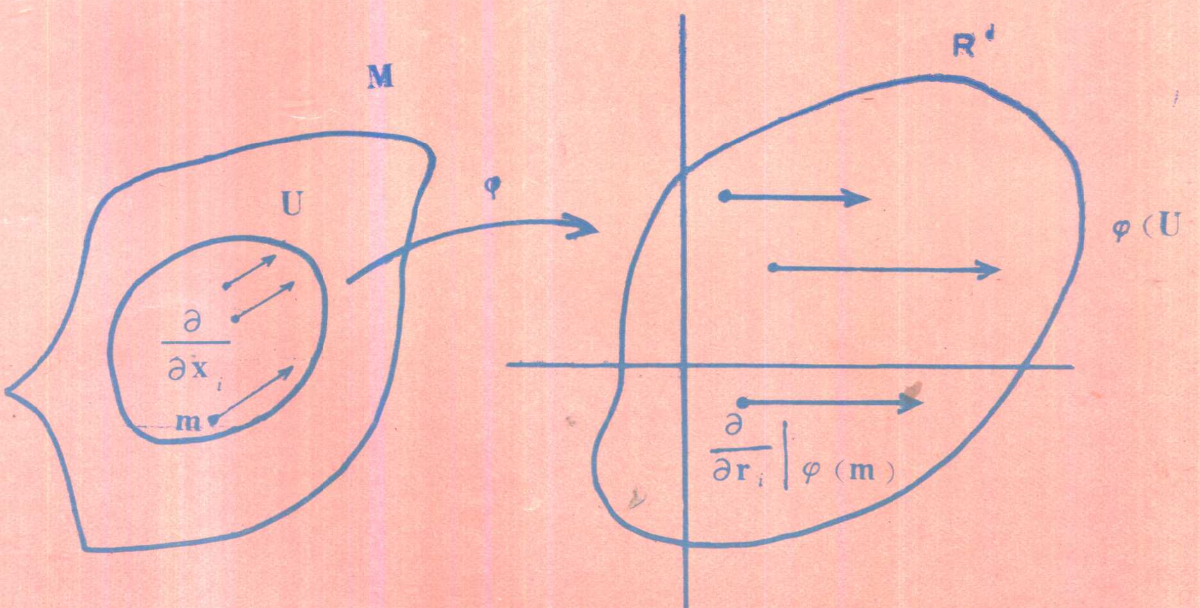
微分流形及其应用

杨万年　主编



重庆大学出版社

序　言

本书是在我们多年来开设有关课程的讲义基础上逐渐形成的。1979年至1981年先后为理论物理研究生讲授《张量分析与黎曼几何》和《纤维丛基础》；1984年用徐森林先生所编的《流形及Stokes定理》一书为应用数学系毕业生开设过选修课；1985年杨万年编写了《微分流形及其应用》的讲义并用它为应用数学研究生开设基础选修课至今。本书是在这本讲义基础上集体修改而成的，其中赵学军修改第一、六、七章，傅强修改第五、八、九章，杨万年修改其余各章并负责全书审定工作。

本书是微分流形的一本入门书籍，主要介绍有限维微分流形上的微积分学及其应用，试图以此向读者介绍微分流形的主要知识和一些基本技巧。限于篇幅本书对微分流形的进一步发展如微分拓扑、微分几何、无穷维微分流形等丰富的内容均未涉及。微分流形与近代物理学及近代力学的发展也是密切相关的。可以毫不夸张地说，学习近代数学、近代物理学及近代科学技术而不具备微分流形的知识将是不可思议的。当前世界正进入非线性与高科技时代，作为当代的大学生、研究生掌握一定的微分流形知识是非常必要的。我们希望此书能为纯粹数学、应用数学、计算数学、物理学、力学及社会经济科学的学生和教师提供一本适用的微分流形教材或参考书。

第一章是微积分学的深入和发展，以统一的方式介绍了从R^n到R^m的映射的微分学，主要是Fréchet可微与Gâteaux可微及几个主要定理的推广与阐述，它们为流形上的微积分学提供了背景和手段。这部分内容具有独立性，并且体系完整、内容丰富，可以作为数学系高年级学生的选修课教材。

第二至五章是流形上的微积分学的主要内容。第二章主要介绍微分流形的定义、映射的可微性、子流形及其在欧氏空间中的嵌入等内容。第三、四章集中介绍微分流形上的纤维丛结构，尤其是切丛、余切丛、张量丛等向量丛结构及向量场、张量场与外微分形式等内容。着重介绍了李导数与外微分等两大微分算子以及它们与拉前算子、拉回算子之间的关系。对大家的Frobenius理论作了较为详细的讨论，并用简短的篇幅介绍了李群、黎曼流形与辛流形等方面的初步内容。第五章是流形上的积分学。

第六至九章是微分流形在几个方面的应用。第六章是有限维临界点理论，主要介绍Sard定理、横截性、Morse引理及Morse不等式，部分所需的拓扑学知识放在附录Ⅰ中。第七章是流形上的微分动力系统的初步介绍。第八章介绍大范围分析在经济均衡理论中的应用。第九章介绍统计学的Riemann几何方法。除第八、九两章外，每章都附有一定数量的习题供读者练习使用。

在本书形成过程中，我们参考了中外作者的大量著作，第九章还参考了韦博成先生当年的讲义（现已出版，见〔27〕），在此一并致谢。

我们衷心感谢重庆大学应用数学系的同事们，感谢朱继生、刘松、段虞荣、陈均平等诸位先生的鼓励和支持，尤其要感谢张世清同志，他为本书提供了不少素材和意见，也要感谢涂光裕同志为本书绘制了全部插图。正是以上各方面的支持和帮助，本书才得以问世。

限于我们的知识水平，书中定有不少错误和不妥之处，敬请读者批评指正。

编　者
于重庆大学应用数学系
一九九一年十二月

目　录

第一章　多元映射的微分学

本章简要介绍一下多元映射的微分学，主要讨论对以后的内容来说非常重要的链式法则、中值不等式、Taylor定理、反函数定理、隐函数定理和秩定理。它不但是研究微分流形上微积分学的必不可少的工具，而且可以从这些最简单的情形中领略到微分流形的背景和含义。流形上的微分学便是把这一章介绍的多元映射的微分学搬到流形上去。因此，为了以后便于推广，这里介绍多元映射的微分学时，尽量采用能够推广的最一般方法。

