

许多实际问题往往牵涉到多方面的因素, 比如理想气体的压强与其温度和体积两个因素有关, 经济生产的产出依赖于资本、劳动、科技等等, 而大语言模型则动辄上百亿、上千亿的参数. 反映在数学上, 它们都是在处理由多个自变量决定的函数, 也就是所谓的多元函数. 本章将沿袭一元函数微分学的思路, 讨论多元函数微分学.

1.1 极限与连续

多元函数的概念

多元函数是指 (实) 自变量多于一个的函数. 如果一个函数 $f: D \rightarrow \mathbb{R}$ 的定义域 D 是 n 维欧氏空间 $\mathbb{R}^n = \{(x_1, \dots, x_n) | x_1, \dots, x_n \in \mathbb{R}\}$ 的一个子集, 则称 f 为 n 元函数. 由于欧氏空间中的点 (x_1, \dots, x_n) 通常简写为 \mathbf{x} 或 \mathbf{r} , 所以常把 $f(x_1, \dots, x_n)$ 写为 $f(\mathbf{x})$ 或 $f(\mathbf{r})$.

与一元函数一样, 作图是理解多元函数的有效方法. 一般我们有两种方式来描绘多元函数的图景, 一种是**函数图像**, 另一种是**等值集**. 比如, 给定一个二元函数 $f(x, y)$, 我们可以在空间中做出一张曲面 $z = f(x, y)$, 它就是**函数 f 的图像**. 但是, 这种做法对于三元函数效果不佳, 因为三元函数的图像需要放在四维空间中, 而这种想象力是大部分人难以企及的. 因此, 也常用等值集来描绘多元函数. 也就是说, 在函数的定义域里把取值相同的点连起来, 从而得到一幅函数取值分布图, 海拔分布图和气温分布图采用的都是这种方法. 它的优势在于只要在函数的定义域里作图, 不需要额外增加维度. 比如对于二元函数 $f(x, y)$, 给定常数 c , **等值集** $f(x, y) = c$ 一般是定义域 xy 平面里的一条曲线, 也称为**等值线**. 因为这个 c 就是函数图像上对应点的竖坐标, 所以等值线也称为**等高线**, 它是平面 $z = c$ 截函数图像 $z = f(x, y)$ 的截痕 (的投影). 当 c 取遍所有可能值时, 就得到了 xy 平面上的一族等值线, 它们完全反映了二元函数 $f(x, y)$ 的取值情况. 对于三元函数 $f(x, y, z)$, 等值集的方法也是适用的, 只不过此时的等值集是空间中的**等值面** $f(x, y, z) = c$.

我们再介绍一下**多元向量值函数**. 我们知道, 空间中的参数曲线是一元向量值函数 $\mathbf{r} = (x(t), y(t), z(t)): I \rightarrow \mathbb{R}^3$, 其中 $I \subset \mathbb{R}$ 是区间. 类似地, 空间中的参数曲面可以看作二元向量值函数 $\mathbf{r} = (x(u, v), y(u, v), z(u, v))$, 若记参数 (u, v) 的变化范围是 $D \subset \mathbb{R}^2$, 则该二元向量值函数是 $\mathbf{r}: D \rightarrow \mathbb{R}^3$. 一般地, 若 $D \subset \mathbb{R}^n$, 则映射 $f: D \rightarrow \mathbb{R}^m$ 就是一个**多元向量值函数**, 它的自变量是 n 维向量、取值是 m 维向量, 此时也可将 f 写为分量的形式

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))$$

其中 $f_1, f_2, \dots, f_m: D \rightarrow \mathbb{R}$ 都是 n 元函数, 称为 f 的**分量函数**. 常见的向量值函数有线性变换、向量场等. 所谓**线性变换**是指由一个矩阵 A 诱导的映射 $\mathcal{A}(\mathbf{x}) = A\mathbf{x}$, 它把向量 \mathbf{x} 变为向量 $A\mathbf{x}$. 所谓**向量场**是流速场、引力场这类对象的数学抽象, 一个分布在 $D \subset \mathbb{R}^n$ 上的向量场可以描述为一个 n 元向量值函数 $\mathbf{v}: D \rightarrow \mathbb{R}^n$, 它在 D 的每个点上指定一个向量.

1.1 极限与连续	1
1.2 微分与导数	9
1.3 Hesse 矩阵	17
1.4 链式法则	19
1.5 Taylor 公式	24
1.6 反函数定理与隐函数定理	28
1.7 曲线与曲面的几何	33
1.8 极值与最值	37
1.9 含参积分的正则性	42

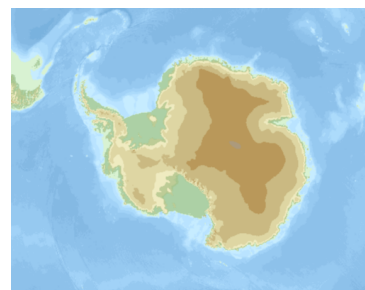


图 1.1. 南极洲及周边海域的等高线图.
https://commons.wikimedia.org/wiki/File:Topographic_map_of_antarctica.png

欧氏空间的拓扑

为了讨论多元函数的微积分, 我们需要讨论它们的定义域 \mathbb{R}^n . 习惯上, 我们用大写字母 (如 A, B) 表示 \mathbb{R}^n 中的点, 用粗体的小写字母 (如 \mathbf{a}, \mathbf{b}) 表示向量, 用不加粗的小写字母 (如 (a_1, a_2, \dots, a_n)) 表示坐标. 方便起见, 我们不区分点、向量、坐标这三个事实上并不相同的概念, 而在行文中自由地选择合适的表达形式.

设 $\mathbf{a} = (a_1, \dots, a_n)$ 和 $\mathbf{b} = (b_1, \dots, b_n)$ 是 \mathbb{R}^n 中的两个向量, 它们可以进行线性运算、内积, 也可以计算范数和距离, 即

- ▶ $\lambda \mathbf{a} = (\lambda a_1, \dots, \lambda a_n), \lambda \in \mathbb{R},$ (数乘)
- ▶ $\mathbf{a} + \mathbf{b} = (a_1 + b_1, \dots, a_n + b_n),$ (加法)
- ▶ $\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + \dots + a_n b_n,$ (内积)
- ▶ $|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{a_1^2 + \dots + a_n^2},$ (范数/模)
- ▶ $\rho(\mathbf{a}, \mathbf{b}) = |\mathbf{a} - \mathbf{b}| = \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2}.$ (距离)

利用距离, 在 \mathbb{R}^n 中也可以引入邻域概念: 称 $U(\mathbf{x}_0; \delta) = \{\mathbf{x} | \rho(\mathbf{x}, \mathbf{x}_0) < \delta\}$ 为 \mathbf{x}_0 点的 δ 邻域, 称 $U^\circ(\mathbf{x}_0; \delta) = U(\mathbf{x}_0; \delta) - \{\mathbf{x}_0\}$ 为 \mathbf{x}_0 点的 δ 去心邻域. 我们知道 $\lim_{j \rightarrow \infty} \mathbf{x}_j = \mathbf{x}$ 或者 $\mathbf{x}_j \rightarrow \mathbf{x}$ 的充要条件是 $\lim_{j \rightarrow \infty} \rho(\mathbf{x}_j, \mathbf{x}) = 0$, 这也可以用邻域的语言描述为

$$\forall \epsilon > 0, \exists N, \forall j \geq N : \mathbf{x}_j \in U(\mathbf{x}; \epsilon).$$

显然, 上述邻域的概念便于我们讨论函数在一点附近的性态. 一般地, 为了讨论多元函数的性质, 我们希望它的整个定义域有比较简单的结构, 通常是以下三种: 开集、闭集、连通集. 开集比较合适于讨论可微这种局部性质, 闭集比较合适讨论最值这种整体性质, 而连通集则比较合适研究函数大范围的变化.

内点与开集

设 E 是 \mathbb{R}^n 的一个子集, P 是 E 中的一个点, 如果存在 P 点一个邻域 $U(P; \delta)$, 使得 $U(P; \delta) \subset E$, 则称 P 是 E 的一个内点. 集合 E 的所有内点的集合记作 $\text{int}(E)$, 称为 E 的内部. 如果 $E = \text{int}(E)$, 则称 E 是开集.

直觉上, 这个“开”字大概指向集合不含它的“边界”那一层皮肤. 不难从下面边界的定义看出这一点.

边界点

设 E 是 \mathbb{R}^n 的一个子集, P 是 \mathbb{R}^n 中的一个点, 如果 P 的任何一个邻域 $U(P; \delta)$ 中既有属于 E 的点也有不属于 E 的点, 则称 P 是 E 的一个边界点. 集合 E 的所有边界点的集合记作 ∂E , 称为 E 的边界.

根据定义, E 和它的余集 E^c 显然有相同的边界.

极限点与闭集

设 E 是 \mathbb{R}^n 的一个子集, P 是 \mathbb{R}^n 中的一个点, 如果存在互不相同的点列 $P_j \in E$, 使得 $P_j \rightarrow P$, 则称 P 是 E 的一个极限点/聚点. 集合

E 的所有极限点的集合记作 E^d , 称为 E 的**导集**. 如果 $E^d \subset E$, 则称 E 是**闭集**.

从定义上看, 这个“闭”字来自于集合关于极限运算的封闭性. 事实上, 这个“闭”字也表示集合应包含那一层“边界”, 这是因为边界点如果不属于 E 那么必然是 E 的极限点. 这暗示了一个重要的事实: E 是**闭集**等价于 E^c 是**开集**. 证明留作练习.

连通集

设 E 是 \mathbb{R}^n 的一个子集, 如果对于 E 中的任意两个点 P, Q , 存在连续曲线 $r(t) : [0, 1] \rightarrow E$, 使得 $r(0) = P$ 且 $r(1) = Q$, 则称 E 是一个**道路连通集**. 如果上述连续曲线是折线段, 则称 E 是**折线连通集**; 如果上述连续曲线是直线段, 则称 E 是**凸集**.

连通的开集称为**开区域**; 开区域并上其边界称为**闭区域**, 它一定是闭集.¹ 就讨论微积分而言, 区域是多元函数比较良好的定义域.

为了将一元函数微积分中的一些重要定理推广到多元函数中, 我们还需要讨论 \mathbb{R}^n 的完备性. 前文已经提及了 \mathbb{R}^n 的 Cauchy 完备性, 此处作出重述.

1: 事实上, 对于任意的集合 E , 总有 $E \cup \partial E$ 是闭集, 此集合称为 E 的**闭包** (closure), 记作 $\text{cl}(E)$.

Cauchy 原理

设 (P_j) 是 \mathbb{R}^n 的任一点列, 则 (P_j) 收敛的充要条件是它是 Cauchy 列, 即它满足 Cauchy 条件:

$$\forall \epsilon > 0, \exists N, \forall j, k \geq N : \rho(P_j, P_k) < \epsilon.$$

除此之外, 我们还有以下几个常用的定理.

致密性定理

欧氏空间中的任意有界点列必有收敛子列.²

2: 欧氏空间中的子集 E 有**界**, 是指它的**直径有限**, 其中直径的定义是

$$\text{diam}(E) := \sup_{P, Q \in E} \rho(P, Q).$$

易见, E 有界等价于其中的向量的范数有界.

证明. 若点列有界, 则点列的所有坐标都是有界的. 根据直线上的致密性定理, 第一个坐标列有收敛子列, 在此子列中又存在子列使其第二个坐标列也收敛, 依次执行有限次, 可得一子列, 它的所有坐标列均收敛, 即它本身作为点列也收敛. \square

闭集套定理

设 (E_j) 是 \mathbb{R}^n 中的一列非空闭集满足 $E_{j+1} \subset E_j$ 和 $\text{diam}(E_j) \rightarrow 0$, 则存在 $P \in \mathbb{R}^n$, 使得 $\{P\} = \cap_j E_j$.

证明. 取 $P_j \in E_j$. 由于 $\text{diam}(E_j) \rightarrow 0$, 故 (P_j) 是 Cauchy 列, 进而可设 $P_j \rightarrow P$. 注意到点列 $(P_{j+k})_{k=0}^{\infty}$ 完全含于 E_j , 且该点列也收敛到 P , 结合 E_j 是闭集, 可知 $P \in E_j$. 故而 $P \in \cap_j E_j$. 又由于 $\text{diam}(E_j) \rightarrow 0$, 所以交集中不可能有除 P 之外的点. \square

有限覆盖定理

欧氏空间中的有界闭集是紧集, 即有界闭集的任意开覆盖必有有限子覆盖.

证明. 设 (V_j) 是有界闭集 E 的开覆盖. 设 Q_0 是覆盖 E 的一个闭方体, 将 Q_0 等分为 2^n 个小的闭方体. 如果 E 不能被有限个 V_j 覆盖, 则必然有一个与 E 相交的小方体不能被有限个 V_j 覆盖, 记该小方体为 Q_1 . 继续把 Q_1 等分为 2^n 份, 又可得到一个更小的方体 Q_2 . 依次可得一系列闭方体 (Q_j) , 它们都不能被有限覆盖. 注意到这列闭方体满足闭集套定理, 所以它们最后收缩到一个点 P . 由于每个 Q_j 中都有 E 中的点, 根据 E 的闭性可知 $P \in E$, 进而存在 J 使得 $P \in V_J$. 但是 V_J 是开集, 所以存在 P 的一个邻域 $U(P)$ 含于 V_J . 注意到 $P \in Q_j$ 且 $\text{diam}(Q_j) \rightarrow 0$, 所以当 j 充分大时, 必有 $Q_j \subset U(P) \subset V_J$, 这意味着 Q_j 可以被一个 V_J 覆盖, 矛盾. \square

重极限与累次极限

多元函数的极限也称为**重极限**, 这是讨论多元函数微分学的第一步.

