17.3 偏导数和方向导数

摘录

1. （梯度与导数矩阵）根据定理17.3.8和引理17.3.5，我们可以知道若有函数$f:E\to\mathbb{R}^m$在某个集合F上的偏导数存在且连续，那么在F中的任意一个内点x_0处，全体方向导数都存在且有公式：

$$D_{(v_1,\dots,v_n)}f(x_0)=\sum_{j=1}^{n}v_j\frac{\partial f}{\partial x_j}(x_0)$$

特别地，如果$f:E\to\mathbb{R}$是一个实值函数，并且定义f在x_0处的**梯度**$\nabla f(x_0)$为n维行向量为$\nabla f(x_0):=\left(\frac{\partial f}{\partial x_1}(x_0),\dots,\frac{\partial f}{\partial x_n}(x_0)\right)$。那么只要$x_0$是某个梯度存在且连续的内点，我们就有公式：

$$D_vf(x_0)=v\cdot\nabla f(x_0)$$

更一般地，如果$f:E\to\mathbb{R}^m$是一个函数，其中$f=(f_1,\dots,f_m)$，而F是满足f的偏导数存在且连续的E的子集，x_0是F的内点，那么根据定理17.3.8我们有：

$$\begin{aligned}f'(x_0)(v_j)_{1\le j\le n}&=\sum_{j=1}^{n}v_j\frac{\partial f}{\partial x_j}(x_0)\\&=\left(\sum_{j=1}^{n}v_j\frac{\partial f_i}{\partial x_j}(x_0)\right)_{i=1}^{m}\end{aligned}$$

考虑到$f'(x_0)$是一个线性映射，于是上面的式子可以改写为：

$$L_{Df(x_0)}(v_j)_{1\le j\le n}$$

其中$Df(x_0)$是如下所示的$m\times n$矩阵：

$$\begin{aligned}Df(x_0)&:=\left(\frac{\partial f_i}{\partial x_j}(x_0)\right)_{1\le i\le m;1\le j\le n}\\&=\begin{pmatrix}\frac{\partial f_1}{\partial x_1}(x_0)&\frac{\partial f_1}{\partial x_2}(x_0)&\cdots&\frac{\partial f_1}{\partial x_n}(x_0)\\\frac{\partial f_2}{\partial x_1}(x_0)&\frac{\partial f_2}{\partial x_2}(x_0)&\cdots&\frac{\partial f_2}{\partial x_n}(x_0)\\\vdots&\vdots&\ddots&\vdots\\\frac{\partial f_m}{\partial x_1}(x_0)&\frac{\partial f_m}{\partial x_2}(x_0)&\cdots&\frac{\partial f_m}{\partial x_n}(x_0)\end{pmatrix}\end{aligned}$$

也即：

$$(D_vf(x_0))^T=(f'(x_0)v)^T=Df(x_0)v^T$$

矩阵$Df(x_0)$有时也被称为f在x_0处的**导数矩阵**或者**微分矩阵**。考虑到梯度的定义，我们同样可以把Df写成：

$$Df(x_0)=\left(\frac{\partial f}{\partial x_1}(x_0)^T,\frac{\partial f}{\partial x_2}(x_0)^T,\cdots,\frac{\partial f}{\partial x_n}(x_0)^T\right)=\begin{pmatrix}\nabla f_1(x_0)\\\nabla f_2(x_0)\\\vdots\\\nabla f_m(x_0)\end{pmatrix}$$