§1.1　多元映射的微分

在初等微积分中，可微函数f：$U\subset R^1\to R^1$在点$x_0\in U$处的导数通常解释为图象在点$(x_0, f(x_0))$处的切线的斜率。这个解释不便于推广到多元映射。为了推广导数（或微分）的概念，把f在点x_0处的微分$Df(x_0)$(或$f'(x_0)$)视为从$U\to R^1$的一个线性映射.f的图象的切线便是线性映射$Df(x_0)$作用在x_0的邻域的像，即映射

$$g:\ U\to R,\ x\to g(x)=f(x_0)+Df(x_0)(x-x_0) \tag{1.1}$$

的像，也称映射g在x_0与f相切。于是f在$x_0\in U$的微分$Df(x_0)$便定义了从R^1到R^1的唯一的线性映射。微分或导数的这种解释便可以推广到多元映射。

定义1.1.1　设$U\subset R^n$为开集，如果映射f、g：$U\subset R^n\to R^m$满足条件

$$\lim_{x\to x_0}\frac{\|f(x)-g(x)\|}{\|x-x_0\|}=0$$

其中$\|\cdot\|$为R^m上的范数，则称f与g在$x_0\in U$相切.

为了方便，今后我们约定$U\subset R^n$皆表示开集。

定理1.1.2　设f：$U\subset R^n\to R^m$，$x_0\in U$，$\mathscr{L}$（R^n，R^m）表示从R^n到R^m的全体线性映射的集合，则至多存在一个映射$L\in\mathscr{L}$（R^n，R^m）使得映射

g_L：$U\subset R^n\to R^m$，$x\to g_L(x)=f(x_0)+L(x-x_0)$在$x_0\in U$与$f$相切.

证明　假设有两个映射L_1、$L_2\in\mathscr{L}$（R^n，R^m）使得g_{L_1}，g_{L_2}均在x_0与f相切，即

$$\lim_{x\to x_0}\frac{\|f(x)-f(x_0)-L_1(x-x_0)\|}{\|x-x_0\|}=0$$

$$\lim_{x\to x_0}\frac{\|f(x)-f(x_0)-L_2(x-x_0)\|}{\|x-x_0\|}=0$$

设$e\in R^n$为满足$\|e\|=1$的任一向量，$x=x_0+\lambda e$，$\lambda\in R$，只要选择λ使得$x\in U$，便有

$$\begin{aligned}\|(L_1-L_2)(e)\|&=\|(L_1-L_2)(\lambda e)\|/\lambda\\&=\|(L_1-L_2)(x-x_0)\|/\|x-x_0\|=\|L_1(x-x_0)-L_2(x-x_0)\|/\|x-x_0\|\\&\leqslant\frac{\|f(x)-f(x_0)-L_1(x-x_0)\|}{\|x-x_0\|}+\frac{\|f(x)-f(x_0)-L_2(x-x_0)\|}{\|x-x_0\|}\end{aligned}$$

当$x\to x_0$时，右端趋于0，从而$\|(L_1-L_2)e\|=0$，由e的任意性，及线性映射范数的定义

$(\forall L\in\mathscr{L}(R^n, R^m)$，$\|L\|=\sup\limits_{\|e\|=1}\|L(e)\|$可知

$$\|L_1-L_2\|=0，即L_1=L_2。\qquad \square$$

定义1.1.3 设f：$U\subset R^n\to R^m$，$x_0\in U$，如果存在线性映射$L\in\mathscr{L}(R^n, R^m)$（由定理1.1.2是唯一的），使得映射$g_L$：$U\subset R^n\to R^m$，$x\to g_L(x)=f(x_0)+L(x-x_0)$在$x_0\in\ddot{U}$与$f$相切，则称$f$在$x_0$是可微的，且称线性映射$L$为$f$在$x_0$的微分，记为$\mathrm{D}f(x_0)=L$。若$f$在$U$中每一点$x_0$都是可微的，则称$f$在$U$上是可微的，映射$\mathrm{D}f$：$U\to\mathscr{L}(R^n, R^m)$，$x\to\mathrm{D}f(x)$称为$f$的微分。如果$\mathrm{D}f$还是连续映射，则称$f$是$C^1$-类的。

微分的上述定义容易理解但不便于应用。为了方便以后的证明，用高阶无穷小给出更为简洁的表示。称映射f：$U\subset R^n\to R^m$在$x_0\in U$可微，即指存在线性映射$\mathrm{D}f(x_0)\in\mathscr{L}(R^n, R^m)$满足如下关系：

$$\begin{cases} f(x_0+h)=f(x_0)+\mathrm{D}f(x_0)h+o(h) \\ \lim\limits_{h\to 0}\dfrac{o(h)}{\|h\|}=0 \end{cases} \tag{1.2}$$

表达式（1·2）与初等微积分中微分的定义一致。

当f的定义域为R的一个开集V时，f：$V\to R^m$成为单元向量值函数。此时f的导数可以象单元实值函数一样来定义；与$\mathscr{L}(R, R)$同构于R（此时，线性映射$L(x)=ax$与实数$a\in R$一一对应）一样，映射φ：$\mathscr{L}(R, R^m)\to R^m$；$\mathrm{D}f(t)\to\mathrm{D}f(t)(1)$，$1\in R$也是线性映射空间$\mathscr{L}(R, R^m)$到$R^m$的一个同构，因此$\mathscr{L}(R, R^m)$标准同构于$R^m$，并且

$$\frac{\mathrm{d}f(t)}{dt}=\mathrm{D}f(t)(1)=\lim_{h\to 0}\frac{f(t+h)-f(t)}{h}$$

