最菜的群友的笔记 – Calcul Différentiel

最菜的群友

2025 年 10 月 6 日

图 1: 作者

0 引言

0.1 关于本拙作

本作品自发表以来，陆续收到了许多同志的肯定与好评. 诚然大家的赞同是具有不可替代的价值，但数学研究的真正进步并不依赖于赞扬，而更需要来自不同角度的质疑与批评. 唯有通过不断的反思与讨论，作品中的不足之处才能被暴露，从而被修改完善. 我也必须坦率地强调，~~ChatGPT和~~我在数学方面的水平十分有限，难免在表述或推理的翻译、转述中存在不严谨甚至错误之处.

在此对这段时间来指出本作品错误的同志们一并感谢.

0.2 注意事项

1. **本人衷心希望您看到这里的时候并非考试前夕，否则我建议您早点睡觉，以便明天以丰沛的精神面对考试；**
2. 本作品不可代替教材；
3. 如有问题请立刻向作者指出，本作品允许一切形式的谩骂、攻击、批评；
4. 本作品可随便转发（反正没几个人看）.

0.3 关于Calcul Différentiel

图 2: 中国航天科技事业的先驱和杰出代表、享誉世界的杰出科学家钱学森同志

法国教育系统的“多元微积分”同国内高等数学的多元函数有较大区别. 后者更偏重计算与应用，强调偏导、全微分、重积分等具体方法；而前者更强调抽象与严谨，把可微性直接定义为“可被线性映射近似”，偏导只是这种结构的坐标体现，并进一步引入Jacobian、Hessian、隐函数定理等更理论化的内容.

注. *按照钱老的人生经历，他应该既接触（其实是指导）过国内的工科教学，也接触过起源于欧洲的偏理论的数学教育. 他说的人再笨也能学会的微积分具体指什么就不得而知了.*

0.4 以及...

Au cours de la rédaction de ces notes, M. Mathieu Charlot a fourni des supports et des documents pédagogiques ; je le remercie pour sa contribution ainsi que pour son travail accompli pendant deux années à Centrale.

图 3: Merci, monsieur.

1 可微性

对于生活中常见的一元实函数，“微分”和“导数”这两个概念基本上无需区分，~~（尤其学物理的根本不管）~~. 但在更一般（更复杂）的情况下，这两个概念并不是完全一样的.

记号规定：整个章节中，E 和 F 都表示有限维实向量空间，U 是 E 的一个开集，$f : U \to F$ 是一个映射. 记 $\|\cdot\|_E$ 和 $\|\cdot\|_F$ 分别为 E 和 F 上的范数.

1.1 微分

在很久很久以前的分析2中，我们已经学过了 $E = \mathbb{R}$ 的情形：f 在 a 处可导，如果极限

$$v = \lim_{\substack{x \to a \\ x \in U \setminus \{a\}}} \frac{f(x) - f(a)}{x - a}$$

存在，此时称其为 f 在 a 处的**导数**，记作 $f'(a)$. 函数 f 在 U 上称为**可导的**，如果它在 U 的每一点都可导；此时，映射

$$f' : U \to F$$

将每个 $x \in U$ 关联到 f 在 x 处的导向量，称为 f 的**导函数**. 利用 Landau 记号，可以写成：

$$f(x) - f(a) = (x - a)v + o(x - a), \quad x \to a.$$

换个方法表述 f 在 a 处可导的含义：存在一个向量 $v \in E$，使得若用向量 $(x - a)v$（正比于 $x - a$）去替代函数增量 $f(x) - f(a)$，则当 $x \to a$ 时，误差相对于 $x - a$ 可忽略不计. 在物理学的语言中，如果 f 描述了 F 中一个点的运动（而 x 表示时间），那么 v 就是瞬时速度.

换句话说，当x距离a足够近的时候，只需要一个线性相关于$x - a$的量就可以描述函数的增量. 这个线性因子就是微分.

这一章的目的，就是将可导性的概念推广到多变量函数，即任意有限维赋范向量空间 E.

定义. 称函数 f 在 U 中一点 a 处是 **可微的**，如果 f 在 a 处的增量可以由一个线性部分表示. 换句话说，存在一个线性映射 $\varphi \in \mathcal{L}(E; F)$，使得

$$f(x) - f(a) \underset{\substack{x \to a \\ x \in U}}{=} \varphi(x - a) + o(x - a).$$

注. 上述定义等价于

$$\lim_{\substack{x \to a \\ x \in U \setminus \{a\}}} \frac{\|f(x) - f(a) - \varphi(x - a)\|_F}{\|x - a\|_E} = 0,$$

也可以写成常见的恶心人的形式，令 $x = a + h$：

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall h \in E \setminus \{0\}, \quad \|h\|_E < \delta \implies \big(a + h \in U \text{ 且 } \|f(a+h) - f(a) - \varphi(h)\|_F < \varepsilon \|h\|_E\big).$$

“注”里面的形式，在证明的时候是比定义更常用的. 一般是先找到φ再证明它是微分.

此外，由于在有限维实向量空间中所有范数都是等价的，所以 f **在** a **的可微性与所选的范数** $\|\cdot\|_E$、$\|\cdot\|_F$ **无关**. 做题的时候要尽量挑对自己有利的范数.

引理. 设 $\varphi \in \mathcal{L}(E;F)$（说明它是线性的），若满足 $\varphi(h) = o(h)$ 当 $h \to 0$ 时，则 φ 是零线性映射.

证明. 任取 $h \in E \setminus \{0\}$，我们要证明$\varphi(h) = 0$.

对于线性映射，应该能够想到把h放缩t倍（t是正实数）. 注意到在h的方向上，

$$\frac{\|\varphi(h)\|_F}{\|h\|_E} = \frac{\|\varphi(th)\|_F}{\|th\|_E}$$

当$t \to 0$的时候，右边会$\to 0$，但左边的分子是h的范数，不一定是0. 所以我们似乎证完了. 作如下变换.

$$\|\varphi(h)\|_F = \frac{\|\varphi(th)\|_F}{\|th\|_E}\|h\|_E,$$

可知 $\varphi(h) = 0$. □

这个引理可以说明，如果函数f可微，那么得出来的“微分”，也就是线性函数φ是唯一的.

假设有两个线性映射 $\varphi_1, \varphi_2 \in \mathcal{L}(E;F)$ 都可微性的定义，即对所有 $x \to a$ 有

$$f(x) - f(a) = \varphi_1(x-a) + o(x-a) = \varphi_2(x-a) + o(x-a).$$

两边相减得

$$(\varphi_1 - \varphi_2)(x-a) = o(x-a).$$

记 $\psi = \varphi_1 - \varphi_2 \in \mathcal{L}(E;F)$（这玩意叫psi），于是有 $\psi(h) = o(h), h \to 0$. 由引理可知，ψ 必为零算子. 因此 $\psi = 0$，即 $\varphi_1 = \varphi_2$.

当我们在一本教材中看唯一性，就说明有定义要来了.

定义. 如果 f 在 a 处可微，满足可微性定义中的唯一线性映射 φ 称为 **f 在 a 处的微分**，记作 $Df(a)$. 对于 $h \in E$，有时写作 $Df(a)h$ 而不是 $Df(a)(h)$.