如果特别考虑f是一个实值函数的情景，那么Df就是∇f。

定义

1. (17.3.1 方向导数) 设E是$\mathbb{R}^n$的子集，$f : E \to \mathbb{R}^m$是一个函数，x_0是E的内点，并设v是$\mathbb{R}^n$中的向量。如果极限

$$\lim_{t\to 0; t>0, x_0+tv\in E} \frac{f(x_0+tv)-f(x_0)}{t}$$

存在，那么我们称**f在x_0处沿方向v可微**，并将上述极限记作$D_v f(x_0)$：

$$D_v f(x_0) := \lim_{t\to 0; t>0} \frac{f(x_0+tv)-f(x_0)}{t}$$

(注：应当将这个概念与可微的定义做比较，由于方向导数的除数是标量因此我们可以直接通过极限去定义它；我们不在边界给出定义，但是如果只是为了定义“朝里”方向上的方向导数，那么也可以在边界上给出方向导数的定义)

2. (17.3.7 偏导数) 设E是$\mathbb{R}^n$的子集，$f : E \to \mathbb{R}^m$是一个函数，x_0是E的内点，并设$1 \le j \le n$。如果极限

$$\lim_{t\to 0; t\neq 0, x_0+te_j\in E} \frac{f(x_0+te_j)-f(x_0)}{t}$$

存在，那么我们定义**f在x_0处关于变量x_j的偏导数**为$\frac{\partial f}{\partial x_j}(x_0)$：

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t\to 0; t\neq 0, x_0+te_j\in E} \frac{f(x_0+te_j)-f(x_0)}{t} = \frac{\mathrm{d}}{\mathrm{d}t} f(x_0+te_j)|_{t=0}$$

(注：偏导数即将函数视为x_j的单变量函数然后求导；如果我们将f写为$f=(f_1,\dots,f_m)$，那么显然有$\frac{\partial f}{\partial x_j}(x_0) = \left(\frac{\partial f_1}{\partial x_j}(x_0),\dots,\frac{\partial f_m}{\partial x_j}(x_0)\right)$；注意区分偏导数与方向导数，这点我们会在习题17.3.2中讨论)

命题

1. (17.3.5 全导数与方向导数？) 设E是$\mathbb{R}^n$的子集，$f : E \to \mathbb{R}^m$是一个函数，x_0是E的内点，并设v是$\mathbb{R}^n$中的向量。如果f在x_0处可微，那么f在x_0处沿方向v也可微，并且

$$D_v f(x_0) = f'(x_0)v$$

(注：这个命题说明了全可微性蕴含了方向可微性，但是反过来这个命题并不成立)

2. (17.3.8 偏导数与全导数？) 设E是$\mathbb{R}^n$的子集，$f : E \to \mathbb{R}^m$是一个函数，F是E的子集，并设x_0是F的内点。如果f在F上的全体偏导数$\frac{\partial f}{\partial x_j}$都存在，并且它们在$x_0$处也都是连续的，那么$f$在$x_0$处是可微的，并且线性变换$f'(x_0) : \mathbb{R}^n \to \mathbb{R}^m$定义有

$$f'(x_0)(v_j)_{1\le j\le n} = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)$$

(注：这给出了另一种确定全导数的方法)

