\mathbf{x}))\equiv 0,\quad \mathbf{x}\in O(\mathbf{x}_0).$$

据$f=(f_1,f_2,...,f_m):O(\mathbf{x}_0)\to O(\mathbf{y}_0)$ 是同胚知$f(O(\mathbf{x}_0))=O(\mathbf{y}_0)$. 因此对任意$\mathbf{y}\in O(\mathbf{y}_0)$ 存在(唯一) $\mathbf{x}\in O(\mathbf{x}_0)$ 使得$\mathbf{y}=f(\mathbf{x})$. 代入上式即得$F(\mathbf{y})=F(f(\mathbf{x}))=0$. 因此$F(\mathbf{y})\equiv 0$ 于开集$V:=O(\mathbf{y}_0)=f(O(\mathbf{x}_0))$. 所以函数组$f_1,f_2,...,f_m$ 在$\mathbf{x}_0$的邻域$O(\mathbf{x}_0)$内为函数无关.

其次设$k=m<n$. 据秩定理, 对于映射$f=(f_1,f_2,...,f_m)$, 存在$\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0)$, 以$f(\mathbf{x}_0)$ 为中心的m维开区间$I\subset\mathbb{R}^m$, $n-m$ 维开区间$J\subset\mathbb{R}^{n-m}$, 以及C^r类同胚$\varphi:O(\mathbf{x}_0)\to I\times J$, 使得

$$f(\varphi^{-1}(u,v))\equiv u,\qquad u\in I,\ v\in J.$$

注意到$O(\mathbf{x}_0)=\varphi^{-1}(I\times J)$ 可知上式蕴含$f(O(\mathbf{x}_0))=I$. 设$F:\mathbb{R}^n\to\mathbb{R}$ 为满足

$$F(f(\mathbf{x}))\equiv 0,\quad \mathbf{x}\in O(\mathbf{x}_0)$$

任一连续函数. 对任意$\mathbf{y}\in I$, 取一点$v\in J$. 则有$\mathbf{x}:=\varphi^{-1}(\mathbf{y},v)\in O(\mathbf{x}_0)$ 从而有$f(\mathbf{x})=f(\varphi^{-1}(\mathbf{y},v))=\mathbf{y}$,

$$F(\mathbf{y})=F(f(\mathbf{x}))=0.$$

这证明了$F(\mathbf{y})\equiv 0$ 于I. 因$I=f(O(\mathbf{x}_0))$ 是开集, 所以函数组$f_1,f_2,...,f_m$ 在$\mathbf{x}_0$的邻域$O(\mathbf{x}_0)$ 内为函数无关.

(b): 设$k<n$. 若$k=0$ 即$\mathrm{rank}f'(\mathbf{x})\equiv 0$, $\mathbf{x}\in U(\mathbf{x}_0)$ 则易见$f'(\mathbf{x})\equiv 0,\mathbf{x}\in U(\mathbf{x}_0)$. 这蕴含$f(\mathbf{x})\equiv f(\mathbf{x}_0)=$常值于$U(\mathbf{x}_0)$ 也即$f_1,f_2,...,f_m$在$U(\mathbf{x}_0)$上均为常数. 此时取连续函数$F(\mathbf{y})=|\mathbf{y}-f(\mathbf{x}_0)|$, 则有$F(f(\mathbf{x}))\equiv 0,\mathbf{x}\in U(\mathbf{x}_0)$, 而对任意开集$V\supset f(U(\mathbf{x}_0))=\{f(\mathbf{x}_0)\}$ 有$F(\mathbf{y})\not\equiv 0,\mathbf{y}\in V$. 所以$f_1,f_2,...,f_m$ 在$U(\mathbf{x}_0)$上为函数相关.

设$1 \le k < m$. 我们只需就$k < \min\{m,n\}$ 的情形给出证明, 因为对$k = n < m$的情形的证明已包含在其中(且相对简单).

将映射$f(\mathbf{x})$的坐标函数和自变量$\mathbf{x}$的分量进行位置调整, 即考虑$\widetilde{f}(\mathbf{x}) = Pf(Q^{-1}\mathbf{x})$ (其中P, Q 为常数矩阵, 可逆), 并将$\widetilde{f}$ 仍记作f, 使得$f(\mathbf{x})$在$\mathbf{x}_0$的Jacobi 矩阵$f'(\mathbf{x}_0)$ 中的**左上角**的主对角$k \times k$ 方阵可逆. 如前, 我们将自变量$\mathbf{x}$划分为两块:

$$\mathbf{x} = (x, y),\ \ \mathbf{x}_0 = (x_0, y_0),\quad \text{其中}\quad x, x_0 \in \mathbb{R}^k,\ \ y, y_0 \in \mathbb{R}^{n-k}.$$

相应地映射f被分解为

$$f(\mathbf{x}) = (\Phi(\mathbf{x}),\, \Psi(\mathbf{x})),\quad \mathbf{x} = (x, y) \in U(\mathbf{x}_0)$$

其中

$$\Phi = (f_1, f_2, ..., f_k),\quad \Psi = (f_{k+1}, f_{k+2}, ..., f_m).$$

在**定理8.43(秩定理)之(c)** 的证明中我们已证明了存在以$u_0 := \Phi(\mathbf{x}_0) = \Phi(x_0, y_0)$为中心的一个$k$ 维开区间$I \subset \mathbb{R}^k$, 以$v_0 = y_0$为中心的一个$n-k$维开区间$J \subset \mathbb{R}^{n-k}$, $\mathbf{x}_0$的一个邻域$O(\mathbf{x}_0) \subset U(\mathbf{x}_0)$, C^r类同胚$\varphi : O(\mathbf{x}_0) \to I \times J$, 以及$C^r$ 类映射$g : I \to \mathbb{R}^{m-k}$ 使得

$$(u_0, v_0) = \varphi(\mathbf{x}_0),$$

$$\Phi(\varphi^{-1}(u, v)) = u,\quad \Psi(\varphi^{-1}(u, v)) = g(u),\quad (u, v) \in I \times J. \tag{$**$}$$

并且这个函数g 实际上是在闭包$\overline{I}$ 上连续[见**定理8.43(秩定理)之(c)** 的证明].

把$(**)$中的第一个关系式代入第二个得到

$$\Psi(\varphi^{-1}(u, v)) = g\Big(\Phi(\varphi^{-1}(u, v))\Big),\quad (u, v) \in I \times J.$$

因$O(\mathbf{x}_0) = \varphi^{-1}(I \times J)$, 故上式蕴含

$$\Psi(\mathbf{x}) = g(\Phi(\mathbf{x})),\quad \mathbf{x} \in O(\mathbf{x}_0)$$

也即

$$(f_{k+1}(\mathbf{x}), f_{k+2}(\mathbf{x}), ..., f_m(\mathbf{x})) = g(f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_k(\mathbf{x})),\quad \mathbf{x} \in O(\mathbf{x}_0). \tag{$***$}$$

因g在闭区间$\overline{I}$上连续, 根据连续映射的延拓定理(见本讲义第七章§7.4), 存在连续映射$G : \mathbb{R}^k \to \mathbb{R}^{m-k}$ 使得$G(u) = g(u)\ \ \forall\, u \in \overline{I}$. 取

$$F(\mathbf{y}) = |(y_{k+1}, y_{k+2}, ..., y_m) - G(y_1, y_2, ..., y_k)|,\quad \mathbf{y} = (y_1, y_2, ..., y_m) \in \mathbb{R}^m.$$

则显然$F:\mathbb{R}^m\to\mathbb{R}$ 连续且

$$F(f(\mathbf{x}))\equiv 0,\quad \mathbf{x}\in O(\mathbf{x}_0).$$

而对任意开集$V\supset f(O(\mathbf{x}_0))$, 因$f(\mathbf{x}_0)\in f(O(\mathbf{x}_0))\subset V$, 故存在$\varepsilon>0$ 使得$B(f(\mathbf{x}_0),\varepsilon)\subset V$. 任取$h=(h_1,h_2,...,h_{m-k})\in\mathbb{R}^{m-k}$ 满足$0<|h|<\varepsilon$, 并取$\mathbf{y}=(y_1,y_2,...,y_m)$ 其中

$$y_j=f_j(\mathbf{x}_0),\ j=1,2,...,k;\quad y_{k+j}=h_j+f_{k+j}(\mathbf{x}_0),\ j=1,2,...,m-k.$$

则$\mathbf{y}=(0,...,0,h_1,h_2,...,h_{m-k})+f(\mathbf{x}_0)\in B(f(\mathbf{x}_0),\varepsilon)\subset V$ 且有

$$\begin{aligned}&F(\mathbf{y})\\&=|h+(f_{k+1}(\mathbf{x}_0),f_{k+2}(\mathbf{x}_0),...,f_m(\mathbf{x}_0))-G(f_1(\mathbf{x}_0),f_2(\mathbf{x}_0),...,f_k(\mathbf{x}_0))|,\\&=|h+(f_{k+1}(\mathbf{x}_0),f_{k+2}(\mathbf{x}_0),...,f_m(\mathbf{x}_0))-g(f_1(\mathbf{x}_0),f_2(\mathbf{x}_0),...,f_k(\mathbf{x}_0))|\\&=|h|>0.\end{aligned}$$

所以函数组$f_1,f_2,...,f_m$ 在$\mathbf{x}_0$的邻域$O(\mathbf{x}_0)$ 内为函数相关.

最后令g_j 为g的第j个坐标函数, $j=1,2,...,m-k$, 则$(***)$可写为

$$f_{k+j}(\mathbf{x})\equiv g_j(f_1(\mathbf{x}),f_2(\mathbf{x}),...,f_k(\mathbf{x})),\quad x\in O(\mathbf{x}_0),\quad j=1,2,...,m-k.\qquad\Box$$

【作业题】

1. 设$\Omega\subset\mathbb{R}^n$为开集, $f:\Omega\to\mathbb{R}^m$ 属于C^1类. 证明$x\mapsto \mathrm{rank}f(x)\ (\equiv\mathrm{rank}f'(x))$ 是下半连续的, 即对任意$x_0\in\Omega$ 存在x_0的一个邻域使得当$x\in U(x_0)$时$\mathrm{rank}f(x)\ge\mathrm{rank}f(x_0)$. [注意, 因$\mathrm{rank}f(x)$是整数值函数, 所以"$\mathrm{rank}f(x)>\mathrm{rank}f(x_0)-\varepsilon\ \ (0<\varepsilon<1)$"与"$\mathrm{rank}f(x)\ge\mathrm{rank}f(x_0)$"是一回事.]

2. 设C^1映射$f:\mathbb{R}^n\to\mathbb{R}^n$满足$f\circ f=f$. 令$S=f(\mathbb{R}^n)$.

证明: $\mathrm{rank}f(x)\equiv$ 常数, $x\in S$. 此外证明对任意$x_0\in S$ 存在$\delta>0$ 使得$\mathrm{rank}f(x)=\mathrm{rank}f(x_0)$ for all $x\in B(x_0,\delta)$.

[提示几个关键词和步骤: 证明$S=\{x\in\mathbb{R}^n\mid f(x)=x\}$; 找出满足$A^2=A$的方阵$A$的秩与迹的关系; 上题结果, 等等.]

§8.9 $\mathbb{R}^n$中的曲面和条件极值理论

很多曲面/曲线是方程或方程组的零点的集合. 例如设a, b, c 不全为零, 则由方程

$$ax + by + cz - d = 0$$

确定的点(x, y, z)的集合便是$\mathbb{R}^3$中的一个平面. 而由方程组

$$\begin{cases} ax + by + cz = 0, \\ x^2 + y^2 + z^2 - R^2 = 0 \end{cases}$$

确定的点(x, y, z)的集合便是$\mathbb{R}^3$中的一个以原点为中心以$R(> 0)$为半径的圆周.

