多元函数微分学

Eliauk

2024 年 8 月 14 日

目录

1 微分

令 E 是 $\mathbb{R}^n$ 中的开集, 函数 $f: E \to \mathbb{R}^m$, $x_0 \in E$, 如果存在线性映射 $L: \mathbb{R}^n \to \mathbb{R}^m$ 使得

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

那么我们说 f 在 x_0 处**可微**, 线性映射 L 被称为 f 在 x_0 处的**微分**. 若 f 在任意 $x_0 \in E$ 处都可微, 那么我们说 f 在 E 中**可微**.

定理 1.1 (微分的唯一性). 沿用上面的记号，设 $L_1:\mathbb{R}^n\to\mathbb{R}^m$ 和 $L_2:\mathbb{R}^n\to\mathbb{R}^m$ 都是线性映射且都为 f 在 x_0 处的微分，那么 $L_1=L_2$.

Proof. 假设 $L_1\neq L_2$, 那么存在非零向量 v 使得 $L_1(v)\neq L_2(v)$, 令 $x=x_0+tv$, 那么 $t\to 0$ 的时候 $x\to x_0$. 又因为

$$\begin{aligned}|L_1(tv)-L_2(tv)| &= \left|\big(f(x)-f(x_0)-L_2(tv)\big)-\big(f(x)-f(x_0)-L_1(tv)\big)\right| \\ &\le |f(x)-f(x_0)-L_2(tv)|+|f(x)-f(x_0)-L_1(tv)|,\end{aligned}$$

所以

$$\lim_{t\to 0}\frac{|L_1(tv)-L_2(tv)|}{|tv|}=0,$$

L_1,L_2 是线性映射表明

$$\lim_{t\to 0}\frac{|t|\,|L_1(v)-L_2(v)|}{|t|\,|v|}=\frac{|L_1(v)-L_2(v)|}{|v|}=0,$$

这与 $L_1(v)\neq L_2(v)$ 矛盾, 所以 $L_1=L_2$. □

我们一般把微分 L 记为 $f'(x_0)$. 有时我们也会用余项的形式叙述微分, 即 f 在 x_0 处可微当且仅当存在线性映射 $L:\mathbb{R}^n\to\mathbb{R}^m$ 使得

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+r(x),$$

其中 $r(x)$ 满足

$$\lim_{x\to x_0}\frac{|r(x)|}{|x-x_0|}=0.$$

此时 $f'(x_0)$ 通常被称为 f 在 x_0 处的**全导数**. 若 f 在 E 中可微, 那么我们可以把 f' 视为映射 $f':E\to\mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m)$, 其将 $x\in E$ 送到线性映射 $f'(x)$. 此外, 我们会用到矩阵范数, 即 $f'(x_0)$ 可以视为 $m\times n$ 矩阵, 那么我们用 $\|f'(x_0)\|$ 表示 $f'(x_0)$ 的算子 2-范数, 即其最大的奇异值. 根据定义, 一个显然的结果是可微必连续.

命题 1.2. $E\subseteq\mathbb{R}^n$ 是开集，若 $f:E\to\mathbb{R}^m$ 在 $x_0\in E$ 处可微，那么 f 在 x_0 处连续.

Proof. f 在 x_0 处可微表明

$$|f(x)-f(x_0)|\le\left\|f'(x_0)\right\||x-x_0|+|r(x)|,$$

于是在 $x\to x_0$ 的时候 $f(x)\to f(x_0)$, 即 f 在 x_0 处连续. □

定理 1.3 (链式法则). 设 $E\subseteq\mathbb{R}^n$ 是开集，函数 $f:E\to\mathbb{R}^m$ 在 $x_0\in E$ 处可微. $F\subseteq\mathbb{R}^m$ 是开集且包含 $f(E)$，函数 $g:F\to\mathbb{R}^k$ 在 $f(x_0)$ 处可微，那么复合函数 $g\circ f:E\to\mathbb{R}^k$ 在 x_0 处可微，并且

$$(g\circ f)'(x_0)=g'(f(x_0))\,f'(x_0).$$

Proof. 根据定义, 我们有

$$\begin{aligned} f(x)-f(x_0) &= f'(x_0)(x-x_0)+r_1(x), \\ g(f(x))-g(f(x_0)) &= g'(f(x_0))(f(x)-f(x_0))+r_2(f(x)), \end{aligned}$$

f 在 x_0 处连续表明 $x \to x_0$ 时 $f(x) \to f(x_0)$, 所以 $x \to x_0$ 时有 $r_1(x) = o(|x-x_0|)$ 以及 $r_2(f(x)) = o(|f(x)-f(x_0)|)$. 因此

$$\begin{aligned} g(f(x))-g(f(x_0)) &= g'(f(x_0))(f(x)-f(x_0))+r_2(f(x)) \\ &= g'(f(x_0))f'(x_0)(x-x_0)+r_2(f(x))+g'(f(x_0))r_1(x), \end{aligned}$$

记 $r(x) = r_2(f(x))+g'(f(x_0))r_1(x)$, 那么

$$|r(x)| \le |r_2(f(x))| + \left\|g'(f(x_0))\right\| |r_1(x)|,$$

所以

$$\lim_{x\to x_0} \frac{|r(x)|}{|x-x_0|} = 0,$$

这就表明 $g \circ f$ 在 x_0 处可微并且 $(g\circ f)'(x_0) = g'(f(x_0))f'(x_0)$. □

2 偏导数和方向导数

令 $E \subseteq \mathbb{R}^n$ 是开集, 函数 $f: E \to \mathbb{R}^m$, $x_0 \in E$. 任取向量 $v \in \mathbb{R}^n$, 如果极限

$$\lim_{t\to 0^+} \frac{f(x_0+tv)-f(x_0)}{t}$$

存在, 那么我们说 f 在 x_0 处**沿着** v **方向可微**, 此时我们将上述极限记为 $D_v f(x_0)$, 称为 f 在 x_0 处沿 v 方向的**方向导数**. 注意, 上述极限中 t 为正向趋于 0, 并且 $D_v f(x_0) \in \mathbb{R}^m$.