注. 当 $E = \mathbb{R}$ 时情况就很简单了，因为任意线性映射 $\varphi \in \mathcal{L}(\mathbb{R};F)$ 都可以写成

$$\varphi(h) = hv, \quad \forall h \in \mathbb{R},$$

其中 $v = \varphi(1)$. 映射 $\varphi \mapsto \varphi(1)$ 给出同构 $\mathcal{L}(\mathbb{R};F) \cong F$，从而微分 $Df(a)$ 与导向量 $f'(a)$ 视为同一对象.

因此，对于一元函数，它的微分跟导数基本上就是一回事.

定义. 我们称映射 $f: U \to F$ 在 U 上是 **可微的**，如果它在 U 的每一点都可微. 在这种情况下，f 的微分是映射

$$Df: U \to \mathcal{L}(E; F),$$

它把每个 $x \in U$ 映射到线性算子 $Df(x)$.

这里一定要注意微分和在某一点上的微分的区别.

谬论. Df实际上就等于可微性定义中的线性映射ϕ.

很显然这是在胡说八道. 我们举个例子说明一下这一大堆名词之间的关系.

我们考虑一个典型的函数$f: x \mapsto \sin x$. 可以证明在$a \in \mathbb{R}$上有

$$f(a+h) = f(a) + h\cos a + o(h)$$

那么按照定义，$Df(a) = \phi_a : h \mapsto h\cos a$. 对于给定的$a$，$\phi_a$是线性的.

此外，很显然f在整个实数轴上面都可微，如果a并非定点，就可以定义在整个实数轴上的微分映射

$$Df: \mathbb{R} \to \mathcal{L}(\mathbb{R}), \quad a \mapsto \phi_a,$$

也就是

$$Df: \mathbb{R} \to \mathcal{L}(\mathbb{R}), \quad a \mapsto (h \mapsto h\cos a).$$

总而言之，某一点的微分和函数的微分映射的关系，大体上相当于导数和导函数的关系. 前者是一个特定的值（对微分来说是特定的线性映射），和选取的点有关；后者是个映射（函数），自变量是选取的点.

命题. 如果映射 f 在 a 处可微，那么它在 a 处一定连续.

证明. 假设 f 在 $a \in U$ 处可微；由于有限维实向量空间 E 上的线性映射 $Df(a)$ 是连续的，所以可推出

$$\lim_{\substack{x \to a \\ x \neq a}} f(x) = f(a).$$

□

现在考虑一个常见情况. 设 E 为两个有限维实向量空间的直积： $E = E_1 \times E_2$. 令

$$f: E_1 \times E_2 \to F$$

是一个双线性映射. 给 E_1 和 E_2 赋范，记作 $\|\cdot\|_{E_1}$ 与 $\|\cdot\|_{E_2}$ 并在 $E_1 \times E_2$ 上取乘积范数$\|\cdot\|_E$. 由于 E 是有限维的，f 连续，因此存在常数 $C \geq 0$ 使得对任意 $h = (h_1, h_2) \in E$ 有

$$\|f(h_1, h_2)\|_F \leq C\|h_1\|_{E_1}\|h_2\|_{E_2} \leq C\|h\|_E^2.$$

图 4: 可导一定连续，可微也一定连续

于是 $f(h) = o(h)$ 当 $h \to 0$. 由下面的展开可见 f 在任意点可微：

$$\begin{aligned} f(a+h) &= f(a_1+h_1, a_2+h_2) \\ &= f(a_1,a_2) + f(h_1,a_2) + f(a_1,h_2) + f(h_1,h_2) \\ &= f(a) + f(h_1,a_2) + f(a_1,h_2) + o(h), \end{aligned}$$

其中 $a = (a_1, a_2)$. 因此 f 在 a 处的微分为线性映射

$$Df(a) : E_1 \times E_2 \to F, \qquad (h_1,h_2) \mapsto f(h_1,a_2) + f(a_1,h_2).$$

注. 乘积范数的定义比较像有限维向量空间的范数，即可以取两个（或者多个）范数的最大值，可以取它们的和，也可以都平方求和再开方. 但是不管怎么取，都会有上面的不等式关系.

注. 这个形式和函数乘积的导数的形式很像，后面就知道了.

1.2 偏导数

定义. 设 $a \in U$ 且 $h \in E \setminus \{0\}$. 称 f 沿方向 h 在 a 处可导（*dérivable*）如果函数

$$\gamma : t \mapsto f(a+th)$$

在 0 处可导，此时称沿 h 的导向量（或导数）为

$$\gamma'(0) = \frac{d}{dt}\bigg|_{t=0} f(a+th)$$

记作 $D_h f(a)$.

注. 由于函数 f 的定义域 U 是 E 的一个开集，存在 $\delta > 0$ 使得以 a 为中心、半径 δ 的开球包含在 U 中；因此函数 γ 在区间 $]-\delta/\|h\|_E, \delta/\|h\|_E[$ 上有定义.

大家不要被有限维实向量空间吓到，不管目标空间多抽象，γ的自变量都是实数，因此γ'是可以被定义的：

$$\gamma'(0) = \frac{d}{dt}\bigg|_{t=0} f(a+th) = \lim_{\delta t\to 0} \frac{\gamma(0+\delta t)-\gamma(0)}{\delta t}$$

h的含义就是导数的方向. 对于一元函数，x轴是一条直线，只有一个方向进行求导. 但对于有限维实向量空间的映射，自变量是一个向量，可以有无数个方向. 不同方向上的导数是不一样的. 例如，在山坡上顺着坡滑下去，路径的坡度（导数）就很大；但同一位置下如果在沿着山腰走，路径的坡度就为0. 一元函数的导数其实就是h取1的情形.

t是一个标量参数，用来缩放方向向量h. 这样就把多变量函数的增量问题变成一个一维的参数问题.

谬论. *那为什么不直接对h求导呢？*

$$\gamma' = \frac{df}{dh}$$

看到你把一个向量加上微分符号并放在分母，一位戴眼镜有胡子的年轻物理老师把刚喝下的珍珠奶茶一口喷了出来.

命题. *假设 f 在 a 处可微. 则对任意 $h \in E \setminus \{0\}$, f 沿 h 在 a 处可导且*

$$Df(a)h = D_h f(a)$$

证明. 固定一个任意方向 $h \in E \setminus \{0\}$. 按照微分的定义，对于$t \to 0$，

$$f(a+th) = f(a) + Df(a)(th) + o(th) = f(a) + t\,Df(a)h + o(t)$$

这里注意Df是线性的. 这正表明映射 $t \mapsto f(a+th)$ 在 0 处可导，其导数为 $Df(a)h$. □

注. *显然，如果 f 沿某非零向量 $h \in E$ 在 a 处可导，则它沿任一与 h 共线的非零向量也可导，并且对任意 $t \in \mathbb{R} \setminus \{0\}$ 有*

$$D_{th}f(a) = t\,D_h f(a).$$

若 f 为实值函数，即值域为$\mathbb{R}$，则可以对 $D_h f(a)$ 归一化为$\|h\|_E^{-1} D_h f(a)$. 这个数仅取决于由 h 的方向. 并称为 f 在 a 处沿 h 的斜率（pente suivant h de f en a）. 该实数等于对同方向同向量的单位向量 e 的 $D_e f(a)$（看到没有，这就是刚才说的h取1），这对应于在 $\mathbb{R}^2$ 中函数 $t \mapsto f(a+te)$ 在点 $(0, f(a))$ 的图形的斜率.

刚才说到，可微一定可导，但**可导不一定可微**.