课后习题

17.3.1 证明引理17.3.5（这与习题17.2.2类似）

> 若f在x_0处可微，那么存在线性映射L使得极限
>
> $$\lim_{x\to x_0;x\in E\setminus\{x_0\}}\frac{\|f(x)-(f(x_0)+L(x-x_0))\|}{\|x-x_0\|}=0$$
>
> 换言之也即对任意的$\varepsilon>0$，存在$\delta>0$使得对任意的$x\in B(x_0,\delta)\setminus\{x_0\}$都有：
>
> $$\frac{\|f(x)-(f(x_0)+L(x-x_0))\|}{\|x-x_0\|}<\varepsilon$$
>
> 于是考虑任意的$v\in\mathbb{R}^n$，对任意的$\varepsilon>0$，根据上面的结论我们知道存在一个$\delta>0$使得对任意的$x\in B(x_0,\delta)\setminus\{x_0\}$都有：
>
> $$\frac{\|f(x)-(f(x_0)+L(x-x_0))\|}{\|x-x_0\|}<\frac{\varepsilon}{\|v\|}$$
>
> 特别地，我们考察满足$x=x_0+tv$（其中$t>0$）的$x\in B(x_0,\delta)\setminus\{x_0\}$（于是显然有$t<\dfrac{\delta}{\|v\|}$）。于是上面的式子可以改写为：
>
> $$\frac{\|f(x_0+tv)-f(x_0)-L(tv)\|}{t\|v\|}<\frac{\varepsilon}{\|v\|}\iff\left\|\frac{f(x_0+tv)-f(x_0)}{t}-Lv\right\|<\varepsilon$$
>
> 换言之即：对任意的$\varepsilon>0$，存在$\dfrac{\delta}{\|v\|}>0$使得对任意的t满足$t>0$且$|t|<\dfrac{\delta}{\|v\|}$有
>
> $$\left\|\frac{f(x_0+tv)-f(x_0)}{t}-Lv\right\|<\varepsilon$$
>
> 从而根据收敛的定义，即有$D_vf(x_0)=\lim\limits_{t\to0;t>0,x_0+tv\in E}\dfrac{f(x_0+tv)-f(x_0)}{t}=Lv$，考虑到$L$就是$f'(x_0)$，于是引理17.3.5得证。

17.3.2 设E是$\mathbb{R}^n$的子集，$f:E\to\mathbb{R}^m$是一个函数，x_0是E的内点，并设$1\leq j\leq n$。证明：$\dfrac{\partial f}{\partial x_j}(x_0)$存在当且仅当$D_{e_j}f(x_0)$与$D_{-e_j}f(x_0)$都存在且互为相反数（也即$D_{e_j}f(x_0)=-D_{-e_j}f(x_0)$），据此进一步得到$\dfrac{\partial f}{\partial x_j}(x_0)=D_{e_j}f(x_0)$

> 分别证明充分必要性。
>
> ---
>
> 若$\dfrac{\partial f}{\partial x_j}(x_0)$存在（我们简记为$L$），则根据极限的定义即对任意的$\varepsilon>0$，存在$\delta>0$使得对任意的$t\in(-\delta,0)\cup(0,\delta)$都有：
>
> $$\left\|\frac{f(x_0+te_j)-f(x_0)}{t}-L\right\|<\varepsilon$$
>
> 从而对任意的$z\in(0,\delta)$，我们有：
>
> $$\begin{cases}\left\|\dfrac{f(x_0+ze_j)-f(x_0)}{z}-L\right\|<\varepsilon & z\in(0,\delta)\\ \left\|\dfrac{f(x_0+(-z)e_j)-f(x_0)}{(-z)}-L\right\|<\varepsilon & -z\in(-\delta,0)\end{cases}$$

特别地，由于对第二个式子化简，我们有：

$$\left\|\frac{f(x_0+z(-e_j))-f(x_0)}{z}-(-L)\right\|<\varepsilon$$

于是第一第二个式子分别对应了极限 $\lim\limits_{t\to 0;t>0}\dfrac{f(x_0+te_j)-f(x_0)}{t}=L$与

$\lim\limits_{t\to 0;t>0}\dfrac{f(x_0+t(-e_j))-f(x_0)}{t}=L$，即$D_{e_j}f(x_0)$与$D_{-e_j}f(x_0)$都存在且互为相反数，从而必要性得证。

若$D_{e_j}f(x_0)$与$D_{-e_j}f(x_0)$都存在且互为相反数（我们分别记为L与$-L$），于是即对任意的$\varepsilon>0$，分别存在$\delta_1,\delta_2>0$使得：

$$\begin{cases}\forall\, t\in(0,\delta_1),\left\|\dfrac{f(x_0+te_j)-f(x_0)}{t}-L\right\|<\varepsilon\\ \forall\, t\in(0,\delta_2),\left\|\dfrac{f(x_0+t(-e_j))-f(x_0)}{t}-(-L)\right\|<\varepsilon\end{cases}$$