引理 2.1. 沿用上面的记号, 如果 f 在 x_0 处可微, 那么对于任意 $v \in \mathbb{R}^n$, 都有 f 在 x_0 处沿着 v 方向可微, 并且

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

Proof. f 在 x_0 处可微表明

$$f(x_0+tv)-f(x_0) = f'(x_0)(tv)+r(x_0+tv),$$

其中 $t \to 0^+$ 时有 $r(x_0+tv) = o(|tv|) = o(t)$, 那么

$$D_v f(x_0) = \lim_{t\to 0^+} \frac{tf'(x_0)(v)+r(x_0+tv)}{t} = f'(x_0)(v).$$ □

引理 2.1 告诉我们全导数 $f'(x_0)$ 可以通过方向导数来表示，即考虑 $\mathbb{R}^n$ 的标准基 $e_1, \ldots, e_n$，那么

$$f'(x_0)(e_i) = D_{e_i} f(x_0),$$

于是对于任意 $v = v_1 e_1 + \cdots + v_n e_n \in \mathbb{R}^n$，有

$$D_v f(x_0) = f'(x_0)(v) = \sum_{i=1}^{n} v_i f'(x_0)(e_i) = \sum_{i=1}^{n} v_i D_{e_i} f(x_0),$$

所以我们只要知道了基向量方向的方向导数 $D_{e_i} f(x_0)$, 就相当于知道了任意方向的方向导数, 这引出了偏导数的定义.

令 $E \subseteq \mathbb{R}^n$ 是开集, 函数 $f : E \to \mathbb{R}^m$, $x_0 \in E$. $e_1, \ldots, e_n$ 是 $\mathbb{R}^n$ 的标准基. 定义 f 在 x_0 处相对于变量 x_i $(1 \leq i \leq n)$ 的**偏导数**为

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

注意, 在偏导数的定义中是 $t \to 0$ 而不是 $t \to 0^+$. 当 $n = 2$ 的时候, 我们一般用 (x, y) 表示 $\mathbb{R}^2$ 中的点, 所以此时偏导数一般记为 $\partial f / \partial x$ 和 $\partial f / \partial y$.

如果 f 在 $x_0 \in E$ 处可微, 引理 2.1 告诉我们

$$D_{e_i} f(x_0) = f'(x_0)(e_i) = -f'(x_0)(-e_i) = -D_{-e_i} f(x_0),$$

所以此时偏导数存在且

$$\frac{\partial f}{\partial x_i}(x_0) = D_{e_i} f(x_0) = -D_{-e_i} f(x_0) = f'(x_0) e_i,$$

此时沿 v 方向的方向导数可以表示为

$$D_v f(x_0) = \sum_{i=1}^{n} v_i \frac{\partial f}{\partial x_i}(x_0).$$

我们也可以写出 $f'(x_0)$ 在标准基下的表示矩阵, 记 $f = (f_1, \ldots, f_m)$, 其中 $f_i : E \to \mathbb{R}$ 是实值函数, 那么按定义可知

$$\frac{\partial f}{\partial x_i}(x_0) = \left(\frac{\partial f_1}{\partial x_i}(x_0), \ldots \frac{\partial f_m}{\partial x_i}(x_0) \right),$$

所以 $f'(x_0)$ 在标准基下的表示矩阵为

$$Jf(x_0) = \left(\frac{\partial f_i}{\partial x_j}(x_0) \right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) & \cdots & \frac{\partial f_2}{\partial x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix} \in M_{m,n}(\mathbb{R}).$$

矩阵 $Jf(x_0)$ 被称为 f 在 x_0 处的 **Jacobi 矩阵**.

对于实值函数 $f:\mathbb{R}^n\to\mathbb{R}$, 我们也把它的 Jacobi 矩阵称作 **梯度**, 记为

$$\nabla f(x)=Jf(x)=\begin{pmatrix}\frac{\partial f}{\partial x_1}(x_0) & \frac{\partial f}{\partial x_2}(x_0) & \cdots & \frac{\partial f}{\partial x_n}(x_0)\end{pmatrix},$$

此时 f 在 x 处沿 $v=v_1e_1+\cdots+v_ne_n$ 方向的方向导数为

$$D_vf(x)=\sum_{i=1}^n v_i\frac{\partial f}{\partial x_i}(x)=v\cdot\nabla f(x).$$

例 2.2. 引理 2.1 的逆命题不成立, 即若 f 在 x_0 的任意方向上都可微, 也不能说明 f 在 x_0 处可微. 考虑 $f:\mathbb{R}^2\to\mathbb{R}$ 为

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

任取 $v=(v_1,v_2)\in\mathbb{R}^2$, 那么

$$D_vf(0,0)=\lim_{t\to0^+}\frac{f(tv_1,tv_2)-f(0,0)}{t}=\lim_{t\to0^+}\frac{v_1^3}{v_1^2+v_2^2}=\frac{v_1^3}{v_1^2+v_2^2},$$

所以 $D_vf(0,0)$ 都存在. 另一方面, 我们有

$$\begin{aligned}\frac{\partial f}{\partial x}(0,0)&=\lim_{t\to0}\frac{f(t,0)-f(0,0)}{t}=\lim_{t\to0}\frac{t^2}{t^2}=1,\\ \frac{\partial f}{\partial y}(0,0)&=\lim_{t\to0}\frac{f(0,t)-f(0,0)}{t}=\lim_{t\to0}\frac{0}{t}=0,\end{aligned}$$