重极限

设 x_0 是多元函数 f 的定义域 D 的极限点. 如果当 x 在定义域 D 内趋于 x_0 时, 函数值 $f(x)$ 趋于 L , 就称函数 f 在点 x_0 处的极限是 L , 记做

$$\lim_{\substack{x \rightarrow x_0 \\ x \in D}} f(x) = L.$$

不引起混淆的情况下, 可省略 $x \in D$, 即写为 $\lim_{x \rightarrow x_0} f(x) = L$.³

3: 对于二元函数, 重极限经常写为以下两种形式:

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x,y), \quad \lim_{\substack{x \rightarrow x_0 \\ y \rightarrow y_0}} f(x,y).$$

也可以用 $\epsilon\delta$ 语言来描述:

$$\forall \epsilon > 0, \exists \delta > 0, \forall x \in U^\circ(x_0; \delta) \cap D : |f(x) - L| < \epsilon.$$

此外, 也可仿照一元函数定义形如 $\lim_{x \rightarrow x_0} f(x) = \infty$, $\lim_{|x| \rightarrow \infty} f(x)$ 甚至

$\lim_{\substack{x \rightarrow \infty \\ y \rightarrow b}} f(x,y)$ 的广义极限.

向量值函数的重极限

设 x_0 是函数 $f : D \rightarrow \mathbb{R}^m$ 的定义域 D 的极限点. 如果存在 $L \in \mathbb{R}^m$ 使得

$$\forall \epsilon > 0, \exists \delta > 0, \forall x \in U(x_0; \delta) \cap D : |f(x) - L| < \epsilon,$$

则称函数 f 在点 x_0 处的极限是 L , 记做

$$\lim_{\substack{x \rightarrow x_0 \\ x \in D}} f(x) = L.$$

容易看到, f 有极限的充要条件是它的所有分量函数 f_1, \dots, f_m 都有极限, 并且

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = \left(\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f_1(\mathbf{x}), \dots, \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f_m(\mathbf{x}) \right).$$

因此, 下面的示例不再讨论向量值函数.

例 1.1.1 证明: $\lim_{(x,y) \rightarrow (2,1)} (x^2 + xy + y^2) = 7$.

证明. 作平移

$$\begin{aligned} x^2 + xy + y^2 - 7 &= (x - 2 + 2)^2 + (x - 2 + 2)(y - 1 + 1) + (y - 1 + 1)^2 - 7 \\ &= (x - 2)^2 + 5(x - 2) + (x - 2)(y - 1) + 4(y - 1) + (y - 1)^2. \end{aligned}$$

对任意 $\epsilon > 0$, 取 $\delta = \min\{\frac{1}{3}, \frac{\epsilon}{10}\}$, 当 $0 < \sqrt{(x-2)^2 + (y-1)^2} < \delta$ 时,

$$\begin{aligned} |x^2 + xy + y^2 - 7| &\leq \delta^2 + 5\delta + \delta^2 + 4\delta + \delta^2 \\ &= 3\delta^2 + 9\delta \leq 10\delta \leq \epsilon. \end{aligned}$$

□

例 1.1.2 设 $f(x, y) = xy \frac{x^2 - y^2}{x^2 + y^2}$, 证明: $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$.

证明. 用极坐标变换, 可得

$$\left| xy \frac{x^2 - y^2}{x^2 + y^2} \right| = |\cos \theta \sin \theta (r^2 \cos^2 \theta - r^2 \sin^2 \theta)| \leq r^2.$$

因此, 当 $x^2 + y^2 < \epsilon$ 时, 有 $|f(x, y)| \leq \epsilon$.

□

与一元函数不同, 在高维空间中点 \mathbf{x} 趋于 \mathbf{x}_0 的方式是多种多样的, 可以沿着直线趋近也可以沿着曲线趋近, 可以从右边趋近也可以从上边趋近. 所以, 函数在一点存在重极限是非常苛刻的, 它要求 \mathbf{x} 以任意方式趋向 \mathbf{x}_0 时, 函数极限都存在且相同. 基于此, 我们常常用**路径法**来判断函数极限不存在.

例 1.1.3 求极限 $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$.

解. 我们沿着直线 $y = kx$ 来求极限, k 是任意常数.

$$\lim_{\substack{(x,y) \rightarrow (0,0) \\ y=kx}} \frac{x^2 - y^2}{x^2 + y^2} = \lim_{x \rightarrow 0} \frac{x^2 - k^2 x^2}{x^2 + k^2 x^2} = \frac{1 - k^2}{1 + k^2}.$$

显然, 极限与直线的斜率 k 有关. 也就是说, 沿着不同的路径函数极限不同, 所以函数在原点的重极限不存在.

□

例 1.1.4 求极限 $\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x+y}$.

解. 仿照上例, 我们沿着直线 $y = kx$ 求极限. 因为函数的定义域是 $x + y \neq 0$, 所以这里的 $k \neq -1$. 此时

$$\lim_{\substack{(x,y) \rightarrow (0,0) \\ y=kx}} \frac{xy}{x+y} = \lim_{x \rightarrow 0} \frac{kx^2}{x+kx} = 0.$$

看似任意方向极限都是 0, 但不能由此断言函数的极限就是 0. 因为并非一定要沿着直线趋近. 比如, 当我们沿着非常靠近 $x + y = 0$ 的曲线 $x + y = x^2$ 逼近时,

$$\lim_{\substack{(x,y) \rightarrow (0,0) \\ x+y=x^2}} \frac{xy}{x+y} = \lim_{x \rightarrow 0} \frac{x(x^2-x)}{x^2} = -1.$$

它与前述极限并不相同, 因此所求极限不存在. \square

上例说明有些情况下直线也不足以说明问题. 举个更简单的例子, 令函数 $f(x, y)$ 在抛物线 $y = x^2$ 上取值为 1, 其余点处取值为 0, 则它在原点处沿任意直线的极限均为 0, 而沿着抛物线的极限是 1. 可见, 重极限存在的条件较高: 无论沿多么复杂的曲线, 极限必须都存在且相同. 一般地, 我们有以下判别法, 证明留给读者.

路径判别法

$\lim_{\substack{x \rightarrow x_0 \\ x \in D}} f(x) = L$ 的充要条件是: 对于任意满足 $\lim_{t \rightarrow 0^+} \mathbf{r}(t) = \mathbf{x}_0$ 的参数曲线 $\mathbf{r} : (0, 1] \rightarrow D - \{\mathbf{x}_0\}$, 成立 $\lim_{t \rightarrow 0^+} f(\mathbf{r}(t)) = L$.

显然, 路径法比较合适用来判断极限不存在.

例 1.1.5 判断二元函数 $f(x, y) = \frac{x^2 y^2}{x^3 + y^3}$ 在原点的重极限是否存在.

解. 沿曲线 $x^3 + y^3 = ky^n (n > 3)$ 趋于原点时有

$$\lim_{\substack{(x,y) \rightarrow 0 \\ x^3+y^3=ky^n}} f(x, y) = \lim_{y \rightarrow 0} \frac{y^2(-y^3 + ky^n)^{2/3}}{ky^n} = \lim_{y \rightarrow 0} \frac{y^4}{ky^n}.$$

因此沿不同曲线极限不同, 极限不存在. \square

关于重极限, 还有一点值得指出. 很多一元函数的极限性质与法则在重极限里也成立, 比如唯一性、保号性、四则运算、换元法、迫敛性等.

例 1.1.6 求极限 $\lim_{(x,y) \rightarrow (0,0)} \frac{\sin(xy)}{x(e^y - 1)}$.

解.

$$\begin{aligned}\lim_{(x,y) \rightarrow (0,0)} \frac{\sin(xy)}{x(e^y - 1)} &= \lim_{(x,y) \rightarrow (0,0)} \frac{\sin(xy)}{xy} \cdot \lim_{(x,y) \rightarrow (0,0)} \frac{y}{e^y - 1} \\ &= \lim_{u \rightarrow 0} \frac{\sin u}{u} \cdot \lim_{y \rightarrow 0} \frac{y}{e^y - 1} = 0.\end{aligned}$$

上面每一步的合理性留给读者自行验证. \square

除了重极限之外, 多元函数还有累次极限的概念. 下面以二元函数为例展开说明.

累次极限

称 $\lim_{x \rightarrow a} \lim_{y \rightarrow b} f(x, y)$ 称为 f 在 (a, b) 处的先 y 后 x 的累次极限. 该累次极限为 L 是指: $\forall x \in U^\circ(a), \lim_{y \rightarrow b} f(x, y) = \phi(x)$ 存在, 并且 $\lim_{x \rightarrow a} \phi(x) = L$. 类似地有先 x 后 y 的累次极限.

仅就定义而言, 累次极限与重极限并无瓜葛. 但如果它们都存在, 则的确相等.

累次极限与重极限

若 $\lim_{x \rightarrow a} \lim_{y \rightarrow b} f(x, y)$ 和 $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$ 均存在, 则它们相等. 进一步, 若 $\lim_{y \rightarrow b} \lim_{x \rightarrow a} f(x, y)$ 也存在, 则两个累次极限可以更序

$$\lim_{x \rightarrow a} \lim_{y \rightarrow b} f(x, y) = \lim_{(x,y) \rightarrow (a,b)} f(x, y) = \lim_{y \rightarrow b} \lim_{x \rightarrow a} f(x, y).$$

证明. 记重极限为 L , 则对于任意 $\epsilon > 0$, 存在 $\delta > 0$, 当 $(x, y) \in U^\circ((a, b); \delta)$ 时, 有 $|f(x, y) - L| < \epsilon$. 记 $\lim_{y \rightarrow b} f(x, y) = \phi(x)$. 于是前式中令 $y \rightarrow b$ 有 $|\phi(x) - L| \leq \epsilon$. 即当 $0 < |x - a| < \delta$ 时有 $|\phi(x) - L| \leq \epsilon$. 因此 $\lim_{x \rightarrow a} \phi(x) = L$. \square

下面展示了 5 个不符合定理条件的示例.

$f(x, y)$	$\lim_{x \rightarrow 0} \lim_{y \rightarrow 0} f(x, y)$	$\lim_{y \rightarrow 0} \lim_{x \rightarrow 0} f(x, y)$	$\lim_{(x,y) \rightarrow (0,0)} f(x, y)$
$\frac{xy}{x^2 + y^2}$	0	0	不存在
$\frac{x^2 - y^2}{x^2 + y^2}$	1	-1	不存在
$x \sin \frac{1}{y}$	不存在	0	0
$y \sin \frac{1}{x}$	0	不存在	0
$x \sin \frac{1}{y} + y \sin \frac{1}{x}$	不存在	不存在	0

多元函数的连续性

为了讨论方便, 我们直接引入向量值函数的连续性.

(向量值) 函数的连续性

设 $f: D \rightarrow \mathbb{R}^m$, 其中 $D \subset \mathbb{R}^n$. 若点 $x_0 \in D$ 满足:

$$\forall \epsilon > 0, \exists \delta > 0, \forall x \in U(x_0; \delta) \cap D: |f(x) - f(x_0)| < \epsilon,$$

则称 f 在点 x_0 处连续. 若 f 在 D 上各点处均连续, 称其在 D 上连续, 记作 $f \in C(D; \mathbb{R}^m)$.

若记 $f = (f_1, \dots, f_m)$, 易知 f 连续的充要条件是它的所有分量函数 f_1, \dots, f_m 均连续.

根据上述定义, 若 $x_0 \in D$ 是 D 的聚点, 则 f 在点 x_0 处连续等价于

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

若 $x_0 \in D$ 不是 D 的聚点, 此时称 $x_0 \in D$ 是 D 的孤立点, 显然函数在定义域的孤立点处必然连续.

与一元函数类似, 多元连续函数经过有限次四则运算和复合运算之后, 是其定义域内的连续函数. 四则运算较为简单, 不再赘述. 此处仅给出复合函数的连续性证明.

复合函数的连续性

设 $F \subset \mathbb{R}^n, G \subset \mathbb{R}^m$. 若 $f \in C(F; G), g \in C(G; \mathbb{R}^l)$, 则 $g \circ f \in C(F; \mathbb{R}^l)$.

证明. 记 $y_0 = f(x_0)$, 则对于任意 $\epsilon > 0$, 存在 $\sigma > 0$, 当 $|y - y_0| < \sigma$ 时, 成立

$$|g(y) - g(y_0)| < \epsilon.$$

对于上述 $\sigma > 0$, 存在 $\delta > 0$, 当 $|x - x_0| < \delta$ 时, 成立

$$|f(x) - f(x_0)| < \sigma.$$

进而, 此时有

$$|g(f(x)) - g(f(x_0))| < \epsilon.$$

这表明符合函数在 x_0 处连续. \square

例 1.1.7 设 $f(x, y) = \phi(x)$, 并且 ϕ 连续, 证明 f 连续.

证明. 设 $p(x, y) = x$ 是投影函数, 它显然连续. 进而 $f(x, y) = \phi(p(x, y))$ 也是连续函数. \square

上例表明, 当我们把一元连续函数看做多元函数时, 该多元函数仍然是连续的. 进而, 不同自变量的一元初等函数经过有限次的四则运算和复合运算之后得到的函数也必然连续. 比如 $f(x, y, z) = (x + z) \sin(xy)$ 就是一个三元连续函数.