这就是向量值函数f在t的普通导数，也称向量的微分。

设$U\subset R^n$为一开子集。为讨论多元实值函数与向量值函数的关系，考虑映射

$$f：U\subset R^n\to R^m$$

如果π^i：$R^m\to R$为到第i个坐标的投影，即$\pi^i(x^1, \cdots, x^m)=x^i$，则映射$f$：$U\subset R^n\to R^m$将由它的坐标函数$f^i=\pi^i f$唯一确定，且

$$f(x)=(f^1(x), \cdots, f^m(x))^T，x\in U\subset R^n$$

反之，在U上任意m个取值于R的实值函数f^1，…，f^m的函数组（$f^1,\cdots,f^m$）都确定了一个映射f：$U\subset R^n\to R^m$，它以函数$f^1,\cdots,f^m$为坐标函数。

定理1.1.4 设f：$U\subset R^n\to R^m$，$f=(f^1, \cdots, f^m)$，其中f^i：$U\to R^1$，$i=1,2,\cdots,m$，均为实值函数，则向量值函数f在$x\in U$可微，当且仅当每个f^i在$x\in U$可微，且$\mathrm{D}f=(\mathrm{d}f^1, \cdots, \mathrm{d}f^m)^T$。 $\square$

证明 由定义1.1.3或表达式（1.2）易证，留作练习。

例 映射f：$R\to R^2$，$x\to(\cos x, \sin 2x)^T$的微分$\mathrm{D}f(x)=(-\sin x, 2\cos 2x)^T$，$x\in R$。

下面介绍微分的一些简单性质：

定理1.1.5 若f：$U\subset R^n\to R^m$在$x_0\in U$可微，则f在x_0连续。

证明 由（1.2）有

$$\lim_{h\to 0} f(x_0+h)=\lim_{h\to 0}[f(x_0)+\mathrm{D}f(x_0)h+o(h)]=f(x_0)$$

性质1.1.6 常值映射在其定义域内任一点均可微，且其微分为0映射，$0\in\mathscr{L}(R^n, R^m)$。

性质1.1.7 线性映射$f=A$：$R^n\to R^m$在任一点$x\in R^n$可微，且$\mathrm{D}f=A$。

上述两性质的证明留作练习。

性质1.1.8（线性性） 设f、g：$U\subset R^n\to R^m$为可微映射，a为实数，则af与$f+g$：$U\subset R^n\to R^m$也是可微的，且

$$\mathrm{D}(af)(x)=a\mathrm{D}f(x),$$
$$\mathrm{D}(f+g)(x)=\mathrm{D}f(x)+\mathrm{D}g(x)。$$

证明 设$x\in U$，$h\in R^n$，使$x+h\in U$，由f、g可微，得

$$f(x+h)=f(x)+\mathrm{D}f(x)(h)+o(h)$$
$$g(x+h)=g(x)+\mathrm{D}g(x)(h)+o(h)$$

两式相加得

$$(f+g)(x+h)=(f+g)(x)+(\mathrm{D}f(x)+\mathrm{D}g(x))(h)+o(h)$$

即
$$\mathrm{D}(f+g)(x)=\mathrm{D}f(x)+\mathrm{D}g(x)$$

类似地，

$$af(x+h)=af(x)+a\mathrm{D}f(x)(h)+ao(h)$$
$$=af(x)+a\mathrm{D}f(x)(h)+o(h)$$

故
$$\mathrm{D}(af)(x)=a\mathrm{D}f(x)$$ □

性质1.1.9（Leibniz法则） 设f、g：$U\subset R^n\to R^m$为可微映射，则$f\cdot g$：$U\subset R^n\to R^m$仍为可微映射且

$$\mathrm{D}(f\cdot g)(x)=(\mathrm{D}g(x))^T\cdot f(x)+(\mathrm{D}f(x))^T\cdot g(x)$$

证明
$$(f\cdot g)(x+h)=f(x+h)\cdot g(x+h)$$
$$=(f(x)+\mathrm{D}f(x)h+o(h))^T(g(x)+\mathrm{D}g(x)h+o(h))$$
$$=f(x)\cdot g(x)+(\mathrm{D}f(x))^T g(x)h+(\mathrm{D}g(x))^T f(x)h+o(h)$$
$$=f(x)\cdot g(x)+[(\mathrm{D}f(x))^T g(x)+(\mathrm{D}g(x))^T\cdot f(x)]h+o(h)$$ □

性质1.1.10（链式法则） 设f：$U\subset R^n\to V\subset R^m$，$g$：$V\to R^p$，$f$在$x_0\in U$和$g$在$y_0=f(x_0)\in V$可微，则复合映射$g\circ f$：$U\to R^p$也在$x_0$可微且

$$\mathrm{D}(g\circ f)(x_0)=\mathrm{D}g(f(x_0))\circ\mathrm{D}f(x_0)$$

证明
$$g\circ f(x_0+h)=g\circ[f(x_0)+\mathrm{D}f(x_0)h+o(h)]$$
$$=g\circ f(x_0)+\mathrm{D}g(f(x_0))[\mathrm{D}f(x_0)h+o(h)]+o[\mathrm{D}f(x_0)h+o(h)]$$
$$=g\circ f(x_0)+\mathrm{D}g(f(x_0))\mathrm{D}f(x_0)h+\mathrm{D}g(f(x_0))(o(h))$$
$$+o[\mathrm{D}f(x_0)h+o(h)]$$

而
$$\frac{\|\mathrm{D}f(x_0)h+o(h)\|}{\|h\|}\leqslant\frac{\|\mathrm{D}f(x_0)\|\|h\|}{\|h\|}+\frac{\|o(h)\|}{\|h\|}$$
$$=\|\mathrm{D}f(x_0)\|+\frac{\|o(h)\|}{\|h\|}\leqslant M$$

$$\|Dg(f(x_0))(o(h))\| \leqslant \|Dg(f(x_0))\|\|o(h)\|$$

所以 $$\frac{\|o(Df(x_0)(h)+o(h))\|}{\|h\|}=\frac{\|o(Df(x_0)(h)+o(h))\|}{\|Df(x_0)(h)+o(h)\|}\times\frac{\|Df(x_0)(h)+o(h)\|}{\|h\|}$$