如果 f 在 $(0,0)$ 处可微, 那么

$$Jf(0,0)=\begin{pmatrix}\frac{\partial f}{\partial x}(0,0) & \frac{\partial f}{\partial y}(0,0)\end{pmatrix}=\begin{pmatrix}1 & 0\end{pmatrix}.$$

此时

$$\begin{aligned}&\lim_{(x,y)\to(0,0)}\frac{\left|f(x,y)-f(0,0)-Jf(0,0)(x,y)^T\right|}{|(x,y)|}\\ =&\lim_{(x,y)\to(0,0)}\frac{x^3/(x^2+y^2)-x}{\sqrt{x^2+y^2}}=\lim_{(x,y)\to(0,0)}\frac{-xy^2}{(x^2+y^2)^{3/2}},\end{aligned}$$

当 $y=x$ 的时候极限值为 $-\sqrt{2}/4$, 当 $y=0$ 的时候极限值为 0, 所以上述极限不存在, 这与 f 在 $(0,0)$ 处可微矛盾. 此外, 该例也告诉我们, 若 $f:E\to\mathbb{R}^m$ 在 $x_0\in E$ 处偏导数都存在, 也不意味着 f 在 x_0 处可微.

令开集 $E \subseteq \mathbb{R}^n$, 函数 $f: E \to \mathbb{R}^m$, 如果 f 在 E 中可微, 并且 $f': E \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ 是连续映射, 那么我们说 f 在 E 中**连续可微**, 记为 $f \in C^1$. 更准确地说, $f \in C^1$ 当且仅当对于每个 $x \in E$, 任取 $\varepsilon > 0$, 都存在 $\delta > 0$, 使得 $|y-x| < \delta$ 的时候有

$$\left\| f'(y) - f'(x) \right\| < \varepsilon.$$

定理 2.3. $f \in C^1$ 当且仅当对于任意 $1 \le i \le m, 1 \le j \le n$, 偏导数 $\partial f_i / \partial x_j : E \to \mathbb{R}$ 都存在且在 E 上连续.

Proof. $(\Rightarrow)$ 若 $f \in C^1$, 那么任取 $x \in E$, 有

$$\frac{\partial f}{\partial x_j}(x) = f'(x)(e_j),$$

其中 $e_j \in \mathbb{R}^n$ 是标准基向量. 那么

$$\begin{aligned}\left| \frac{\partial f}{\partial x_j}(y) - \frac{\partial f}{\partial x_j}(x) \right| &= \left| f'(y)(e_j) - f'(x)(e_j) \right| = \left| \big(f'(y) - f'(x)\big)(e_j) \right| \\ &\le \left\| f'(y) - f'(x) \right\| |e_j| = \left\| f'(y) - f'(x) \right\|,\end{aligned}$$

而 f' 连续, 所以 $\partial f / \partial x_j : E \to \mathbb{R}^m$ 连续, 故每个分量 $\partial f_i / \partial x_j$ 连续.

$(\Leftarrow)$ 只需证明 $m = 1$ 的情况即可. 因为如果 $m = 1$ 的时候结论成立, 那么对于 $1 \le i \le m$, $\partial f_i / \partial x_j$ 都连续表明 $f_i' : E \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$ 连续, 进而 $f' : E \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) = \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})^m$ 连续. 故下面假设 $m = 1$. 任取 $x \in E$, 我们先说明 f 在 x 处可微, 此时 Jacobi 矩阵必须为

$$Jf(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix},$$

故需要证明任取 ε, 都存在 $\delta > 0$, 使得 $|y - x| < \delta$ 的时候有

$$\left| f(y) - f(x) - Jf(x)(y-x)^T \right| < \varepsilon \, |y - x| .$$

$\partial f / \partial x_j(x)$ 连续表明存在 $\delta_j > 0$ 使得 $|y - x| < \delta_j$ 的时候有

$$\left| \frac{\partial f}{\partial x_j}(y) - \frac{\partial f}{\partial x_j}(x) \right| < \varepsilon,$$

取 $\delta = \min\{\delta_1, \ldots, \delta_n\}$. 设 $y - x = (v_1, \ldots, v_n) = v_1 e_1 + \cdots + v_n e_n \in \mathbb{R}^n$, 那么 $|y - x| < \delta$ 表明

$|v_j| < \delta$. 此时

$$\begin{aligned}\left|f(y)-f(x)-Jf(x)(y-x)^T\right| &= \left|f(x+v_1e_1+\cdots+v_ne_n)-f(x)-\sum_{j=1}^n v_j\frac{\partial f}{\partial x_j}(x)\right| \\ &= \left|\sum_{j=1}^n[f(x+w_j)-f(x+w_{j-1})]-\sum_{j=1}^n v_j\frac{\partial f}{\partial x_j}(x)\right| \\ &\le \sum_{j=1}^n\left|f(x+w_j)-f(x+w_{j-1})-v_j\frac{\partial f}{\partial x_j}(x)\right|,\end{aligned}$$

其中 $w_0=0$, $w_j=v_1e_1+\cdots+v_je_j$. 记函数 $g_j:[0,v_j]\to\mathbb{R}$ 为

$$g_j(t)=f(x+w_{j-1}+te_j),$$

那么 g_j 在 $[0,v_j]$ 中连续且在 $(0,v_j)$ 中可微, 根据中值定理, 存在 $\theta_j\in(0,v_j)$ 使得

$$g_j(v_j)-g_j(0)=g_j'(\theta_j)v_j=v_j\frac{\partial f}{\partial x_j}(x+w_{j-1}+\theta_je_j),$$

此时 $\left|x+w_{j-1}+\theta_je_j-x\right|<|y-x|<\delta$, 所以

$$\begin{aligned}\left|f(x+w_j)-f(x+w_{j-1})-v_j\frac{\partial f}{\partial x_j}(x)\right| &= \left|v_j\frac{\partial f}{\partial x_j}(x+w_{j-1}+\theta_je_j)-v_j\frac{\partial f}{\partial x_j}(x)\right| \\ &< \varepsilon\left|v_j\right|\le\varepsilon\left|y-x\right|,\end{aligned}$$

因此

$$\left|f(y)-f(x)-Jf(x)(y-x)^T\right|<n\varepsilon\left|y-x\right|,$$

这就表明 f 在 $x\in E$ 处可微.