例如，令

$$A = \{x \in \mathbb{R}^2 \mid x_1 > 0 \text{ 且 } 0 < x_2 < x_1^2\},$$

并定义函数

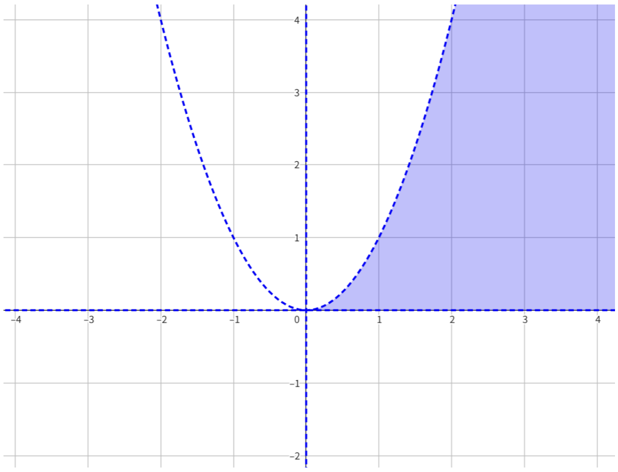


图 5: 集合A. 横轴为x_1，纵轴为x_2，下同

$$f(x)=\begin{cases}1, & x\in A,\\ 0, & x\notin A.\end{cases}$$

对任意 $h=(h_1,h_2)\in\mathbb{R}^2\setminus\{(0,0)\}$，考察 $\varphi_h:\mathbb{R}\to\mathbb{R}$ 所给的函数 $\varphi_h(t)=f(th)$. 在 0 的邻域内 φ_h 通常为零函数. 更精确地说：

- 若 $h_1>0$ 且 $h_2>0$, 则 φ_h 在区间 $]-\infty,\ h_2/h_1^2[$ 这一块恒为 0.
- 否则$\varphi_h\equiv 0$

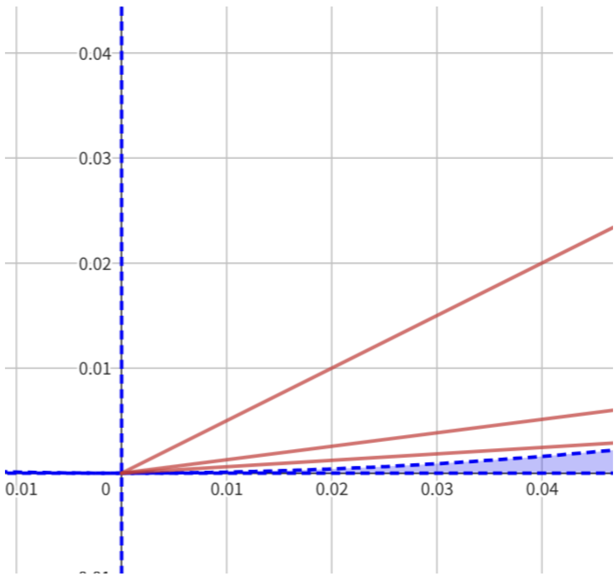


图 6: 不管方向怎么取，0附近的一段都不属于A

因此对任意方向 h，f 在 0 点沿 h 可导且导数为 0，即 $D_hf(0)=0$. 然而

$$\lim_{t\to 0^+} f\left(t,\tfrac{1}{2}t^2\right)=1\neq 0=f(0),$$

所以 f 在 0 处不连续，因而不可能在0可微.

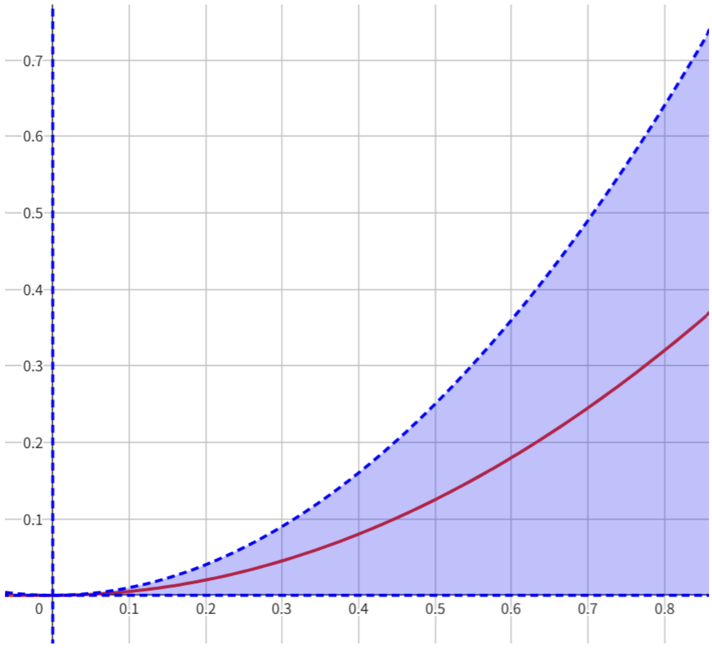


图 7: 但我要是这么穿过去，它还连续吗？

再考虑这样一个函数$f:\mathbb{R}^2\to\mathbb{R}$

$$f(x)=\begin{cases}\dfrac{x_1^3+x_2^3}{x_1^2+x_2^2}, & x\neq 0,\\ 0, & x=0.\end{cases}$$

该函数是齐次的，即对任意$\lambda\in\mathbb{R}$，$f(\lambda x)=\lambda f(x)$. 所以对任意 $h\in\mathbb{R}^2\setminus\{0\}$ 有

$$\gamma(t)=f(0+th)=tf(h),$$

因此 γ 为线性函数，求导之后得到$f(h)$故 f 在 0 点沿任意方向 h 可导，且 $D_hf(0)=f(h)$.

如果f可微，那么微分应当等于偏导数，也就是说 $Df(0)h=D_hf(0)=f(h)$从而$Df(0)=f$. 但f并不线性，所以假设是不成立的. f不可微. 即使f事实上是连续的.

综上所述，不可微性可以通过不连续性或导数的非线性举出反例.

定义. *若 f 在 a 处可微且微分为零映射，则称点 a 为 f 的临界点.*

命题. *设函数 f 为实值函数（值域为实数域即$F=\mathbb{R}$），并且在点 $a\in U$ 可微. 若 f 在 a 处有局部极值，则 a 必为 f 的临界点.*

证明. 只讨论最大值的情况，最小值的情况把 f 换为 $-f$就行. 把定义域 U 缩小到 a 的某一小邻域，从而 f 在 a 取得最大值. 对于一维情形 $E=\mathbb{R}$，我们去年已经学过了：

$$t<a\implies\frac{f(t)-f(a)}{t-a}\geq 0,\qquad t>a\implies\frac{f(t)-f(a)}{t-a}\leq 0,$$

因此 $f'(a)=0$. 对一般情形，任取 $h\in E\setminus\{0\}$，将其转化为我们喜欢的一维情况：考虑一维函数 $\gamma:t\mapsto f(a+th)$. 函数可微一定可导，因此 γ 在 0 处可导，并且在 0 处取得最大值（因为f在a取最大值$f(a)$就是$\gamma(0)$），所以 $\gamma'(0)=0$，且$Df(a)h=D_hf(a)=\gamma'(0)=0$. 因此由于$h$是任取的所以$Df(a)=0$. □

命题. 设 $F=\mathbb{R}^p$，将f写成各分量形式：$f=(f_1,\dots,f_p)$. 映射 f 在点 a 可微当且仅当其各分量函数 $f_1,\dots,f_p$ 在 a 可微. 在此情况下，对任意 $h\in E$ 有

$$Df(a)h=\big(Df_1(a)h,\ \dots,\ Df_p(a)h\big).$$

证明. 设 $\varphi=(\varphi_1,\dots,\varphi_p)$ 为 $E\to\mathbb{R}^p$ 的线性映射. 根据可微的极限定义，

$$\lim_{h\to 0,h\neq 0}\frac{f(a+h)-f(a)-\varphi(h)}{\|h\|_E}=0$$

当且仅当对每个 $j\in\{1,\dots,p\}$ 都有

$$\lim_{h\to 0,h\neq 0}\frac{f_j(a+h)-f_j(a)-\varphi_j(h)}{\|h\|_E}=0.$$

由此等价性可推出命题结论. □

其实就是没什么可证的...但这一性质比较重要，它定义了我们平时常见的偏导数.