研究由方程或方程组定义的曲面/曲线与研究隐函数问题几乎是一回事. 换言之，隐函数定理(以及函数组相关性)也可以用几何的语言即用曲线或曲面的语言来描述. 为看清这点, 让我们进一步考察具有一般性的例子：

【例】 设$1 \le k < n$, 映射$F = (F_1, F_2, ..., F_{n-k})$在一个开集$\Omega \subset \mathbb{R}^n$内是$C^r$光滑的$(r \ge 1)$且集合

$$S = \{\mathbf{x} \in \Omega \,|\, F(\mathbf{x}) = 0\} \neq \emptyset.$$

假设F 在点$\mathbf{x}_0 \in S$ 处是满秩的, 即$\mathrm{rank}F'(\mathbf{x}_0) = n - k$. 调整自变量$\mathbf{x}$的分量位置后我们得到$\mathbf{x}$ 的分解

$$\mathbf{x} = (x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k}, \quad \mathbf{x}_0 = (x_0, y_0) \quad 满足 \quad \det F'_y(x_0, y_0) \neq 0.$$

由局部隐函数定理知, 存在$\delta_0 > 0, \eta_0 > 0$ 使得方程$F(\mathbf{x}) = 0$决定了唯一的光滑的局部隐函数：

$$x \mapsto y = f(x) \in B(y_0, \eta_0) \subset \mathbb{R}^{n-k}, x \in B(x_0, \delta_0) \subset \mathbb{R}^k,$$
$$f(x_0) = y_0; \quad B(x_0, \delta_0) \times B(y_0, \eta_0) \subset \Omega.$$

为了得到某种标准形式，我们做些技术处理. 取一个以$y_0 = (y_{01}, y_{02}, ..., y_{0,n-k})$为中心开区间

$$J(y_0) := \prod_{j=1}^{n-k} (y_{0j} - \eta, y_{0j} + \eta) \subset B(y_0, \eta_0).$$

由f的连续性和$f(x_0) = y_0$, 可以选取一个以$x_0 = (x_{01}, x_{02}, ..., x_{0k})$为中心的开区间$I(x_0)$,

$$I(x_0) = \prod_{i=1}^{k} (x_{0i} - \delta, x_{0i} + \delta) \subset B(x_0, \delta_0) \quad 使得 \quad f(I(x_0)) \subset J(y_0).$$

令$U(\mathbf{x}_0) = I(x_0) \times J(y_0)$, 则$U(\mathbf{x}_0) \subset \Omega$, 且由局部隐函数的唯一性和蕴含关系$x \in I(x_0) \Longrightarrow f(x) \in J(y_0)$ 有

$$S \cap U(\mathbf{x}_0) = \{(x, f(x)) \mid x \in I(x_0)\}.$$

换言之, $S \cap U(\mathbf{x}_0)$是一段显式曲线($k = 1$)或一块显式曲面($k \geq 2$). 由**命题8.41(光滑曲面的表示)**我们知道可以构造一个C^r类微分同胚$\varphi : U(\mathbf{x}_0) \to (-1, 1)^n$使得

$$\varphi(S \cap U(\mathbf{x}_0)) = (-1, 1)^k \times \{0\}, \qquad \varphi(\mathbf{x}_0) = 0. \qquad \Box$$

基于以上分析我们引入曲面及其维数的定义.

【$\mathbb{R}^n$中的曲面(子流形)的定义】设$k \in \{1, 2, ..., n\}, S \subset \mathbb{R}^n$.

(1) 称S 称为一个k维连续曲面, 如果对每一点$x_0 \in S$, 存在$\mathbb{R}^n$中的邻域$U(x_0) \ni x_0$ 使得$S \cap U(x_0)$ 与$(-1, 1)^k$ 同胚, 即存在同胚映射$\psi : S \cap U(x_0) \to (-1, 1)^k$. 此时也称$t \mapsto x(t) := \psi^{-1}(t), t \in (-1, 1)^k$, 是曲面$S$ 在$S \cap U(x_0)$ 上的一个参数表示. $\mathbb{R}^n$中的k维连续曲面也称为$\mathbb{R}^n$中的k维子流形. 当$k = 1$时, 1维连续曲面也称为连续曲线或1 维流形.

(2) 设r为正整数或$r = \infty$. 称S 为一个k维C^r类曲面, 如果对每一点$x_0 \in S$, 存在$\mathbb{R}^n$中的邻域$U(x_0) \ni x_0$ 和一个C^r类微分同胚$\varphi : U(x_0) \to (-1, 1)^n$, 使得$\varphi$把$S \cap U(x_0)$恰好映成$(-1, 1)^n$的$k$维子区间且把$x_0$ 映为原点, 即

$$\varphi(S \cap U(x_0)) = (-1, 1)^k \times \{0\}, \qquad \varphi(x_0) = 0.$$

此时也称映射$t \mapsto x(t) := \varphi^{-1}(t, 0), t \in (-1, 1)^k$, 是曲面$S$ 在$S \cap U(x_0)$ 上的一个C^r的参数表示. $\mathbb{R}^n$中的k维C^r类的曲面($r \geq 1$)也称为$\mathbb{R}^n$中的k维光滑子流形. 当$k = 1$时, 1维C^r类的曲面也称C^r类的曲线. $\Box$

【几点说明】

1. 比较定义中的(1),(2)看出, 光滑曲面的定义比连续曲面的定义要"苛刻"得多, 其主要原因是在曲面上一般无法"直接定义沿曲面的微分", 所以需要通过将曲面上一点的邻域扩充为全空间中的一个邻域来考察曲面在该点的"光滑行为". 此外不难看出光滑k维曲面一定是连续k维曲面.

2. 上述关于曲面的定义中没有考虑曲面带边的情形. 例如对于一段包含端点的不自交的曲线S, 除去S的两个端点外, S上的任何一点x_0 都具有定义中的性

质(对应于$k=1$), 而S的端点则不具有上述性质. 又例如$\mathbb{R}^3$中的单位球面的上半球面$S=\{(x,y,z)\in\mathbb{R}^3\,|\,x^2+y^2+z^2=1, z\ge 0\}$包含了它的边, 即包含球面的大圆周$\Gamma=\{(x,y,z)\in\mathbb{R}^3\,|\,x^2+y=1, z=0\}$. 这时$S_+=\{(x,y,z)\in\mathbb{R}^3\,|\,x^2+y^2+z^2=1, z>0\}$中的每一点具有定义中的性质(对应于$k=2$), 因此$S_+$是一个光滑2 维曲面. 但$S$的边$\Gamma$中的点就不具有上述性质. 关于带边的曲面以及流形和带边流形, 将在第13章欧空间中的流形中学习.

有些曲面/曲线是不带边的, 例如圆周是不带边的曲线, 完整的球面、轮胎是不带边的二维曲面. 对于带边的曲面, 本章只考虑曲面的边以外点, 这样的点也叫做曲面的内点.

3. 在上述曲面的定义中, 维数k 是一个公共数, 即粗略地说, S的每一点x_0的邻域$S\cap U(x_0)$都是k维的. 此外, 如果S含于$\mathbb{R}^n$的某个开集Ω内，则还可以要求邻域$U(x_0)$也含于Ω内.

4. 当$k=n$时, 定义中的0 不出现. 此时, $\mathbb{R}^n$ 以及任何与$\mathbb{R}^n$ 同胚的集合, 都是$\mathbb{R}^n$中的n维曲面.

5. 以上定义也可对$k=0$建立, 即可以定义零维曲面. 但这是比较平凡的，因为按此定义, 零维曲面便是一些**孤立点**. 特别, 当S还是连通集时, S作为零维曲面便是一个单点集. □

根据上述定义和上面的**例** 立即得到

【命题8.45.】设映射$F=(F_1,F_2,...,F_{n-k})$在一个开集$\Omega\subset\mathbb{R}^n$内属于$C^r$类, $r\ge 1$. 假设F 的零点的集合非空且F 在零点集上满秩, 即

$$S=\{x\in\Omega\,|\,F(x)=0\}\ \text{非空}, \qquad \mathrm{rank}F'(x)=n-k\quad \forall\, x\in S.$$

则由上面分析的结果和曲面维数的定义可知S便是$\mathbb{R}^n$中的一个k维C^r类曲面，并且可以要求S中的每点x的相应的邻域$U(x)$ 都含于Ω内. □

下面是$k=n-1$的特殊情形—— 超曲面. 【提醒注意: 如果不刻意区别行向量与列向量, 则数值函数的微分等于梯度.】

【例】(1) 设$\Omega\subset\mathbb{R}^3$ 为开区域, 函数$F:\Omega\to\mathbb{R}$ 属于C^r类($r\ge 1$). 假设

$$S=\{(x,y,z)\in\Omega\,|\ F(x,y,z)=0\}\ \text{非空}, \qquad \nabla F(x,y,z)\ne 0\quad \forall\,(x,y,z)\in S.$$

则S 便是$\mathbb{R}^3$ 中的一个C^r类的2维曲面.

(2) 一般地设$n \ge 2$, 设$\Omega \subset \mathbb{R}^n$ 为开区域, 函数$F : \Omega \to \mathbb{R}$ 属于C^r类$(r \ge 1)$. 假设

$$S = \{x \in \Omega \mid F(x) = 0\} \text{ 非空}, \qquad \nabla F(x) \ne 0 \qquad \forall\, x \in S.$$

则S 便是$\mathbb{R}^n$ 中的一个C^r类的超曲面($n-1$维流形). $\square$

【例】设$n \ge 2$. 称集合

$$\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n \mid |x| = 1\}$$

为$\mathbb{R}^n$中的单位球面. 令$F(x) = |x|^2 - 1$. 则F 是$\mathbb{R}^n$上的C^∞光滑函数且$\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n \,|\, F(x) = 0\}$. 因

$$\nabla F(x) = 2x \ne 0 \qquad \forall\, x \in \mathbb{S}^{n-1}$$

故$\mathbb{S}^{n-1}$是$\mathbb{R}^n$ 中的一个光滑的超曲面($n-1$维流形). $\square$

【例】令

$$S = \{(x, y) \in \mathbb{R}^2 \mid y^2 - x^2 = 0\}.$$

易见S是两条交叉的直线:

$$S = \{(x, y) \in \mathbb{R}^2 \,|\, y = x\} \cup \{(x, y) \in \mathbb{R}^2 \,|\, y = -x\}.$$

试通过证明交叉点$(0, 0) \in S$ 不具有曲面(子流形)定义中的性质, 来证明S不是流形.

【证】反证法. 假设存在$k \in \{0, 1, 2\}$ 使得S 是一个k维流形. 因S 没有孤立点, 故k必不等于0, 也即只能是$k = 1$ 或$k = 2$.

Case 1: $k = 1$. 据流形的定义知存在$\mathbb{R}^2$中的邻域$U((0, 0)) \ni (0, 0)$ 使得$S \cap U((0, 0))$ 与$(-1, 1)$ 同胚, 即存在同胚映射$\psi : S \cap U((0, 0)) \to (-1, 1)$. 为分析方便，我们取一个小圆盘$B = B((0, 0); \delta) \subset U((0, 0))$. 则易见$S \cap B$为一个连通集（实为道路连通集). 于是由$\psi$的连续性可知$\psi(S \cap B)$是$(-1, 1)$中的一个连通集. 因$\mathbb{R}$中的连通集是区间，故$\psi(S \cap B)$ 是一个实数区间(显然它不是单点集!). 易见集合$S \cap B$等于原点$(0, 0)$和四条互不相交的直线段的并:

$$S \cap B = \{(0, 0)\} \cup A_1 \cup A_2 \cup A_3 \cup A_4$$

其中

$$A_1 = \{(x, y) \in B \mid y = x,\, x > 0\}, \quad A_3 = \{(x, y) \in B \mid y = -x, x > 0\},$$

$$A_2 = \{(x,y) \in B \mid y = x,\, x < 0\}, \quad A_4 = \{(x,y) \in B \mid y = -x,\, x < 0\}.$$

易见A_1, A_2, A_3, A_4都是连通集且互不相交. 因$\psi : S \cap B \to \psi(S \cap B)$ 是连续单满射, 故得到分解

$$\psi(S \cap B) = \{t_0\} \cup I_1 \cup I_2 \cup I_3 \cup I_4 \subset (-1,1)$$

其中

$$t_0 = \psi(0,0), \quad I_i = \psi(A_i), \quad i = 1,2,3,4$$

由于$\{t_0\}, I_1, I_2, I_3, I_4$**互不相交** 并且每个$I_i$都是非空区间（因$A_i$连通并且$\psi$连续), 故不难看出, 即便$t_0$落在$I_1, I_2, I_3, I_4$中的某个$I_i$的端点上, 也不能使集合$\{t_0\} \cup I_1 \cup I_2 \cup I_3 \cup I_4$成为连通集(即区间). 这就与$\psi(S \cap B) = \{t_0\} \cup I_1 \cup I_2 \cup I_3 \cup I_4$是区间矛盾！

Case 2: $k = 2$. 据流形的定义知存在$\mathbb{R}^2$中的邻域$U((0,0)) \ni (0,0)$ 使得$S \cap U((0,0))$与$(-1,1)^2$ 同胚, 即存在同胚映射$\psi : S \cap U((0,0)) \to (-1,1)^2$.