然后我们说明 $f':E\to\mathrm{Hom}(\mathbb{R}^n,\mathbb{R})$ 连续. 由于

$$f'(x)=\begin{pmatrix}\frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_n}(x)\end{pmatrix},$$

所以

$$\left\|f'(y)-f'(x)\right\|\le\sum_{j=1}^n\left|\frac{\partial f}{\partial x_j}(y)-\frac{\partial f}{\partial x_j}(x)\right|,$$

所以 f' 连续. □

3 高阶导数

令 $E \subseteq \mathbb{R}^n$ 是开集, 函数 $f : E \to \mathbb{R}^m$, 偏导数 $\partial f/\partial x_i : E \to \mathbb{R}^m$ 在 E 中可微, 那么我们定义 f 的**二阶偏导数**为

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

如果所有的 $\partial f/\partial x_i \in C^1$, 即所有的二阶偏导数 $\partial^2 f/\partial x_i \partial x_j$ 都连续, 那么我们说 f **二阶连续可微**, 记为 $f \in C^2$. 类似的, 我们可以定义高阶偏导数和 C^n.

例 3.1. 一般而言, $\partial^2 f/\partial x_i \partial x_j \neq \partial^2 f/\partial x_j \partial x_i$. 考虑 $f : \mathbb{R}^2 \to \mathbb{R}$ 为

$$f(x,y) = \begin{cases} xy^3/(x^2+y^2) & (x,y) \neq (0,0), \\ 0 & (x,y) = (0,0). \end{cases}$$

容易算得

$$\frac{\partial f}{\partial x} = \begin{cases} \dfrac{y^3(y^2-x^2)}{(x^2+y^2)^2} & (x,y) \neq (0,0), \\ 0 & (x,y) = (0,0), \end{cases} \quad \frac{\partial f}{\partial y} = \begin{cases} \dfrac{x(y^4+3x^2y^2)}{(x^2+y^2)^2} & (x,y) \neq (0,0), \\ 0 & (x,y) = (0,0), \end{cases}$$

故

$$\begin{aligned} \frac{\partial^2 f}{\partial x \partial y}(0,0) &= \lim_{t\to 0} \frac{\partial f/\partial y(t,0) - \partial f/\partial y(0,0)}{t} = 0, \\ \frac{\partial^2 f}{\partial y \partial x}(0,0) &= \lim_{t\to 0} \frac{\partial f/\partial x(0,t) - \partial f/\partial x(0,0)}{t} = 1, \end{aligned}$$

此时 $\partial^2 f/\partial x \partial y(0,0) \neq \partial^2 f/\partial y \partial x(0,0)$.

定理 3.2. 令 $E \subseteq \mathbb{R}^n$ 是开集，$f : E \to \mathbb{R}^m$ 在 E 中二阶连续可微，那么对于任意 $1 \leq i, j \leq n$，有

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

Proof. 任取 $x = x_1e_1 + \cdots + x_ne_n \in E$, 考虑

$$\Delta = f(x+te_i+se_j) - f(x+te_i) - f(x+se_j) + f(x),$$

由于 E 是开集, 我们总是可以令 t, s 足够小使得 Δ 有意义. 令

$$g(u) = f(x+ue_i+se_j) - f(x+ue_i),$$

对 $g : [0, t] \to \mathbb{R}$ 使用中值定理, 存在 $\theta \in (0, t)$ 使得

$$\begin{aligned}\Delta &= g(t) - g(0) = g'(\theta)t \\ &= t\left[\frac{\partial f}{\partial x_i}(x + \theta e_i + s e_j) - \frac{\partial f}{\partial x_i}(x + \theta e_i)\right],\end{aligned}$$

令

$$h(u) = \frac{\partial f}{\partial x_i}(x + \theta e_i + u e_j),$$

$f \in C^2$ 保证了 $h : [0, s] \to \mathbb{R}$ 在 $(0, s)$ 中可微, 再次使用中值定理, 存在 $\eta \in (0, s)$ 使得

$$\Delta = tsh'(\eta) = ts\frac{\partial^2 f}{\partial x_j \partial x_i}(x + \theta e_i + \eta e_j).$$

固定 t, 令 $s \to 0$, 有

$$\frac{\partial f}{\partial x_j}(x + te_i) - \frac{\partial f}{\partial x_j}(x) = \lim_{s \to 0} \frac{\Delta}{s} = t\frac{\partial^2 f}{\partial x_j \partial x_i}(x + \theta e_i),$$

然后令 $t \to 0$, 就得到了

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \lim_{t \to 0} \frac{\partial f/\partial x_j(x + te_i) - \partial f/\partial x_i(x)}{t} = \frac{\partial^2 f}{\partial x_j \partial x_i}(x). \qquad \square$$

4 压缩映射原理

令 (X, d) 是度量空间, $f : X \to X$ 是映射, 如果存在 $c \in (0, 1)$, 使得对于任意 $x, y \in X$ 都有

$$d(f(x), f(y)) \le cd(x, y),$$

那么我们说 f 是一个**压缩映射**. 显然压缩映射一定是连续映射, 并且压缩映射与自身的复合也是压缩映射.

定理 4.1 (压缩映射原理). (X, d) 是度量空间，$f : X \to X$ 是压缩映射，那么 f 最多只有一个不动点. 此外，如果 (X, d) 是完备度量空间，那么 f 有唯一的不动点.