定义. 设 $E=\mathbb{R}^n$. 记 $(e_1,\dots,e_n)$ 为标准基，坐标为 $x_1,\dots,x_n$. 对 $i\in\{1,\dots,n\}$，若 f 沿基向量 e_i 在点 $a=(a_1,\dots,a_n)$ 可导，则称

$$\frac{\partial f}{\partial x_i}(a)=\frac{d}{dt}\bigg|_{t=0}f(a+te_i)=\frac{d}{dt}\bigg|_{t=0}f(a_1,\dots,a_{i-1},a_i+t,a_{i+1},\dots,a_n)$$

为 f 在 a 处的第 i 个偏导数.

这表明，偏导数的意义就是在固定其它坐标不变的情况下，对某一坐标方向上，函数的瞬时变化率. 众所周知，正交标准基（甚至不用正交标准）的各个元素是相互独立的，对其中一个方向求偏导和其他方向没有关系.

某位喜欢喝珍珠奶茶的物理老师在讲热力学的时候经常使用 $\frac{\partial U}{\partial P}\big|_T$ 这样的符号，它的意思就是：假设压强和温度相互独立的基础上，在“压强-温度空间”中求内能对压强的偏导.

命题. 设 $E=\mathbb{R}^n$ 且 $F=\mathbb{R}^p$. 若映射 $f=(f_1,\dots,f_p)$ 在点 a 可微，则线性映射 $Df(a)$ 在标准基下的矩阵为

$$\begin{pmatrix}\partial_1 f_1(a) & \partial_2 f_1(a) & \cdots & \partial_n f_1(a)\\ \partial_1 f_2(a) & \partial_2 f_2(a) & \cdots & \partial_n f_2(a)\\ \vdots & \vdots & \ddots & \vdots\\ \partial_1 f_p(a) & \partial_2 f_p(a) & \cdots & \partial_n f_p(a)\end{pmatrix}.$$

定义. 在命题的假设下，这一大块矩阵称为 f 在点 a 的雅可比矩阵（*matrice jacobienne*）或雅可比矩阵.

证明. 记$b_E=(e_1,\dots,e_n)$ 为 $\mathbb{R}^n$ 的标准基. 设想线性映射 $Df(a)$ 在两个标准基下的矩阵. 根据线性映射的矩阵的定义，

$$Mat_{b_F,b_E}(Df(a))=Mat_{b_F}(Df(a)b_E)$$

也就是说，矩阵的第i列应该是$(Df(a)e_i)$在b_F下的坐标. 这一列的第j行上的元素应该是$Df(a)e_i$的第j个坐标，即$Df_j(a)e_i$. 它是a点f_j的微分在e_i上的作用，也就等于f_j在a点处在e_i方向的偏导数，即：$(f_j(a)e_i=D_{e_i}f_j(a)=\partial_i f_j(a)$. 换句话说，矩阵的第$i$列第$j$行元素为$\partial_i f_j(a)$. 证毕.

□

注. *微分符号写D还是写d其实无所谓.*

通过证明不难看出，雅可比矩阵的第 i 列是向量 $\partial_i f(a)$ 在 $\mathbb{R}^p$ 的标准基下的矩阵表示. 对偶地，它的第 j 行是线性泛函 $df_j(a)$ 在 $\mathbb{R}^n$ 的标准基下的矩阵表示.

考虑 $p=1$ 即 $F=\mathbb{R}$ 的情况，记 $(x_1,\dots,x_n)$ 为 $\mathbb{R}^n$ 的坐标. 如果 f 在点 a 可微，则对于 $h=(h_1,\dots,h_n)\in\mathbb{R}^n$，线性泛函 $df(a)$ 满足:

$$df(a)h=(\partial_1 f(a)\ \dots\ \partial_n f(a))\begin{pmatrix}h_1\\ \vdots\\ h_n\end{pmatrix}=\sum_{i=1}^n \partial_i f(a)\,h_i,$$

对于任意向量，x_i的作用是提取它的第i个坐标. 在点a上取任意方向向量h有

$$x_i(a+h)-x_i(a)=x_i(h)=h_i$$

$$x_i(a+h)=x_i(a)+h_i$$

因此

$$dx_i(a)h=h_i=x_i(h)$$

这个式子表明，x_i在任一点上的微分等于它自己，所以我们有

$$df(a)h=\sum_{i=1}^n \partial_i f(a)\,h_i=\sum_{i=1}^n \partial_i f(a)\,dx_i(a)h$$

省略h和a：

$$df=\sum_{i=1}^n \frac{\partial f}{\partial x_i}\,dx_i.$$

除此之外，雅可比矩阵在多元函数的计算用处很大，在力学、工程学也有很多应用，这里就不展开了（反正我也不会）.

命题. 设 $E=\mathbb{R}^n$. 假设f的偏导数 $\partial_1 f,\dots,\partial_n f$ 在开集 U 上有定义且在 U 的某点 a 处连续. 则 f 在点 a 可微.

证明. 首先，只需考虑$a=0$的情况，否则可以将 f 替换为 $x\mapsto f(a+x)$ 并把 U 替换为 $\{x-a;\ x\in U\}$. 给定 F 的一组基后，只需证明 f 在该基下的各分量在 a 可微; 因此不妨假设 $F=\mathbb{R}$（相当于只讨论一个分量）. 若 f 在 0 可微，则其微分必为从 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的线性映射

$$x\mapsto \partial_1 f(0)x_1+\cdots+\partial_n f(0)x_n.$$

设余项

$$r(x)=f(x)-f(0)-\sum_{i=1}^{n}\partial_i f(0)\cdot x_i.$$

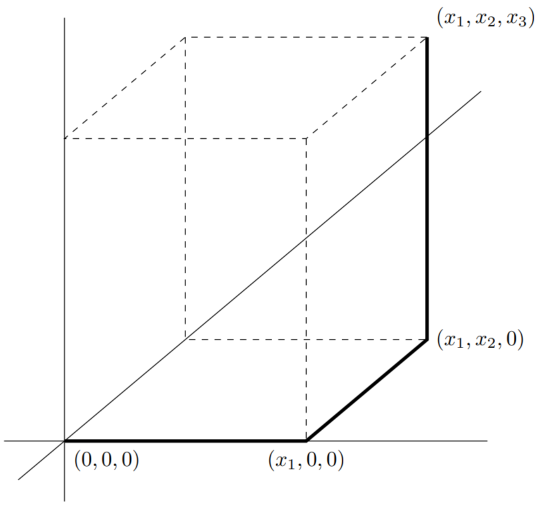
要证明 f 在 0 可微等价于验证 $r(x)=o(\|x\|)$ 当 $x\to 0$ 时成立.