于是我们令$\delta:=\min(\delta_1,\delta_2)$，从而对任意的$t\in(-\delta,0)\cup(0,\delta)$，我们有：

$$\begin{cases}\text{if}\, t\in(0,\delta)\implies t\in(0,\delta_1),\left\|\dfrac{f(x_0+te_j)-f(x_0)}{z}-L\right\|<\varepsilon\\ \text{if}\, t\in(-\delta,0)\implies -t\in(0,\delta_2),\left\|\dfrac{f(x_0+(-t)(-e_j))-f(x_0)}{(-t)}-(-L)\right\|<\varepsilon\end{cases}$$

再化简第二行的式子，有：

$$\begin{cases}\text{if}\, t\in(0,\delta)\implies t\in(0,\delta_1),\left\|\dfrac{f(x_0+te_j)-f(x_0)}{z}-L\right\|<\varepsilon\\ \text{if}\, t\in(-\delta,0)\implies -t\in(0,\delta_2),\left\|\dfrac{f(x_0+te_j)-f(x_0)}{t}-L\right\|<\varepsilon\end{cases}$$

从而根据极限的定义我们有 $\lim\limits_{t\to 0;t\neq 0,x_0+te_j\in E}\dfrac{f(x_0+te_j)-f(x_0)}{t}=L$，于是结论得证。

综上，于是题目结论得证。

17.3.3 设$f:\mathbb{R}^2\to\mathbb{R}$是具有如下定义的函数：当$(x,y)\neq(0,0)$时，$f(x,y):=\dfrac{x^3}{x^2+y^2}$；当$(x,y)=(0,0)$时，$f(x,y):=0$。证明：$f$在(0，0)处不可微，尽管$f$在$(0,0)$处沿着任何方向$v\in\mathbb{R}^2$都可微。解释为什么这不与定理17.3.8相矛盾

不妨使用反证法去证明f在(0，0)处不可微。我们设f是在$(0,0)$处可微的且导数为$L:\mathbb{R}^2\to\mathbb{R}$，从而对任意的$\varepsilon>0$，都存在$\delta>0$使得当$(x,y)\neq 0\in\mathbb{R}^2$满足$\|(x,y)-0\|<\delta$（也就是$x^2+y^2<\delta^2$）的时候都有：

$$\frac{\|f(x,y)-(f(0,0)+L((x,y)-(0,0)))\|}{\|(x,y)-(0,0)\|}<\varepsilon\iff\frac{\left|\dfrac{x^3}{x^2+y^2}-L(x,y)\right|}{\sqrt{x^2+y^2}}<\varepsilon$$

特别地，由于L是一个线性变换，因此存在一个1×2矩阵$A=(a\quad b)$使得：

$$L(x,y)=(a\quad b)\begin{pmatrix}x\\y\end{pmatrix}=ax+by$$

于是我们的反证假设即存在两个实数a,b使得对任意的$\varepsilon>0$，存在$\delta>0$使得：

$$\frac{\left|\frac{x^3}{x^2+y^2}-ax-by\right|}{\sqrt{x^2+y^2}}<\varepsilon \tag{1}$$

对所有的满足$\sqrt{x^2+y^2}<\delta$的(x,y)成立。假如b是一个非零实数，则我们考虑$\varepsilon=|b|/2$的情况，此时考虑点$(0,\delta/2)$代入(1)式：

$$\frac{|b\delta/2|}{\delta/2}=|b|>|b|/2$$

与"对所有的满足$\sqrt{x^2+y^2}<\delta$的(x,y)成立(1)式"的假设矛盾，因此b只能是0。然后类似地假如a是一个非零实数，则我们考虑$\varepsilon=|1-a|/2$的情况，在上面$b=0$的前提下将点$(\delta/2,0)$代入(1)式：

$$\frac{|\delta/2-a\delta/2|}{|\delta/2|}=|1-a|>|1-a|/2$$

同样与"对所有的满足$\sqrt{x^2+y^2}<\delta$的(x,y)成立(1)式"的假设矛盾，因此a同样也只能是0。但是如果a,b都为0也会导出矛盾，这是因为：