因映射$\psi^{-1} : (-1,1)^2 \to S \cap U((0,0)) \subset \mathbb{R}^2$ 是连续单射, 故由Brouwer 区域不变定理知ψ^{-1} 是开映射. 因此$S \cap U((0,0)) = \psi^{-1}((-1,1)^2)$ 是$\mathbb{R}^2$的开集. 但显然S没有内点, 从而$S \cap U((0,0))$没有内点, 这就与“$S \cap U((0,0))$是$\mathbb{R}^2$的开集”矛盾. □

♠ **切空间**. 曲线或曲面的切空间概念已在本章开头做了简介. 现在我们对光滑曲面重述切空间的定义:

【定义(切空间)】设$S \subset \mathbb{R}^n$是一个k维光滑曲面$(1 \le k \le n)$. 设$x_0 \in S$. 则

$$TS_{x_0} = \{\gamma'(0) \mid \text{存在}\delta > 0 \text{ 和可微映射 } \gamma : (-\delta,\delta) \to S \text{ 使得 } \gamma(0) = x_0\}. \qquad \square$$

切空间的这种定义, 由于采用了运动学语言的描述, 比较直观。缺点是从它不易直接看出切空间TS_{x_0}是线性空间(稍后我们将证明TS_{x_0}是一个线性空间). 例如若$\xi_1 = \gamma_1'(0) \in TS_{x_0}, \xi_2 = \gamma_2'(0) \in TS_{x_0}$, 如何看出$\xi_1 + \xi_2 \in TS_{x_0}$? 眼下容易证明$TS_{x_0}$关于数乘是封闭的, 也即对任意$\xi \in TS_{x_0}$和任意$\lambda \in \mathbb{R}$ 都有$\lambda\xi \in TS_{x_0}$. 事实上由TS_{x_0}的定义知存在$\delta > 0$ 和可微曲线$\gamma : (-\delta,\delta) \to S$ 满足$\gamma(0) = x_0$ 使得$\gamma'(0) = \xi$. 取$\eta = \frac{\delta}{1+|\lambda|} > 0$. 考虑可微曲线$(-\eta,\eta) \ni \tau \mapsto \gamma_1(\tau) := \gamma(\lambda\tau) \in S$. 我们有$\gamma_1(0) = \gamma(0) = x_0$, $\gamma_1'(0) = \gamma'(0)\lambda = \lambda\xi$. 所以$\lambda\xi = \gamma_1'(0) \in TS_{x_0}$.

现在我们从另一角度看k维光滑曲面的切空间：设$x_0 \in S$, $U(x_0) \ni x_0$ 是$\mathbb{R}^n$中的邻域, 而设$\varphi : U(x_0) \to (-1,1)^n$是$k$维$C^r$类曲面定义中的$C^r$类同胚. 也即$\varphi(x_0) = 0$

且φ把$S\cap U(x_0)$恰好映成$(-1,1)^n$的k维子区间, 即$\varphi(S\cap U(x_0))=(-1,1)^k\times\{0\}$. 由$\varphi$是微分同胚可知逆映射$\varphi^{-1}:(-1,1)^n\to U(x_0)$ 的Jacobi矩阵

$$(\varphi^{-1})'(t,s)=\Big((\varphi^{-1})'_t(t,s),\,(\varphi^{-1})'_s(t,s)\Big),\quad (t,s)\in(-1,1)^k\times(-1,1)^{n-k}$$

处处可逆. 这蕴含

$$\mathrm{rank}(\varphi^{-1})'_t(t,s)\equiv k,\quad \mathrm{rank}(\varphi^{-1})'_s(t,s)\equiv n-k,\quad (t,s)\in(-1,1)^k\times(-1,1)^{n-k}.$$

特别有

$$\mathrm{rank}\varphi^{-1}(t,0)=\mathrm{rank}(\varphi^{-1})'_t(t,0)\equiv k,\quad t\in(-1,1)^k.$$

令

$$x(t)=\varphi^{-1}(t,0),\quad t\in(-1,1)^k.$$

则$x:(-1,1)^k\to S\cap U(x_0)$ 是C^r类映射且

$$x(0)=x_0,\quad \mathrm{rank}\,x(t)=k\qquad \forall\,t\in(-1,1)^k.$$

注意: 以上表明映射$t\mapsto x(t)\in S\cap U(x_0)$, $t\in(-1,1)^k$, 给出了k维光滑曲面S 在x_0 附近的一个参数表示. 由微分展开$x(t)-x_0=x'(0)t+o(|t|)$ 可知在这个参数表示下, 曲面S 在x_0 的切平面(或切线)方程为

$$x-x_0=x'(0)t,\quad t\in\mathbb{R}^k.$$

基于以上分析, 我们可以引入光滑曲面切空间的另一定义, 即证明下面的命题成立:

【命题8.46(切空间的表示).】设$S\subset\mathbb{R}^n$是一个k维光滑曲面$(1\le k\le n)$, $x_0\in S$, $(-1,1)^k\ni t\mapsto x(t)\in S\cap U(x_0)$是光滑曲面定义中所说的$S$ 在点$x_0\in S$ 的邻域内的任意一个光滑的参数形式. 则总有

$$TS_{x_0}=\{x'(0)t\mid t\in\mathbb{R}^k\}.$$

这表明

(1) 切空间TS_{x_0} 是一个k维实线性空间.

(2) TS_{x_0}的上述表示不依赖于光滑参数$(-1,1)^k\ni t\mapsto x(t)$的选择.

【证】沿用上面记号$\varphi:U(x_0)\to(-1,1)^n, x(t)=\varphi^{-1}(t,0)$等等, 并记

$$\widetilde{T}S_{x_0}=\{x'(0)t\mid t\in\mathbb{R}^k\}.$$

要证明$TS_{x_0}=\widetilde{T}S_{x_0}$.

任取$\xi\in TS_{x_0}$. 由定义, 存在$\delta>0$ 和可微曲线$(-\delta,\delta)\ni\tau\mapsto\gamma(\tau)\in S$ 满足$\gamma(0)=x_0$, $\gamma'(0)=\xi$. 取$0<\eta<\delta$充分小使得$\gamma(\tau)\in S\cap U(x_0)$ for all $\tau\in(-\eta,\eta)$. 作分解$\varphi(x)=(\varphi_1(x),\varphi_2(x))=\mathbb{R}^k\times\mathbb{R}^{n-k},x\in U(x_0)$. 因$\varphi(S\cap U(x_0))=(-1,1)^k\times\{0\}$, 故当$x\in S\cap U(x_0)$时有$\varphi(x)=(\varphi_1(x),0)$. 于是由$\gamma(\tau)\in S\cap U(x_0)$ 知$\varphi(\gamma(\tau))=\varphi_1(\gamma(\tau)),0)$ 从而由$x(t)=\varphi^{-1}(t,0)$得到

$$\gamma(\tau)=\varphi^{-1}(\varphi_1(\gamma(\tau),0))=x(\varphi_1(\gamma(\tau))),\quad \tau\in(\eta,\eta).$$

于是由复合映射的微分法有

$$\gamma'(\tau)=x'(\varphi_1(\gamma(\tau)))\varphi_1'(\gamma(\tau))\gamma'(\tau),\quad \tau\in(-\eta,\eta).$$

代入$\tau=0$并注意$\gamma(0)=x_0$, $\varphi_1(x_0)=0$, 即知

$$\xi=\gamma'(0)=x'(\varphi_1(x_0))\varphi_1'(x_0)\gamma'(0)=x'(0)\varphi_1'(x_0)\xi\in\widetilde{T}S_{x_0}.$$

这证明了$TS_{x_0}\subset\widetilde{T}S_{x_0}$.

反之对任意$\xi\in\widetilde{T}S_{x_0}$, 则存在$t\in(-1,1)^k$ 使得$\xi=x'(0)t$. 令$\gamma(\tau)=x(\tau t),\tau\in(-1,1)$, 则$\gamma$可微且$\gamma(\tau)\in S$ for all $\tau\in(-1,1)$. 由复合映射的微分法有

$$TS_{x_0}\ni\gamma'(0)=\gamma'(\tau)|_{\tau=0}=x'(\tau t)t|_{\tau=0}=x'(0)t=\xi.$$

这又证明了$\widetilde{T}S_{x_0}\subset TS_{x_0}$. 所以$\widetilde{T}S_{x_0}=TS_{x_0}$.

从等式$TS_{x_0}=\widetilde{T}S_{x_0}$易见$TS_{x_0}$是一个线性空间, 并由$\operatorname{rank} x'(0)=k$, 即Jacobi矩阵$x'(0)$满秩, 可知$TS_{x_0}$ 的维数$=k$. $\square$

【注】从切空间TS_{x_0}的定义和上述命题我们看到: TS_{x_0}只由x_0的邻域$S\cap U(x_0)$决定, 与$S\cap U(x_0)$之外的信息无关.

• **由方程或方程组定义的曲面的切空间**.

【命题8.47.】设开集$\Omega\subset\mathbb{R}^n$, 函数$F:\Omega\to\mathbb{R}$ 是光滑的(即属于C^r类, $r\ge 1$). 假设

$$S=\{\mathbf{x}\in\Omega\,|\,F(\mathbf{x})=0\}\ \text{非空且}\ \nabla F(x)\neq 0\quad\forall\mathbf{x}\in S.$$

则S 是一块光滑的超曲面(即S的维数$=n-1$) 且对任意$x_0\in S$ 有

$$TS_{\mathbf{x}_0}=\{\xi\in\mathbb{R}^n\mid\langle\nabla F(\mathbf{x}_0),\xi\rangle=0\}.$$

【**命题8.48.**】设开集$\Omega \subset \mathbb{R}^n$, 映射$F = (F_1, F_2, ..., F_{n-k}) : \Omega \to \mathbb{R}^{n-k}$是光滑的. 假设

$$S = \{\mathbf{x} \in \Omega \,|\, F(\mathbf{x}) = 0\} \text{ 非空 且 } \mathrm{rank} F(\mathbf{x}) = n - k \quad \forall \mathbf{x} \in S.$$

则S是一个k维光滑曲面且对任意$\mathbf{x}_0 \in S$ 有

$$TS_{\mathbf{x}_0} = \{\xi \in \mathbb{R}^n \mid F'(\mathbf{x}_0)\xi = 0\}.$$

【两个命题的证明】易见第一个命题是第二个命题的特殊情形(对应于$k = n - 1$), 其中注意当F是**数值函数**时, $F'(\mathbf{x}) = (\nabla F(\mathbf{x}))^\tau$, $F'(\mathbf{x})\xi = \langle \nabla F(\mathbf{x}), \xi \rangle$.

因此只需证第二个命题. 首先由第二个命题的假设和**命题8.45** 知S 是一个k维光滑曲面.

设$\mathbf{x}_0 \in S$. 调整自变量$\mathbf{x}$的分量的位置后我们得到$\mathbf{x}$的分解

$$\begin{aligned} &\mathbf{x} = (x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k}, \quad \mathbf{x}_0 = (x_0, y_0) \in \mathbb{R}^k \times \mathbb{R}^{n-k}, \\ &F'(\mathbf{x}_0) = \Big(F'_x(x_0, y_0) \;\; F'_y(x_0, y_0)\Big) \quad \text{满足} \quad \det F'_y(x_0, y_0) \neq 0. \end{aligned}$$

由局部隐函数定理知, 存在$\delta > 0, \eta > 0$ 使得方程$F(\mathbf{x}) = F(x, y) = 0$决定了唯一的光滑的局部隐函数

$$x \mapsto y = f(x) \in B(y_0, \eta) \subset \mathbb{R}^{n-k}, \; x \in B(x_0, \delta) \subset \mathbb{R}^k; \quad f(x_0) = y_0.$$

此外由隐函数的微分法和$f(x_0) = y_0$有

$$f'(x_0) = -[F'_y(x_0, y_0)]^{-1} F'_x(x_0, y_0).$$

记

$$\mathrm{Ker}(F'(\mathbf{x}_0)) = \{\xi \in \mathbb{R}^n \mid F'(\mathbf{x}_0)\xi = 0\}.$$

要证明$TS_{\mathbf{x}_0} = \mathrm{Ker}(F'(\mathbf{x}_0))$.

对任意$\xi \in TS_{\mathbf{x}_0}$, 由切空间的定义, 存在$\delta > 0$ 和可微曲线$(-\delta, \delta) \ni t \mapsto \gamma(t) \in S$ 满足$\gamma(0) = \mathbf{x}_0$, $\gamma'(0) = \xi$. 因γ位于曲面S上, 即$F(\gamma(t)) \equiv 0$, 故由复合求导有

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} F(\gamma(t))\Big|_{t=0} = F'(\gamma(0))\dot{\gamma}(0) = F'(\mathbf{x}_0)\xi.$$

所以$\xi \in \mathrm{Ker}(F'(\mathbf{x}_0))$. 这证明了$TS_{\mathbf{x}_0} \subset \mathrm{Ker}(F'(\mathbf{x}_0))$.

反之对任意$\xi \in \mathrm{Ker}(F'(\mathbf{x}_0))$, 做分解

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \quad \xi_1 \in \mathbb{R}^k, \quad \xi_2 \in \mathbb{R}^{n-k}.$$

则有

$$\begin{aligned}
& 0 = F'(\mathbf{x}_0)\xi = \Big(F'_x(x_0,y_0) \;\; F'_y(x_0,y_0)\Big) \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = F'_x(x_0,y_0)\xi_1 + F'_y(x_0,y_0)\xi_2, \\
& \Longrightarrow \qquad \xi_2 = -[F'_y(x_0,y_0)]^{-1}F'_x(x_0,y_0)\xi_1 = f'(x_0)\xi_1, \\
& \Longrightarrow \qquad \xi = \begin{pmatrix} \xi_1 \\ f'(x_0)\xi_1 \end{pmatrix}.
\end{aligned}$$

取$\delta > 0$ 充分小使得$x_0 + t\xi_1 \in B(x_0, \delta)$ for all $t \in (-\delta, \delta)$. 则有

$$F(x_0 + t\xi_1, f(x_0 + t\xi_1)) \equiv 0, \quad t \in (-\delta, \delta).$$

这给出一条位于S 上的光滑曲线:

$$(-\delta, \delta) \ni t \mapsto \gamma(t) := (x_0 + t\xi_1,\, f(x_0 + t\xi_1)) \in S; \quad \gamma(0) = (x_0, y_0) = \mathbf{x}_0.$$

按列向量计算有

$$\gamma'(t) = \begin{pmatrix} \xi_1 \\ f'(x_0 + t\xi_1)\xi_1 \end{pmatrix} \quad \Longrightarrow \quad \gamma'(0) = \begin{pmatrix} \xi_1 \\ f'(x_0)\xi_1 \end{pmatrix} = \xi.$$

由切空间的定义知$\xi = \gamma'(0) \in TS_{\mathbf{x}_0}$. 因此$\mathrm{Ker}(F'(\mathbf{x}_0)) \subset TS_{\mathbf{x}_0}$. 所以$\mathrm{Ker}(F'(\mathbf{x}_0)) = TS_{\mathbf{x}_0}$. □

【条件极值理论】

所谓条件极值问题, 就是研究函数在一些约束下的极值问题, 或等价地, 研究函数限制在其定义域的某个子集上的极值问题. 例如研究当容器的容量限定后, 何时容器的表面材料最省. 条件极值理论的重点是针对这种子集的特性给出求解条件极值的一些可操作的方法. 首先引进一般定义.