取任意 $\varepsilon>0$. 我们在 $\mathbb{R}^n$ 上使用范数 $\|\cdot\|_1$（注意看，选对自己有利的范数）. 由偏导的连续性，存在 $\delta>0$ 使得对每个 $i\in\{1,\dots,n\}$ 都有

$$\|x\|_1\le\delta\Longrightarrow|\partial_i f(x)-\partial_i f(0)|\le\varepsilon.$$

考虑任意满足 $\|x\|_1\le\delta$ 的点 $x=(x_1,\dots,x_n)$. 将 $f(x)-f(0)$ 写成逐次分量增加的差分和:

$$f(x)-f(0)=\sum_{i=1}^{n}\big(f(x_1,\dots,x_i,0,\dots,0)-f(x_1,\dots,x_{i-1},0,\dots,0)\big).$$



$$\begin{aligned}f(x)-f(0)=\ &f(x_1,0,0)-f(0,0,0)\\&+f(x_1,x_2,0)-f(x_1,0,0)\\&\quad+f(x_1,x_2,x_3)-f(x_1,x_2,0)\end{aligned}$$

图 8: 沿着各个坐标轴的方向，一次一次加上去

若 $x_i\neq 0$，对一元函数 $t\mapsto f(x_1,\dots,x_{i-1},t,0,\dots,0)$ ，将其看成一元函数，则根据拉格朗日中值定理，

$$\begin{aligned} & f(x_1,\ldots,x_i,0,\ldots,0)-f(x_1,\ldots,x_{i-1},0,\ldots,0) \\ & = \frac{d}{dt}f(x_1,\ldots,x_{i-1},t,0,\ldots,0)\bigg|_{t=c_i}\cdot x_i \\ & = \partial_i f(x_1,\ldots,x_{i-1},c_i,0,\ldots,0)\cdot x_i. \end{aligned}$$

其中 c_i 是区间 $(0,x_i)$ 中的某点. 若 $x_i=0$，则取 $c_i=0$. 因此余项可以写为

$$r(x)=\sum_{i=1}^{n}\big(\partial_i f(x_1,\ldots,x_{i-1},c_i,0,\ldots,0)-\partial_i f(0)\big)\cdot x_i.$$

由此前对 $\partial_i f$ 的连续性讨论，对每个 i 在 $\|x\|_1\le\delta$ 时都有

$$\big|\partial_i f(x_1,\ldots,x_{i-1},c_i,0,\ldots,0)-\partial_i f(0)\big|\le\varepsilon.$$

因此

$$|r(x)|\le\sum_{i=1}^{n}\big|\partial_i f(x_1,\ldots,x_{i-1},c_i,0,\ldots,0)-\partial_i f(0)\big|\cdot|x_i|\le\varepsilon\sum_{i=1}^{n}|x_i|=\varepsilon\|x\|_1.$$

由此 $r(x)=o(\|x\|_1)$ 当 $x\to 0$. 因此 f 在 0 可微，在任一点 a 处也可微. □

将$\mathcal{C}^1$类的定义推广到线性空间：

定义. *若 f 可微，且映射*

$$Df:U\longrightarrow\mathcal{L}(E;F)$$

是连续的，则称 f 为 $\mathcal{C}^1$ 类函数.

推论. *设 $E=\mathbb{R}^n$. 若 f 的偏导数 $\partial_1 f,\ldots,\partial_n f$ 在 U 上存在并且连续，则映射 f 属于 C^1 类.*

1.2′ 回顾

开始下一节之前，我们来换个角度思考一下以上这几个知识点.

考虑函数$f(x)=x^2$. 很显然，它并不是一个线性映射. 如图，这样的非线性映射将尺子上均匀的刻度映射为不均匀的.

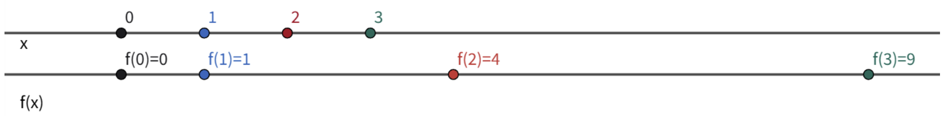


图 9: 非线性映射$y=x^2$

我们放大一下，观察2周围的刻度，并将2和$f(2)=4$对齐：

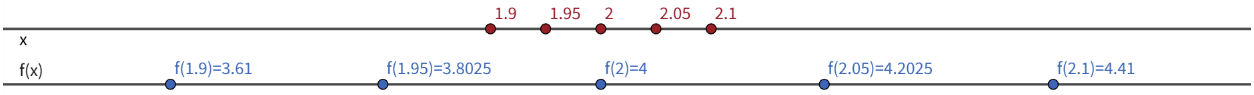


图 10: f对2附近的刻度的作用

由图10可知，对于2周围的均匀刻度，f 将它们映射为看起来比较均匀（误差很小）的均匀刻度，并且间距扩大为原来的4倍. 因此，如果 x 在2周围，则有：

$$f(x) - f(2) \approx 4(x-2)$$

或者

$$f(x) - f(2) = 4(x-2) + o(x-2)$$

如果定义线性映射 $\phi : x \mapsto 4x$，则

$$f(x) - f(2) = \phi(x-2) + o(x-2)$$

不难看出这个 ϕ 就是微分定义里面的那个线性映射，也就是 f 在2处的微分. 这个4源于 $y = x^2$ 在2处的斜率，也就是一元函数的导数.

一维的情况是很好理解的，那么二维如何考虑？考察函数

$$f(x,y) = (f_1(x,y), f_2(x,y)) = (x^2 - y^2, 3xy)$$

很显然它不是一个线性映射. 我们知道，线性映射会将二维空间的网格映射为另一种网格. 但这一非线性映射只会得到扭曲的不均匀网格. 但如果放大，考虑 $A(1,1)$ 周围的点，再考察这些点的像.

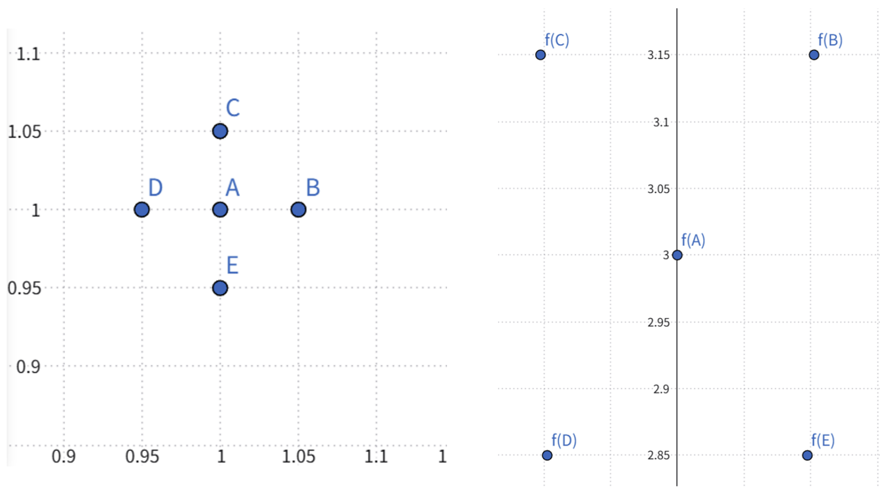


图 11: A附近的若干网格点，映射后似乎仍然很规律

注意到这些像仍然很规律，因此在小尺度下观察，f似乎可以看作是线性的. 注意到 B 是 A 向右一个单位，$f(B)$ 大约是 $f(A)$ 向右两个单位再向上三个单位； C 是 A 向上