$$\leqslant M\frac{\|o(Df(x_0)(h)+o(h))\|}{\|Df(x_0)(h)+o(h)\|}$$

从而 $$Dg(f(x_0))(o(h))+o(Df(x_0)(h)+o(h))=o(h)$$

故 $$D(g\circ f)(x_0)=Dg(f(x_0))\circ Df(x_0)$$ □

性质1.1.11（逆映射的微分） 设U、$V\subset R^n$为开集，f：$U\to V$为一同胚，f^{-1}：$V\to U$为f的逆同胚。如果f在点x_0可微，且$Df(x_0)$是R^n到R^n的线性同胚，则f^{-1}在$y_0=f(x_0)$可微，且$Df^{-1}(y_0)$是$Df(x_0)$的逆映射，即

$$Df^{-1}(y_0)=[Df(x_0)]^{-1}.$$

亦即$Df(x)$是线性空间R^n上的线性同构。

证明 $y_0=f(x_0)$，令$t=f^{-1}(y_0+h)-f^{-1}(y_0)$，则

$$f^{-1}(y_0+h)=t+f^{-1}(y_0)=t+x_0$$

从而 $$h=f(x_0+t)-y_0=f(x_0+t)-f(x_0)$$

由于f在x_0可微，即

$$f(x_0+t)-f(x_0)=Df(x_0)(t)+o(t)$$

又由于$Df(x_0)$是线性同胚，因此$[Df(x_0)]^{-1}$存在 且$\|(Df(x_0))^{-1}(h)\|\leqslant M\|h\|$，$M$为某常数。并考虑

$$\begin{aligned}&\|f^{-1}(y_0+h)-f^{-1}(y_0)-[Df(x_0)]^{-1}(h)\|\\=&\|t-[Df(x_0)]^{-1}(f(x_0+t)-f(x_0))\|\\=&\|t-[Df(x_0)]^{-1}[Df(x_0)(t)+o(t)]\|\\=&\|[Df(x_0)]^{-1}(Df(x_0)(t)-Df(x_0)(t)-o(t))\|\\=&\|(Df(x_0))^{-1}\|\|o(t)\|\end{aligned}$$

其中 $t=f^{-1}(y_0+h)-f^{-1}(y_0)$

下面证明$\lim\limits_{h\to 0}\frac{\|o(t)\|}{\|h\|}=0$

事实上，因为$[Df(x_0)]^{-1}(h)=t+[Df(x_0)]^{-1}(o(t))$，而

$$\|(Df(x_0))^{-1}(o(t))\|\leqslant M\|o(t)\|\leqslant M\varepsilon\|t\|\leqslant\frac{1}{2}\|t\|$$

（取ε充分小）于是

$$\|[Df(x_0)]^{-1}(h)\|\geqslant\|t\|-\frac{1}{2}\|t\|=\frac{1}{2}\|t\|$$

即 $\|t\|\leqslant 2M\|h\|$，故$\|o(t)\|\leqslant\varepsilon\|t\|\leqslant 2\varepsilon M\|h\|$，

从而有 $$\lim_{h\to 0}\frac{\|o(t)\|}{\|h\|}=0$$ □

定理1.1.12（微积分基本定理）

（1） 设g：$[a,b]\to R^m$连续，则映射f：$[a,b]\to R^m$，$f(t)=\int_a^t g(s)ds$是可微的且$Df=g$.

（2） 设f：$[a,b]\to R^m$连续，且在$[a,b]$内可微，则$f(b)-f(a)=\int_a^b f'(s)\mathrm{d}s$.

证明 （1） $t_0\in[a,b]$，积分作为一个映射是线性的，连续的。考虑

$$\|f(t_0+h)-f(t_0)-g(t_0)h\| =\left\|\int_{t_0}^{t_0+h}(g(s)-g(t_0))\mathrm{d}s\right\|$$
$$\leqslant |h|\cdot\sup\{\|g(s)-g(t_0)\| \mid t_0\leqslant s\leqslant t_0+h\}$$

由g在x_0的连续性可知，当$h\to 0$时，上式右端趋于0，从而由定义知f在x_0可微，且$\mathrm{D}f=g$。

（2） 令$h(t)=\int_a^t f'(s)\mathrm{d}s-f(t)$

则由（1），在$[a,b]$有$h'(t)=0$，在$[a,b]$上$h(t)$是连续的，若能证$h(t)=h(a)$，$\forall t\in[a,b]$，则有$h(b)=h(a)$，结论得证。

设对某$t\in[a,b]$有$h(t)\neq h(a)$，则由Hahn-Banach定理，存在$\alpha\in(R^n)^*=R^n$使

$$(\alpha\circ h)(t)\neq(\alpha\circ h)(a)$$

因α为线性映射，h在$[a,b]$可微，从而$\alpha\circ h$在$[a,b]$可微，且

$$\mathrm{D}(\alpha\circ h)(t)=\alpha\circ\mathrm{D}h(t)=0$$

即在$[a,b]$内，$\alpha\circ h$为常值映射，由$h(t)$在$[a,b]$上连续知，$\alpha\circ h$在$[a,b]$上也为常值映射，这与$(\alpha\circ h)(t)\neq(\alpha\circ h)(a)$矛盾，故$h(t)=h(a)$，$\forall t\in[a,b]$。 □

下列性质是函数估值的基本工具之一。

定理1.1.13 设f：$U\subset R^n\to R^m$为C^1-映射，$x,y\in U$，映射c：$[0,1]\to U$是U中连接x，y的C^1-弧，即在$(0,1)$上c是C^1的，且$c(0)=x$，$c(1)=y$，则有

$$f(y)-f(x)=\int_0^1\mathrm{D}f(c(t))(c'(t))\mathrm{d}t$$

若U是凸的，且$c(t)=(1-t)x+ty$，则

$$f(y)-f(x)=\left[\int_0^1\mathrm{d}f((1-t)x+ty)\mathrm{d}t\right](y-x)$$

证明 设$g(t)=(f\circ c)(t)$，则$g'(t)=\mathrm{D}f(c(t))(c'(t))$.