【定义】设$\Omega \subset \mathbb{R}^n$为开集, 函数$f: \Omega \to \mathbb{R}$, 以及集合$S \subset \Omega$.

(1) 如果存在x_0的一个邻域$U(x_0) \subset \Omega$ 使得 (图示)

$$f(x) \ge f(x_0) \qquad \forall\, x \in S \cap U(x_0)$$

$$[\quad f(x) \le f(x_0) \qquad \forall x \in S \cap U(x_0) \quad]$$

则称x_0 为$f|_S$的一个**局部**极小[大] 值点同时称为$f(x_0)$为$f|_S$在的一个**局部**极小[大] 值.

(2) 如果

$$f(x) \ge f(x_0) \qquad \forall x \in S$$

$$[\quad f(x) \le f(x_0) \qquad \forall x \in S \quad]$$

则称x_0 为$f|_S$的一个**整体**最小[大] 值点同时称$f(x_0)$为$f|_S$的一个**整体**最小[大] 值.

(3) 类似地可以定义$f|_S$的局部**严格**极小[大] 值点和极小[大] 值以及整体**严格**最小[大] 值点和最小[大] 值. □

在以上定义中, f称为目标函数, 集合S称为f的约束集.

在很多问题中, 约束集S是由一组方程给出的:

$$\begin{cases} F_1(x_1, x_2, ..., x_n) = 0 \\ F_2(x_1, x_2, ..., x_n) = 0 \\ .... \\ F_m(x_1, x_2, ..., x_n) = 0 \end{cases}$$

其中$m \le n$. 多数情况下是$m < n$ (**即方程的个数<未知量的个数**)且

$$F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{pmatrix}$$

满足一定条件, 以保证这个约束非空, 也即S 是由$F(x) = 0$决定的非空曲面:

$$S = \{x \in \Omega \,|\; F(x) = 0\} \ne \emptyset.$$

在这类条件极值问题中, $F = (F_1, F_2, ..., F_m)^\tau$称为约束函数. 如写成紧凑形式, 则此类条件极值问题可以笼统地表为

$$\min_{F(x)=0} f(x) = ? \quad 或 \quad \max_{F(x)=0} f(x) = ?$$

条件极值的求解方法:

(I) **代入法**. 设$m < n$. 如果从约束方程$F(x) = 0$ 可以解出隐函数, 例如解出

$$\begin{cases} x_{m+1} = g_1(x_1, x_2, ..., x_m) \\ x_{m+2} = g_2(x_1, x_2, ..., x_m) \\ .... \\ x_n = g_{n-m}(x_1, x_2, ..., x_m) \end{cases}$$

则有恒等式(至少是局部恒等式)

$$F\Big(x_1, x_2, ..., x_m, g_1(x_1, x_2, ..., x_m), g_2(x_1, x_2, ..., x_m), ..., g_{n-m}(x_1, x_2, ..., x_m)\Big) \equiv 0$$

$(x_1, x_2, ..., x_m)$属于某个区域G. 将$x_{m+i} = g_i(x_1, x_2, ..., x_m)\ (i = 1, 2, ..., n-m)$ 代入$f(x_1, x_2, ..., x_m, x_{m+1}, ..., x_n)$中便化成了求复合函数

$$(x_1, x_2, ..., x_m) \mapsto f\Big(x_1, x_2, ..., x_m, g_1(x_1, x_2, ..., x_m), g_2(x_1, x_2, ..., x_m), ..., g_{n-m}(x_1, x_2, ..., x_m)\Big)$$

在G 内的无条件极值(即普通极值)问题.

【例】求平面上的一点(a, b) 到曲线$y = (x-1)^2$的最短距离.[6]

【解】此问题是求最小值

$$\min\{\sqrt{(x-a)^2 + (y-b)^2}\ \Big|\ y = (x-1)^2\} = \min_{y=(x-1)^2} \sqrt{(x-a)^2 + (y-b)^2}.$$

因函数$u \mapsto \sqrt{u}$在$[0, +\infty)$ 严格单调增加, 故

$$\min_{y=(x-1)^2} \sqrt{(x-a)^2 + (y-b)^2} = \sqrt{\min_{y=(x-1)^2} (x-a)^2 + (y-b)^2}.$$

所以只需求解最小值问题

$$\min_{y=(x-1)^2} [(x-a)^2 + (y-b)^2].$$

此问题的目标函数为$f(x, y) = (x-a)^2 + (y-b)^2$，约束方程为$F(x, y) = y - (x-1)^2 = 0$. 将$y = (x-1)^2$代入$f(x, y)$ 便化为求复合函数

$$\varphi(x) = (x-a)^2 + \Big((x-1)^2 - b\Big)^2, \quad x \in \mathbb{R}$$

的普通的最小值问题. 因

$$\lim_{x\to\pm\infty} \varphi(x) = +\infty$$

[6]这里“最短”二字从逻辑上来说可以省略, 因为“距离”本身已经是用下确界定义的了. 但是习惯上有时还是用这种说法.

故φ 确有最小值. 因而其最小值点必定是临界点. 求解临界点方程：

$$0=\varphi'(x)=2(x-a)+2\Big((x-1)^2-b\Big)2(x-1).$$

上式右边是一个实系数一元三次多项式, 故它有实根且至多有3个实根. 计算函数$\varphi(x)$在这些实根处的值, 它们中的最小者即为所求的最短距离.

(II) **Lagrange 乘子法**(需注意：此法只是给出了条件极值存在的必要条件, 应用时还要结合其他条件).

首先给出取得条件极值的必要条件及其几何描述:

【命题8.49(条件极值的必要条件及其几何描述).】设开集$\Omega\subset\mathbb{R}^n$, 目标函数$f\in C^1(\Omega)$, $S\subset\Omega$ 是一块光滑曲面. 如果$x_0\in S$ 是$f|_S$的局部极值点, 则有

$$TS_{x_0}\subset\{\xi\in\mathbb{R}^n\mid f'(x_0)\xi=0\}$$

其中TS_{x_0}是曲面S在点x_0的切空间. 如令$N=H_{f(x_0)}=\{x\in\Omega\,|\,f(x)=f(x_0)\}$ 为f的以$f(x_0)$为高度的等高集, 则当$f'(x_0)\neq 0$ (从而切空间TN_{x_0} 有意义)时, 上式可写为

$$TS_{x_0}\subset TN_{x_0}.$$

【注】当x_0是列向量时, 数值函数f的微分$f'(x_0)$ 是行向量. 在不计转置的意义下, 它实际上等于梯度$\nabla f(x_0)$, 因而数$f'(x_0)\xi$也可以理解为内积$\langle\nabla f(x_0),\xi\rangle$.

【证】由切空间的定义, 对任意$\xi\in TS_{x_0}$, 存在$\delta>0$和可微曲线$\gamma:(-\delta,\delta)\to S$满足$\gamma(0)=x_0$和$\gamma'(0)=\xi$. 因$x_0$ 是$f|_S$的一个极值点, 故函数$(-\delta,\delta)\ni t\mapsto f(\gamma(t))$在$t=0$取得极值. 据无条件极值的必要条件和复合函数求导法则有

$$0=\frac{\mathrm{d}}{\mathrm{d}t}f(\gamma(t))\Big|_{t=0}=f'(\gamma(t))\gamma'(t)\Big|_{t=0}=f'(x_0)\xi.$$

这证明了$TS_{x_0}\subset\{\xi\in\mathbb{R}^n\mid f'(x_0)\xi=0\}$.

现在设$f'(x_0)\neq 0$. 由f属于C^1类知$x\mapsto f'(x)$ 连续. 因此存在x_0的一个邻域$U(x_0)\subset\Omega$使得$f'(x)\neq 0$ for all $x\in U(x_0)$. 据**命题8.47**知等高集$\widetilde{N}:=N\cap U(x_0)$ 是一块$n-1$维的C^1 曲面(即C^1 超曲面)并且

$$T\widetilde{N}_{x_0}=\{\xi\in\mathbb{R}^n\,|\,f'(x_0)\xi=0\}.$$

但$\widetilde{N} = N \cap U(x_0)$, 故曲面$N$ 与$\widetilde{N}$在x_0点有相同的切空间. 因此

$$TS_{x_0} \subset \{\xi \in \mathbb{R}^n \mid f'(x_0)\xi = 0\} = T\widetilde{N}_{x_0} = TN_{x_0}. \qquad \square$$

现在设上述命题中的约束集S由光滑约束函数$F = (F_1, F_2, ..., F_m)^\tau$给出$(m < n)$, 即

$$S = \{x \in \Omega \mid F(x) = 0\}, \qquad x_0 \in S$$

且假定F在S中的每一点都是满秩的, 即rank$F'(x) = m$ for all $x \in S$. [也即S是$\mathbb{R}^n$中的一个$n - m$维光滑曲面.] 则由**命题8.48** 知

$$TS_{x_0} = \{\xi \in \mathbb{R}^n \mid F'(x_0)\xi = 0\}.$$

于是上述几何关系可写为

$$\{\xi \in \mathbb{R}^n \mid F'(x_0)\xi = 0\} \subset \{\xi \in \mathbb{R}^n \mid f'(x_0)\xi = 0\}.$$

这等于说两个线性齐次方程组

$$\begin{cases} F'(x_0)\xi = 0 \\ f'(x_0)\xi = 0 \end{cases} \qquad \text{and} \qquad F'(x_0)\xi = 0$$

的解的集合相同. 据线性代数知这两个系数矩阵的秩相同:

$$\text{rank}\begin{pmatrix} F'(x_0) \\ f'(x_0) \end{pmatrix} = \text{rank}F'(x_0).$$

因此行向量$f'(x_0)$等于$F'(x_0)$中的行向量的线性组合, 即存在行向量$\lambda \in \mathbb{R}^m$ 使得

$$f'(x_0) = \lambda F'(x_0).$$

即

$$f'(x_0) = (\lambda_1, \lambda_2, ..., \lambda_m)\begin{pmatrix} {F_1}'(x_0) \\ {F_2}'(x_0) \\ \vdots \\ {F_m}'(x_0) \end{pmatrix} = \sum_{i=1}^m \lambda_i {F_i}'(x_0)$$

如用梯度表示, 则这一关系可表示为

$$\nabla f(x_0) = \sum_{i=1}^m \lambda_i \nabla F_i(x_0)$$

也即
$$\frac{\partial f}{\partial x_j}(x_0)=\sum_{i=1}^{m}\lambda_i\frac{\partial F_i}{\partial x_j}(x_0),\quad j=1,2,...,m.$$
引进Lagrange 函数:
$$L(x,\lambda)=f(x)-\sum_{i=1}^{m}\lambda_iF_i(x),\quad (x,\lambda)\in\Omega\times\mathbb{R}^m.$$
我们有

【定理8.50(条件极值的Lagrange 乘子法(件极值的必要条件).】

设开集$\Omega\subset\mathbb{R}^n$, 目标函数$f\in C^1(\Omega)$, 约束函数$F=(F_1,F_2,...,F_m)^\tau(m<n)$属于$C^1(\Omega)$且$F$在$S=\{x\in\Omega\,|\,F(x)=0\}$上满秩, 即$\mathrm{rank}F'(x)=m$ for all $x\in S$. 如果$x_0\in S$是$f|_S$的局部极值点, 则必存在唯一的$\lambda=(\lambda_1,\lambda_2,...,\lambda_m)\in\mathbb{R}^m$ 使得(x_0,λ) 是Lagrange函数$L(x,\lambda)$的临界点, 即
$$\nabla L(x_0,\lambda)=0$$
也即
$$\begin{cases}\frac{\partial L}{\partial x_j}(x_0,\lambda)=0,\quad j=1,2,...,n;\\ \frac{\partial L}{\partial \lambda_i}(x_0,\lambda)=0,\quad i=1,2,...,m\end{cases}$$
即
$$\begin{cases}\frac{\partial f}{\partial x_j}(x_0)-\sum_{i=1}^{m}\lambda_i\frac{\partial F_i}{\partial x_j}(x_0)=0,\quad j=1,2,...,n;\\ F_i(x_0)=0,\quad i=1,2,...,m.\end{cases}$$
这个λ 称为(相应于x_0)的Lagrange乘子. [当$m=1$, λ也称为Lagrange乘数.]