一个单位，$f(C)$ 大约是 $f(A)$ 向左两个单位再向上三个单位. 运用线性代数的直觉，我们可以定义线性映射：

$$Mat(u) = \begin{pmatrix} 2 & -2 \\ 3 & 3 \end{pmatrix}$$

满足：

$$f(x) - f(A) \approx u(x - A)$$

事实上，

$$f(x) - f(A) = u(x - A) + o(x - A)$$

这个近似矩阵是怎么找到的？观察图像，f 将横坐标间距为1的点 D, A, B 变成了横坐标间距为2、纵坐标间距为3的点. 因此，$f = (f_1, f_2)$ 中，f_1 将横坐标的刻度扩大为2倍并映射到横坐标 f_2 将横坐标的刻度扩大为3倍并映射到纵坐标：

$$\frac{\partial f_1}{\partial x} = 2x, \quad \frac{\partial f_2}{\partial x} = 3y$$

代入 A 的坐标就得到2、3.

继续观察图像，对于纵坐标间距为1的点 E, A, C，f_1 将纵坐标的刻度扩大为2倍并映射到横坐标，但E, A, C 是纵坐标从小到大，$f(E), f(A), f(C)$ 横坐标从大到小，方向反过来了，所以这个系数应当为-2； f_2 将纵坐标的刻度扩大为3倍并映射到纵坐标：

$$\frac{\partial f_1}{\partial y} = -2y, \quad \frac{\partial f_2}{\partial y} = 3x$$

代入 A 的坐标就得到-2、3.

因此整个矩阵可以写成：

$$\mathrm{Mat}(u) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(x, y) & \frac{\partial f_1}{\partial y}(x, y) \\ \frac{\partial f_2}{\partial x}(x, y) & \frac{\partial f_2}{\partial y}(x, y) \end{pmatrix}.$$

这就是雅可比矩阵的意义.

1.3 梯度

若 E 是一个欧几里得空间，内积记作 $(\cdot \mid \cdot)$，则任一线性泛函（注意这玩意叫forme linéaire） φ 在 E 上可以由唯一的向量 $v \in E$ 表示，满足

$$\forall h \in E \qquad \varphi(h) = (v \mid h)$$

这个定理在线性代数里面已经看过了. 其实也没有那么复杂，在d维的有限维实线性空间，设$x = (x_1, \ldots, x_d)$线性泛函可以写作如下形式：

$$\varphi(x) = \varphi_1 x_1 + \cdots + \varphi_d x_d$$

因此规定向量$v = (\varphi_1, \ldots, \varphi_d)$即可.

定义. 设 E 为一个欧几里得空间，且 f 为值域为实数域的映射. 若映射 f 在 U 中某点 a 可微，则称 f 在 a 处的梯度（*gradient*）（记作 $\nabla f(a)$ 或 $\operatorname{grad} f(a)$）为代表 f 在 a 处微分的向量，满足

$$\forall h \in E \qquad df(a)h = \big(\nabla f(a) \mid h\big).$$

若 f 在整个 U 上可微，则梯度是一个向量场：

$$\nabla f : U \longrightarrow E, \qquad x \mapsto \nabla f(x).$$

很容易验证，若 $(e_1, \ldots, e_n)$ 为 E 的一组正交单位基，则

$$\nabla f(a) = \sum_{i=1}^{n} \partial_i f(a)\, e_i.$$

这也是比较常用的形式.

例如在 $\mathbb{R}^n$ 上取通常的内积，记作 $(\,\cdot \mid \cdot\,)$. 则

$$f : \mathbb{R}^n \setminus \{0\} \longrightarrow \mathbb{R}, \qquad x \mapsto \|x\|_2$$

在每个非零点 a 处可微. 取固定 $a \in \mathbb{R}^n \setminus \{0\}$.利用展开式

$$\sqrt{1+t} = 1 + \tfrac{1}{2}t + O(t^2) \quad (t \to 0),$$

对 $f(a+h) = \|a+h\|_2 = \big((a+h \mid a+h)\big)^{1/2}$ 作变形：

$$\begin{aligned}
f(a+h) &= \big(\|a\|_2^2 + 2(a \mid h) + \|h\|_2^2\big)^{1/2} \\
&= \|a\|_2 \Big(1 + 2\frac{(a \mid h)}{\|a\|_2^2} + \frac{\|h\|_2^2}{\|a\|_2^2}\Big)^{1/2} \\
&= \|a\|_2 \Big(1 + \frac{(a \mid h)}{\|a\|_2^2} + O(\|h\|_2^2)\Big) \\
&= f(a) + \|a\|_2^{-1}(a \mid h) + o(\|h\|_2).
\end{aligned}$$

因此 f 在 a 可微，且

$$df(a) = \|a\|_2^{-1}\big(a \mid \cdot\big).$$

由此得到梯度场

$$\nabla f(x) = \frac{x}{\|x\|_2}, \qquad x \in \mathbb{R}^n \setminus \{0\}.$$

事实上，梯度表示的是某一点处函数增大最快的方向. 譬如，如果(x, y)表示位置，$z(x, y)$是海拔高度函数，那么梯度的方向就是坡最陡的方向. 此外还可以证明梯度垂直于等高线，推广到一般情况下的高维线性空间，则梯度正交于函数的等值面.

对于上面的例子，f将向量映射为它的模长，模长增长最快的方向很显然是向量本身的方向，所以∇f是向量方向上的单位矢量. 此外，不难看出f的等值面为以$\|x\|$为半径的球面，∇f显然是正交于球面的.

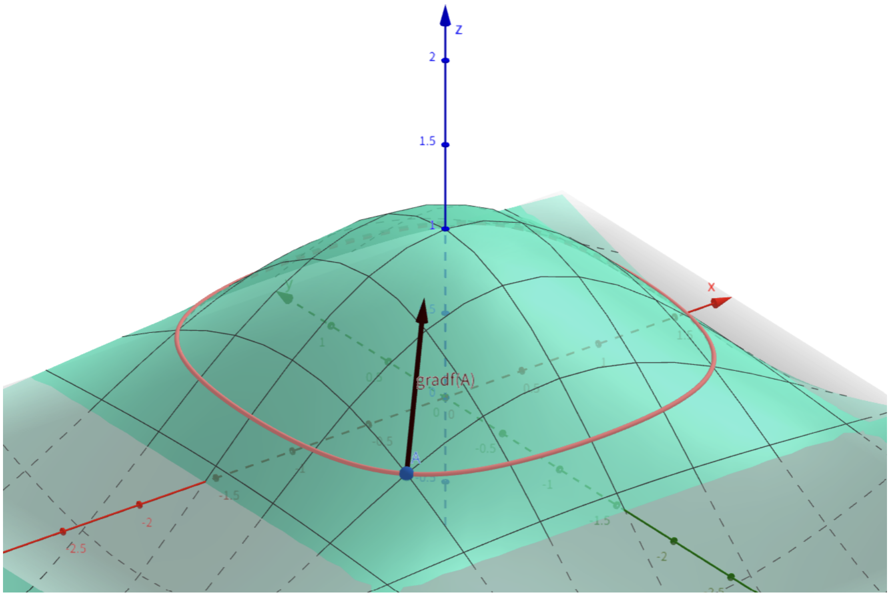


图 12: 梯度表示函数增大最快的方向，且垂直于等高线.