现在介绍多元连续函数的整体性质. 我们知道闭区间上的连续函数具有最值性、介值性、一致连续性, 它们的确可以推广到多元连续函数上. 但请注意, 对于向量值函数而言, 它们未必有意义, 留给读者思考.

连续函数的整体性质

- ▶ **最值性.** 有界闭集上的连续函数一定有界, 并且能取到最大值和最小值.
- ▶ **介值性.** 连通集上的连续函数, 能取到任何两个函数值之间的所有值, 即它的值域是连续区间.
- ▶ **一致连续性.** 有界闭集上的连续函数一定一致连续.

证明. (1) 连续函数局部有界, 依据有限覆盖性, 可得连续函数在紧集上有界. 若上确界达不到, 则可取 (x_j) , 使得 $f(x_j) \rightarrow \sup f$. 利用致密性定理不妨假设 $x_j \rightarrow x_0$, 则 $f(x_0) = \sup f$, 矛盾.

(2) 在定义域中任取两点 x_0, x_1 , 取连续曲线 $r \in C([0, 1])$ 连接它们. 注意到 $f(r(t))$ 的连续性, 成立介质性.

(3) 反证法. 设 $(x'_j), (x''_j)$ 满足 $\rho(x'_j, x''_j) \rightarrow 0$ 和 $|f(x'_j) - f(x''_j)| \geq \epsilon_0 > 0$. 根据致密性定理, 不妨假设 $(x'_j), (x''_j)$ 均收敛至 x_0 , 这将导致 $|f(x_0) - f(x_0)| \geq \epsilon_0 > 0$. \square

1.2 微分与导数

微分与偏导数的概念

我们来回顾一下一元函数的相关内容. 设 $f(x)$ 是一元函数, 微分学的核心是研究它的增量

$$\Delta f_a(h) = f(h+a) - f(a).$$

函数 f 在 a 处可微的含义是指存在实数 L 使得

$$\Delta f_a(h) = Lh + o(h), \quad (h \rightarrow 0).$$

此处的 L 就是函数 f 在 a 点的导数 $f'(a)$. 我们把线性近似 Lh 称为函数 f 在 a 处的微分, 记作 df_a .

本节的目标是将一元函数的上述理论推广至多元函数.

为了便于理解, 我们先来研究多元函数的微分学, 向量值函数的微分学之后再讨论.

设 $f: D \rightarrow \mathbb{R}$, 其中 D 是 \mathbb{R}^n 一个开集. 我们希望讨论函数在 $a \in D$ 附近的性态. 仿照一元函数, 记自变量对点 a 的增量为 $h = \Delta x = x - a$, 进而定义 f 在 a 处的增量函数为

$$\Delta f_a(h) = f(h+a) - f(a).$$

之所以假定 D 是开集, 就是为了保证此处的 h 可以取遍所有方向. 事实上, 因为 $a \in D$ 是内点, 所以存在 $\delta > 0$, 使得对于所有 $|h| < \delta$, 总有 $a+h \in D$.

微分的核心思想是线性近似. 因此, 我们需要用一个 h 的线性函数来近似 $\Delta f_a(h)$. 注意到欧氏空间 \mathbb{R}^n 上的线性函数 ω 必然可写为如下形式

$$\omega(x) = k \cdot x = k_1 x_1 + k_2 x_2 + \cdots + k_n x_n,$$

其中 $\mathbf{k} = (k_1, k_2, \dots, k_n) \in \mathbb{R}^n$ 是个常向量, 因此下述的多元函数微分概念是非常自然的.

多元函数的微分

设 f 是定义在开集 $D \subset \mathbb{R}^n$ 上的函数, $\mathbf{a} \in D$. 若存在 $\mathbf{L} \in \mathbb{R}^n$ 使得

$$\Delta f_{\mathbf{a}}(\mathbf{h}) = \mathbf{L} \cdot \mathbf{h} + o(\mathbf{h}), \quad (\mathbf{h} \rightarrow 0)$$

则称 f 在 \mathbf{a} 点可微, 称线性函数 $\mathbf{L} \cdot \mathbf{h}$ 为 f 在 \mathbf{a} 点的微分, 记作 $df_{\mathbf{a}}$, 即

$$df_{\mathbf{a}}(\mathbf{h}) = \mathbf{L} \cdot \mathbf{h}.$$

这里需要对 $o(\mathbf{h})$ 略作说明. 设 $\alpha(\mathbf{h})$ 是一个函数, 则表达式 $\alpha(\mathbf{h}) = o(\mathbf{h})$ 的含义是

$$\lim_{\mathbf{h} \rightarrow 0} \frac{\alpha(\mathbf{h})}{|\mathbf{h}|} = 0.$$

习惯上记 $\rho = |\Delta \mathbf{x}| = |\mathbf{h}|$, 故而 $o(\mathbf{h})$ 也可写为 $o(\rho)$. 由于 $o(\rho) = o(1)$, 因此可微时必有 $\lim_{\mathbf{h} \rightarrow 0} \Delta f_{\mathbf{a}}(\mathbf{h}) = 0$, 即可微必连续.

下面我们来考察向量 $\mathbf{L} = (l_1, \dots, l_n)$ 究竟为何. 注意到

$$\Delta f_{\mathbf{a}}(\mathbf{h}) - \mathbf{L} \cdot \mathbf{h} = o(\mathbf{h}),$$

所以

$$\lim_{\mathbf{h} \rightarrow 0} \frac{\Delta f_{\mathbf{a}}(\mathbf{h}) - \mathbf{L} \cdot \mathbf{h}}{|\mathbf{h}|} = 0.$$

根据极限的路径法, 既然上述极限为 0, 那么 \mathbf{h} 沿着任意路径趋于 0 时, 上述极限应当都是 0. 记 \mathbf{e}_j 是 \mathbb{R}^n 中第 j 个单位坐标向量, 取 $\mathbf{h} = h\mathbf{e}_j$, 则应有

$$\lim_{h \rightarrow 0} \frac{\Delta f_{\mathbf{a}}(h\mathbf{e}_j) - hl_j}{|h|} = 0.$$

这等价于

$$\lim_{h \rightarrow 0} \frac{\Delta f_{\mathbf{a}}(h\mathbf{e}_j) - hl_j}{h} = 0.$$

进而, 可微时必有

$$l_j = \lim_{h \rightarrow 0} \frac{f(h\mathbf{e}_j + \mathbf{a}) - f(\mathbf{a})}{h} = \left. \frac{d}{dx_j} \right|_{x_j=a_j} f(a_1, \dots, a_{j-1}, x_j, a_{j+1}, \dots, a_n).$$

因此, l_j 是函数 f 在 \mathbf{a} 处沿着 x_j -轴方向的变化率, 称之为**偏导数**.

偏导数

设 f 是定义在开集 $D \subset \mathbb{R}^n$ 上的函数, $\mathbf{a} \in D$. 若下述极限存在

$$\lim_{h \rightarrow 0} \frac{f(h\mathbf{e}_j + \mathbf{a}) - f(\mathbf{a})}{h},$$

则称其极限为 f 在 \mathbf{a} 处对 x_j 的偏导数, 记作 $\frac{\partial f}{\partial x_j}|_{\mathbf{a}}$. 若 f 在 \mathbf{a} 处的各个偏导数都存在, 则称它在 \mathbf{a} 处可 (偏) 导.

值得指出, $\frac{\partial f}{\partial x_j}$ 的常见记号还有很多, 比如 $f_{x_j}, \partial_j f, D_j f, \nabla_j f$, 甚至 f_j .⁴

例 1.2.1 设 $f(x, y) = x^y$, 求 $f_x(2, 3)$ 和 $f_y(2, 3)$.

解. 根据定义, 有

$$\begin{aligned} f_x(2, 3) &= \left. \frac{d}{dx} \right|_{x=2} f(x, 3) = \left. \frac{d}{dx} \right|_{x=2} x^3 = 12, \\ f_y(2, 3) &= \left. \frac{d}{dy} \right|_{y=3} f(2, y) = \left. \frac{d}{dy} \right|_{y=3} 2^y = 8 \ln 2. \end{aligned}$$

□

根据前面的分析, 立得下述结论.

可微必可导

若函数 f 在 \mathbf{a} 处可微, 则其在 \mathbf{a} 处必可导, 此时还成立

$$df_{\mathbf{a}}(\mathbf{e}_j) = \left. \frac{\partial f}{\partial x_j} \right|_{\mathbf{a}}.$$

为了更清晰的表示微分, 我们再介绍一些符号.

- 称 $(\partial_1 f(\mathbf{a}), \dots, \partial_n f(\mathbf{a}))$ 为函数 f 在点 \mathbf{a} 处的**梯度**, 记作 $\left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{a}}$ 或 $\text{grad } f$, 也写为 $\partial f(\mathbf{a}), Df(\mathbf{a}), \nabla f(\mathbf{a})$. 后文再解释称其为梯度的缘由.
- 欧氏空间的坐标 x_j 也可以看作函数, 它把一个点变为它的第 j 个坐标, 称之为坐标函数. 注意到

$$(\Delta x_j)_{\mathbf{a}}(\mathbf{h}) = x_j(\mathbf{h} + \mathbf{a}) - x_j(\mathbf{a}) = h_j = \mathbf{e}_j \cdot \mathbf{h},$$

因此 x_j 处处可微, 并且 $(dx_j)_{\mathbf{a}}(\mathbf{h}) = \mathbf{e}_j \cdot \mathbf{h}$. 由于右端与 \mathbf{a} 无关, 所以我们略去下标, 记为

$$dx_j(\mathbf{h}) = \mathbf{e}_j \cdot \mathbf{h}.$$

进而, 如果 f 可微, 则

$$df_{\mathbf{a}}(\mathbf{h}) = \sum_{j=1}^n l_j h_j = \sum_{j=1}^n l_j dx_j(\mathbf{h}) = \sum_{j=1}^n \left. \frac{\partial f}{\partial x_j} \right|_{\mathbf{a}} dx_j(\mathbf{h}).$$

记 $d\mathbf{x} = (dx_1, \dots, dx_n)^T$ 为自变量的(向量值)增量函数, 即

$$d\mathbf{x}(\mathbf{h}) = (dx_1(\mathbf{h}), \dots, dx_n(\mathbf{h}))^T = \mathbf{h}.$$

那么, 有

$$df_{\mathbf{a}}(\mathbf{h}) = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{a}} \cdot d\mathbf{x}(\mathbf{h}).$$

因此, 一般函数的微分是梯度与自变量的增量的内积. 常常表示为下述略去 \mathbf{a} 和 \mathbf{h} 的形式.

4: 符号 $\frac{\partial f}{\partial x_j}$ 是一个整体记号, 不能看成 ∂f 和 ∂x_j 的商. 姑且不论 ∂f 与 ∂x_j 如何定义, 单纯从符号系统而言, 如果可以看做商, 那么

$$\partial f = \frac{\partial f}{\partial x_j} \partial x_j = \frac{\partial f}{\partial x_k} \partial x_k.$$

这明显不合理. 但是我们可以把 $\frac{\partial f}{\partial x_j}$ 看做偏微分算子 $\frac{\partial}{\partial x_j}$ 作用在函数 f 上, 所以也可以写做 $\frac{\partial}{\partial x_j} f$.

$$df = \sum_{j=1}^n \frac{\partial f}{\partial x_j} dx_j = \frac{\partial f}{\partial \mathbf{x}} \cdot d\mathbf{x}.$$

与一元函数不同, 对多元函数而言, 可导并不蕴含可微. 原因很简单, 可导仅仅意味着沿着坐标轴的方向极限存在, 并不能得到沿着任何路径极限存在. 因此反例极易构造: 设二元函数 $f(x, y)$ 在两个坐标轴上均为 0, 在其它地方均为 1. 显然, $f_x(0, 0) = f_y(0, 0) = 0$, 但由于 f 在原点不连续, 故而不可微.

虽然可导未必可微, 但如果导函数连续, 则必然可微.

连续可导必可微

设 f 在 \mathbf{a} 的某个邻域上均可导, 若其所有偏导数均在 \mathbf{a} 点连续, 则 f 在 \mathbf{a} 点可微.

证明. 为了书写方便, 不妨假设维数为 2 且 $\mathbf{a} = (0, 0)$. 此时, 我们需要讨论的是

$$\Delta f_{(0,0)}(\mathbf{h}) = f(x, y) - f(0, 0),$$

其中自变量的增量为 $\mathbf{h} = (x, y) - (0, 0) = (x, y)$.