【证】Lagrange乘子λ的存在性已在上面证明了. 下证λ的唯一性. 设$\widetilde{\lambda}=(\widetilde{\lambda}_1,\widetilde{\lambda}_2,...,\widetilde{\lambda}_m)$也是相应于$x_0$的Lagrange乘数. 则有
$$\sum_{i=1}^{m}(\widetilde{\lambda}_i-\lambda_i)\frac{\partial F_i}{\partial x_j}(x_0)=0,\quad j=1,2,...,n.$$
也即
$$(\widetilde{\lambda}-\lambda)F'(x_0)=0.$$
由假设$F'(x_0)\in\mathbb{R}^{m\times n}$ $(m<n)$满秩可知$\widetilde{\lambda}-\lambda=0$. 此即$\widetilde{\lambda}=\lambda$. □

【命题8.51.】设开集$\Omega\subset\mathbb{R}^n$, 目标函数$f\in C^1(\Omega)$, 约束函数$F=(F_1,F_2,...,F_m)_\tau$ $(m<n)$属于$C^1(\Omega,\mathbb{R}^m)$ 且F在非空集合$S=\{x\in\Omega\,|\,F(x)=0\}$上满秩, 即$\mathrm{rank}F'(x)=m$ for all $x\in S$. 假设S 是紧集, 则$f|_S$ 必有最小值和最大值.

设$\mathbf{a}, \mathbf{b} \in S$ 分别是$f|_S$的最小值点和最大值点. 则存在$\mu = (\mu_1, \mu_2, ..., \mu_m), \nu = (\nu_1, \nu_2, ..., \nu_m) \in \mathbb{R}^m$使得$(\mathbf{a}, \mu), (\mathbf{b}, \nu)$ 是Lagrange函数$L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i F_i(x)$的临界点, 即$(\mathbf{a}, \mu), (\mathbf{b}, \nu)$ 满足临界点方程

$$\begin{cases} \frac{\partial f}{\partial x_j}(x) - \sum_{i=1}^m \lambda_i \frac{\partial F_i}{\partial x_j}(x) = 0, \quad j = 1, 2, ..., n; \\ F_i(x) = 0, \quad i = 1, 2, ..., m. \end{cases}$$

【证】留为**作业题1**. $\square$

【例】设$A \in \mathbb{R}^{n\times n}$ 为对称阵, 设λ_1, λ_n 分别是A的最小特征值和最大特征值. 则有

$$\min_{|x|=1} x^\tau Ax = \lambda_1, \quad \max_{|x|=1} x^\tau Ax = \lambda_n.$$

【证】将此问题与条件极值问题的一般陈述对照, 有

$$\Omega = \mathbb{R}^n, \quad f(x = x^\tau Ax, \quad F(x) = |x|^2 - 1, \quad S = \{x \in \mathbb{R}^n \,|\, F(x) = 0\}.$$

因$F'(x) = 2x^\tau \neq 0$ for all $x \in S$, 也即$\text{rank}F'(x) \equiv 1$ (满秩) for all $x \in S$, 故S 是一个光滑曲面(实际上S 是单位球面). 显然S是紧集(非空).因此$f|_S$ 有最大值最小值：存在$\mathbf{a}, \mathbf{b} \in S$ 使得

$$f(\mathbf{a}) = \min_{|x|=1} f(x), \quad f(\mathbf{b}) = \min_{|x|=1} f(x).$$

据Lagrange 乘数法(见上一命题), 存在$\mu, \nu \in \mathbb{R}$使得$(\mathbf{a}, \mu), (\mathbf{b}, \nu)$ 是Lagrange函数$L(x, \lambda) = f(x) - \lambda F(x)$的临界点, 即$(\mathbf{a}, \mu), (\mathbf{b}, \nu)$ 满足临界点方程

$$\begin{cases} \frac{\partial f}{\partial x_j}(x) - \lambda \frac{\partial F}{\partial x_j}(x) = 0, \quad j = 1, 2, ..., n; \\ F(x) = 0. \end{cases}$$

即

$$\begin{cases} f'(x) - \lambda F'(x) = 0, \\ F(x) = 0. \end{cases} \quad \text{i.e.} \quad \begin{cases} \langle x, Ax\rangle' - \lambda(|x|^2 - 1)' = 0, \\ |x|^2 = 1. \end{cases}$$

计算：

$$\langle x, Ax\rangle' = (Ax)^\tau (x)' + x^\tau (Ax)' = (Ax)^\tau + x^\tau A = 2x^\tau A,$$

$$(|x|^2 - 1)' = (|x|^2)' = 2x^\tau.$$

所以临界点方程为

$$\begin{cases} x^\tau A = \lambda x^\tau \\ |x|^2 = 1. \end{cases} \quad \text{i.e.} \quad \begin{cases} Ax = \lambda x \\ |x|^2 = 1. \end{cases}$$

这表明Lagrange函数$L(x,\lambda) = f(x) - \lambda F(x)$的临界点$(x,\lambda)$中的$\lambda$必是$A$的特征值, 而$x$是从属于$\lambda$的特征向量(已归一化). 因$(\mathbf{a},\mu),(\mathbf{b},\nu)$ 是$L(x,\lambda)$的临界点, 故μ,ν 都是A的特征值而$\mathbf{a},\mathbf{b}$分别是从属于μ,ν的特征向量(已单位化). 设λ_1,λ_n 分别是A的最小和最大特征值, 则由

$$f(\mathbf{a}) = \mathbf{a}^\tau A\mathbf{a} = \mu\mathbf{a}^\tau\mathbf{a} = \mu, \quad f(\mathbf{b}) = \mathbf{b}^\tau A\mathbf{b} = \mu\mathbf{b}^\tau\mathbf{b} = \nu$$

可知$f(\mathbf{a}) = \mu \ge \lambda_1$, $f(\mathbf{b}) = \nu \le \lambda_n$. 而易见对于$A$的任意特征值$\lambda$ 和从属于λ的特征向量ξ 有(假定$|\xi| = 1$) $f(\xi) = \xi^\tau A\xi = \lambda\xi^\tau\xi = \lambda$. 因此

$$\mu = f(\mathbf{a}) \le \lambda \le f(\mathbf{b}) = \nu.$$

分别取$\lambda = \lambda_1, \lambda = \lambda_n$ 即知$f(\mathbf{a}) = \mu = \lambda_1, f(\mathbf{b}) = \nu = \lambda_n$. 这就证明了上述结论. □

【**例**】设$\beta_i > 0$, $i = 1,2,...,n; \sum_{i=1}^n \beta_i = 1$. 则对任意$a_1 \ge 0, a_2 \ge 0, ..., a_n \ge 0$ 有

$$a_1^{\beta_1}a_2^{\beta_2}\cdots a_n^{\beta_n} \le \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_n a_n.$$

【**证**】先考虑约束极值问题:

$$\max_{x_i\ge 0, \beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n = A} x_1^{\beta_1}x_2^{\beta_2}\cdots x_n^{\beta_n} = ?$$

然后应用于本问题. 细节留为**作业题2**. 注意, 本题中对应的开集、约束函数和曲面S分别为

$$\Omega = \{x = (x_1,x_2,...,x_n)\,|\,x_i > 0,\ i = 1,2,...,n\},\ \ F(x) = \beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n - A,$$

$$S = \{x \in \Omega\,|\,F(x) = 0\} \qquad (\text{注意}S\ \text{不含边}).$$

为了应用条件极值理论, 需先证明上述条件最大值点可以在S中取到！也即需先证明

$$\max_{x_i\ge 0, \beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n = A} x_1^{\beta_1}x_2^{\beta_2}\cdots x_n^{\beta_n} = \max_{x\in S} x_1^{\beta_1}x_2^{\beta_2}\cdots x_n^{\beta_n}.$$

与无条件极值类似，我们将在条件极值的必要条件基础上考察相应的Hesse 矩阵的定性, 以期获得条件极值的充分条件.

【**定理8.52.**】设开集$\Omega \subset \mathbb{R}^n$, 目标函数$f \in C^2(\Omega)$, 约束函数$F = (F_1,F_2,...,F_m)^\tau$ $(m < n)$ 属于$C^2(\Omega,\mathbb{R}^m)$ 且F在$S = \{x \in \Omega\,|\ F(x) = 0\}$上满秩, 即$\mathrm{rank}F'(x) = m$ for all $x \in S$. 设$x_0 \in S$ 满足f 在S上取得极值的必要条件, 即存在$\lambda = (\lambda_1, \lambda_2, ..., \lambda_m) \in \mathbb{R}^m$ 使得(x_0, λ) 是Lagrange函数

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i F_i(x)$$

的一个临界点, 即$\nabla L(x_0, \lambda) = 0$, 即$(x_0, \lambda)$ 满足方程

$$\begin{cases} \frac{\partial f}{\partial x_j}(x_0) - \sum_{i=1}^m \lambda_i \frac{\partial F_i}{\partial x_j}(x_0) = 0, \quad j = 1, 2, ..., m; \\ \\ F_i(x_0) = 0, \quad i = 1, 2, ..., m. \end{cases}$$

考察映射$x \mapsto L(x, \lambda)$在x_0的Hesse矩阵

$$H_L(x_0) = \Big(\frac{\partial^2 L(x_0, \lambda)}{\partial x_i \partial x_j}\Big)_{n \times n}.$$

我们有

(1) 若$H_L(x_0)$在在切空间$TS_{x_0} = \{\xi \in \mathbb{R}^n \,|\, F'(x_0)\xi = 0\}$ 上是正定的，即

$$\xi^\tau H_L(x_0)\xi > 0 \qquad \forall \xi \in TS_{x_0} \setminus \{0\}$$

则x_0是$f|_S$的一个严格局部极小值点.

若$H_L(x_0)$在在切空间$TS_{x_0} = \{\xi \in \mathbb{R}^n \,|\, F'(x_0)\xi = 0\}$ 上是负定的, 即

$$\xi^\tau H_L(x_0)\xi < 0 \qquad \forall \xi \in TS_{x_0} \setminus \{0\}$$

则x_0是$f|_S$的一个严格局部极大值点.

(2) 若x_0 是$f|_S$的一个局部极小(大)值点, 则Hesse 矩阵$H_L(x_0)$在在切空间TS_{x_0} 上是半正定的(半负定的), 即

$$\xi^\tau H_L(x_0)\xi \ge 0 \ \ (\le 0\,) \qquad \forall \xi \in TS_{x_0}.$$

(3) 若$H_L(x_0)$在在切空间TS_{x_0} 上是不定的,也即二次型$\xi \mapsto \xi^\tau H_L(x_0)\xi$ 在TS_{x_0}上变号, 则x_0 不是$f|_S$的极值点.

【证】由假设知S 是$n-m$维C^2类曲面. 故存在$\mathbb{R}^n$中的x_0的邻域$U(x_0)$ 和一个C^2类同胚$\varphi = (\varphi_1, \varphi_2, ..., \varphi_n)^\tau : U(x_0) \to (-1, 1)^n$, 使得$\varphi(x_0) = 0$ 且$\varphi(S \cap U(x_0)) =$

$(-1,1)^{n-m}\times\{0\}$. 令$x(t)=\varphi^{-1}(t,0), t\in(-1,1)^{n-m}$. 则映射$t\mapsto x(t)$ 是$(-1,1)^{n-m}$到$S\cap U(x_0)$的单满射且属于C^2类. 此外有

$$x(0)=x_0,\quad \operatorname{rank} x'(t)=n-m \qquad \forall\, t\in(-1,1)^{n-m}.$$

特别有$\operatorname{rank} x'(0)=n-m$. 令$A=x'(0)$. 则据切空间的定义有

$$TS_{x_0}=\{At\mid t\in\mathbb{R}^{n-m}\}.$$

考察Hesse 矩阵$H_L(x_0)$在线性子空间TS_{x_0}上的行为, 也即考察二次型$\xi\mapsto\xi^\tau H_L(x_0)\xi$在$TS_{x_0}$上的限制:

$$\xi^\tau H_L(x_0)\xi=(At)^\tau H_L(x_0)At=t^\tau A^\tau H_L(x_0)At,\quad t\in\mathbb{R}^{n-m}.$$

从这一关系式易见

$H_L(x_0)$在TS_{x_0} 上是正定(负定)的$\Longleftrightarrow$ $A^\tau H_L(x_0)A$是正定(负定)矩阵.

$H_L(x_0)$在TS_{x_0} 上是半正定(半负定)的$\Longleftrightarrow$ $A^\tau H_L(x_0)A$是半正定(半负定)矩阵.