Proof. 我们先说明 f 如果有不动点, 那么一定唯一. 若 $x, y \in X$ 使得 $f(x) = x$ 以及 $f(y) = y$, 那么

$$d(x, y) = d(f(x), f(y)) \le cd(x, y),$$

这就表明 $d(x, y) = 0$, 故 $x = y$.

现在假设 X 是完备度量空间. 任取 $x_0 \in X$, 令序列 $x_n = f(x_{n-1})$, 那么

$$\begin{aligned} d(x_{n+1}, x_n) &= d(f(x_n), f(x_{n-1})) \le c d(x_n, x_{n-1}) \\ &\le c^2 d(x_{n-1}, x_{n-2}) \le \cdots \le c^n d(x_1, x_0), \end{aligned}$$

所以对于 $m > n$, 有

$$\begin{aligned} d(x_m, x_n) &\le d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \\ &\le (c^{m-1} + c^{m-2} + \cdots + c^n) d(x_1, x_0) \\ &< \frac{c^n}{1-c} d(x_1, x_0) \to 0, \end{aligned}$$

故 $\{x_n\}$ 是 Cauchy 列, 从而收敛, 设 $x_n \to x$. 又因为 f 是连续映射, 所以

$$f(x) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x,$$

即 x 是 f 唯一的不动点. □

5 反函数定理

我们先介绍两个中值定理.

定理 5.1 (实值函数的微分中值定理). 设 $E \subseteq \mathbb{R}^n$ 为凸开集，$f : E \to \mathbb{R}$ 是可微函数，那么任给 $x, y \in E$，存在 $\xi = x + \theta(y - x) \in E, \theta \in (0, 1)$ 使得

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

Proof. 考虑曲线 $\gamma : [0, 1] \to E$ 为 $\gamma(t) = x + t(y - x)$. 那么复合映射 $f \circ \gamma : [0, 1] \to \mathbb{R}$ 可微, 且

$$(f \circ \gamma)'(t) = f'(\gamma(t))\gamma'(t) = f'(\gamma(t))(y - x).$$

根据中值定理, 存在 $\theta \in (0, 1)$ 使得

$$f \circ \gamma(1) - f \circ \gamma(0) = (f \circ \gamma)'(\theta),$$

即

$$f(y) - f(x) = f'(\xi)(y - x).$$ □

对于向量值函数而言, 没有这么好的结果, 但是下述结论依旧是有用的.

定理 5.2 (拟微分中值定理). 设 $E \subseteq \mathbb{R}^n$ 是凸开集，$f : E \to \mathbb{R}^m$ 可微，那么任取 $x, y \in E$，存在 $\xi = x + \theta(y - x) \in E, \theta \in (0, 1)$ 使得

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

Proof. 令曲线 $\gamma : [0, 1] \to E$ 为 $\gamma(t) = x + t(y - x)$. 考虑函数 $\varphi : [0, 1] \to \mathbb{R}$ 为

$$\varphi(t) = f(\gamma(t)) \cdot (f(y) - f(x)),$$

这里 $\cdot$ 表示向量的点乘. 那么 φ 在 $(0, 1)$ 中可微, 根据中值定理, 存在 $\theta \in (0, 1)$ 使得

$$|f(y) - f(x)|^2 = \varphi(1) - \varphi(0) = \varphi'(\theta) = \big(f'(\gamma(\theta))(y - x)\big) \cdot (f(y) - f(x)),$$

所以

$$|f(y) - f(x)|^2 \le \left|\varphi'(\theta)\right| \le \left|f'(\gamma(\theta))(y - x)\right| |f(y) - f(x)| ,$$

故

$$|f(y) - f(x)| \le \left\| f'(\xi) \right\| |y - x| . \qquad \square$$

推论 5.3. $E \subseteq \mathbb{R}^n$ 是凸开集，如果 $f : E \to \mathbb{R}^m$ 在任意 $x \in E$ 处的微分 $f'(x) \equiv 0$，那么 f 是常值映射，即存在 $y_0 \in \mathbb{R}^m$ 使得 $f(x) \equiv y_0$.

定理 5.4 (反函数定理). 令 $E \subseteq \mathbb{R}^n$ 是开集，函数 $f : E \to \mathbb{R}^n$ 且 $f \in C^1$，对于某个 $x_0 \in E$，记 $y_0 = f(x_0)$，如果 $f'(x_0)$ 可逆，那么

(1) 存在包含 x_0 的开集 U 和包含 y_0 的开集 V，使得 $f(U) = V$ 并且 $f : U \to V$ 是双射.

(2) 设 $g : V \to U$ 是 $f : U \to V$ 的逆映射，那么 $g \in C^1$ 并且对于任意 $y \in V$ 有

$$g'(y) = (f'(g(y)))^{-1}.$$

Proof. (1) 令 $\lambda = 1/ (2 \|f'(x_0)\|)$, f' 连续表明存在开球 $U \subseteq E$ 使得 $x \in U$ 时有

$$\left\| f'(x) - f'(x_0) \right\| \le \lambda .$$

令 $V = f(U)$, 我们先说明 $f : U \to V$ 是双射, 再说明 V 是开集.

对于任意 $y \in \mathbb{R}^n$, 定义函数 $\varphi_y : E \to \mathbb{R}^n$ 为

$$\varphi_y(x) = x + f'(x_0)^{-1}(y - f(x)),$$

这样定义是因为 $y = f(x)$ 当且仅当 $\varphi_y(x) = x$, 即 φ_y 有不动点, 那么 x 的唯一性便可以由压缩映射原理保证. 下面先说明 φ_y 是 $U \to \mathbb{R}^n$ 的压缩映射. 由于

$$\varphi_y'(x) = I - f'(x_0)^{-1} f'(x) = f'(x_0)^{-1}\big(f'(x_0) - f'(x)\big),$$

所以

$$\|\varphi_y'(x)\| \le \|f'(x_0)^{-1}\| \, \|f'(x_0) - f'(x)\|,$$

所以 $x \in U$ 时有

$$\|\varphi_y'(x)\| \le \lambda \|f'(x_0)^{-1}\| = \frac{1}{2},$$

根据定理 5.2, 所以对于任意 $x, x' \in U$, 有

$$\left|\varphi_y(x') - \varphi_y(x)\right| \le \frac{1}{2}\left|x' - x\right|, \tag{1}$$

故 $\varphi_y : U \to \mathbb{R}^n$ 是压缩映射. 回到函数 $f : U \to V$, 任取 $y \in V = f(U)$, 设 $y = f(x)$, 那么 $\varphi_y(x) = x$, 即 x 是 φ_y 的不动点, 根据压缩映射原理, 这样的 x 是唯一的, 所以 $f : U \to V$ 是双射.