因为偏导数只能体现函数在水平和垂直方向的变化率, 所以为了考察函数从 $(0, 0)$ 到 (x, y) 的增量, 我们借助点 $(x, 0)$ 作为桥梁来表示增量:

$$\Delta f_{(0,0)}(\mathbf{h}) = (f(x, y) - f(x, 0)) + (f(x, 0) - f(0, 0)).$$

记 $g(t) = f(x, t)$, 那么第一个括号可以看成 $g(y) - g(0)$. 根据 Lagrange 微分中值定理, 存在介于 0 和 y 之间的 η , 使得

$$f(x, y) - f(x, 0) = g(y) - g(0) = g'(\eta)y = f_y(x, \eta)y.$$

类似地, 第二个括号可以写为

$$f(x, 0) - f(0, 0) = f_x(\xi, 0)x,$$

其中 ξ 介于 0 和 x 之间. 所以

$$\Delta f_{(0,0)}(\mathbf{h}) = f_x(\xi, 0)x + f_y(x, \eta)y.$$

因为偏导数连续, 所以当 $\mathbf{h} = (x, y) \rightarrow (0, 0)$ 时, 成立

$$f_x(\xi, 0) = f_x(0, 0) + o(1), \quad f_y(x, \eta) = f_y(0, 0) + o(1).$$

进而

$$\Delta f_{(0,0)}(\mathbf{h}) = f_x(0, 0)x + f_y(0, 0)y + o(1)x + o(1)y.$$

注意到 $|x|, |y| \leq |\mathbf{h}|$, 所以 $o(1)x = o(\mathbf{h})$, $o(1)y = o(\mathbf{h})$. 从而

$$\Delta f_{(0,0)}(\mathbf{h}) = \text{grad } f(0, 0) \cdot \mathbf{h} + o(\mathbf{h}).$$

这表明函数在 $(0, 0)$ 点可微. □

注意, 偏导连续并非可微的必要条件, 比如 $f(x) = x^2 \sin(1/x)$ 在原点补充定义为 0 之后, 在原点是可微的, 但导数不连续.

下面再给出几个计算偏导数和微分的示例.

例 1.2.2 设 $f(x, y, z) = \sin yz + (x-1)z + e^{xz^2}$, 求 $f_z(1, 0, 1)$.

解. 可以用定义直接计算.

$$f_z(1, 0, 1) = \left. \frac{d}{dz} \right|_{z=1} f(1, 0, z) = \left. \frac{d}{dz} \right|_{z=1} e^{z^2} = 2ze^{z^2} \Big|_{z=1} = 2e.$$

也可以先求偏导函数, 得

$$f_z = y \cos yz + (x-1) + 2xze^{xz^2}.$$

再代入 $(1, 0, 1)$, 可得相同结果. \square

例 1.2.3 设 $f(x, y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}}, & x^2+y^2 \neq 0, \\ 0, & x^2+y^2 = 0 \end{cases}$, 求它在原点处的两个偏导数.

解. 此处需要使用定义, 我们用极限的语言来写:

$$f_x(0, 0) = \lim_{\Delta x \rightarrow 0} \frac{f(\Delta x, 0) - f(0, 0)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{0 - 0}{\Delta x} = 0.$$

同理可得 $f_y(0, 0) = 0$. \square

下面我们给出几个可微性分析的示例.

例 1.2.4 求 $f(x, y) = e^x \sin y$ 的微分.

解. 因为 $f_x = e^x \sin y$ 和 $f_y = e^x \cos y$ 都连续, 所以函数可微, 并且 $df = e^x \sin y dx + e^x \cos y dy$. \square

例 1.2.5 讨论 $f(x, y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}}, & x^2+y^2 \neq 0, \\ 0, & x^2+y^2 = 0 \end{cases}$ 在原点处的可微性.

解. 我们已经知道 $f_x(0, 0) = f_y(0, 0) = 0$. 根据可微与可导的关系, 如果函数在原点可微, 那么必有 $df_{(0,0)} = 0$, 即

$$f(x, y) - f(0, 0) = 0 \cdot dx + 0 \cdot dy + o(\rho) = o(\rho).$$

然而

$$\lim_{(x,y) \rightarrow (0,0)} \frac{f(x, y) - f(0, 0)}{\rho} = \lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2}.$$

我们知道这个极限不存在, 所以函数在原点不可微. \square

微分与偏导数的几何意义

偏导数有明显的几何意义. 以二元函数 $f(x, y)$ 为例, 用平面 $y = y_0$ 去截函数图像 $z = f(x, y)$, 会得到一条曲线 $(x, y_0, f(x, y_0))$. 这条曲线在 $x = x_0$ 处的切线关于 x 轴的斜率就是偏导数 $f_x(x_0, y_0)$. 同样, 偏导数 $f_y(x_0, y_0)$ 是平面 $x = x_0$ 截函数图像所得曲线在 $y = y_0$ 处的切线关于 y 轴的斜率.

微分的几何意义是函数图像的切平面. 仍以二元函数为例. 设函数 $f(x, y)$ 可微, 记 $z_0 = f(x_0, y_0)$. 作函数 $z = z(x, y)$ 使得

$$z - z_0 = f_x(x_0, y_0) \cdot (x - x_0) + f_y(x_0, y_0)(y - y_0).$$

由于右端正是函数 f 在 (x_0, y_0) 处的微分, 因此

$$z(x, y) - z_0 = f(x, y) - f(x_0, y_0) + o(\rho)$$

即

$$z(x, y) - f(x, y) = o(\rho), \quad (x, y) \rightarrow (x_0, y_0).$$

因此, 我们称 $z = z(x, y)$ 的图像为函数 f 的图像在 (x_0, y_0, z_0) 处的切平面是合理的.

另一方面, 在 f 的图像上取两条曲线 $(x, y_0, f(x, y_0))$ 和 $(x_0, y, f(x_0, y))$. 利用偏导数的几何意义可知两条曲线的切向量分别为 $T_x = (1, 0, f_x(x_0, y_0))$ 和 $T_y = (0, 1, f_y(x_0, y_0))$. 因此切平面的法向量可以写为 $\mathbf{n} = T_x \times T_y = (-f_x(x_0, y_0), -f_y(x_0, y_0), 1)$, 所以切平面方程是

$$-f_x(x_0, y_0) \cdot (x - x_0) - f_y(x_0, y_0)(y - y_0) + (z - z_0) = 0.$$

这再次表明前述切平面的合理性.

方向导数

偏导数反映的是函数沿坐标轴方向的变化率, 然而世界是自由的, 它并不局限于几个固定的方向, 因此我们有必要来讨论函数沿任一方向的变化率.

方向导数

函数 f 在点 \mathbf{a} 处沿单位向量 \mathbf{l} 的方向导数定义为

$$\left. \frac{\partial f}{\partial \mathbf{l}} \right|_{\mathbf{a}} = \left. \frac{d}{ds} \right|_{s=0} f(\mathbf{a} + s\mathbf{l}) = \lim_{s \rightarrow 0} \frac{f(\mathbf{a} + s\mathbf{l}) - f(\mathbf{a})}{s}.$$

显然 $\partial f / \partial \mathbf{e}_j = \partial f / \partial x_j$, 所以方向导数是一个比偏导数更一般的概念. 但有些文献把方向导数定义为 $s \rightarrow 0^+$ 时的极限, 此时, 偏导数并非方向导数的特例.

与偏导数一样, 也可以用函数图像来给出方向导数的几何意义. 过函数图像 $z = f(x, y)$ 上的点 $P(x_0, y_0, f(x_0, y_0))$ 做一张竖直的平面, 使之平行于方向 \mathbf{l} . 这张平面与函数图像的交线在点 P 处关于 \mathbf{l} 在 xy 平面上的投影轴线的斜率就是函数的方向导数.

从定义来看, 方向导数的计算并不容易. 但我们可以用函数的梯度对其进行便捷地计算.

方向导数计算公式

如果函数 f 可微, 那么 f 沿单位向量 \boldsymbol{l} 的方向导数是

$$\frac{\partial f}{\partial \boldsymbol{l}} = \nabla f \cdot \boldsymbol{l} = df(\boldsymbol{l}).$$

证明. 根据可微性, 有

$$\Delta f(\boldsymbol{h}) = \nabla f \cdot \boldsymbol{h} + o(\boldsymbol{h}).$$

因此, 当 $\boldsymbol{h} = s\boldsymbol{l}$ 时, $o(\boldsymbol{h}) = o(s)$, 进而有

$$\lim_{s \rightarrow 0} \frac{\Delta f(s\boldsymbol{l})}{s} = \lim_{s \rightarrow 0} \frac{\nabla f \cdot s\boldsymbol{l}}{s} = \nabla f \cdot \boldsymbol{l}.$$

这意味着方向导数是存在的, 并且成立需证公式. \square

此公式表明, 虽然方向导数看似比偏导数含有更多的信息量, 但对于可微函数而言, 用偏导数足矣.

记夹角 $\langle \nabla f, \boldsymbol{l} \rangle = \theta$, 那么 $\partial f / \partial \boldsymbol{l} = |\nabla f| \cos \theta$. 所以, 当 \boldsymbol{l} 与 ∇f 同向时, 方向导数达到最大值 $|\nabla f|$; 当 \boldsymbol{l} 与 ∇f 反向时, 方向导数达到最小值 $-|\nabla f|$; 当 \boldsymbol{l} 与 ∇f 正交时, 方向导数为 0.

梯度方向是函数增加最快的方向; 负梯度方向是函数减少最快的方向; 垂直于梯度方向函数变化率为零; 梯度的模反映了这种变化率的大小.

从函数图像的角度也能理解这一点. 由于图像 $z = f(x, y)$ 的法向量是 $\boldsymbol{n} = (-f_x, -f_y, 1)$, 因此函数图像的切平面与 xy 平面的夹角正切恰好为 $|\nabla f|$. 换言之, 梯度的模就是函数图像切平面的“坡度”, 这也是 ∇f 被称为梯度的原因. 因此, 梯度的模越大, 函数图像越陡, 函数值变化越快, 等高线越密集.

例 1.2.6 求函数 $f(x, y) = x \sin y$ 在点 $P(1, \frac{\pi}{3})$ 处增加最快的方向, 并求函数在该点处沿 $\boldsymbol{l} = (1, 2)$ 的方向的方向导数.

解. 函数 f 可微, 可以利用上述定理. 梯度 $\nabla f = (\sin y, x \cos y)$, 代入点 P 得 $\nabla f(P) = (\frac{\sqrt{3}}{2}, \frac{1}{2})$. 这就是函数在该点增加最快的方向. 下面计算沿 \boldsymbol{l} 的方向导数. 注意到 \boldsymbol{l} 并非单位向量, 把它单位化为 $\boldsymbol{l}_0 = \frac{1}{\sqrt{5}}(1, 2)$. 于是 $\frac{\partial f}{\partial \boldsymbol{l}_0}|_P = \nabla f(P) \cdot \boldsymbol{l}_0 = \frac{\sqrt{3}+2}{2\sqrt{5}}$. \square

习惯上, 即使 \boldsymbol{l} 不是单位向量, 仍然把 $\partial f / \partial \boldsymbol{l}_0$ 写做 $\partial f / \partial \boldsymbol{l}$.

还有一事值得一提. 由于梯度是函数增加最快的方向, 所以它给我们提供了一种最小化函数的方法, 也就是**梯度下降法**. 假设我们希望寻找函数 f 的最小值. 随意取定一点 \boldsymbol{x}_0 作为起点, 选取步长 $\eta > 0$, 执行迭代

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j - \eta \nabla f(\boldsymbol{x}_j).$$

若 η 选取恰当, 那么 $f(\boldsymbol{x}_j)$ 就能快速减小, 在较好情形下它们能收敛到函数的最小值, 实现函数最小化. 这就是梯度下降法.⁵



图 1.2. 云南哈尼梯田, 田埂就是等高线. 田埂密集处山势陡峭, 梯度大; 稀疏处山势平缓, 梯度小. 图源: 视觉中国

5: 如果将迭代方式改为

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j + \eta \nabla f(\boldsymbol{x}_j),$$

那么函数值随之增大, 称为**梯度上升法**.

向量值函数的微分

现在我们简要介绍向量值函数的微分. 方便起见, 假设所讨论的向量值函数为 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, 它的分量形式为

$$f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))^T.$$

仿照前面的可微性定义, 我们给出向量值函数的可微性.

若存在矩阵 J 使得

$$\Delta f_a(\mathbf{h}) = J\mathbf{h} + o(\mathbf{h}), \quad (\mathbf{h} \rightarrow 0)$$

则称 f 在点 a 处可微. 这意味着向量值函数 f 在 a 附近的增量几乎是一个线性变换 $J\mathbf{h}$, 该线性变换就称为函数 f 在 a 点的微分, 记作 df_a . 此处的矩阵 J 称为函数 f 在 a 点的 Jacobi 矩阵, 记为 $J_f(a)$ 或 $f'(a)$.

如果用分量来表示, 则易见函数 f 可微等价于其所有分量函数 f_1, \dots, f_m 都可微, 并且可微时有

$$df_a = \begin{pmatrix} df_1 \\ df_2 \\ \vdots \\ df_m \end{pmatrix}_a = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}_a \begin{pmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{pmatrix}.$$

因此, Jacobi 矩阵为 $J_f(a) = \left(\frac{\partial f_j}{\partial x_k} \Big|_a \right)$, 据此, 我们记

$$J_f = \left(\frac{\partial f_j}{\partial x_k} \right) =: \frac{\partial f}{\partial \mathbf{x}}.$$

我们沿用多元函数偏导数的记号, 记

$$\frac{\partial f}{\partial x_j} = f_{x_j} = (\partial_j f_1, \dots, \partial_j f_m)^T = \partial_j f = D_j f = f_j,$$

则

$$\frac{\partial f}{\partial \mathbf{x}} = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right).$$

利用 Jacobi 矩阵, 向量值函数的微分可写为

$$df = J_f d\mathbf{x} = \frac{\partial f}{\partial \mathbf{x}} d\mathbf{x}.$$

若 $m = 1$ 即 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 则其 Jacobi 矩阵就是函数的梯度 $J_f = (\partial_1 f, \dots, \partial_n f)$. 若 $f: \mathbb{R} \rightarrow \mathbb{R}$ 是一元函数, 则其 Jacobi 矩阵 $J_f = f'_x = f'(x)$ 就是一元函数的导数.