又由定理1.1.12有$g(1)-g(0)=\int_0^1 g'(t)\mathrm{d}t$，即

$$f(y)-f(x)=\int_0^1\mathrm{d}f(c(t))(c'(t))\mathrm{d}t$$ □

定理1.1.14（中值定理） 设$U\subset R^n$是凸集，f：$U\subset R^n\to R^m$是C^1-的，则$\forall x$，$y\in U$有

$$\|f(y)-f(x)\|\leqslant\sup\|\mathrm{D}f((1-t)x+ty)\|\cdot\|y-x\| \tag{1.3}$$

又若$\|\mathrm{D}f(x)\|$在U上一致有界M，则$\forall x$，$y\in U$有

$$\|f(y)-f(x)\|\leqslant M\|y-x\|$$

若$m=1$，即f：$U\subset R^n\to R$为实值函数，则

$$\|f(y)-f(x)\|=\mathrm{D}f(\xi)(y-x) \tag{1.4}$$

其中ξ为连接x，y的直线上的一点。 □

注意，上述中值定理中，只有当$m=1$即f为实值函数时，（1.4）才成立。对于一般的映射（$m>1$），只能得出（1.3），而（1.4）一般不成立。例如，f：$R^2\to R^2$由下式定义：

$y=f(x)$，$x=(x_1, x_2)$，$y=(y_1, y_2)$，$y_1=x_1^2$，$y_2=x_2^3$考虑点$(0, 0)$与$(1, 1)$，则

$f(1, 1)-f(0, 0)=(1, 1)$而$Df(\tau)=(2\tau, 3\tau^2)$，故此时（1.4）相当于解方程组

$$1=2\tau,\quad 1=3\tau^2$$

显然此方程组无解，因此无论$\tau\in(0, 1)$何值，公式（1.4）都不成立。

§1.2　方向导数及微分的矩阵表示

定义1.2.1　设f：$U\subset R^n\to R^m$，U为R^n的开子集，e是R^n中一个单位向量，如果

$$\frac{d}{dt}f(x_0+te)\bigg|_{t=0}=\lim_{t\to 0}\frac{f(x_0+te)-f(x_0)}{t}$$

存在，则称f沿方向e可导，并称上述导数为f沿e的方向导数，记为$D_ef(x)$。方向导数与前一节所讲的微分之间有如下关系：

定理1.2.2　设f在x_0可微，则f在x_0处对任意方向e的方向导数均存在且

$$d_ef(x_0)=Df(x_0)e \tag{1.5}$$

证明　设$U\subset R^n$是含x_0的开集且f：$U\to R^m$在x_0可微。任给R^n内的单位向量e，取$\varepsilon>0$使得对每个$t\in(-\varepsilon, \varepsilon)$，$x_0+te\in U$。定义$c$：$(-\varepsilon, \varepsilon)\to U$使得$c(t)=x_0+te$。显然$c$在$t=0$可微且$Dc(0)=e$。又由假设$f$在$x_0=c(0)$可微，据链式法则$f\circ c$在$t=0$可微且

$$D(f\circ c)(0)=Df(c(0))(Dc(0))=Df(x_0)e$$

另一方面，由方向导数的定义

$$D_ef(x_0)=\lim_{t\to 0}\frac{f\circ c(t)-f\circ c(0)}{t}=D(f\circ c)(0)$$

故　$$D_ef(x_0)=Df(x_0)(e)$$ □

上述性质是普通微积分学中方向导数与梯度的关系的推广。这一推广不但建立了映射的方向导数与其微分的关系，而且也可为计算$Df(x_0)e$提供方便。例如，计算从R^n到R^m的二次齐次多项式$f(y)=A(y,y)$的微分，其中$y\in R^n$，$A\in\mathscr{L}^2(R^n,R^m)$（其中$\mathscr{L}^2(R^n,R^m)\equiv\mathscr{L}(R^n,R^n;R^m)$是从$R^n\times R^n$到$R^m$的二重线性映射的集合）。则由链式法则和Leibniz法则有

$$Df(x_0)e=\frac{d}{dt}f(x_0+te)\bigg|_{t=0}\frac{d}{dt}A(x_0+te,\ x_0+te)\bigg|_{t=0}$$

$$=A(x_0,e)+A(e,x_0)$$

如果进一步假定A是对称的，则$Df(x_0)e=2A(x_0,e)$。

在应用中，称所有方向导数都存在的函数是Gâteaux可微的，称上一节定义的可微性是Fréchet可微。如果函数$f(x)$在x_0Gâteaux可微且对一切单位向量e，方向导数$d_ef(x_0)$均相等，则称它为$f(x)$在x_0的Gâteaux导数，记为$d_Gf(x_0)$。Fréchet可微比Gâteaux可微要强些，显然

f 在x处Fréchet可微，则必在x处Gâteaux可微，反之则不然，但若加以限制，二者也可以是一致的。

定义1.2.3 设f：$U \subset R^n \to R^m$，如果f是Gâteaux可微的，且其Gâteaux微分$d_G f(x)$在$\mathscr{L}(R^n, R^m)$中连续，则称f是C^1-Gateaux的。

定理1.2.4 如果f：$U \subset R^n \to R^m$是C^1-Gâteaux的，则f在U中Fréchet可微，且两个微分值相等，即$d_G f(x_0) = Df(x_0)$。

证明 因f是C^1-Gâteaux的，因此$\forall \varepsilon > 0$，$\exists \delta > 0$，使当$\|h\| < \delta$时，恒有$\|d_G f(x_0+h) - d_G f(x_0)\| < \varepsilon$。

下证当$0 < \|h\| < \delta$时，恒有

$$\|f(x_0+h) - f(x_0) - d_G f(x_0)h\| \leqslant \varepsilon \|h\|$$

不失一般性，假设$f(x_0+h) - f(x_0) - d_G f(x_0)h \neq 0$（否则结论自然成立），由Hahn-Banach定理，存在$\alpha \in (R^m)^* = R^m$，使$\|\alpha\| = 1$，且

$$\alpha(f(x_0+h) - f(x_0) - d_G f(x_0)h) = \|f(x_0+h) - f(x_0) - d_G f(x_0)h\| \tag{1.6}$$