接下来说明 $V = f(U)$ 是开集. 任取 $y = f(x) \in V$, 由于 U 是开集, 所以存在 $r > 0$ 使得开球 $B_r(x) \subseteq U$, 我们可以让 r 足够小使得闭球 $\bar{B}_r(x) \subseteq U$. 下面我们说明 $B_{\lambda r}(y) \subseteq V$, 从而表明 V 是开集. 对于 $y' \in B_{\lambda r}(y)$, 即 $|y' - y| < \lambda r$. $y \in V$ 当且仅当存在 $x' \in U$ 使得 $y' = f(x')$, 这启发我们构造 $\varphi_{y'}$ 的压缩映射. 注意到

$$\left|\varphi_{y'}(x) - x\right| = \left|f'(x_0)^{-1}\big(y' - f(x)\big)\right| < \|f'(x_0)^{-1}\| \lambda r = \frac{r}{2},$$

所以对于任意 $x' \in \bar{B}_r(x) \subseteq U$, 有

$$\begin{aligned}\left|\varphi_{y'}(x') - x\right| &\le \left|\varphi_{y'}(x') - \varphi_{y'}(x)\right| + \left|\varphi_{y'}(x) - x\right| \\ &< \frac{1}{2}\left|x' - x\right| + \frac{r}{2} \le r,\end{aligned}$$

这表明 $\varphi_{y'}(x') \in B_r(x)$, 也即 $\varphi_{y'}\big(\bar{B}_r(x)\big) \subseteq B_r(x)$. $\bar{B}_r(x)$ 作为 $\mathbb{R}^n$ 的闭子集是完备的, 所以 $\varphi_{y'} : \bar{B}_r(x) \to B_r(x)$ 是压缩映射, 故存在唯一的不动点 $x' \in \bar{B}_r(x)$, 因此 $y' = f(x') \in f\big(\bar{B}_r(x)\big) \subseteq f(U) = V$. 这就证明了 V 是开集.

(2) 任取 $y = f(x) \in V$, 我们需要说明 g 在 y 处可微. 注意到 $f'(x_0)\big[I - \varphi_y'(x)\big] = f'(x)$, 由于 $\|\varphi_y'(x)\| \le 1/2$, 所以 $I - \varphi_y'(x)$ 可逆, 所以 $f'(x)$ 可逆. 对于 $y' = f(x') \in V$, 有

$$\begin{aligned}g(y') - g(y) - f'(x)^{-1}(y' - y) &= x' - x - f'(x)^{-1}(f(x') - f(x)) \\ &= -f'(x)^{-1}\big[f(x') - f(x) - f'(x)(x' - x)\big].\end{aligned} \tag{2}$$

根据 (1) 式, 我们有

$$\left|\varphi_y(x') - \varphi_y(x)\right| = \left|x' - x - f'(x_0)^{-1}(y' - y)\right| \le \frac{1}{2}\left|x' - x\right|,$$

所以

$$\left\| f'(x_0)^{-1} \right\| \left| y' - y \right| \geq \left| f'(x_0)^{-1}(y' - y) \right| \geq \frac{1}{2} \left| x' - x \right|, \tag{3}$$

结合 (2) 式, 得到

$$\frac{\left| g(y') - g(y) - f'(x)^{-1}(y' - y) \right|}{|y' - y|} \leq \frac{\left\| f'(x)^{-1} \right\|}{\lambda} \frac{|f(x') - f(x) - f'(x)(x' - x)|}{|x' - x|},$$

当 $y' \to y$ 的时候, (3) 表明 $x' \to x$, 而 f 在 x 处可微, 所以上式表明 g 在 y 处可微, 并且

$$g'(y) = f'(x)^{-1} = f'(g(y))^{-1}.$$

由于 f', g 以及矩阵求逆映射是连续的, 所以 $g'(y) = f'(g(y))^{-1}$ 是连续映射, 故 $g \in C^1$. □

推论 5.5. 令 $E \subseteq \mathbb{R}^n$ 是开集，$f : E \to \mathbb{R}^n$ 连续可微，如果对于任意 $x \in E$，$f'(x)$ 都可逆，那么 f 是开映射，即对于任意开集 $W \subseteq E$，$f(W)$ 也是开集.

Proof. 任取 $f(x) \in f(W)$, 那么 f 在 x 的某个邻域 $U \subseteq W$ 上是 $U \to f(U)$ 的双射, 且 $f(U) \subseteq f(W)$ 是开集, 所以 $f(W)$ 是开集. □

反函数定理告诉我们, 对于方程组

$$f_i(x_1, \ldots, x_n) = y_i \quad (1 \leq i \leq n),$$

如果 f 连续可微且 $f'(x_0)$ 可逆, 那么只要将 $(y_1, \ldots, y_n)$ 和 $(x_1, \ldots, x_n)$ 分别限制在 $f(x_0)$ 和 x_0 的足够小的邻域中, 这个方程组总是有唯一解, 并且解连续可微.