为进一步分析, 我们来证明$\forall 0<\varepsilon<1$ 存在$\delta_\varepsilon>0$ 使得$B(x_0,\delta_\varepsilon)\subset\Omega\cap U(x_0)$且$S\cap B(x_0,\delta_\varepsilon)\subset\varphi^{-1}((-\varepsilon,\varepsilon)^{n-m}\times\{0\})$. 事实上由$\varphi(x_0)=0$ 和φ的连续性存在$\delta_\varepsilon>0$使得$B(x_0,\delta_\varepsilon)\subset\Omega\cap U(x_0)$ 且当$x\in S\cap B(x_0,\delta_\varepsilon)$ 也即当$x\in S$满足$|x-x_0|<\delta_\varepsilon$时$|\varphi(x)|=|\varphi(x)-\varphi(x_0)|<\varepsilon$. 因$\varphi(x)\in(-1,1)^{n-m}\times\{0\}$, 故这蕴含$|\varphi_i(x)|<\varepsilon, i=1,2,...,n-m; \varphi_i(x)=0, i=n-m+1,...,n$. 因此$\varphi(x)\in(-\varepsilon,\varepsilon)^{n-m}\times\{0\}$. 这证明了$\varphi(S\cap B(x_0,\delta_\varepsilon))\subset(-\varepsilon,\varepsilon)^{n-m}\times\{0\}$, 也即等价地, $S\cap B(x_0,\delta_\varepsilon)\subset\varphi^{-1}((-\varepsilon,\varepsilon)^{n-m}\times\{0\})$.

以上说明：每个$x\in S\cap B(x_0,\delta_\varepsilon)$皆可表示成$x=x(t)=\varphi^{-1}(t,0), t\in(-\varepsilon,\varepsilon)^{n-m}$.

注意到$x\in S$ 蕴含$L(x,\lambda)=f(x)$, 并且由假设有

$$\nabla_x L(x_0,\lambda)=\nabla f(x_0)-\sum_{i=1}^m\lambda_i\nabla F_i(x_0)=0.$$

因此对函数$x\mapsto L(x,\lambda)$ 应用二阶Taylor 公式可知当$x\in S\cap B(x_0,\delta_\varepsilon)$ 时

$$\begin{aligned}&f(x)-f(x_0)=L(x,\lambda)-L(x_0,\lambda)\\&=\langle\nabla_x L(x_0,\lambda),x-x_0\rangle+\frac{1}{2}(x-x_0)^\tau H_L(x_0)(x-x_0)+o(|x-x_0|^2)\\&=\frac{1}{2}(x-x_0)^\tau H_L(x_0)(x-x_0)+o(|x-x_0|^2).\end{aligned}$$

由上面分析, 每个$x \in S \cap B(x_0, \delta_\varepsilon)$ 可表示为$x = x(t) = \varphi^{-1}(t, 0), t \in (-\varepsilon, \varepsilon)^{n-m}$. 于是将$x - x_0 = x(t) - x(0) = At + o(|t|)$ 代入上式得到

$$f(x) - f(x_0) = \frac{1}{2} t^\tau A^\tau H_L(x_0) At + o(|t|^2), \quad t \in (-\varepsilon, \varepsilon)^{n-m}.$$

(1) 假设$H_L(x_0)$ 在TS_{x_0} 上是正定的, 也即$A^\tau H_L(x_0)A$是正定矩阵. 则存在$\alpha > 0$ 使得$t^\tau A^\tau H_L(x_0)At \ge \alpha |t|^2$ for all $t \in \mathbb{R}^{n-m}$. 取$0 < \varepsilon < 1$ 充分小使得

$$|o(|t|^2)| \le \frac{\alpha}{4} |t|^2 \qquad \forall\, t \in (-\varepsilon, \varepsilon)^{n-m}.$$

则有对任意$x \in S \cap B(x_0, \delta_\varepsilon)$ 有

$$f(x) \ge f(x_0) + \frac{\alpha}{2}|t|^2 + o(|t|^2) \ge f(x_0) + \frac{\alpha}{4}|t|^2, \quad t \in (-\varepsilon, \varepsilon)^{n-m}.$$

但是注意到: $x \ne x_0$ 等价于$x(t) \ne x(0)$也即等价于$\varphi^{-1}(t, 0) \ne \varphi^{-1}(0, 0)$, 也即等价于$t \ne 0$, 故从上式得到严格不等式:

$$x \in S \cap B(x_0, \delta_\varepsilon) \quad \text{and} \quad x \ne x_0 \quad \Longrightarrow \quad f(x) > f(x_0).$$

因此x_0 是$f|_S$的一个局部严格极小值点.

同法可证: 若$H_L(x_0)$ 在TS_{x_0} 上是负定的, 则x_0 是$f|_S$的一个局部严格极大值点.

(2) 假设x_0 是$f|_S$的一个局部极小值点. 则存在$0 < \varepsilon < 1$ 充分小使得

$$0 \le f(x(t)) - f(x_0) = \frac{1}{2} t^\tau A^\tau H_L(x_0) At + o(|t|^2), \quad t \in (-\varepsilon, \varepsilon)^{n-m}.$$

也即

$$t^\tau A^\tau H_L(x_0) At \ge o(|t|^2), \quad t \in (-\varepsilon, \varepsilon)^{n-m}.$$

对任意$h \in \mathbb{R}^{n-m}$, 当$0 < s << 1$ 时有$sh \in (-\varepsilon, \varepsilon)^{n-m}$ 从而(取$t = sh$) 有

$$h^\tau A^\tau H_L(x_0) Ah \ge \frac{o(s^2|h|^2)}{s^2} \to 0 \quad (s \to 0^+).$$

这证明了$h^\tau A^\tau H_L(x_0)Ah \ge 0$ for all $h \in \mathbb{R}^{n-m}$. 所以$A^\tau H_L(x_0)A$ 是半正定阵, 也即$H_L(x_0)$在TS_{x_0} 上是半正定的.

同法可证: 若x_0 是$f|_S$的一个局部极大值点, 则$H_L(x_0)$在TS_{x_0} 上是半负定的.

(3) 假设$H_L(x_0)$ 在TS_{x_0} 上是不定的, 即$H_L(x_0)$ 在TS_{x_0} 上既非半正定也非半负定. 则由结论(2)可知x_0 不是$f|_S$的极值点. $\square$

【例】设$n \geq 2$, $F \in C^1(\mathbb{R}^n, \mathbb{R})$ 满足

$$S = \{x \in \mathbb{R}^n \mid F(x) = C\}$$

是非空的紧集, 其中C为常数. 则F在$\mathbb{R}^n$上必有最小值**或**最大值, 因而F 在$\mathbb{R}^n$上有临界点, 即存在$x_0 \in \mathbb{R}^n$使得$\nabla F(x_0) = 0$.

【证】因S 是紧集, 故可以取$0 < R < \infty$ 充分大使得S含于闭球$K := \overline{B}(0; R)$中. 取$\mathbf{a}, \mathbf{b} \in K$ 使得

$$F(\mathbf{a}) = \min_{x \in K} F(x), \quad F(\mathbf{b}) = \max_{x \in K} F(x).$$

则由$S \subset K$有

$$F(\mathbf{a}) \leq C \leq F(\mathbf{b})$$

并且对每个$x \in K^c = \mathbb{R}^n \setminus K$ 有$F(x) - C \neq 0$. 但$n \geq 2$ 蕴含K^c是连通集. 因此$F(x) - C$在K^c上不变号. 不妨设$F(x) - C > 0$ for all $x \in K^c$. 即$F(x) > C$ for all $x \in K^c$. 由此有：对任意$x \in \mathbb{R}^n$, 若$x \in K$ 则$F(x) \geq F(\mathbf{a})$; 若$x \in K^c$ 则$F(x) > C \geq F(\mathbf{a})$. 这证明了

$$F(x) \geq F(\mathbf{a}) \qquad \forall x \in \mathbb{R}^n.$$

因此$\mathbf{a}$是F 在$\mathbb{R}^n$上的最小值点，即

$$F(\mathbf{a}) = \min_{x \in \mathbb{R}^n} F(x) \qquad \text{hence} \quad \nabla F(\mathbf{a}) = 0.$$

同理可证若$F(x) - C < 0$ for all $x \in K^c$, 则$\mathbf{b}$是F 在$\mathbb{R}^n$上的最大值点，即

$$F(\mathbf{b}) = \max_{x \in \mathbb{R}^n} F(x) \qquad \text{hence} \quad \nabla F(\mathbf{b}) = 0.$$

综上可知：存在$x_0 \in \{\mathbf{a}, \mathbf{b}\}$ 使得$\nabla F(x_0) = 0$. $\square$

【作业题】

1. 证明**命题8.51.**

2. 设$\beta_i > 0$, $i = 1, 2, ..., n$; $\sum_{i=1}^n \beta_i = 1$. 试利用条件极值理论证明: 对任意$a_1 \geq 0, a_2 \geq 0, ..., a_n \geq 0$ 有

$$a_1^{\beta_1} a_2^{\beta_2} \cdots a_n^{\beta_n} \leq \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_n a_n.$$

3. 陈书习题**13.5.1**, 见第三册pp.336-338.

4. 设$n\ge 2$, $F\in C^2(\mathbb{R}^n,\mathbb{R})$. 假设

$$S=\{x\in\mathbb{R}^n\,|\,F(x)=0\}$$

是非空的紧集且$\nabla F(x)\neq 0$ for all $x\in S$. 由S是紧集, 易见存在$\varepsilon>0$ 使得

$$\nabla F(x)\neq 0\qquad \forall\, x\in\{x\in\mathbb{R}^n\,|\,\mathrm{dist}(x,S)\le\varepsilon\}.$$

令

$$\Omega=\{x\in\mathbb{R}^n\,|\,\mathrm{dist}(x,S)<\frac{\varepsilon}{2}\},\quad \Omega_0=\{x\in\mathbb{R}^n\,|\,\mathrm{dist}(x,S)<\varepsilon\}.$$

则Ω,Ω_0 均为有界开集且

$$S\subset\Omega,\quad \overline{\Omega}\subset\Omega_0,\quad \overline{\Omega_0}=\{x\in\mathbb{R}^n\,|\,\mathrm{dist}(x,S)\le\varepsilon\}.$$

令

$$\mathbf{n}(x)=\frac{\nabla F(x)}{|\nabla F(x)|},\qquad f_t(x)=x+t\mathbf{n}(x),\quad x\in\overline{\Omega_0}$$

其中t 是实数.

(1) 证明：存在$\delta>0$ 使得对任意$t\in(-\delta,\delta)$, 映射$f_t:\Omega\to f_t(\Omega)$ 是C^1 同胚.

(2) 设δ 是(1)中的δ. 令

$$S_t=\{x+t\mathbf{n}(x)\,|\,x\in S\}.$$

证明：对任意$t\in(-\delta,\delta)$, S_t是C^1类的$n-1$维曲面.

(3) 设δ 是(1)中的δ. 令

$$\eta=\min\left\{\delta,\ \frac{\min\limits_{x\in S}|\nabla F(x)|}{1+\max\limits_{x\in\overline{\Omega}}\|H_F(x)\|}\right\}.$$

证明：当$t\in(-\eta,\eta)$时,

$$S_t\subset\{x\in\mathbb{R}^n\ |\,\mathrm{dist}(x,S)=|t|\}=S_t\cup S_{-t}\,;$$

而当$t\in(-\eta,\eta)\setminus\{0\}$时$S_t$ 有如下几何表示:

$$S_t=\{x\in\mathbb{R}^n\ |\,\exists\, x_*\in S\ \text{ s.t. }\ \mathrm{dist}(x,S)=|x-x_*|=|t|\ \text{ and }\ t\langle x-x_*,\mathbf{n}(x_*)\rangle>0\,\}.$$

(4) 设η 是(3)中的η. 证明：当$t_1,t_2\in(-\eta,\eta)$且$|t_1|\neq|t_2|$时, $S_{t_1}\cap S_{t_2}=\emptyset$.

【注：在(3)的证明中要用到F 的二阶Taylor 公式(涉及Hesse 矩阵), 而证明几何表示时, 要用到条件极值理论.]

习题课

1. (学生独立做2分钟) 举例或否证：存在可微函数$f:\mathbb{R}^2\to\mathbb{R}$ 满足

$$\frac{\partial}{\partial x}f(x,y)=xy,\quad \frac{\partial}{\partial y}f(x,y)=y^2\qquad \forall\,(x,y)\in\mathbb{R}^2.$$

[考察二阶偏导数换序问题, 注意f的偏导函数属于C^∞, 因此$f\in C^\infty$, 等等]

2. 设$n\ge 2$, $\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n$ 是$\mathbb{R}^n$中的单位向量, 设$\Omega\subset\mathbb{R}^n$ 为开集, $f:\Omega\to\mathbb{R}$ 可微. 证明: 若$\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n$ 线性无关, 则存在常数$C>0$ 使得

$$|\nabla f(x)|\le C\sqrt{\sum_{i=1}^n(D_{\mathbf{v}_i}f(x))^2}\qquad \forall\,x\in\Omega$$

并确定一个这样的C. 进一步证明: 当$\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n$ 两两正交时有

$$|\nabla f(x)|\equiv\sqrt{\sum_{i=1}^n(D_{\mathbf{v}_i}f(x))^2},\quad x\in\Omega.$$

这里

$$D_{\mathbf{v}}f(x)=\frac{\mathrm{d}}{\mathrm{dt}}(f(x+t\mathbf{v}))\Big|_{t=0}$$

表示f 在点x 处沿方向$\mathbf{v}$的方向导数.