令$\varphi(t) = \alpha(f(x_0+th))$，$0 \leqslant t \leqslant 1$，则$\varphi'(t) = \alpha(d_G f(x_0+th)h)$，由中值公式可知，存在$0 < \theta < 1$，使$\varphi(1) - \varphi(0) = \varphi'(\theta)$，即

$$\alpha(f(x_0+h) - f(x_0)) = \alpha(d_G f(x_0+\theta h)h) \tag{1.7}$$

由（1.6）、（1.7）（注意$\|\theta h\| < \delta$），得

$$\begin{aligned}\|f(x_0+h) - f(x_0) - d_G f(x_0)h\| &= \alpha(f(x_0+h) - f(x_0) - d_G f(x_0)h) \\ &= \alpha(d_G f(x_0+\theta h) - d_G f(x_0))h \\ &\leqslant \|\alpha\| \cdot \|d_G f(x_0+\theta h) - d_G f(x_0)\| \cdot \|h\| \\ &= \|d_G f(x_0+\theta h) - d_G f(x_0)\| \cdot \|h\| \leqslant \varepsilon \|h\|\end{aligned}$$

故f在x_0处F-可微，且$Df(x_0) = d_G f(x_0)$。 □

下面介绍偏导数的概念。

设$e_i = (0, \cdots 0, 1, 0, \cdots, 0)^T$为$R^n$的标准单位向量（$1 \leqslant i \leqslant n$），$e_1, \cdots, e_n$为$R^n$的标准基。

定义1.2.5 设f：$U \to R^m$是定义在开集$U \subset R^n$上的映射，如果映射$f(x+te_i)$在$t=0$都是可微的，即

$$\lim_{t \to 0} \frac{f(x+te_i) - f(x)}{t} = \left.\frac{df(x+te_i)}{dt}\right|_{t=0}$$

存在，则称此极限值为f在$x_0 \in U$关于第i个变量的偏导数，记为$d_i f(x)$或$\frac{\partial f}{\partial x^i}(x)$。

显然，如果f在x_0是Gâteaux可微的，则所有$d_i f(x)$，$1 \leqslant i \leqslant n$都存在。事实上，偏导数是一种特殊的方向导数。

定理1.2.6 设$U \subset R^n$，f：$U \to R^m$，$x \in U$，如果f在x处Fréchet可微，则有

(i) f在x的所有偏导数存在，且

$$d_i f(x) = Df(x)e_i, \quad i = 1, \cdots, n。 \tag{1.8}$$

(ii) 设$h = (h^1, \cdots, h^n)^T \in R^n$，则

$$\mathrm{D}f(x)h = \sum_{i=1}^{m} \mathrm{d}_i f(x) h^i \tag{1.9}$$

证明 (i)是显然的．令 $h = \sum_{i=1}^{n} h^i e_i$，再应用(1.8)即得到(ii)。

反过来，也有

定理1.2.7 若 $\mathrm{d}_i f: U \to \mathscr{L}(R^1, R^n)$，$i=1, \cdots, n$ 均存在且连续，则f是Fréchet可微的，且

$$\mathrm{D}f(x) = (\mathrm{d}_1 f(x), \cdots, \mathrm{d}_n f(x)) \tag{1.10}$$

证明 由定义1.2.1有

$$f(x^1+h^1, \cdots, x^i+h^i, \cdots, x^n+h^n) - f(x^1, \cdots, x^n) - \sum_{i=1}^{n} \mathrm{d}_i f(x) h^i$$

$$= f(x^1+h^1, \cdots, x^n+h^n) - f(x^1, x^2+h^2, \cdots, x^n+h^n) - \mathrm{d}_1 f(x) h^1$$

$$+ \sum_{i=2}^{n-1} [f(x^1, \cdots, x^{i-1}, x^i+h^i, x^{i+1}+h^{i+1}, \cdots, x^n+h^n) - f(x^1,$$

$$\cdots, x^i, \cdots, x^n+h^n) - \mathrm{d}_i f(x) h^i] + f(x^1, \cdots, x^{n-1}, x^n+h^n) - f(x^1, \cdots, x^n)$$

$$- \mathrm{d}_n f(x) h^n = [\mathrm{d}_1 f(x^1+t_1 h^1, x^2+h^2, \cdots, x^n+h^n) - \mathrm{d}_1 f(x)] h^1$$

$$+ \sum_{i=2}^{n-1} [\mathrm{d}_i f(x^1, \cdots, x^{i-1}, x^i+t_i h^i, x^{i+1}+h^{i+1}, \cdots, x^n+h^n) - \mathrm{d}_i f(x)] h^i$$

$$+ [\mathrm{d}_n f(x^1, \cdots, x^{n-1}, x^n+t_n h^n) - \mathrm{d}_n f(x^1, \cdots, x^n)] h^n$$

于是有

$$\|f(x+h) - f(x) - \sum_{i=1}^{n} \mathrm{d}_i f(x) h^i\| \leqslant \sum_{i=1}^{n} \sup \|\mathrm{d}_i f(\xi_i) - \mathrm{d}_i f(x)\| \|h\|$$

其中 $\|h\| = \sum_{i=1}^{n} \|h^i\|$，$\xi_i = (x^1, \cdots, x^{i-1}, x^i + t_i h^i, x^{i+1}, \cdots, x^n)$，$0 < t_i < 1$．由偏导数的连续性，当$h \to 0$时$\xi_i \to x$，上式右端趋于0，从而左端趋于0，这说明$\mathrm{D}f(x)$存在，且

$$\mathrm{D}f(x)h = \sum_{i=1}^{n} \mathrm{d}_i f(x) h^i$$

所以有$\mathrm{D}f(x) = (\mathrm{d}_1 f(x), \cdots, \mathrm{d}_n f(x))$。 □

下面，用偏导数推导函数微分的矩阵表示，将映射f：$U \subset R^n \to R^m$写成分量形式$f(x^1, \cdots, x^n) = (f^1(x^1, \cdots, x^n), \cdots, f^m(x^1, \cdots, x^n))^T$。

定理1.2.8 设映射f：$U \subset R^n \to R^m$在$x \in U$处Fréchet可微，则偏导数$\mathrm{d}_i f(x)$有如下矩阵表示

$$\left(\frac{\partial f^1(x)}{\partial x^i},\ \cdots,\ \frac{\partial f^m(x)}{\partial x^i}\right)^T \tag{1.11}$$