利用 Jacobi 矩阵, 可以对向量值函数作局部线性近似. 比如可微函数 $f(x, y, z) = (x^2 y + z, y^2 x z, x + z y)$ 在 $(1, 1, 1)$ 处的 Jacobi 矩阵是

$$J_x(1, 1, 1) = \begin{pmatrix} 2xy & x^2 & 1 \\ y^2 z & 2yxz & y^2 x \\ 1 & z & y \end{pmatrix}_{(1,1,1)} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

因此在 $(1, 1, 1)$ 附近

$$f(x, y, z) \approx f(1, 1, 1) + \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x-1 \\ y-1 \\ z-1 \end{pmatrix}.$$

利用线性代数的特征分解理论, 右端的变换是容易理解的, 它是三个方向上的伸缩变换.

1.3 Hesse 矩阵

如果函数的偏导数可以继续求偏导, 则可得到二阶偏导数. 一般地, 一个 n 元函数 $f(\mathbf{x}) = f(x_1, \dots, x_n)$ 的 k 阶偏导数定义为

$$\frac{\partial}{\partial x_{i_k}} \left(\dots \frac{\partial}{\partial x_{i_3}} \left(\frac{\partial}{\partial x_{i_2}} \left(\frac{\partial f}{\partial x_{i_1}} \right) \right) \right).$$

它有几种常见的简写

$$\frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}}, \quad \partial_{i_k} \dots \partial_{i_1} f, \quad D_{i_k} \dots D_{i_1} f, \quad \nabla_{i_k} \dots \nabla_{i_1} f,$$

$$\partial_{i_k \dots i_1}^k f, \quad D_{i_k \dots i_1}^k f, \quad \nabla_{i_k \dots i_1}^k f, \quad f_{x_{i_1} \dots x_{i_k}}, \quad f_{i_1 \dots i_k}.$$

若相邻的偏导相同, 常合并为一项, 比如

$$\frac{\partial^4 f}{\partial z \partial x \partial x \partial y} = \frac{\partial^4 f}{\partial z \partial x^2 \partial y} = f_{yx^2z}.$$

有时也在右上角加一定数量的撇以强调导数的阶数, 比如 f''_{xy}, f'''_{xzz} . 阶数超过一阶的偏导数统称为**高阶偏导数**.

高阶导数的复杂度非常高, 原因在于逻辑上求导的次序并不能交换, 比如 $f_{x_1 x_2 x_3}$ 和 $f_{x_3 x_2 x_1}$ 未必相等. 然而, 随意尝试一些初等函数之后, 你也许会坚信这是可以交换的! 下面我们就来讨论这一点. 显然只需要关注二阶导数即可, 更高阶导数的次序交换可以通过若干次的相邻导数交换来实现, 因此高阶导数中最重要的一个概念是 Hesse 矩阵.

Hesse 矩阵

函数 f 在 \mathbf{a} 点的所有二阶偏导数构成的方阵称为 f 在 \mathbf{a} 点的 **Hesse 矩阵**, 记作 $H_f(\mathbf{a})$ 或 $\text{Hess} f(\mathbf{a})$, 也可写为 $\frac{\partial^2 f}{\partial \mathbf{x}^2}, \partial^2 f, \nabla^2 f(\mathbf{a})$ 或 $D^2 f(\mathbf{a})$. 换言之,

$$\text{Hess} f(\mathbf{a}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \bigg|_{\mathbf{a}} \right).$$

Hesse 矩阵也可视为梯度函数 ∇f 的 Jacobi 矩阵 $J_{\nabla f}$.

二阶导数可交换与 Hesse 矩阵对称是一致的. 这种对称性被发现与函数的光滑性有关. 为此, 我们先来阐述一下多元函数的光滑性概念.

- 设 Ω 是 \mathbb{R}^n 中的开集, Ω 上具有 k 阶连续偏导数的函数构成的集合记作 $C^k(\Omega)$, 称其中的函数在 Ω 上 C^k **光滑**.

- ▶ 若 Ω 不是开集, 那么 $f \in C^k(\Omega)$ 是指: 存在一个包含 Ω 的开集 U 使得 $f \in C^k(U)$.
- ▶ 根据可微性定理, 若 $f \in C^k(\Omega)$, 则它的所有低于 k 阶的偏导函数都是可微的, 因此 C^k 光滑函数也称为 k 阶连续可微函数.

下面的定理也被称为 Clairaut-Schwarz-Jordan 定理.⁶

混合偏导交换定理

若 f 二阶连续可微, 则 f 的 Hesse 矩阵是对称阵.

证明. 只要对每一对 (x_i, x_j) 证明 $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ 即可. 因此不妨仅考虑二元函数 $f(x, y)$, 并且证明它在原点处成立 $f_{xy}(0, 0) = f_{yx}(0, 0)$.

根据定义, 知

$$\begin{aligned} f_{xy}(0, 0) &= \lim_{y \rightarrow 0} \frac{f_x(0, y) - f_x(0, 0)}{y} \\ &= \lim_{y \rightarrow 0} \frac{1}{y} \left(\lim_{x \rightarrow 0} \frac{f(x, y) - f(0, y)}{x} - \lim_{x \rightarrow 0} \frac{f(x, 0) - f(0, 0)}{x} \right) \\ &= \lim_{y \rightarrow 0} \lim_{x \rightarrow 0} \frac{f(x, y) - f(0, y) - f(x, 0) + f(0, 0)}{xy}, \end{aligned}$$

同理, 有

$$f_{yx}(0, 0) = \lim_{x \rightarrow 0} \lim_{y \rightarrow 0} \frac{f(x, y) - f(x, 0) - f(0, y) + f(0, 0)}{xy}.$$

显然要做的是交换累次极限次序. 根据交换定理, 只要证明相应的重极限存在. 事实上, 我们有

$$\begin{aligned} \frac{f(x, y) - f(x, 0) - f(0, y) + f(0, 0)}{xy} &\xrightarrow[\text{微分中值定理}]{\text{对 } x \text{ 用}} \frac{f_x(\xi, y) - f_x(\xi, 0)}{y} \\ &\xrightarrow[\text{微分中值定理}]{\text{对 } y \text{ 用}} f_{xy}(\xi, \eta). \end{aligned}$$

结合 f_{xy} 在 $(0, 0)$ 点的连续性, 可知重极限存在, 进而两个累次极限相等. \square

可以发现, 上述证明仅仅用到了 $f_{xy}(0, 0), f_{yx}(0, 0)$ 存在, 并且 f_{xy} 在 $(0, 0)$ 连续. 此外, 上述证明也给我们提供了偏导不可交换函数的构造思路: 使重极限不存在. 事实上, 取

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

易得 $f_{xy}(0, 0) = -1, f_{yx}(0, 0) = +1$.

Clairaut 定理的明显推论是, 若 $f \in C^k(\Omega)$, 则其 k 阶以内的偏导数均可交换次序. 鉴于此, 通常 (尤其是进行计算时) 我们都默认函数充分光滑!

6: Clairaut 的目标是解微分方程 $df = P(x, y)dx + Q(x, y)dy$. 在那个年代, 讨论方程的级数解 $f(x, y) = \sum_{n,m=0}^{\infty} c_{nm}x^n y^m$ 是合理的. Clairaut 观察到 $f_{xy} = f_{yx}$ 必然导致 $P_y = Q_x$, 并且指出这是方程可解的充分条件. 换言之, Clairaut 注意到级数解具有偏导可交换性. Schwarz 给出了首个严格意义的证明. 现代的通行证明归功于 Jordan.

1.4 链式法则

一元函数的微分法有四则运算法（线性性、Leibniz 性）、链式法则、反函数求导法，它们也可以推广到多元函数的情形。多元函数微分的四则运算法则较为简单，此处不做过多介绍。本节主要介绍多元函数链式法则，反函数求导法之后再提。

作为热身，我们先来考察最简单的情形——函数沿着曲线的导数。欧氏空间中的参数曲线一般写为 $\mathbf{r}(t) = (x_1(t), \dots, x_n(t))$ ，它是一个向量值函数。一个 n 元函数 f 沿着曲线 $\mathbf{r}(t)$ 的导数定义为 $\frac{d}{dt}f(\mathbf{r}(t))$ 。可以发现，偏导数、方向导数都是沿曲线导数的特例。根据我们从方向导数那里获得的经验，对可微函数而言，沿曲线的导数大概也不会比偏导数含有更多的信息量。

多元函数微分的四则运算

$$\begin{aligned} d(f \pm g) &= df \pm dg, \\ d(fg) &= gdf + fdg, \\ d\left(\frac{g}{f}\right) &= \frac{f dg - g df}{f^2}. \end{aligned}$$

链式法则 (I)/沿曲线的导数

设 n 元函数 f 在 \mathbf{x}_0 点可微，曲线 $\mathbf{r}(t) = (x_1(t), \dots, x_n(t))$ 在 t_0 点可微，且 $\mathbf{r}(t_0) = \mathbf{x}_0$ ，则复合函数 $g(t) = f \circ \mathbf{r}(t)$ 在 t_0 点可微且

$$\left. \frac{dg}{dt} \right|_{t_0} = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \left. \frac{d\mathbf{r}}{dt} \right|_{t_0} = \sum_{j=1}^n \left. \frac{\partial f}{\partial x_j} \right|_{\mathbf{x}_0} \left. \frac{dx_j}{dt} \right|_{t_0}.$$

证明. 我们需要分析

$$\begin{aligned} \Delta g_{t_0}(\Delta t) &= g(t_0 + \Delta t) - g(t_0) = f(\mathbf{r}(t_0 + \Delta t)) - f(\mathbf{r}(t_0)) \\ &= \Delta f_{\mathbf{x}_0}(\mathbf{r}(t_0 + \Delta t) - \mathbf{r}(t_0)) \\ &= \Delta f_{\mathbf{x}_0}(\Delta \mathbf{r}_{t_0}(\Delta t)). \end{aligned}$$

为了更清晰的看出结构关系，我们把 $\Delta \mathbf{r}_{t_0}(\Delta t)$ 简写为 $\Delta \mathbf{r}_{t_0}$ 。根据 f 的可微性，有

$$\Delta f_{\mathbf{x}_0}(\mathbf{h}) = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \cdot \mathbf{h} + \varepsilon(\mathbf{h})|\mathbf{h}|,$$

其中 $\lim_{\mathbf{h} \rightarrow 0} \varepsilon(\mathbf{h}) = \varepsilon(0) = 0$ 。于是

$$\Delta g_{t_0}(\Delta t) = \Delta f_{\mathbf{x}_0}(\Delta \mathbf{r}_{t_0}) = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \cdot \Delta \mathbf{r}_{t_0} + \varepsilon(\Delta \mathbf{r}_{t_0})|\Delta \mathbf{r}_{t_0}|.$$

结合 $\Delta \mathbf{r}_{t_0} \rightarrow 0 (\Delta t \rightarrow 0)$ 和 $\lim_{\mathbf{h} \rightarrow 0} \varepsilon(\mathbf{h}) = \varepsilon(0) = 0$ ，可知

$$\varepsilon(\Delta \mathbf{r}_{t_0}) = o(1), \quad (\Delta t \rightarrow 0).$$

再根据 \mathbf{r} 的可微性，我们有

$$\Delta \mathbf{r}_{t_0} = \left. \frac{d\mathbf{r}}{dt} \right|_{t_0} \Delta t + o(\Delta t).$$

故而

$$\begin{aligned}
 \Delta g_{t_0}(\Delta t) &= \frac{\partial f}{\partial \mathbf{x}} \Big|_{\mathbf{x}_0} \cdot \Delta \mathbf{r}_{t_0} + o(1) |\Delta \mathbf{r}_{t_0}| \\
 &= \frac{\partial f}{\partial \mathbf{x}} \Big|_{\mathbf{x}_0} \cdot \left(\frac{d\mathbf{r}}{dt} \Big|_{t_0} \Delta t + o(\Delta t) \right) + o(1) \left| \frac{d\mathbf{r}}{dt} \Big|_{t_0} \cdot \Delta t + o(\Delta t) \right| \\
 &= \frac{\partial f}{\partial \mathbf{x}} \Big|_{\mathbf{x}_0} \cdot \frac{d\mathbf{r}}{dt} \Big|_{t_0} \Delta t + o(\Delta t) + o(1) O(\Delta t) \\
 &= \frac{\partial f}{\partial \mathbf{x}} \Big|_{\mathbf{x}_0} \cdot \frac{d\mathbf{r}}{dt} \Big|_{t_0} \Delta t + o(\Delta t).
 \end{aligned}$$

这说明 g 在 t_0 处可微且成立链式求导公式. \square

- ▶ 如果用 $\mathbf{x}(t)$ 表示曲线, 那么可以得到更体现“链式”的公式

$$\frac{df(\mathbf{x}(t))}{dt} = \frac{\partial f}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt}.$$

但一定要注意, 此表达式中右侧两个 \mathbf{x} 的含义并不相同, 前者是 \mathbb{R}^n 的自变量, 后者是曲线的分量函数.

- ▶ **外函数可微的条件不能降为可偏导.** 事实上, f 的可偏导性只反映了 f 在坐标轴方向上的光滑性, 如果曲线 $\mathbf{r}(t)$ 根本不沿着轴向, 则必然无法获知 f 沿着此曲线的可微性.