命题. 设 f 在点 $a \in U$ 可微，且 a 不是 f 的临界点（即 $\nabla f(a) \neq 0$）. 那么 f 在 a 的梯度给出 f 在 a 处最大斜率的方向，并且该斜率的值为 $\|\nabla f(a)\|_E$.

证明. 令
$$e = \frac{\nabla f(a)}{\|\nabla f(a)\|_E}.$$
对任一范数为 1 的向量 $h \in E$，由柯西不等式有
$$D_h f(a) = (\nabla f(a) \mid h) \leq \|\nabla f(a)\|_E \|h\|_E = \|\nabla f(a)\|_E = (\nabla f(a) \mid e) = D_e f(a).$$
因此导数在 e 的方向取到最大值. □

1.4 Hadamard 判据

命题. 设 $a \in U$. f 在 a 可微当且仅当存在一个在 a 处连续的映射 $\Phi : U \to \mathcal{L}(E; F)$，使得
$$\forall x \in U \qquad f(x) - f(a) = \Phi(x)(x - a).$$
若此条件成立，则 $\Phi(a) = Df(a)$.

证明. 固定 F 的一组基. 仍然只需讨论一个分量就够了，因此可假设 $F = \mathbb{R}$. 在 E 上选取内积 $(\cdot \mid \cdot)$ 及相应的范数 $\|\cdot\|$，这样就能构建 $\mathcal{L}(E; \mathbb{R})$ 和 E 的同构（竖着一条和横着一条的同构）. 于是命题中的条件可以写成
$$\forall x \in U \qquad f(x) - f(a) = (\Phi(x) \mid x - a).$$

假设 f 在 a 可微，令
$$r(x) = f(x) - f(a) - (\nabla f(a) \mid x - a),$$

则根据可微性的定义有 $r(x)=o(x-a)$（当 $x\to a$）. 定义映射 Φ 如下：

$$\forall x\in U \qquad \Phi(x)=\begin{cases}\nabla f(a), & x=a,\\ \nabla f(a)+r(x)\dfrac{x-a}{\|x-a\|^2}, & x\neq a.\end{cases}$$

则命题的条件显然成立. 此外，

$$\|\Phi(x)-\Phi(a)\|=\|\Phi(x)-\nabla f(a)\|=\frac{|r(x)|}{\|x-a\|}\xrightarrow[x\to a]{}0,$$

即 Φ 在 a 连续.

反过来，若 $\Phi:U\to E$ 在 a 连续且满足命题条件，由柯西不等式

$$|(\Phi(x)-\Phi(a)\mid x-a)|\le\|\Phi(x)-\Phi(a)\|\,\|x-a\|$$

得

$$f(x)-f(a)-(\Phi(a)\mid x-a)=o(x-a),$$

因为 $\Phi(x)-\Phi(a)=o(1)$.于是 f 在 a 可微，且 $\nabla f(a)=\Phi(a)$. □

Hadamard判据和可微性的定义是等价的，但存在一定区别. 可微性的要求在点a存在一个固定的的线性近似+小量误差，而Hadamard形式把常数项写成一个随x变化但在a连续的算子.

例如，令 $f(x)=x^2$, $a=1$. 则 $\varphi(h)=f'(1)h=2h$， $r(x)=x^2-1-2(x-1)=(x-1)^2=o(x-1)$. 按 Hadamard 构造

$$\Phi(x)=\begin{cases}2, & x=1,\\ 2+\dfrac{(x-1)^2}{x-1}=x+1, & x\neq 1.\end{cases}$$

于是 $x^2-1=(x-1)(x+1)=(x-1)\Phi(x)$，且 Φ 在 1 连续.

这体现出，微分的本质是在函数的某一定点上取切线. 而Hadamard判据则是在函数的任一动点到这个点上连线. 当动点靠近选取的定点时，如果函数可微，则连线会趋向和切线重合. （见图10）

1.5 例题

下一章一块说.

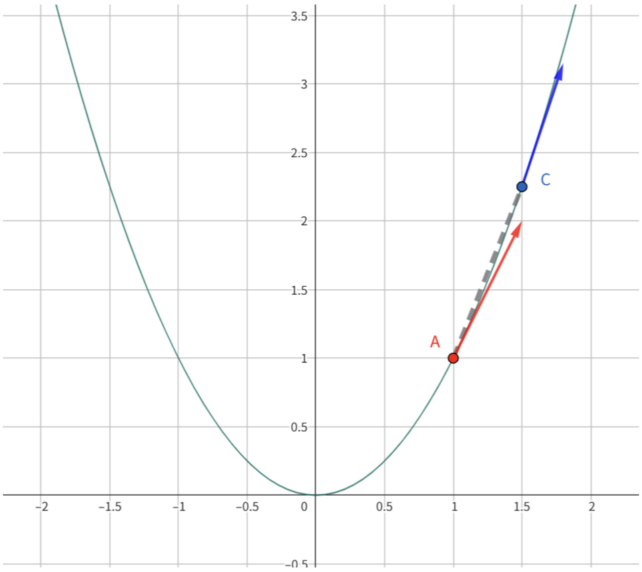


图 13: C趋近于A时，切向会重合

2 可微映射的复合

我们记 E, F, G 为有限维实向量空间.

2.1 复合映射微分：链式法则

命题. 设 U 是 E 的开集，f 是从 U 到 F 的映射，V 是 F 的开集，g 是从 V 到 G 的映射. 假设 $f(U) \subset V$，并记 $g \circ f : U \to G$, $x \mapsto g(f(x))$. 设 $a \in U$，并令 $b = f(a)$. 如果 f 在 a 处可微，且 g 在 b 处可微，则复合映射 $g \circ f$ 在 a 处可微，并且

$$D(g \circ f)(a) = Dg(b) \circ Df(a).$$

证明. 使用我们刚说过的准则.

由于 f 在 a 处可微且 g 在 b 处可微，存在映射 $\Phi : U \to \mathcal{L}(E; F)$ 和 $\Psi : V \to \mathcal{L}(F; G)$，分别在 a 和 b 处连续，使得

$$\forall x \in U \quad f(x) - f(a) = \Phi(x)(x - a) \qquad \text{且} \qquad \forall y \in V \quad g(y) - g(b) = \Psi(y)(y - b).$$

由此可得

$$\forall x \in U \quad (g \circ f)(x) - (g \circ f)(a) = (\Psi(f(x)) \circ \Phi(x))(x - a).$$

而映射 $U \to \mathcal{L}(E; G)$, $x \mapsto \Psi(f(x)) \circ \Phi(x)$ 在 a 处显然是连续的，因此复合映射 $g \circ f$ 在 a 处可微，其微分为

$$\Psi(b) \circ \Phi(a) = Dg(b) \circ Df(a).$$

□

由此显然可知，若 f 与 g 可微，则$g \circ f$ 可微.