6 隐函数定理

我们先规定一些记号. 若点 $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ 和 $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$, 记

$$(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbb{R}^{n+m}.$$

任意线性映射 $L : \mathbb{R}^{n+m} \to \mathbb{R}^m$ 可以分成两个部分 $L_1 : \mathbb{R}^n \to \mathbb{R}^m$ 和 $L_2 : \mathbb{R}^m \to \mathbb{R}^m$, 定义为

$$L_1(x) = L(x, 0), \quad L_2(y) = L(0, y).$$

于是

$$L(x, y) = L_1(x) + L_2(y).$$

引理 6.1. 如果 $L:\mathbb{R}^{n+m}\to\mathbb{R}^m$ 是线性映射，L_2 可逆. 那么对于每个 $x\in\mathbb{R}^n$，存在唯一的 $y\in\mathbb{R}^m$ 使得 $L(x,y)=0$. 并且

$$y=-L_2^{-1}(L_1(x)).$$

Proof. $L(x,y)=0$ 表明 $L_1(x)+L_2(y)=0$, 所以 L_2 可逆表明 $y=-L_2^{-1}(L_1(x))$ 是唯一解. □

定理 6.2 (隐函数定理). 令 $E\subseteq\mathbb{R}^{n+m}$ 是开集，$f:E\to\mathbb{R}^m$ 是连续可微映射，如果某个点 $(x_0,y_0)\in E$ 使得 $f(x_0,y_0)=0$ 且 $f_2'(x_0,y_0):\mathbb{R}^m\to\mathbb{R}^m$ 可逆，那么存在 (x_0,y_0) 的邻域 $U\subseteq\mathbb{R}^{n+m}$ 和 x_0 的邻域 $W\subseteq\mathbb{R}^n$，使得：

(1) 对于每个 $x\in W$，存在唯一的 y 满足 $(x,y)\in U$ 以及 $f(x,y)=0$.

(2) 记上述 $y=g(x)$，那么函数 $g:W\to\mathbb{R}^m$ 是连续可微映射，并且使得

$$f(x,g(x))=0,$$

g 在 $x\in W$ 处的微分为

$$g'(x)=-(f_2'(x,g(x)))^{-1}f_1'(x,g(x)).$$

我们把 g 称为 f 定义的**隐函数**.

Proof. 定义 $F:E\to\mathbb{R}^{n+m}$ 为

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

那么 F 是连续可微映射. 其在 (x_0,y_0) 处的 Jacobi 矩阵为

$$JF(x_0,y_0)=\begin{pmatrix} & I_n & \\ Jf_1(x_0,y_0) & & Jf_2(x_0,y_0)\end{pmatrix},$$

其中 Jf_1,Jf_2 分别表示微分 $f_1'(x_0,y_0),f_2'(x_0,y_0)$ 的表示矩阵. $f_2'(x_0,y_0)$ 可逆表明矩阵 $Jf_2(x_0,y_0)$ 可逆, 所以 $JF(x_0,y_0)$ 可逆. 对 F 应用反函数定理, 存在包含 (x_0,y_0) 的开集 $U\subseteq\mathbb{R}^{n+m}$, 包含 $(x_0,0)=(x_0,f(x_0,y_0))$ 的开集 $V\subseteq\mathbb{R}^{n+m}$, 使得 $F:U\to V$ 是双射. 令

$$W=\{x\in\mathbb{R}^n\mid(x,0)\in V\},$$

那么 $x_0\in W$, 且 V 是开集表明 W 是开集.

(1) 对于每个 $x\in W$, $(x,0)\in V$, 所以存在唯一的 $(x',y)\in U$ 使得 $(x,0)=F(x',y)=(x',f(x',y))$, 这表明 $x'=x$, $f(x',y)=0$, 故 $f(x,y)=0$, 这就证明了 (1).

(2) 任取 $x\in W$, 那么 $f(x,g(x))=0$, 所以 $F(x,g(x))=(x,0)$. 记 $G:V\to U$ 是 F 的逆映射, 于是

$$(x,g(x))=G(x,0),$$

由于 G 连续可微, 所以 g 连续可微.

记 $\Phi(x)=(x,g(x))$, 那么 $f(\Phi(x))=0$, 求 x 处的微分, 根据链式法则, 所以

$$f'(x,g(x))\Phi'(x)=f'(\Phi(x))\Phi'(x)=0,$$

任取 $v\in\mathbb{R}^n$, 有

$$\Phi'(x)(v)=(v,g'(x)(v)),$$

所以

$$f'(x,g(x))\big(v,g'(x)(v)\big)=0,$$

根据引理 6.1, 就有

$$g'(x)(v)=-(f_2'(x,y))^{-1}f_1'(x,y)(v).$$

□

例 6.3. 取 $n=3$, $m=2$. 考虑映射 $f:\mathbb{R}^{3+2}\to\mathbb{R}^2$ 为

$$f(x_1,x_2,x_3,y_1,y_2)=(2e^{y_1}+x_1y_2-4x_2+3,\ y_2\cos y_1-6y_1+2x_1-x_3),$$

取 $(x_0,y_0)=(3,2,7,0,1)$, 那么 $f(x_0,y_0)=0$. Jacobi 矩阵为

$$Jf(x_0,y_0)=\begin{pmatrix}1&-4&0&2&3\\2&0&-1&-6&1\end{pmatrix},$$

所以 $f_1'(x_0,y_0)$ 和 $f_2'(x_0,y_0)$ 的表示矩阵分别为

$$Jf_1(x_0,y_0)=\begin{pmatrix}1&-4&0\\2&0&-1\end{pmatrix},\quad Jf_2(x_0,y_0)=\begin{pmatrix}2&3\\-6&1\end{pmatrix}.$$