由于偏导数本质上就是限制在某一个变量上的导数, 因此根据上述链式法则, 我们有

链式法则 (II)

如果 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 可微, 且 $\mathbf{x}: \mathbb{R}^m \rightarrow \mathbb{R}^n$ 可导, 则复合函数 $g(t) = f(\mathbf{x}(t))$ 可导, 且

$$\frac{\partial g}{\partial t} = \frac{\partial f}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial t}, \quad \frac{\partial g}{\partial t_k} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial t_k}.$$

例 1.4.1 设 $f(u, v, t) = uv + \sin t$, $z = f(t^2, \sin t, e^t)$. 求 $\frac{dz}{dt}$.

解. 此处 f 的第三个自变量和 z 的自变量的符号都是 t . 更改函数自变量的符号并不会改变函数, 为了记号不混淆, 我们把 f 的第三个自变量改为 w , 即 $f(u, v, w) = uv + \sin w$. 这样, 根据链式法则, 有

$$\begin{aligned}
 z_t &= f_u u_t + f_v v_t + f_w w_t = v \cdot 2t + u \cdot \cos t + \cos w \cdot e^t \\
 &= 2t \sin t + t^2 \cos t + e^t \cos e^t.
 \end{aligned}$$

\square

例 1.4.2 设 $z = \frac{\ln t}{u+v}$, 其中 $u = e^t, v = \sin t$. 求 $\frac{dz}{dt}$.

解. 根据陈述, z 是一元函数, 不是三元函数. 为了书写严格, 我们引

入三元函数 $f(u, v, w) = \frac{\ln w}{u+v}$, 于是 z 是复合函数 $z(t) = f(e^t, \sin t, t)$. 进而

$$z_t = f_u u_t + f_v v_t + f_w w_t = -\frac{\ln w}{(u+v)^2} e^t - \frac{\ln w}{(u+v)^2} \cos t + \frac{1}{w(u+v)}.$$

将 $u = e^t, v = \sin t, w = t$ 带入即可, 不再赘述. \square

本例比较容易出现符号混乱的问题. 比如三元函数用相同符号 $z(u, v, t) = \frac{\ln t}{u+v}$, 则链式法则会写为

$$\frac{dz}{dt} = \frac{\partial z}{\partial u} \frac{du}{dt} + \frac{\partial z}{\partial v} \frac{dv}{dt} + \frac{\partial z}{\partial t}.$$

注意此时左右两侧的 z 并不是同一个函数! 如果用下标表示导数, 上式变得更为恼人:

$$z_t = z_u u_t + z_v v_t + z_t.$$

因此, 尽量避免这一写法. 但如果不会引起混淆, 通常也并不引入新的函数名!

例 1.4.3 设 $f(x, y)$ 连续可偏导, $\phi(t) = f(t, f(t, t))$. 求 $\frac{d\phi}{dt}$.

解. 根据链式法则 $\phi_t = f_x + f_y[f(t, t)]_t = f_x + f_y[f_x + f_y] = f_x + f_x f_y + f_y^2$. \square

例 1.4.4 设 $z = uv^2$, 其中 $u = \sin(x^2 + y), v = e^{x+y}$. 求 z_x, z_y .

解. 引入函数 $f(u, v) = uv^2$, 进而 $z(x, y) = f(\sin(x^2 + y), e^{x+y})$ 就是一个复合函数, 根据链式法则, 有

$$\begin{aligned} z_x &= z_u u_x + z_v v_x = v^2 \cos(x^2 + y) \cdot 2x + 2uve^{x+y} \\ &= 2xe^{2(x+y)} \cos(x^2 + y) + 2 \sin(x^2 + y) e^{2(x+y)}, \\ z_y &= z_u u_y + z_v v_y = v^2 \cos(x^2 + y) + 2uve^{x+y} \\ &= e^{2(x+y)} \cos(x^2 + y) + 2 \sin(x^2 + y) e^{2(x+y)}. \end{aligned}$$

\square

例 1.4.5 设 $z = f(x, y)$ 其中 $x = u^2 + v^2, y = uv$. 如果 f 可微并且 $f_x(5, 2) = 3, f_y(5, 2) = -1$, 试在 $u = 1, v = 2$ 处计算 z_u 和 z_v 的值.

解. 根据链式法则, 有

$$\begin{aligned} z_u &= f_x x_u + f_y y_u = f_x \cdot 2u + f_y \cdot v, \\ z_v &= f_x x_v + f_y y_v = f_x \cdot 2v + f_y \cdot u. \end{aligned}$$

当 $u = 1, v = 2$ 时 $x = 5, y = 2$, 代入上式可得

$$\begin{aligned} z_u(1, 2) &= 3 \cdot 2 + (-1) \cdot 2 = 4, \\ z_v(1, 2) &= 3 \cdot 2 \cdot 2 + (-2) \cdot 1 = 10. \end{aligned}$$

□

例 1.4.6 设 $u = f(x, y, z)$, 其中 $y = \phi(x, t), t = \psi(x, z)$. 求 u_x, u_z .

解. 本例使用阿拉伯数字表示偏导数. 先把复合之后的函数写清楚

$$u = f(x, \phi(x, \psi(x, z)), z).$$

于是有

$$u_x = f_1 \cdot 1 + f_2 \cdot [\phi(x, \psi(x, z))]_x + f_3 \cdot 0.$$

注意, 第二项不能将 $[\phi(x, \psi(x, z))]_x$ 写为 $\phi_x(x, \psi(x, z))$, 后者是求完偏导数再复合, 而前者是复合之后的新函数求偏导数.

由于 $[\phi(x, \psi(x, z))]_x = \phi_1 + \phi_2 \psi_1$, 故而 $u_x = f_1 + f_2[\phi_1 + \phi_2 \psi_1]$. 类似可得 $u_z = f_2[\phi_2 \psi_2] + f_3$. □

例 1.4.7 设 f 是 \mathbb{R}^2 上的可微函数, 满足 $y f_x = x f_y$. 证明: f 是径向函数, 即存在一元函数 ϕ 使得 $f(x, y) = \phi(\sqrt{x^2 + y^2})$.

证明. 只要证明 $f(x, y)$ 的值仅与向量 (x, y) 的长度有关, 而与方向无关. 因此, 只要在极坐标 (r, θ) 下证明 f 与 θ 无关即可. 令 $g(r, \theta) = f(r \cos \theta, r \sin \theta)$, 则

$$g_\theta = f_x(r \cos \theta)_\theta + f_y(r \sin \theta)_\theta = -f_x y + f_y x = 0.$$

因此可定义 $\phi(r) = f(r, 0)$, 进而有 $f(x, y) = \phi(\sqrt{x^2 + y^2})$. □

例 1.4.8 设 $w = f(x + y + z, xyz)$ 其中 $f \in C^2(\mathbb{R}^2)$, 求 w_{xz} .

解. 此处未出现 f 的自变量符号, 明确起见, 我们记 $f = f(u, v)$. 这样

$$w_x = f_u u_x + f_v v_x = f_u + f_v \cdot yz.$$

此处 f_u 和 f_v 仍应视作复合函数. 进而

$$\begin{aligned} w_{xz} &= (f_u + f_v \cdot yz)_z = (f_u)_z + (f_v)_z \cdot yz + f_v \cdot (yz)_z \\ &= f_{uu} u_z + f_{uv} v_z + f_{vu} u_z \cdot yz + f_{vv} v_z \cdot yz + f_v \cdot y \\ &= f_{uu} + f_{uv} xy + f_{vu} yz + f_{vv} xy^2 z + f_v y. \end{aligned}$$

因为 f 是 C^2 函数, 根据 Clairaut 定理有 $f_{uv} = f_{vu}$, 所以也可以写做 $w_{xz} = f_{uu} + x(y+z)f_{uv} + xy^2 z f_{vv} + y f_v$.

如果用阿拉伯数字代表求导的变量名, 则也可把结果写为 $w_{xz} = f_{11} + x(y+z)f_{12} + xy^2 z f_{22} + y f_1$.⁷ □

7: 为了避免混淆, 也可写为 $f_{k,j}$, 比如 $f_{1,11}$ 是关于第 1 个和第 11 个变量的二阶混合偏导数.

例 1.4.9 设 $u(x, y) \in C^2(\mathbb{R}^2)$, 并满足拉普拉斯 (Laplace) 方程 $\Delta u = u_{xx} + u_{yy} = 0$.⁸ 试将方程转化为极坐标的形式. 如果 u 是径向函数, 也就是说存在一元函数 $f(r)$ 使得 $u = f(\sqrt{x^2 + y^2})$, 试求 u .

8: 一般地, n 维欧氏空间 \mathbb{R}^n 中的拉普拉斯算子是

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_n^2}.$$

满足拉普拉斯方程 $\Delta u = 0$ 的函数叫做调和函数.

解. 极坐标下 $x = r \cos \theta, y = r \sin \theta$. 令 $v(r, \theta) = u(r \cos \theta, r \sin \theta) = u(x, y)$. 根据链式法则, 有

$$\begin{aligned} u_x &= v_r r_x + v_\theta \theta_x, \\ u_{xx} &= v_{rr} r_x^2 + 2v_{r\theta} \theta_x r_x + v_{\theta\theta} \theta_x^2 + v_r r_{xx} + v_\theta \theta_{xx}; \\ u_y &= v_r r_y + v_\theta \theta_y, \\ u_{yy} &= v_{rr} r_y^2 + 2v_{r\theta} \theta_y r_y + v_{\theta\theta} \theta_y^2 + v_r r_{yy} + v_\theta \theta_{yy}. \end{aligned}$$

为此, 需要计算 r, θ 对 x, y 的偏导数. 因为 $r = \sqrt{x^2 + y^2}, \theta = \arctan \frac{y}{x}$, 所以

$$\begin{aligned} r_x &= \frac{x}{r}, & r_{xx} &= -\frac{y^2}{r^3}, & \theta_x &= -\frac{y}{r^2}, & \theta_{xx} &= \frac{2xy}{r^4}; \\ r_y &= \frac{y}{r}, & r_{yy} &= \frac{x^2}{r^3}, & \theta_y &= \frac{x}{r^2}, & \theta_{yy} &= -\frac{2xy}{r^4}. \end{aligned}$$

代入可得

$$u_{xx} + u_{yy} = v_{rr} + \frac{1}{r} v_r + \frac{1}{r^2} v_{\theta\theta}.$$

若 u 是径向函数, 则 $v(r, \theta) = f(r)$. 从而 $\Delta u = 0$ 等价于 $f_{rr} r + f_r = 0$. 这是可降阶方程, 解得 $f(r) = C_1 \ln r + C_2$, 进而 $u = \frac{1}{2} C_1 \ln(x^2 + y^2) + C_2$. \square

与链式法则 (I) 的证明类似, 我们可以得到向量值函数的链式法则.

链式法则 (III)/一阶微分的形式不变性

若 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 和 $g: \mathbb{R}^m \rightarrow \mathbb{R}^l$ 均可微, 则它们的复合 $g \circ f$ 也可微, 且

$$d(g \circ f) = dg \circ df, \quad \frac{\partial g \circ f}{\partial x} = \frac{\partial g}{\partial y} \frac{\partial f}{\partial x}.$$

第二个公式表明复合函数的 Jacobi 矩阵等于内外函数的 Jacobi 矩阵的乘积, 也常写为

$$J_{g \circ f} = J_g J_f.$$

现在我们来解释为什么上述链式法则被称为一阶微分的形式不变性. 事实上, 它表明换元不会改变你的微分计算! 设 f 就是自变量, 那么可视 $g = g(f)$ 为 f 的函数, 此时

$$dg(f) = J_g df.$$

然后对 f 进行换元, 也就是令 $f = f(x)$, 那么上式右端可写为 $J_g df = J_g J_f dx$. 根据链式法则, 右端正是 $J_{g \circ f} dx$. 因此,

$$dg(f) = dg(f(x)).$$

也就是说, 无论我们把 f 当作因变量还是当作自变量, 求出来的微分是一致的. 下面我们用这一观点来重新计算两个前面的例题.

例 1.4.10 设 $z = uv^2$, 其中 $u = \sin(x^2 + y), v = e^{x+y}$. 求 z_x, z_y .

解. 利用一阶微分的形式不变性, 有

$$\begin{aligned} dz &= v^2 du + 2uv dv = v^2 d \sin(x^2 + y) + 2uv de^{x+y} \\ &= v^2(2x \cos(x^2 + y) dx + \cos(x^2 + y) dy) + 2uv(e^{x+y} dx + e^{x+y} dy) \\ &= (2xv^2 \cos(x^2 + y) + 2uve^{x+y}) dx + (v^2 \cos(x^2 + y) + 2uve^{x+y}) dy. \end{aligned}$$

代入 u, v 即得 z_x 和 z_y . \square

例 1.4.11 设 $u = f(x, y, z)$, 其中 $y = \phi(x, t), t = \psi(x, z)$. 求 u_x, u_z .