对任意的$0<\varepsilon<1$，我们单独考察所有形如$(\sigma,0)$的点，代入(1)式有：

$$\frac{\|f(x,y)-(f(0,0)+L((x,y)-(0,0)))\|}{\|(x,y)-(0,0)\|}=\frac{|\sigma|}{\sqrt{\sigma^2}}=1>\varepsilon \tag{2}$$

这和导数定义要求的"对任意$\varepsilon>0$，都存在$\delta>0$使得对任意的(x,y)满足$\|(x,y)-(0,0)\|<\delta$都有(2)式小于ε"矛盾（因为无论δ多小你总是能找到小于δ的σ）。

另一方面，f在$(0,0)$处沿着任何方向$v\in\mathbb{R}^2$都可微。设$v:=(a,b)$，根据方向导数的定义我们有：

$$D_{(a,b)}f(0,0)=\lim_{t\to0;t>0}\frac{\frac{t^3a^3}{t^2a^2+t^2b^2}}{t}=\frac{a^3}{a^2+b^2}$$

因此f是在$(0,0)$处沿着任意方向可微的。

（事实上也可以先证明这一点再利用引理17.3.5来证明f在$(0,0)$不是全可微的，这里仅提供思路）

之所以不与定理17.3.8矛盾，是因为f在$(0,0)$处关于变量x,y的偏导存在但是不连续，这一点可以尝试从方向$(a,2a)$逼近$(0,0)$取计算得到：

$$\lim_{a\to0}\frac{\partial f}{\partial x}(a,2a)=\frac{13}{25}\neq1 \qquad \lim_{a\to0}\frac{\partial f}{\partial y}(a,2a)=-\frac{4}{25}\neq0$$

因此f不是符合定理17.3.8要求的函数，自然也不能应用定理17.3.8。

17.3.4 设$f:\mathbb{R}^n\to\mathbb{R}^m$是一个可微函数，并且对于所有的$x\in\mathbb{R}^n$都有$f'(x)=0$。证明：$f$是一个常值函数（提示：你可能会用到一元函数的平均值定理（推论10.2.9）或者微积分基本定理。但是需要注意这些定理并不能对多元函数进行直接类比，不建议使用第一性原理来解题）；另一个更难的挑战是，尝试将定义域$\mathbb{R}^n$换成$\mathbb{R}^n$的一个连通开子集Ω后再证明这个命题