【证】由复合函数求导法则有

$$D_{\mathbf{v}_i}f(x)=\langle\nabla f(x),\mathbf{v}_i\rangle,\quad i=1,2,...,n.$$

注意$\mathbf{v}_i,\nabla f(x)$ 皆为列向量, 以上可以写成矩阵形式:

$$\begin{pmatrix} D_{\mathbf{v}_1}f(x)\\ D_{\mathbf{v}_2}f(x)\\ \vdots\\ D_{\mathbf{v}_n}f(x)\end{pmatrix}=\begin{pmatrix}\mathbf{v}_1^\tau\nabla f(x)\\ \mathbf{v}_2^\tau\nabla f(x)\\ \vdots\\ \mathbf{v}_n^\tau\nabla f(x)\end{pmatrix}=\begin{pmatrix}\mathbf{v}_1^\tau\\ \mathbf{v}_2^\tau\\ \vdots\\ \mathbf{v}_n^\tau\end{pmatrix}\nabla f(x)=:A\nabla f(x).$$

设$\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n$ 线性无关, 则因它们都属于$\mathbb{R}^n$, 故它们是$\mathbb{R}^n$的一组基, 也即矩阵A可逆. 因此

$$\nabla f(x)=A^{-1}\begin{pmatrix} D_{\mathbf{v}_1}f(x)\\ D_{\mathbf{v}_2}f(x)\\ \vdots\\ D_{\mathbf{v}_n}f(x)\end{pmatrix}$$

从而有

$$|\nabla f(x)| \leq \|A^{-1}\| \left| \begin{pmatrix} D_{\mathbf{v}_1} f(x) \\ D_{\mathbf{v}_2} f(x) \\ \vdots \\ D_{\mathbf{v}_n} f(x) \end{pmatrix} \right| = \|A^{-1}\| \sqrt{\sum_{i=1}^{n} (D_{\mathbf{v}_i} f(x))^2}, \quad x \in \Omega.$$

因此可以取$C = \|A^{-1}\|$.

最后设$\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ 两两正交. 则A 是正交阵因此A作为线性变换是保范数的, 即$|Ay| = |y|, y \in \mathbb{R}^n$. 于是有

$$\left| \begin{pmatrix} D_{\mathbf{v}_1} f(x) \\ D_{\mathbf{v}_2} f(x) \\ \vdots \\ D_{\mathbf{v}_n} f(x) \end{pmatrix} \right| = |A \nabla f(x)| = |\nabla f(x)|$$

即

$$\sqrt{\sum_{i=1}^{n} (D_{\mathbf{v}_i} f(x))^2} = |\nabla f(x)|, \quad x \in \Omega.$$

□

3 (行列式的导数). 设$\Omega \subset \mathbb{R}^n$ 为开集, 向量值函数

$$\mathbf{a}_j(x) = \begin{pmatrix} a_{1j}(x) \\ a_{2j}(x) \\ \vdots \\ a_{nj}(x) \end{pmatrix}, \quad j = 1, 2, ..., n, \quad t \in I.$$

在I 上可微. 则对于由它们组成的行列式的偏导数有

$$\begin{aligned}
&D_j \det\Big(\mathbf{a}_1(x) \ \ \mathbf{a}_2(x) \ \ \mathbf{a}_3(x) \ \ ... \ \ \mathbf{a}_n(x)\Big) \\
&= \det\Big(D_j \mathbf{a}_1(x) \ \ \mathbf{a}_2(x) \ \ \mathbf{a}_3(x) \ \ ... \ \ \mathbf{a}_n(x)\Big) \\
&+ \det\Big(\mathbf{a}_1(x) \ \ D_j \mathbf{a}_2(x) \ \ \mathbf{a}_3(x) \ \ ... \ \ \mathbf{a}_n(x)\Big) \\
&+ \cdots \\
&+ \det\Big(\mathbf{a}_1(x) \ \ \mathbf{a}_2(x) \ \ ... \ \ \mathbf{a}_{n-1}(x) \ \ D_j \mathbf{a}_n(x)\Big) \\
&= \sum_{k=1}^{n} \det\Big(\mathbf{a}_1(x) \ \ ... \ \ \mathbf{a}_{k-1}(x) \ \ D_j \mathbf{a}_k(x) \ \ \mathbf{a}_{k+1}(x) \ \ ... \ \ \mathbf{a}_n(x)\Big), \quad x \in \Omega
\end{aligned}$$

$j=1,2,...,n$, 其中$D_j=\frac{\partial}{\partial x_j}$.

【证】回忆行列式的定义:

$$\det(\mathbf{a}_1(x)\ \ \mathbf{a}_2(x)\ \ \mathbf{a}_3(x)\ \ ...\ \ \mathbf{a}_n(x))=\det\begin{pmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x)\\ a_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x)\\ \cdots\cdots & \cdots\cdots & \cdots & \\ a_{n1}(x) & a_{n2}(x) & \cdots & a_{nn}(x)\end{pmatrix}$$
$$=\sum_{\sigma\in \mathrm{S}_n}\operatorname{sgn}(\sigma)a_{\sigma(1)1}(x)a_{\sigma(2)2}(x)\cdots a_{\sigma(n)n}(x)$$

其中σ是$\{1,2,...,n\}$到$\{1,2,...,n\}$ 的一一对应, 即置换, S_n 是这种σ的全体, $\operatorname{sgn}(\sigma)$ 是置换σ的符号[即若σ是偶置换, 则$\operatorname{sgn}(\sigma)=1$, 若$\sigma$是奇置换, 则$\operatorname{sgn}(\sigma)=-1$。由函数乘积的导数计算法则有

$$\begin{aligned}
&D_j\det(\mathbf{a}_1(x)\ \ \mathbf{a}_2(x)\ \ \mathbf{a}_3(x)\ \ ...\ \ \mathbf{a}_n(x))\\
&=\sum_{\sigma\in \mathrm{S}_n}\operatorname{sgn}(\sigma)D_j\Big(a_{\sigma(1)1}(x)a_{\sigma(2)2}(x)\cdots a_{\sigma(n)n}(x)\Big)\\
&=\sum_{\sigma\in \mathrm{S}_n}\operatorname{sgn}(\sigma)\Big\{D_ja_{\sigma(1)1}(x)a_{\sigma(2)2}(x)\cdots a_{n\sigma(n)}(x)\\
&+a_{\sigma(1)1}(x)D_ja_{\sigma(2)2}(x)\cdots a_{\sigma(n)n}(x)+\cdots+a_{\sigma(1)1}(x)\cdots a_{\sigma(n-1)n}(x)D_ja_{\sigma(n)n}(x)\Big\}\\
&=\sum_{\sigma\in \mathrm{S}_n}\operatorname{sgn}(\sigma)D_ja_{\sigma(1)1}(x)a_{\sigma(2)2}(x)\cdots a_{\sigma(n)n}(x)\\
&+\sum_{\sigma\in \mathrm{S}_n}\operatorname{sgn}(\sigma)a_{\sigma(1)1}(x)D_ja_{\sigma(2)2}(x)\cdots a_{\sigma(n)n}(x)+\cdots\\
&+\sum_{\sigma\in \mathrm{S}_n}\operatorname{sgn}(\sigma)a_{\sigma(1)1}(x)\cdots a_{\sigma(n-1)n}(x)D_ja_{\sigma(n)n}(x)\\
&=\det\begin{pmatrix} D_ja_{11}(x) & a_{12}(x) & \cdots & a_{1n}(x)\\ D_ja_{21}(x) & a_{22}(x) & \cdots & a_{2n}(x)\\ \cdots\cdots & \cdots\cdots & \cdots & \\ D_ja_{n1}(x) & a_{n2}(x) & \cdots & a_{nn}(x)\end{pmatrix}+\det\begin{pmatrix} a_{11}(x) & D_ja_{12}(x) & \cdots & a_{1n}(x)\\ a_{21}(x) & D_ja_{22}(x) & \cdots & a_{2n}(x)\\ \cdots\cdots & \cdots\cdots & \cdots & \\ a_{n1}(x) & D_ja_{n2}(x) & \cdots & a_{nn}(x)\end{pmatrix}\\
&+\cdots+\det\begin{pmatrix} a_{11}(x) & a_{12}(x) & \cdots & D_ja_{1n}(x)\\ a_{21}(x) & a_{22}(x) & \cdots & D_ja_{2n}(x)\\ \cdots\cdots & \cdots\cdots & \cdots & \\ a_{n1}(x) & a_{n2}(x) & \cdots & D_ja_{nn}(x)\end{pmatrix}\\
&=\det\Big(D_j\mathbf{a}_1(x)\ \ \mathbf{a}_2(x)\ \ \mathbf{a}_3(x)\ \ ...\ \ \mathbf{a}_n(x)\Big)+\det\Big(\mathbf{a}_1(x)\ \ D_j\mathbf{a}_2(x)\ \ \mathbf{a}_3(x)\ \ ...\ \ \mathbf{a}_n(x)\Big)
\end{aligned}$$

$$
\begin{aligned}
&+\cdots+\det\Big(\mathbf{a}_1(x)\ \ \mathbf{a}_2(x)\ \ ...\ \ \mathbf{a}_{n-1}(x)\ \ D_j\mathbf{a}_n(x)\Big)\\
&=\sum_{k=1}^{n}\det\Big(\mathbf{a}_1(x)\ \ ...\ \ \mathbf{a}_{k-1}(x)\ \ D_j\mathbf{a}_k(x)\ \ \mathbf{a}_{k+1}(x)\ \ ...\ \ \mathbf{a}_n(x)\Big).
\end{aligned}
$$

□

4 (Hadamard 恒等式). 设$\Omega\subset\mathbb{R}^n$ 为开集, $f=(f_1,f_2,...,f_n)^{\tau}\in C^2(\Omega,\mathbb{R}^n)$. 设

$$
f'(x)=J_f(x)=\left(\frac{\partial f_i}{\partial x_j}(x)\right)_{n\times n}=\begin{pmatrix}\frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & ... & \frac{\partial f_1}{\partial x_n}(x)\\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & ... & \frac{\partial f_2}{\partial x_n}(x)\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ \frac{\partial f_n}{\partial x_1}(x) & \frac{\partial f_n}{\partial x_2}(x) & ... & \frac{\partial f_n}{\partial x_n}(x)\end{pmatrix}.
$$

令

$$
A_{i,j}(x)=(-1)^{i+j}\det\Big(f'(x)\Big|_{\text{去掉第}i\text{ 行和第}j\text{列}}\Big)
$$

证明

$$
\sum_{j=1}^{n}\frac{\partial}{\partial x_j}A_{i,j}(x)\equiv 0,\quad i=1,2,...,n.
$$

【证】(by 孙宗汉). 当$n=2$ 时直接计算可知成立. 设$n\ge 3$. 只证$i=1$的情形. 令

$$
D_j=\frac{\partial}{\partial x_j},\ j=1,2,...,n;\qquad g(x)=\begin{pmatrix}f_2(x)\\ f_3(x)\\ \vdots\\ f_n(x)\end{pmatrix}.
$$

则

$$
A_{1,j}(x)=(-1)^{1+j}\det\Big(\ D_1g\quad ...\quad D_{j-1}g\quad D_{j+1}g\quad ...\quad D_ng\ \Big).
$$

因此

$$
\sum_{j=1}^{n}\frac{\partial}{\partial x_j}A_{1,j}(x)=\sum_{j=1}^{n}(-1)^{1+j}D_j\det\Big(\ D_1g\quad ...\quad D_{j-1}g\quad D_{j+1}g\quad ...\quad D_ng\ \Big).
$$

而由行列式的求偏导数公式(见上一题) 有

$$
\begin{aligned}
&D_j\det\Big(D_1g\quad ...\quad D_{j-1}g\quad D_{j+1}g\quad ...\quad D_ng\Big)\\
&=\sum_{k=1}^{j-1}\det\Big(D_1g\quad ...\quad D_jD_kg\quad ...\quad D_{j+1}g\quad ...\quad D_ng\Big)\\
&+\sum_{k=j+1}^{n}\det\Big(D_1g\quad ...\quad D_{j-1}g\quad ...\quad D_jD_kg\quad ...\quad D_ng\Big).
\end{aligned}
$$

利用行列式交换列时变号的性质有

$$
\begin{aligned}
&\det\Big(D_1g \quad ... \quad D_jD_kg \quad ... \quad D_{j+1}g \quad ... \quad D_ng\Big)\\
&=(-1)^{k-1}\det\Big(D_jD_kg \quad D_1g \quad ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_ng\Big)\\
&k=1,2,...,j-1(\geq 1);
\end{aligned}
$$

$$
\begin{aligned}
&\det\Big(D_1g \quad ... \quad D_{j-1}g \quad ... \quad D_jD_kg \quad ... \quad D_ng\Big)\\
&=(-1)^{k-2}\det\Big(D_jD_kg \quad D_1g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_ng\Big)\\
&k=j+1,j+2,...,n.
\end{aligned}
$$