解. 直接计算微分

$$\begin{aligned} du &= f_x dx + f_y dy + f_z dz = f_x dx + f_y d\phi + f_z dz \\ &= f_x dx + f_y [\phi_x dx + \phi_t dt] + f_z dz \\ &= f_x dx + f_y [\phi_x dx + \phi_t (\psi_x dx + \psi_z dz)] + f_z dz \\ &= f_x dx + f_y [\phi_x dx + \phi_t (\psi_x dx + \psi_z dz)] + f_z dz \\ &= [f_x + f_y(\phi_x + \phi_t \psi_x)] dx + [f_y \phi_t \psi_z + f_z] dz. \end{aligned}$$

可以发现直接用微分较为便捷, 不必烦恼变量之间关联. \square

1.5 Taylor 公式

Taylor 公式可以说是一元函数微分学的顶峰, 它可以近乎完美的描述光滑函数的局部形态. 本节我们来推导多元函数的 Taylor 公式. 简单起见, 我们重点推导原点处的 Taylor 公式, 也就是 Maclaurin 公式, 其它点处的 Taylor 公式完全类似.

设 n 元函数 f 在原点的某邻域 U 上足够光滑, 取定点 $\mathbf{h} \in U$. 取直线 $\mathbf{x}(t) = t\mathbf{h}$, 则 $g(t) = f(\mathbf{x}(t))$ 就是定义在 $[0, 1]$ 上的光滑函数. 根据一元函数的 Lagrange 型 Taylor 公式, 有

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + \cdots + \frac{1}{k!}g^{(k)}(0) + \frac{1}{(k+1)!}g^{(k+1)}(\tau)$$

其中 $\tau \in (0, 1)$. 为了把此式表达为 f , 我们需要计算 $g^{(n)}(t)$. 记 $\mathbf{h} = (h_1, \dots, h_n)$, 有

$$\begin{aligned} g'(t) &= \sum_{j=1}^n f_j(t\mathbf{h})h_j, \\ g''(t) &= \sum_{j=1}^n \sum_{k=1}^n f_{jk}(t\mathbf{h})h_k h_j, \\ g^{(k)}(t) &= \sum_{j_1, \dots, j_k} f_{j_1 \dots j_k}(t\mathbf{h})h_{j_1} \cdots h_{j_k}. \end{aligned}$$

随着阶数升高, 项数急剧增加. 为此我们先割一茬韭菜 (大部分情形它们已经够用):

► $k = 1$:

$$f(\mathbf{h}) = f(0) + \nabla f(\tau\mathbf{h}) \cdot \mathbf{h}.$$

► $k = 2$:

$$f(\mathbf{h}) = f(0) + \nabla f(0) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T H_f(\tau \mathbf{h}) \mathbf{h}.$$

Lagrange 中值定理

设 f 在凸区域 U 上一阶连续可微, 则对于任意 $\mathbf{x}_0, \mathbf{x}_1 \in U$, 存在 $\tau \in (0, 1)$ 使得

$$f(\mathbf{x}_0) - f(\mathbf{x}_1) = \nabla f(\mathbf{x}_\tau) \cdot (\mathbf{x}_0 - \mathbf{x}_1)$$

其中 $\mathbf{x}_\tau = (1 - \tau)\mathbf{x}_0 + \tau\mathbf{x}_1$.

二阶 Taylor 公式

设 f 在凸区域 U 上二阶连续可微, 则对于任意 $\mathbf{x}_0, \mathbf{x} \in U$, 存在 $\tau \in (0, 1)$ 使得

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \Delta \mathbf{x} + \frac{1}{2} (\Delta \mathbf{x})^T H_f(\mathbf{x}_\tau) (\Delta \mathbf{x})$$

其中 $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_0$, $\mathbf{x}_\tau = (1 - \tau)\mathbf{x}_0 + \tau\mathbf{x}$. 进而, 当 $\Delta \mathbf{x} \rightarrow 0$ 时, 有

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \Delta \mathbf{x} + \frac{1}{2} (\Delta \mathbf{x})^T H_f(\mathbf{x}_0) (\Delta \mathbf{x}) + o(|\Delta \mathbf{x}|^2).$$

证明. 只需说明 Peano 型 Taylor 公式成立即可. 事实上, 因为 $f \in C^2$, 所以 $H_f \in C^0$. 因此

$$(\Delta \mathbf{x})^T H_f(\mathbf{x}_\tau) (\Delta \mathbf{x}) = (\Delta \mathbf{x})^T H_f(\mathbf{x}_0) (\Delta \mathbf{x}) + (\Delta \mathbf{x})^T o(1) (\Delta \mathbf{x}).$$

注意, 此处的 $o(1)$ 是个矩阵. 利用矩阵乘法与模的关系, 可知

$$|(\Delta \mathbf{x})^T o(1) (\Delta \mathbf{x})| \leq |(\Delta \mathbf{x})^T| |o(1)| |\Delta \mathbf{x}| = o(1) |\Delta \mathbf{x}|^2.$$

此处最后一个 $o(1)$ 是函数. □

现在我们来表示一般阶数的 Taylor 公式. 我们需要引入一些新的记号. 对于向量 $\mathbf{h} = (h_1, \dots, h_n)$, 记

$$D_{\mathbf{h}} = \sum_{j=1}^n h_j D_j, \quad D_{\mathbf{h}} f = \sum_{j=1}^n h_j D_j f = \sum_{j=1}^n h_j f_j.$$

容易发现, 对于 $g(t) = f(t\mathbf{h})$, 成立

$$g^{(k)}(t) = [(D_{\mathbf{h}})^k f](t\mathbf{h}).$$

n 阶 Maclaurin 公式

设 $f \in C^k(U(0))$, 则对任意 $\mathbf{h} \in U(0)$, 存在 $\tau \in (0, 1)$ 使得

$$f(\mathbf{h}) = f(0) + D_{\mathbf{h}} f(0) + \dots + \frac{1}{(k-1)!} D_{\mathbf{h}}^{(k-1)} f(0) + \frac{1}{k!} D_{\mathbf{h}}^k f(\tau \mathbf{h}).$$

当 $\mathbf{h} \rightarrow 0$ 时成立

$$f(\mathbf{h}) = f(0) + D_{\mathbf{h}}f(0) + \cdots + \frac{1}{k!} D_{\mathbf{h}}^k f(0) + o(|\mathbf{h}|^k).$$

上述表达式虽然简单, 但其实是压缩包, 我们可以稍作展开. 根据多项式展开定理

$$D_{\mathbf{h}}^k = \sum_{j_1 + \cdots + j_n = k} \frac{k!}{j_1! \cdots j_n!} \cdot h_1^{j_1} \cdots h_n^{j_n} \cdot D_1^{j_1} \cdots D_n^{j_n}.$$

注意, 此处我们默认了偏导交换定理适用!

记 $\alpha = (j_1, \cdots, j_n)$, 称之为**多重指标**. 定义

$$\begin{aligned} |\alpha| &= j_1 + \cdots + j_n, \\ \alpha! &= j_1! \cdots j_n!, \\ \mathbf{h}^\alpha &= h_1^{j_1} \cdots h_n^{j_n}, \\ D^\alpha &= D_1^{j_1} \cdots D_n^{j_n}. \end{aligned}$$

那么 Maclaurin 公式可表示成

$$f(\mathbf{h}) = \sum_{j=0}^k \sum_{|\alpha|=j} \frac{D^\alpha f(0)}{\alpha!} \mathbf{h}^\alpha + \sum_{|\alpha|=k+1} \frac{D^\alpha f(\tau \mathbf{h})}{\alpha!} \mathbf{h}^\alpha.$$

注意到

$$\frac{D^\alpha f(0)}{\alpha!} \mathbf{h}^\alpha = \frac{f_{x_1^{j_1} \cdots x_n^{j_n}}(0)}{j_1! \cdots j_n!} h_1^{j_1} \cdots h_n^{j_n},$$

Taylor 公式的这一表示可以看作 f 按照变量 x_1, x_2, \cdots, x_n 逐个展开的模样.

上述各种 Taylor 公式通常用于理论计算. 具体计算时往往使用下述的唯一性定理.

唯一性

若 $\sum_{|\alpha| \leq k} c_\alpha \mathbf{h}^\alpha = o(|\mathbf{h}|^k) (\mathbf{h} \rightarrow 0)$ 其中 $c_\alpha \in \mathbb{R}$, 则 $c_\alpha = 0$.

证明. 对于任意非零向量 \mathbf{v} , 由条件知

$$\sum_{|\alpha| \leq k} c_\alpha \cdot (t\mathbf{v})^\alpha = o(|t\mathbf{v}|^k) = o(t^k), \quad t \rightarrow 0.$$

左端按 t 的幂此合并

$$\sum_{j=0}^k \left(\sum_{|\alpha|=j} c_\alpha \mathbf{v}^\alpha \right) t^j = o(t^k).$$

视上式左端为 t 的函数, 根据一元函数 Taylor 公式的唯一性, 有

$$\sum_{|\alpha|=j} c_{\alpha} \mathbf{v}^{\alpha} = 0.$$

它是 h_1, \dots, h_n 的多项式, 由 \mathbf{v} 的任意性可知 $c_{\alpha} = 0$. □

唯一性表明, 只要我们能将函数写为 Peano 型 Taylor 公式的样子, 那么它必然就是 Peano 型 Taylor 公式.

例 1.5.1 试写出 $f(x, y) = \ln(1 + x + y)$ 的三阶 Maclaurin 公式.

解. 当然可以用公式硬算各阶偏导数, 再写出 Maclaurin 公式. 我们换一个思路. 根据 $\ln(1 + t) = t - \frac{1}{2}t^2 + \frac{1}{3}t^3 + o(t^3)$ 可得

$$\begin{aligned} \ln(1 + x + y) &= (x + y) - \frac{1}{2}(x + y)^2 + \frac{1}{3}(x + y)^3 + o((x + y)^3) \\ &= (x + y) - \frac{1}{2}(x + y)^2 + \frac{1}{3}(x + y)^3 + o(\rho^3). \end{aligned}$$

显然, 这简单很多. □

例 1.5.2 设 $f(x, y) = \frac{e^x}{1-xy}$, 求 $f_{x^{100}y^{38}}(0, 0)$.

解. 利用幂级数, 当 $|xy| < 1$ 时, 有

$$\begin{aligned} f(x, y) &= \sum_{j=0}^{\infty} \frac{1}{j!} x^j \cdot \sum_{k=0}^{\infty} x^k y^k = \sum_{j,k=0}^{\infty} \frac{1}{j!} x^{j+k} y^k \\ &= \sum_{j,k=0}^{138} \frac{1}{j!} x^{j+k} y^k + o(\rho^{138}) \end{aligned}$$

由于 $f_{x^{100}y^{38}}(0, 0)$ 与 138 阶 Taylor 多项式的相应的导数一致, 故

$$f_{x^{100}y^{38}}(0, 0) = \left[\frac{1}{62!} x^{62+38} y^{38} \right]_{x^{100}y^{38}} (0, 0) = \frac{1}{62!} 100! \cdot 38!.$$

当然也可以对幂级数逐项求导. □

Taylor 公式的一个典型应用是近似计算, 下面给出一个示例.

例 1.5.3 估计 $1.08^{3.96}$.

证明. 利用 $f(x, y) = x^y$ 在 $(1, 4)$ 的 Taylor 公式, 存在 $1 < \xi < 1.08$ 和 $3.96 < \eta < 4$, 使得

$$\begin{aligned} f(1.08, 3.96) &= 1^4 + (4, 0) \cdot (0.08, -0.04) \\ &\quad + \frac{1}{2}(0.08, -0.04) \begin{pmatrix} f_{xx}(\xi, \eta) & f_{xy}(\xi, \eta) \\ f_{yx}(\xi, \eta) & f_{yy}(\xi, \eta) \end{pmatrix} \begin{pmatrix} 0.08 \\ -0.04 \end{pmatrix} \\ &= 1.32 + \text{error}. \end{aligned}$$

此处的误差 error 就是二阶项的值, 它估计比较繁琐. 粗糙地看, 可以将各个二阶导数都放大到可计算的值, 比如 $f_{xx} = y(y-1)x^{y-2}$ 在 (ξ, η) 的取值可以控制为

$$|f_{xx}(\xi, \eta)| = |\eta(\eta-1)\xi^{\eta-2}| \leq 4(4-1) \cdot 1.08^{4-2} \leq 14.$$

这样, 至少可以做到 $|\text{error}| \leq 0.4$. 事实上, $\text{error} = 0.036 \dots$. \square

最后我们简要说明向量值函数的 Taylor 公式. 一般情况下, Lagrange 型 Taylor 公式并不成立. 比如 $f(t) = (\cos t, \sin t)$, 一个预期的中值定理应是

$$f(\alpha) - f(\beta) = f'(\tau)(\alpha - \beta).$$

但当 $\alpha = 0, \beta = 2\pi$ 时, 上式不可能成立. 虽然如此, 我们仍然可以用 Jacobi 矩阵来控制函数的变化.

中值不等式

设 $f \in C^1(U; \mathbb{R}^m)$, 其中 $U \subset \mathbb{R}^n$ 是凸区域. 若 $\|J_f\| \leq M$, 则

$$|f(x_0) - f(x_1)| \leq M|x_0 - x_1|, \quad x_0, x_1 \in U.$$

证明. 令

$$g(x) = (f(x_0) - f(x_1)) \cdot (f(x) - f(x_1)).$$

对函数 g 使用 Lagrange 中值定理, 得

$$\begin{aligned} |f(x_0) - f(x_1)|^2 &= g(x_0) = g(x_0) - g(x_1) \\ &= \nabla g(x_\tau) \cdot (x_0 - x_1) \\ &= (f(x_0) - f(x_1))^T J_f(x_\tau)(x_0 - x_1) \\ &\leq |f(x_0) - f(x_1)| \cdot M \cdot |x_0 - x_1|. \end{aligned}$$

\square

1.6 反函数定理与隐函数定理

一元函数微分学里有一个重要的反函数定理: 如果 $f' \neq 0$, 则 f 存在反函数 g , 且 $g'(f(x)) = 1/f'(x)$. 本节的目标是把它推广到向量值函数. 显然, 所考虑到向量值函数的值域和定义域的维数应当一致, 否则很难想象它有光滑性很好的反函数.