> （思路不是本人想的， 转述别人的证明）
>
> 我们直接来证明对$\mathbb{R}^n$的一个连通开子集Ω的结论，对于$\mathbb{R}^n$也是通用的。为了完成这个证明，我们需要完成两个部分结论的证明：
>
> 我们设f具有$f(x)=(f_1(x),\dots,f_m(x))$的形式。
>
>> 结论1：设$a,b\in\Omega$，若对任意的$t\in[0,1]$都有$a+t(b-a)\in\Omega$，则有$f(a)=f(b)$。
>>
>> 证明：
>>
>> 对任意的$1\leq i\leq m$，我们定义函数$g:[0,1]\to\mathbb{R}$有：
>>
>> $$g(t):=f_i(a+t(b-a))$$
>>
>> 注意到对任意的$t_0\in(0,1)$有：
>>
>> $$g'(t_0)=\lim_{t\to0}\frac{f(a+(t+t_0)(b-a))-f(a+t(b-a))}{t}$$
>>
>> 由方向导数定义此极限的左极限为$D_{b-a}f(a+t_0(b-a))$，右极限为$-D_{a-b}f(a+t_0(b-a))$。而由于$f'(a+t_0(b-a))=0$，因此根据引理17.3.5我们有$D_{b-a}f(a+t_0(b-a))=-D_{a-b}f(a+t_0(b-a))=0$，也即$g$在$t_0$处可微且导数为0。
>>
>> 此时根据平均值定理我们有：
>>
>> $$0=\frac{g(0)-g(1)}{0-1}\iff f_i(a)=f_i(b)$$
>>
>> 由于i是任意的因此这也表明了有$f(a)=f(b)$，结论得证。
>
> 对任意的$a,b\in\Omega$，若对任意的$t\in[0,1]$都有$a+t(b-a)\in\Omega$，则我们称a,b之间存在直线道路；若存在Ω中的有限点列$(p_i)_{i=0}^{l}$满足$p_0=a$，$p_l=b$且对任意的$1\leq i\leq l$都有p_{i-1},p_i之间存在直线道路，则我们称两个点$a,b\in\Omega$之间存在有限折线道路，并记为$a\sim b$。
>
> 由于我们已经证明了结论1，因此不难发现通过归纳法我们可以得证对任意的$a,b\in\Omega$若有$a\sim b$则有$f(a)=f(b)$。由于Ω是非空的因此我们设$x_0\in\Omega$，然后定义集合Γ为：
>
> $$\Gamma:=\{p\in\Omega:p\sim x_0\}$$
>
> 首先Γ必然是非空的（$x_0\in\Gamma$），其次对任意的$p\in\Gamma$由于$p\sim x_0$因此有$f(x_0)=f(p)$，最后我们显然可以证明$\sim$是一个等价关系（所以它满足自反性，对称性，传递性公理）。为了证明f是一个常值函数，我们希望能够证明$\Gamma=\Omega$，于是尝试证明结论2。
>
>> 结论2：Γ是是Ω的一个又开又闭的子集。
>>
>> 证明：
>>
>> 先证明Γ是一个开集，于是我们需要证明对任意的$p\in\Gamma$存在$r>0$使得$B(p,r)\subseteq\Gamma$。
>>
>> 首先由于Ω是一个开集，因此存在一个$r>0$使得$B(p,r)\subseteq\Omega$。此时对任意的$q\in B(p,r)$，我们可以注意到显然p,q之间存在直线道路，进而有$p\sim q$，从而根据$\sim$的传递性我们有$q\sim x_0\implies q\in\Gamma$，从而有$B(p,r)\subseteq\Gamma$。
>>
>> 然后证明Γ是一个闭集，于是我们需要证明对任意的$(p_i)_{i=0}^{\infty}$是Γ中的收敛点列（设其收敛于p），都有$p\in\Gamma$。

> > 一方面，由于Ω是一个开集，因此存在一个$r>0$使得$B(p,r)\subseteq\Omega$；另一方面，因为点列$(p_i)_{i=0}^{\infty}$收敛因此存在$N\geq 0$使得对任意的$i\geq N$都有$\|p_i-p\|<r$。于是我们有$p_N\in B(p,r)$，类似于开集证明中的讨论我们知道有$p_N\sim p$，同理利用传递性公理我们有$p\sim x_0\implies p\in\Gamma$。
> >
> > 综上，于是结论得证。
>
> 由于Γ是非空的又开又闭集，若$\Gamma\neq\Omega$则根据不连通集的定义可以得到Ω是不连通的，这和Ω是连通开集的前提矛盾，因此只能有$\Gamma=\Omega$。再结合上面的讨论于是我们证明了"对任意的$p\in\Omega$有$f(x_0)=f(p)$"，也即f是一个常值函数，题目结论得证。
>
> ---
>
> 题外话：这里其实也证明了$\mathbb{R}^n$的任意一个连通开子集Ω都是道路连通的（两点之前存在有限折线道路那么我们很容易分段列出它们之间的道路函数）

本节相关跳转

[实分析 10.2 局部最大值、局部最小值以及导数]

[实分析 11.9 微积分的两个基本定理]

[实分析 17.2 多元微积分中的导数]