在 $E=F=\mathbb{R}$ 的情形下，f,g 及 $g\circ f$ 都是单变量函数，这个推论说明：若 f 和 g 可导，则 $g\circ f$ 可导；计算 $g\circ f$ 的导数（值在向量空间 G 中）：

$$\begin{aligned}
(g\circ f)'(x) &= D(g\circ f)(x)(1) && \text{（导数转微分）}\\
&= Dg\big(f(x)\big)\big(Df(x)(1)\big) && \text{（链式法则）}\\
&= Dg\big(f(x)\big)\big(f'(x)\big) && \text{（微分转导数）}\\
&= f'(x)\, Dg\big(f(x)\big)(1) && \text{（利用微分的线性）}\\
&= f'(x)\, g'\big(f(x)\big). && \text{（微分转导数）}
\end{aligned}$$

这是在向量空间 G 中的等式；它也表明向量 $(g\circ f)'(x)$ 和 $g'(f(x))$ 是共线的.

现在固定 E,F,G 的基，从而构造 $E\simeq\mathbb{R}^n$, $F\simeq\mathbb{R}^p$, $G\simeq\mathbb{R}^q$（记 $n=\dim E$, $p=\dim F$, $q=\dim G$）；记 $f=(f_1,\dots,f_p)$ 和 $g=(g_1,\dots,g_q)$. 微分用雅可比矩阵代替，记作 $f'(x)$ 等；则有

$$(g\circ f)'(x) = g'(f(x))\, f'(x).$$

（注意：上式的$f'(x)$要放在后面，否则根本乘不了）

也就是说：

$$\begin{pmatrix} \partial_1(g_1\circ f)(a) & \cdots & \partial_n(g_1\circ f)(a)\\ \vdots & & \vdots\\ \partial_1(g_q\circ f)(a) & \cdots & \partial_n(g_q\circ f)(a)\end{pmatrix} = \begin{pmatrix} \partial_1 g_1(b) & \cdots & \partial_p g_1(b)\\ \vdots & & \vdots\\ \partial_1 g_q(b) & \cdots & \partial_p g_q(b)\end{pmatrix} \begin{pmatrix} \partial_1 f_1(a) & \cdots & \partial_n f_1(a)\\ \vdots & & \vdots\\ \partial_1 f_p(a) & \cdots & \partial_n f_p(a)\end{pmatrix}$$

或者

$$\forall\, i\in[1,n]\quad \forall\, k\in[1,q],\quad \partial_i(g_k\circ f)(a) = \sum_{j=1}^{p} \partial_j g_k(b)\, \partial_i f_j(a).$$

在允许一些**“符号的滥用”**的前提下，可以给出上式一个更方便的写法. 设 $(x_1,\dots,x_n)$、$(y_1,\dots,y_p)$ 和 $(z_1,\dots,z_q)$ 分别是 $\mathbb{R}^n$、$\mathbb{R}^p$ 和 $\mathbb{R}^q$ 上的标准坐标；并约定记 $y_j(x)=f_j(x)$、$z_k(y)=g_k(y)$ 以及 $z_k(x)=(g_k\circ f)(x)$. 于是可以写成

$$\partial_i f_j(a) = \frac{\partial y_j}{\partial x_i}(a),\qquad \partial_j g_k(b) = \frac{\partial z_k}{\partial y_j}(b),\qquad \partial_i(g_k\circ f)(a) = \frac{\partial z_k}{\partial x_i}(a),$$

并且省略点 a 与 b 的标记后，还能给出更简洁的形式：

$$\forall\, i\in[1,n]\quad \forall\, k\in[1,q],\quad \frac{\partial z_k}{\partial x_i} = \sum_{j=1}^{p} \frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial x_i}.$$

进而

$$dz_k = \sum_{j=1}^{p} \frac{\partial z_k}{\partial y_j}\, dy_j = \sum_{j=1}^{p} \frac{\partial z_k}{\partial y_j}\left(\sum_{i=1}^{n} \frac{\partial y_j}{\partial x_i}\, dx_i\right) = \sum_{i=1}^{n}\left(\sum_{j=1}^{p} \frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial x_i}\right) dx_i\,;$$

注. “符号的滥用（*abus d'écriture*）”是*MHD*常用的表达，它表示不严谨、不规范的符号使用，即便这些符号很好记. 因此，上面的公式只能作为记忆公式，不要用于严格证明，~~尤其是不要在他判卷的时候用~~.

对这种又长又复杂的式子，最好的办法是实操. 参见TD03的第二题.

2.2 可微映射的一般性质

命题. 设 f_1 和 f_2 是从 $U \subset E$ 到 F 的映射，并在点 $a \in U$ 处可微，且 λ_1 与 λ_2 为实数. 则映射 $\lambda_1 f_1 + \lambda_2 f_2$ 在 a 处可微，且其微分为

$$\lambda_1 \, df_1(a) + \lambda_2 \, df_2(a).$$

证明. 考虑映射$f : U \to F \times F, \quad x \mapsto \big(f_1(x),\, f_2(x)\big)$. 显然 f 在 a 处可微，其微分为

$$df(a) : E \to F \times F, \qquad h \mapsto \big(df_1(a)h,\, df_2(a)h\big).$$

另一方面，定义线性映射 $\varphi : F \times F \to F, \quad (y_1, y_2) \mapsto \lambda_1 y_1 + \lambda_2 y_2$. 由于 φ 是线性的，故可微；并且其在任一点处的微分等于 φ 本身（往回翻一翻）. 因此 $\lambda_1 f_1 + \lambda_2 f_2 = \varphi \circ f$ 在 a 处可微，且其微分为

$$\varphi \circ df(a) = \lambda_1 \, df_1(a) + \lambda_2 \, df_2(a).$$

□

这一结论说明，从 U 到 F 的 C^1 类映射构成 F^U 的一个线性子空间，记作 $C^1(U; F)$. 映射

$$C^1(U; F) \longrightarrow C\big(U; \mathcal{L}(E; F)\big), \qquad f \mapsto Df$$

是线性的.

命题. （*formule de Leibniz*）设 F_1, F_2 与 G 为有限维向量空间，且有双线性映射

$$F_1 \times F_2 \to G, \qquad (y_1, y_2) \mapsto y_1 \cdot y_2.$$

设 $f_1 : U \to F_1$ 与 $f_2 : U \to F_2$ 在点 $a \in U$ 可微. 映射

$$f_1 \cdot f_2 : U \to G, \qquad x \mapsto f_1(x) \cdot f_2(x)$$

在 a 处可微，其微分为

$$d(f_1 \cdot f_2)(a) : E \to G, \qquad h \mapsto df_1(a)h \cdot f_2(a) + f_1(a) \cdot df_2(a)h.$$

证明. 仍然考虑映射

$$f : U \to F_1 \times F_2, \qquad x \mapsto \big(f_1(x),\, f_2(x)\big),$$

在 a 处可微. 定义

$$g : F_1 \times F_2 \to G, \qquad (y_1, y_2) \mapsto y_1 \cdot y_2.$$

由于 g 为双线性映射，故可微；在点 $b = (b_1, b_2)$ 处的微分为

$$dg(b) : F_1 \times F_2 \to G, \qquad (u_1, u_2) \mapsto g(u_1, b_2) + g(b_1, u_2).$$

取 $b = f(a) = (f_1(a), f_2(a))$，代入上式并复合 f 与 g，可得 $f_1 \cdot f_2 = g \circ f$ 在 a 处可微，且其微分为

$$d(g \circ f)(a) = dg\big(f(a)\big) \circ df(a),$$

即对任意 $h \in E$，

$$d(f_1 \cdot f_2)(a)h = g\big(df_1(a)h,\, f_2(a)\big) + g\big(f_1(a),\, df_2(a)h\big).$$

得证. □