从而，由定理1.2.7微分$\mathrm{D}f(x)$有如下矩阵表示

$$A=\frac{\partial(f^1,\ \cdots,\ f^m)}{\partial(x^1,\ \cdots,\ x^n)}=\begin{pmatrix}\dfrac{\partial f^1}{\partial x^1} & \cdots & \dfrac{\partial f^1}{\partial x^n}\\ \vdots & & \vdots\\ \dfrac{\partial f^m}{\partial x^1} & \cdots & \dfrac{\partial f^m}{\partial x^n}\end{pmatrix},$$

亦即$\mathrm{D}f(x)$的矩阵表示即为f在x处的Jacobi矩阵。

当$h\in R^n$时它们满足下列关系

$$\mathrm{D}f(x)(h)=A\cdot h \tag{1.12}$$

事实上，由向量值函数的微分定义有

$$\begin{aligned}\mathrm{d}_i f(x)&=\lim_{t\to 0}\frac{f(x+te_i)-f(x)}{t}\\ &=\lim_{t\to 0}\left(\frac{f^1(x+te_i)-f^1(x)}{t},\ \cdots,\ \frac{f^m(x+te_i)-f^m(x)}{t}\right)^T\\ &=\left(\frac{\partial f^1}{\partial x^i},\ \cdots,\ \frac{\partial f^m}{\partial x^i}\right)^T\end{aligned}$$

这就是公式（1.11）。由定理1.2.7，公式（1.12）是显然的。

推论1.2.9 当$n=1$时，f：$(a,\ b)\to R^m$即为普通的向量值函数，当$t\in(a,\ b)$时，$f(t)=(f^1(t),\ \cdots,\ f^m(t))^T$其微分$\mathrm{D}f(t)$的矩阵表示为：

$$\left(\frac{\mathrm{d}f^1}{\mathrm{d}t},\ \cdots,\ \frac{\mathrm{d}f^m}{\mathrm{d}t}\right)^T \tag{1.13}$$

推论1.2.10 当$m=1$时，f：$U\subset R^n\to R^1$为普通的n元实值函数，当$x\in U$，$f(x)=f(x^1,\ \cdots,\ x^n)$在$x$处的微分$\mathrm{D}f(x)$的矩阵表示为：

$$\left(\frac{\partial f}{\partial x^1},\ \cdots,\ \frac{\partial f}{\partial x^n}\right)=\mathrm{grad}f(x) \tag{1.14}$$

推论1.2.11（链式法则的矩阵表示） 函数f，g的假设条件同性质1.1.10，设$x\in U\subset R^n$，$y=f(x)\in V\subset R^m$，$z=g(y)\in R^p$，则链式法则

$$\mathrm{D}(g\circ f)(x)=\mathrm{D}g(f(x))\circ\mathrm{D}f(x)$$

可用相应的矩阵表示为：

$$\frac{\partial(g^1\circ f,\ \cdots,\ g^p\circ f)}{\partial(x^1,\ \cdots,\ x^n)}=\frac{\partial(g^1,\ \cdots,\ g^p)}{\partial(y^1,\ \cdots,\ y^m)}\bigg|_{y=f(x)}\frac{\partial(f^1,\ \cdots,f^m)}{\partial(x^1,\cdots,x^n)}(x) \tag{1.15}$$

事实上，用分量表示有

$$\frac{\partial(g\circ f)^j}{\partial x^i}=\sum_{k=1}^{m}\frac{\partial g^j(f(x))}{\partial y^k}\cdot\frac{\partial f^k(x)}{\partial x^i},\quad i=1,\cdots,n;\ j=1,\cdots,p$$

再按矩阵乘法整理即得到(2.11)。 □

例 设f：$R^2\to R^3$，$f(x,y)=(x^2,\ x^3y,\ x^4y^2)^T$，则微分$\mathrm{D}f(x,\ y)$的矩阵表示为：

$$\begin{pmatrix} \frac{\partial f^1}{\partial x}, & \frac{\partial f^1}{\partial y} \\ \frac{\partial f^2}{\partial x}, & \frac{\partial f^2}{\partial y} \\ \frac{\partial f^3}{\partial x}, & \frac{\partial f^3}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x & 0 \\ 3x^2y & x^3 \\ 4x^3y^2, & 2x^4y \end{pmatrix}$$

其中$f^1(x, y)=x^2$，$f^2(x, y)=x^3y$，$f^3(x,y)=x^4y^2$。

值得注意的是，只要映射$f=(f^1, \cdots, f^m)$中的每个函数f^i在x_0处的偏导数存在就可以形成其Jacobi矩阵，但不能因此得出f在x_0可微的结论，其Jacobi矩阵也不能说成是f的微分。仅当映射f在x_0可微时，才可说Jacbi矩阵是其微分的表示。有一个充分条件说，当f的所有偏导数在x_0连续则f在x_0可微，从而其Jacobi矩阵也就是其微分的矩阵表示，这也是定理1.2.7的结论。

§1.3 高阶微分与Taylor公式

与一元函数微分学一样，可以引进高阶微分的概念。设f是从开集$U\subset R^n$到R^m的可微映射，在每点$x\in U$均有Fréchet微分$\mathrm{D}f(x)$。这样微分$\mathrm{D}f(x)$又可以看成U上的线性映射。如果$\mathrm{D}f$：$U\to\mathscr{L}(R^n, R^m)$仍在$x$处可微，则说$f$在$x$处二阶Fréchet可微，并记其微分为：

$$\mathrm{D}^2f(x)=\mathrm{D}(\mathrm{D}f(x))\in\mathscr{L}(U, \mathscr{L}(R^n, R^m))$$

如此下去，一般地f在x处的k阶微分为

$$\mathrm{D}^kf(x)=\mathrm{D}(\mathrm{D}^{k-1}f)(x),$$

它是$\mathscr{L}(U, \underbrace{\mathscr{L}(R^n, \cdots\mathscr{L}(R^n, R^m))}_{k-1\text{重}})$的元素。

为了刻划空间$\mathscr{L}(R^n, \mathscr{L}(R^n, \cdots, \mathscr{L}(R^n, R^m)))$，引入多重线性映射的概念（在以后引进张量时，还要用到多重线性映射）。