故 $f_2'(x_0,y_0)$ 可逆. 这表明在 $(3,2,7)$ 的某个邻域 W 上存在连续可微函数 $g:W\to\mathbb{R}^2$, 其满足 $g(3,2,7)=(0,1)$ 以及 $f(x,g(x))=0$, 并且

$$Jg(x_0)=-(Jf_2(x_0,y_0))^{-1}Jf_1(x_0,y_0)=-\frac{1}{20}\begin{pmatrix}1&-3\\6&2\end{pmatrix}\begin{pmatrix}1&-4&0\\2&0&-1\end{pmatrix}=\begin{pmatrix}\frac14&\frac15&-\frac{3}{20}\\-\frac12&\frac65&\frac{1}{10}\end{pmatrix}.$$

例 6.4 (隐式曲面)**.** 设 $f:\mathbb{R}^3\to\mathbb{R}$ 连续可微, 对于 $c\in\mathbb{R}$, 记

$$S=f^{-1}(c)=\{(x,y,z)\in\mathbb{R}^3\mid f(x,y,z)=c\},$$

S 称为 f 在 c 处的水平集. 如果对于任意 $(x,y,z)\in S$, 均有 $\nabla f(x,y,z)\neq 0$, 那么 S 称为 f 确定的**隐式曲面**.

设 $(x_0,y_0,z_0)\in S$, 不妨设 $\partial f/\partial z(x_0,y_0,z_0)\neq 0$, 根据隐函数定理, 在 (x_0,y_0,z_0) 附近, 存在连续可微函数 g 使得

$$f(x,y,g(x,y))=c,$$

此时 S 可以用 $z=g(x,y)$ 表示, 并且

$$\frac{\partial g}{\partial x}(x,y)=-\frac{\partial f/\partial x(x,y,z)}{\partial f/\partial z(x,y,z)},\quad \frac{\partial g}{\partial y}(x,y)=-\frac{\partial f/\partial y(x,y,z)}{\partial f/\partial z(x,y,z)}.$$

7 秩定理

秩定理可以视为反函数定理的进一步推广，秩定理表明一个连续可微映射 f 在点 x 附近的行为类似于线性映射 $f'(x)$.

首先我们简单介绍一下线性代数中的投影映射. 设 $P: V \to V$ 是线性变换, 如果 $P^2 = P$, 那么我们说 P 是**投影变换**. 假设 V 是有限维向量空间, 我们有 $V = \ker P + \operatorname{im} P$. 任取 $P(x) \in \ker P \cap \operatorname{im} P$, 那么 $P(x) = P^2(x) = 0$, 所以 $\ker P \cap \operatorname{im} P = 0$, 因此 $V = \ker P \oplus \operatorname{im} P$. 给定 V 的一个子空间 U, 设子空间 W 使得 $V = U \oplus W$, 那么任意 $x \in V$ 可以唯一写为 $x = x_1 + x_2$, 其中 $x_1 \in U$, $x_2 \in W$. 定义 $P_U : V \to V$ 为 $P_U(x) = x_1$, 显然 P 是线性变换并且 $\operatorname{im} P_U = U$, 并且 $P_U^2(x) = P_U(x_1) = x_1 = P_U(x)$, 所以此时 P_U 称为 V **在 U 上的投影**.

定理 7.1 (秩定理). 设 $E \subseteq \mathbb{R}^n$ 是开集，$f : E \to \mathbb{R}^m$ 是连续可微映射，并且对于任意 $x \in E$，线性映射 $f'(x)$ 的秩恒为 r，此时 $0 \le r \le \min(m, n)$. 给定 $x_0 \in E$，记 $S = \operatorname{im} f'(x_0)$，$P_S$ 表示 $\mathbb{R}^m$ 在 S 上的投影. 那么存在连续可微的双射 $G : U \to V$ (并且逆映射也连续可微)，其中 U 是包含 x_0 的开集，V 是 $\mathbb{R}^n$ 的开集，使得 $x \in V$ 时有

$$f(G^{-1}(x)) = f'(x_0)(x) + \varphi(f'(x_0)(x)),$$

其中 φ 是 S 的开集 $f'(x_0)(V)$ 到 $\ker P_S = S^{\perp}$ 的连续可微映射.

Proof. 如果 $r = 0$, 根据推论 5.3, 可以取 U 为 x_0 处的一个开球, $V = U$, $G(x) = x$, $\varphi(0) = f(x_0)$, 那么 f 在 U 上为常值映射, 故 $f(G^{-1}(x)) = x_0 = f'(x_0)(x) + \varphi(f'(x_0)(x))$.

下面假设 $r > 0$. 定义 $G : E \to \mathbb{R}^n$ 为

$$G(x) = x + f'(x_0)^{\dagger}(f(x) - f'(x_0)x),$$

$f'(x_0)^{\dagger}$ 表示 $f'(x_0)$ 的 Moore-Penrose 广义逆, 其满足 $f'(x_0)f'(x_0)^{\dagger} = P_S$ 以及 $f'(x_0)^{\dagger}f'(x_0) = P_W$, $W = \ker f'(x_0)^{\perp}$. 此时 $G'(x_0) = I_n$ 为 $\mathbb{R}^n$ 上的恒等映射. 根据反函数定理, 存在包含 x_0 的开集 U 和 $\mathbb{R}^n$ 的开集 V 使得 $G : U \to V$ 是连续可微的双射. 此外, 通过缩小 U, V, 我们可以使得 V 是凸开集.

任取 $x \in V$, 那么

$$x = G(G^{-1}(x)) = G^{-1}(x) + f'(x_0)^{\dagger} f(G^{-1}(x)) - f'(x_0)^{\dagger} f'(x_0)(G^{-1}(x)),$$

将 $f'(x_0)$ 作用于上式两边, 所以

$$P_S f(G^{-1}(x)) = f'(x_0)(x) - f'(x_0)(G^{-1}(x)) + f'(x_0)(G^{-1}(x)) = f'(x_0)(x).$$

□

参考文献

[1] Axler S. Linear Algebra Done Right. Springer Nature; 2024.