所以

$$
\begin{aligned}
&D_j\det\Big(D_1g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_ng\Big)\\
&=\sum_{k=1}^{j-1}(-1)^{k-1}\det\Big(D_jD_kg \quad D_1g \quad ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_ng\Big)\\
&-\sum_{k=j+1}^{n}(-1)^{k-1}\det\Big(D_jD_kg \quad D_1g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_ng\Big).
\end{aligned}
$$

代入上式计算:

$$
\begin{aligned}
&\sum_{j=1}^{n}\frac{\partial}{\partial x_j}A_{1,j}(x)\\
&=\sum_{j=1}^{n}(-1)^{1+j}D_j\det\Big(\ D_1g \ ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_ng\ \Big)\\
&=\sum_{j=1}^{n}(-1)^{j}\sum_{k=1}^{j-1}(-1)^{k}\det\Big(D_jD_kg \quad D_1g \ ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_ng\Big)\\
&-\sum_{j=1}^{n}(-1)^{j}\sum_{k=j+1}^{n}(-1)^{k}\det\Big(D_jD_kg \quad D_1g \ ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_ng\Big)\\
&=\sum_{j=2}^{n}(-1)^{j}\sum_{k=1}^{j-1}(-1)^{k}\det\Big(D_jD_kg \quad D_1g \ ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_ng\Big)\\
&-\sum_{j=1}^{n-1}(-1)^{j}\sum_{k=j+1}^{n}(-1)^{k}\det\Big(D_jD_kg \quad D_1g \ ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_ng\Big)\\
&=\sum_{j=2}^{n}\sum_{k=1}^{j-1}(-1)^{j+k}\det\Big(D_jD_kg \quad D_1g \ ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_ng\Big)\\
&-\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}(-1)^{j+k}\det\Big(D_jD_kg \quad D_1g \ ... \quad D_{j-1}g \quad D_{j+1}g \quad ... \quad D_{k-1}g \quad D_{k+1}g \quad ... \quad D_ng\Big)\\
&:=I_1-I_2.
\end{aligned}
$$

对于第二项I_2, 计算

$$
\begin{aligned}
&I_2\\
&=\sum_{j=1}^{n}\sum_{k=1}^{n}(-1)^{k+j}\mathbf{1}_{\{1\le j<k\le n\}}\det\Big(D_jD_kg\ \ D_1g\ \ ...\ \ D_{j-1}g\ \ D_{j+1}g\ \ ...\ \ D_{k-1}g\ \ D_{k+1}g\ \ ...\ \ D_ng\Big)\\
&=\sum_{k=1}^{n}\sum_{j=1}^{n}(-1)^{k+j}\mathbf{1}_{\{1\le k<j\le n\}}\det\Big(D_kD_jg\ \ D_1g\ \ ...\ \ D_{k-1}g\ \ D_{k+1}g\ \ ...\ \ D_{j-1}g\ \ D_{j+1}g\ \ ...\ \ D_ng\Big)\\
&=\sum_{j=2}^{n}\sum_{k=1}^{j-1}(-1)^{k+j}\det\Big(D_kD_jg\quad D_1g\quad ...\quad D_{k-1}g\quad D_{k+1}g\quad ...\quad D_{j-1}g\quad D_{j+1}g\quad ...\quad D_ng\Big).
\end{aligned}
$$

注意到g属于C^2类蕴含偏导数可换序:

$$D_kD_jg=D_jD_kg$$

因此与将上式与I_1 比较看出$I_2=I_1$. 所以

$$\sum_{j=1}^{n}\frac{\partial}{\partial x_j}A_{1,j}(x)=I_1-I_2=0.$$

同法可证$\sum\limits_{j=1}^{n}\frac{\partial}{\partial x_j}A_{i,j}(x)=0,i=2,3,...,n.$ $\square$

5(次调和函数). 设$\Omega\subset\mathbb{R}^n$ 为开集, 实值函数$u\in C^2(\Omega)$ 称为在Ω 上为次调和的如果

$$\Delta u(x)\ge 0\qquad\forall\, x\in\Omega.$$

其中$\Delta=\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\cdots+\frac{\partial^2}{\partial x_n^2}$ 称为Laplace 算子，即

$$\Delta u(x)=\frac{\partial^2u}{\partial x_1^2}(x)+\frac{\partial^2u}{\partial x_2^2}(x)+\cdots+\frac{\partial^2u}{\partial x_n^2}(x).$$

设$\Omega\subset\mathbb{R}^n$ 为有界开集, $u\in C^2(\Omega)\cap C(\overline{\Omega})$ 且在Ω 上是次调和的. 证明u在$\overline{\Omega}$上的最大值可以在边界$\partial\Omega$上达到, 即

$$\max_{x\in\overline{\Omega}}u(x)=\max_{x\in\partial\Omega}u(x).$$

【证】先设u 是严格次调和的, 即

$$\Delta u(x)>0\qquad\forall\, x\in\Omega.$$

此时证明u在$\overline{\Omega}$上的最大值必在$\partial\Omega$ 上达到. 设$x_0\in\overline{\Omega}$ 为u的最大值点. 若$x_0\in\Omega$, 则由函数取得极值的必要条件知u 在x_0 的Hesse 矩阵$H_u(x_0)$ 是半负定的, 从而

$$0<\Delta u(x_0)=\mathrm{trace}H(x_0)\le 0$$

矛盾。所以必有$x_0 \in \partial\Omega$.

对于一般情形, 考虑对u的逼近序列:

$$u_k(x) = u(x) + \frac{|x|^2}{2nk}, \quad x \in \overline{\Omega}, \quad k \in \mathbb{N}.$$

易见

$$\Delta u_k(x) = \Delta u + \frac{1}{2nk}\Delta|x|^2 = \Delta u + \frac{1}{k} \geq \frac{1}{k} > 0, \quad x \in \Omega.$$

所以u_k在$\overline{\Omega}$上的最大值点x_k 属于边界$\partial\Omega, k = 1, 2, 3, \dots$. 因$\partial\Omega$ 是紧集(因Ω 有界), 故存在子列x_{k_j} 和$x_0 \in \partial\Omega$, 使得$x_{k_j} \to x_0\,(j \to \infty)$. 于是由$u$ 连续得到: 对任意$x \in \overline{\Omega}$ 有

$$u(x) = \lim_{j\to\infty} u_{k_j}(x) \leq \lim_{j\to\infty} u_{k_j}(x_0) = u(x_0).$$

所以这个$x_0 \in \partial\Omega$ 是u在$\overline{\Omega}$上的一个最大值点. $\Box$

6(调和函数). 设$\Omega \subset \mathbb{R}^n$ 为开集, 实值函数$u \in C^2(\Omega)$ 称为在Ω 上为调和的如果

$$\Delta u(x) = 0 \qquad \forall\, x \in \Omega.$$

设$\Omega \subset \mathbb{R}^n$ 为有界开集, $u \in C^2(\Omega) \cap C(\overline{\Omega})$ 且u在Ω 上是调和的. 则u在$\overline{\Omega}$上的最大值点和最小值点都可以在边界$\partial\Omega$上达到, 即

$$\max_{x\in\overline{\Omega}} u(x) = \max_{x\in\partial\Omega} u(x), \quad \min_{x\in\overline{\Omega}} u(x) = \min_{x\in\partial\Omega} u(x).$$

【证】正是因为u和$-u$ 都是次调和的. $\Box$

7(调和函数). 设$\Omega \subset \mathbb{R}^n$ 为开集, 实值函数$f, g \in C^2(\Omega) \cap C(\overline{\Omega})$ 在Ω 上都是调和函数且$f|_{\partial\Omega} = g|_{\partial\Omega}$. 证明$f = g$ 于$\overline{\Omega}$.

【证】令$u(x) = f(x) - g(x)$. 则$u \in C^2(\Omega) \cap C(\overline{\Omega})$ 在Ω 上是调和函数且$u(x) = 0$ for all $x \in \partial\Omega$。因此

$$\max_{x\in\overline{\Omega}} u(x) = \max_{x\in\partial\Omega} u(x) = 0, \quad \min_{x\in\overline{\Omega}} u(x) = \min_{x\in\partial\Omega} u(x) = 0.$$

所以$u \equiv 0$ 于$\overline{\Omega}$. $\Box$

8. 设$\Omega\subset\mathbb{R}^n$ 是有界开集, $a_1,a_2,...,a_n\in\mathbb{R}$为常数, 实函数$u\in C^2(\Omega)\cap C(\overline{\Omega})$ 满足

$$\Delta u(x)+\sum_{j=1}^{n}a_j\frac{\partial u}{\partial x_j}(x)\le u(x),\qquad x\in\Omega.$$

证明：(1) 若$u|_{\partial\Omega}\ge 0$, 则$u\ge 0$ 于$\overline{\Omega}$.

(2) 若$u|_{\partial\Omega}>0$, 则$u>0$ 于$\overline{\Omega}$.

【证】(1)：反证法：假设存在$x_*\in\overline{\Omega}$ 使得$u(x_*)<0$. 令$u(x_0)=\min\limits_{x\in\overline{\Omega}}u(x), x_0\in\overline{\Omega}$. 则$u(x_0)\le u(x_*)<0$. 因$u|_{\partial\Omega}\ge 0$, 故这蕴含$x_0\in\Omega$. 由极值原理(取得极值的必要条件)知$\frac{\partial u}{\partial x_j}(x_0)=0, j=1,2,...,n$ 且Hesse 矩阵$H_u(x_0)$ 半正定, 从而有

$$\Delta u(x_0)=\mathrm{trace}H_u(x_0)\ge 0,$$

$$\Delta u(x_0)+\sum_{j=1}^{n}a_j\frac{\partial u}{\partial x_j}(x_0)=\Delta u(x_0)\ge 0.$$

另一方面由题中的偏微分不等式却有

$$\Delta u(x_0)+\sum_{j=1}^{n}a_j\frac{\partial u}{\partial x_j}(x_0)\le u(x_0)<0$$

矛盾. 这矛盾证明了(1) 成立.

(2): 对于常数$c>0$, 考虑正值函数

$$\varphi(x)=c\exp\Big(\frac{1}{\sqrt{n}}(x_1+x_2+\cdots+x_n)\Big)\quad \text{if}\quad \sum_{j=1}^{n}a_j\ge 0;$$

$$\varphi(x)=c\exp\Big(\frac{-1}{\sqrt{n}}(x_1+x_2+\cdots+x_n)\Big)\quad \text{if}\quad \sum_{j=1}^{n}a_j< 0.$$

则易见

$$\frac{\partial}{\partial x_j}\varphi(x)=\frac{\pm 1}{\sqrt{n}}\varphi(x),\quad \frac{\partial^2}{\partial x_j^2}\varphi(x)=\frac{1}{n}\varphi(x),\quad j=1,2,...,n.$$

因此

$$\sum_{j=1}^{n}a_j\frac{\partial\varphi}{\partial x_j}(x)=\Big(\pm\sum_{j=1}^{n}a_j\Big)\frac{\varphi(x)}{\sqrt{n}}\ge 0\quad \forall x\in\mathbb{R}^n,$$

$$\Delta\varphi(x)=\varphi(x)\qquad \forall x\in\mathbb{R}^n.$$

由此可见在Ω 内有

$$\Delta(u-\varphi)+\sum_{j=1}^{n}a_j\frac{\partial(u-\varphi)}{\partial x_j}$$

$$
\begin{aligned}
&= \Delta u - \Delta\varphi + \sum_{j=1}^{n} a_j \frac{\partial u}{\partial x_j} - \sum_{j=1}^{n} a_j \frac{\partial \varphi}{\partial x_j} \\
&\le \Delta u - \varphi + \sum_{j=1}^{n} a_j \frac{\partial u}{\partial x_j} \le u - \varphi.
\end{aligned}
$$

即

$$
\Delta(u-\varphi) + \sum_{j=1}^{n} a_j \frac{\partial (u-\varphi)}{\partial x_j} \le u - \varphi \quad 于 \quad \Omega.
$$

如果我们能选取合适的$c > 0$ 使得$(u-\varphi)\Big|_{\partial\Omega} \ge 0$, 则由(1) 即知$u - \varphi \ge 0$于$\overline{\Omega}$ 从而$u \ge \varphi > 0$于$\overline{\Omega}$.

下面寻找合适的$c > 0$使得$(u-\varphi)\Big|_{\partial\Omega} \ge 0$.

令

$$
b = \min_{x\in\partial\Omega} u(x).
$$

则$b > 0$. 取定一点$x_0 \in \Omega$ 令$M = \mathrm{diam}(\overline{\Omega}) + |x_0|$, 取$c > 0$ 使得

$$
ce^M \le b \qquad \text{i.e.} \qquad 0 < c \le be^{-M}.
$$

则对任意$x \in \partial\Omega$ 有

$$
\frac{1}{\sqrt{n}}|x_1 + x_2 + \cdots + x_n| \le |x| \le |x - x_0| + |x_0| \le \mathrm{diam}(\overline{\Omega}) + |x_0| = M
$$

因此

$$
当x \in \partial\Omega 时 \quad u(x) - \varphi(x) \ge b - ce^M \ge 0.
$$

□