设$E_1, \cdots, E_k$和F是线性空间，映射A：$E_1\times E_2\times\cdots\times E_k\to F$称为$k$重线性映射，如果$A(x_1, \cdots, x_k)$，$x_i\in E_i$，$i=1, \cdots, k$，当任意固定其中$k-1$个变元时，作为剩下的一个变元的映射是线性的，即$A(x_1, \cdots, \lambda x_i+\mu\widetilde{x}_i, \cdots, x_k)=\lambda A(x_1, \cdots, x_i, \cdots, x_k)+\mu A(x_1, \cdots, \widetilde{x}_i, \cdots, x_k)$。

记$\mathscr{L}(E_1, \cdots, E_k; F)$为从$E_1\times\cdots\times E_k$到$F$的连续$k$重线性映射的集合，当$E_1=\cdots=E_k=E$时，记为$\mathscr{L}^k(E; F)$；空间$\mathscr{L}(E_1, \cdots, E_k; F)$的范数定义为

$$\|A\|=\sup\{\|A(x_1, \cdots, x_k)\| \mid \|x_1\|=\cdots=\|x_k\|=1\}$$

定理1.3.1 $\mathscr{L}(E_1, \mathscr{L}(E_2, \cdots, \mathscr{L}(E_K; F)))$等距同构于$\mathscr{L}(E_1, E_2, \cdots, E_k; F)$。

证明 设$A\in\mathscr{L}(E_1, \mathscr{L}(E_2, \cdots, (\mathscr{L}E_k; F)))$，定义$\widetilde{A}\in\mathscr{L}(E_1, \cdots E_k; F)$为$\widetilde{A}(x_1, \cdots, x_k)=A(x_1)(x_2)\cdots(x_k)$，则$A\to\widetilde{A}$是一对一的和线性的，且$\|\widetilde{A}\|=\|A\|$。 □

特别地，当$E_1=E_2=\cdots=E_k=R^n$时，$\mathscr{L}(R^n, \mathscr{L}(R^n, \cdots \mathscr{L}(R^n, R^m)))$与$\mathscr{L}^k(R^n, R^m)=$

$\mathscr{L}(R^n, \cdots, R^n; R^m)$ 等距同构。

现在来定义高阶微分：

定义1.3.2 设 $f: U\subset R^n\to R^m$ 是 $k-1$ 阶Fréchet可微的(简称 $k-1$ 阶 F-可微)即 $D^{k-1}f(x)$ 在 U 上有定义，如果 $D^{k-1}f: U\to \mathscr{L}^{k-1}(R^n, R^m)$ 在 $x\in U$ 仍然Fréchet可微，则称 f 在 x 处 k 阶 F-可微，记 $D^{k-1}f$ 在 x 处的 F-微分为 $D^kf(x)\in\mathscr{L}^k(R^n, R^m)$，称它为 f 在 x 处的 k 阶 F-微分。如果 $D^kf(x)$ 在 U 上还是连续的，则称 f 在 U 上 k 阶连续 F-可微，或称 f 是 C^k-的，记为 $f\in C^k(U, R^m)$。

下面不加证明地列出 n 阶微分的性质。

定理1.3.3(线性性) 设 f、g: $U\subset R^n\to R^m$ 是 k 次可微映射，a 为一实数，则 af 与 $f+g$: $U\subset R^n\to R^m$ 也是 k 次可微的，且

$$D^k(af)=aD^kf,\quad D^k(f+g)=D^kf+D^kg$$

定理1.3.4 设 f^i: $U\subset R^n\to R$, $1\leqslant i\leqslant m$ 是 r 次可微映射，则 $f=(f^1, \cdots, f^m)$: $U\subset R^n\to R^m$ 在 x 处是 r 次可微的，且 $D^rf=(D^rf^1, \cdots, D^rf^m)$。

定理1.3.5 设 f: $U\subset R^n\to V\subset R^m$ 与 g: $V\subset R^m\to R^p$ 均为 C^k-的，则复合映射 $g\circ f$: $U\subset R^n\to R^p$ 也是 C^k-的。

定理1.3.6 如果 f: $U\subset R^n\to R^m$ 的所有偏导数 d_if: $U\to\mathscr{L}(R, R^m)$, $i=1,\cdots, m$ 均为 C^{k-1} 的，则 f 是 C^k-的。

定理1.3.7 高阶微分与方向导数有如下关系，设 f: $U\subset R^n\to R^m$, $h_1,\cdots, h_k\in R^n$, 则

$$D^kf(x)(h_1, \cdots, h_k)=\frac{\partial}{\partial t_k}\cdots\frac{\partial}{\partial t_1}\left\{f\left(x+\sum_{i=1}^k t_ih_i\right)\right\}\Bigg|_{t_1=\cdots=t_k=0} \tag{1.16}$$

定理1.3.8 设 f: $U\subset R^n\to R^m$ 在 $x\in U$ 处 k 阶可微，$f(x)=(f^1(x), \cdots, f^m(x))^T$，则在 R^n 的标准基下 $D^kf(x)$ 的矩阵表示的分量为：

$$\frac{\partial^kf^j}{\partial x^{i_1}\cdots\partial x^{i_n}}, \text{其中} 1\leqslant j\leqslant m,\ 0\leqslant i_1,\cdots,i_n\leqslant k,\ i_1+i_2+\cdots+i_n=k \tag{1.17}$$

由此可推知，当所有的 k 阶偏导数都存在且连续时，f 是 C^k-的。

例1 $m=1$ 时 $D^2f(x)$ 的矩阵表示为：

$$A=\begin{pmatrix}\dfrac{\partial^2f}{\partial x^1\partial x^1} & \cdots & \dfrac{\partial^2f}{\partial x^1\partial x^n}\\ \vdots & & \\ \dfrac{\partial^2f}{\partial x^n\partial x^1} & & \dfrac{\partial^2f}{\partial x^n\partial x^n}\end{pmatrix}$$

它与 $D^2f(x)$ 满足如下

$$D^2f(x)(h)(k)=h^TAk,\qquad h,\ k\in R^n$$

有时也简记为 $A\cdot h\cdot k$。

例2 f: $U\subset R^n\to R^m$, $x\in U$, $D^2f(x)$ 可表示为