让我们从线性变换建立直觉. 设线性变换 $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \mathcal{A}(x) = Ax$, 其中 A 是 n 阶方阵. 我们知道 \mathcal{A} 可逆的充要条件是矩阵 A 可逆 (即 $\det(A) \neq 0$). 对于一个光滑性较好的向量值函数 $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, 它有局部线性化

$$f(x) - f(a) \approx J_f(a)(x - a).$$

换言之, 在 a 附近, f 几乎就是一个表示矩阵为 $J_f(a)$ 的线性变换. 所以我们很自然的可以猜测: 如果 $J_f(a)$ 可逆, 则 f 在 a 附近可逆. 这就是高维的反函数定理.

反函数定理

设 $U \subset \mathbb{R}^n$ 是开集, $f \in C^1(U; \mathbb{R}^n)$. 若在 $a \in U$ 处成立 $\det J_f(a) \neq 0$, 则存在含 a 的开集 V 和含 $f(a)$ 的开集 W , 使得 $f|_V : V \rightarrow W$ 有连续可微的反函数 $g : W \rightarrow V$, 且 $J_g(f(x)) = [J_f(x)]^{-1}$.

证明. 通过平移并乘以 $J_f(a)^{-1}$, 我们不妨假设 $a = 0, f(0) = 0, J_f(0) = I_n$. 下面逐步证明定理.

- 首先说明 f 在原点附近是单射. 记 $\delta(x) = f(x) - x$, 则 $J_\delta(0) = 0$. 根据函数的连续可微性, 存在 $r > 0$, 当 $|x| \leq r$ 时, 成立 $\|J_\delta(x)\| < 1/2$. 因此, 当 $|x'|, |x''| \leq r$ 时成立

$$|\delta(x') - \delta(x'')| \leq \frac{1}{2}|x' - x''|.$$

进而

$$\frac{1}{2}|x' - x''| \leq |f(x') - f(x'')| \leq \frac{3}{2}|x' - x''|.$$

此式称为 f 的**双 Lipschitz 性**. 若记 B_ρ 为半径为 ρ 的球心在原点的开球, 则 $f|_{B_r} : B_r \rightarrow B_{3r/2}$ 是个单射.

- 其次说明 $0 \in f(B_r)$ 是 $f(B_r)$ 的内点. 我们需要寻找一个半径 $s \in (0, 3r/2)$, 使得对任一 $y_* \in B_s$, 方程 $f(x) = y_*$ 在 B_r 上有解. 不动点法是求解这类方程的经典方法. 注意到函数

$$\phi(x) = y_* - f(x) + x$$

的不动点就是方程的解, 因此考虑迭代点列 $x_{n+1} = \phi(x_n), x_0 = 0$. 注意到在 B_r 内成立 $\|J_\phi(x)\| = \|J_\delta(x)\| < 1/2$, 所以此时有

$$|\phi(x)| \leq |y_*| + \frac{1}{2}|x|.$$

取 $s = r/3$, 则 $\phi(B_r) \subset B_{5r/6}$, 进而整个迭代数列均落在 $B_{5r/6}$ 内, 于是

$$|x_{n+1} - x_n| = |\phi(x_n) - \phi(x_{n-1})| \leq \frac{1}{2}|x_n - x_{n-1}|.$$

所以 (x_n) 是压缩点列, 必收敛于不动点 x_* . 显然 $|x_*| \leq 5r/6 < r$.

记 $W = B_{r/3}, V = f^{-1}(W) \cap B_r$, 则 $f|_V : V \rightarrow W$ 是双射. 不难证明 W 是开集 (留作练习).

- 最后说明反函数 $g : W \rightarrow V$ 的可微性. 任取 $y_0 \in W$, 记 $x_0 = g(y_0) \in V$. 根据 f 的可微性, 有

$$f(x) - f(x_0) = J_f(x_0)(x - x_0) + o(x - x_0), \quad (x \rightarrow x_0).$$

代入 $x = g(y), x_0 = g(y_0)$, 得

$$y - y_0 = J_f(x_0)(g(y) - g(y_0)) + o(g(y) - g(y_0)), \quad (g(y) - g(y_0) \rightarrow 0).$$

注意到 f 的双 Lipschitz 性意味着 g 的双 Lipschitz 性. 从而, $g(y) - g(y_0) \rightarrow 0$ 等价于 $y - y_0 \rightarrow 0$, 且 $o(g(y) - g(y_0)) = o(y - y_0)$, 因此

$$y - y_0 = J_f(x_0)(g(y) - g(y_0)) + o(y - y_0), \quad (y - y_0 \rightarrow 0).$$

故而

$$g(y) - g(y_0) = J_f(x_0)^{-1}(y - y_0) + o(y - y_0), \quad (y - y_0 \rightarrow 0).$$

这表明 g 在 y_0 可微, 且 $J_g(y_0) = J_f(x_0)^{-1}$. 由于逆矩阵是矩阵元素的有理运算, 因此导数的连续性是显然的.

□

反函数定理是坐标变换的理论依据. 比如 (x, y) 是一个平面坐标系, 又 C^1 函数 $(x, y) = (x(u, v), y(u, v))$ 满足

$$\det \begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix} \neq 0,$$

那么局部而言, (u, v) 和 (x, y) 是一一对应, 换言之 (u, v) 可以当作 xy -平面上的局部 (曲线) 坐标系, 并且反函数 $(u, v) = (u(x, y), v(x, y))$ 满足

$$\begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} = \begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix}^{-1}.$$

这可以借助平面直角坐标与极坐标的变换关系来理解.

下面我们讨论隐函数, 为了方便理解, 我们详细介绍二维情形.

隐函数定理 (I)

设 $F(x, y)$ 在 $P_0(x_0, y_0)$ 的某个邻域内具有连续偏导数, 且 $F(x_0, y_0) = 0$, $F_y(x_0, y_0) \neq 0$. 那么在 P_0 点的某个邻域上, 方程 $F(x, y) = 0$ 能确定一个可微的隐函数 $f(x)$ 使得

$$F(x, f(x)) = 0, \quad f(x_0) = y_0; \quad \frac{df}{dx} = -\frac{F_x(x, f(x))}{F_y(x, f(x))}.$$

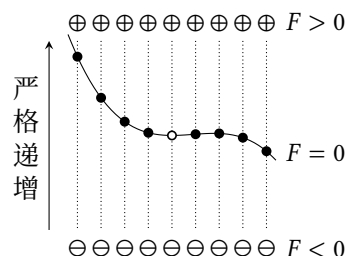


图 1.3. 当 $F_y(x_0, y_0) > 0$ 时, 函数 $F(x, y)$ 在 (x_0, y_0) 附近的取值情况. 可以看出, 中间存在一条零值线, 它就是函数 $f(x)$ 的图像.

证明. 如图所示, 条件 $F_y \neq 0$ 保证了函数在 (x_0, y_0) 附近沿着 y 轴方向的严格单调性, 进而保证了在附近每条平行于 y 轴的线段上存在唯一的零点, 零点的连线就是隐函数的图像. 因此隐函数的存在性是简单的, 但为了便于推广到高维, 我们用反函数定理来证明这一点.

直觉上, 如果 $F(x, y) = 0$ 是一条光滑曲线, 则对接近 0 的常数 c , $F(x, y) = c$ 也应该是光滑曲线. 进而向量值函数

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

应当是一个很好的“压平”映射. 由于它把 $F(x, y) = 0$ 压成了 x 轴, 所以如果 F 是一个双射, 则 x 轴的原像就是所求隐函数的图像.

注意到 $\det(J_F)(x_0, y_0) = F_y(x_0, y_0) \neq 0$, 由反函数定理可知, F 是从 (x_0, y_0) 附近到 $(x_0, 0)$ 附近到一个双射. 若记反函数为 G , 则它必可写为

$$G(x, y) = (x, G(x, y)).$$

于是 $F(x, G(x, 0)) = 0$ 即 $f(x) = G(x, 0)$ 就是所求反函数. 由于 G 连续可微, 故而 f 连续可微.

对 $F(x, f(x)) = 0$ 两边求导, 利用链式法则可得

$$F_x(x, f(x)) + F_y(x, f(x)) \cdot f'(x) = 0.$$

由于 $F_y \neq 0$, 即得 $f'(x) = -\frac{F_x(x, f(x))}{F_y(x, f(x))}$. □

如果 F 和 f 有更高阶的可微性, 那么

$$\begin{aligned} 0 &= [F(x, f(x))]_{xx} \\ &= [F_x(x, f(x))]_x + [F_y(x, f(x)) \cdot f'(x)]_x \\ &= [F_x(x, f(x))]_x + [F_y(x, f(x))]_x \cdot f'(x) + F_{yy}(x, f(x)) \cdot f''(x) \\ &= F_{xx} + F_{xy}f' + [F_{yx} + F_{yy}f'] \cdot f' + F_y f''. \end{aligned}$$

将 $f' = -F_x/F_y$ 代入有隐函数的二阶导数公式

$$\frac{d^2 f}{dx^2} = -\frac{F_{xx}F_y^2 - 2F_{xy}F_xF_y + F_{yy}F_x^2}{F_y^3}.$$

例 1.6.1 计算椭圆曲线 $y^2 = x^3 + 17$ 在点 $(2, 5)$ 处的切线和曲率.

解. 记 $F(x, y) = y^2 - x^3 - 17$. 因为 $F_y(2, 5) = 2y|_{y=5} = 10 \neq 0$, 根据隐函数定理, 方程 $F(x, y) = 0$ 在点 $(2, 5)$ 附近可以确定隐函数 $y = f(x)$ 且

$$f'(x) = -\frac{F_x}{F_y} = -\frac{-3x^2}{2y} = \frac{3x^2}{2f(x)}.$$

所以 $f'(2) = \frac{12}{10} = \frac{6}{5}$, 进而切线方程是 $y - 5 = \frac{6}{5}(x - 2)$.

为了计算曲率还要求 f'' . 直接计算可得 $F_x(2, 5) = -12, F_y(2, 5) = 10, F_{xx}(2, 5) = -12, F_{xy} = 0, F_{yy} = 2$. 代入隐函数的二阶导数公式有 $f''(2) = \frac{114}{125}$. 下面换一种方法, 根据 $f'(x) = 3x^2/(2f(x))$, 得

$$f''(x) = \frac{3x}{f(x)} - \frac{3x^2 f'(x)}{2f^2(x)}.$$

代入 $f'(2) = \frac{6}{5}$ 可得

$$f''(2) = \frac{3 \cdot 2}{5} - \frac{3 \cdot 2^2 \cdot \frac{6}{5}}{2 \cdot 5^2} = \frac{114}{125}.$$

于是, 曲率为

$$K = \frac{|f''(2)|}{(1 + [f'(2)]^2)^{3/2}} = \frac{114}{61\sqrt{61}}.$$

□

通常, 在具体计算时, 我们会通过逐步对 $F(x, f(x)) = 0$ 求导得到 f 的各阶导数, 而并不会直接应用它的导数公式.

隐函数定理 (I) 可以直接推广到高维, 只要把证明中的 F, x, y 换成向量 $F, \mathbf{x}, \mathbf{y}$ 即可.

隐函数定理 (II)

设 $F \in C^1(\mathbb{R}^n \times \mathbb{R}^N; \mathbb{R}^N)$ 满足 $F(\mathbf{x}_0, \mathbf{y}_0) = 0$ 和 $\det(\partial F / \partial \mathbf{y})(\mathbf{x}_0, \mathbf{y}) \neq 0$,

则存在 C^1 的局部隐函数 $y = f(x)$, 使得

$$F(x, f(x)) = 0, \quad f(x_0) = y_0; \quad \frac{\partial f}{\partial x} = -\left(\frac{\partial F}{\partial y}\right)^{-1} \frac{\partial F}{\partial x}.$$

上述定理也称为隐函数组定理. 事实上, 如果写成分量的形式, 那么它表明方程组

$$\begin{cases} F_1(x_1, \dots, x_n, y_1, \dots, y_N) = 0 \\ F_2(x_1, \dots, x_n, y_1, \dots, y_N) = 0 \\ \dots \\ F_N(x_1, \dots, x_n, y_1, \dots, y_N) = 0 \end{cases}$$

在条件

$$\frac{\partial(F_1, \dots, F_N)}{\partial(y_1, \dots, y_N)} = \det\left(\frac{\partial F_i}{\partial y_k}\right) \neq 0$$

下可以确定局部隐函数

$$\begin{cases} y_1 = f_1(x_1, \dots, x_n), \\ y_2 = f_2(x_1, \dots, x_n), \\ \dots \\ y_N = f_N(x_1, \dots, x_n). \end{cases}$$

这些隐函数的导数通常也通过直接计算获得. 比如 $F(x, y, z) = 0$ 确定了隐函数 $z = f(x, y)$, 那么对 $F(x, y, f(x, y)) = 0$ 求导数可得

$$0 = [F(x, y, f(x, y))]_x = F_x \cdot (x)_x + F_y \cdot (y)_x + F_z \cdot (f)_x = F_x + F_z f_x,$$

$$0 = [F(x, y, f(x, y))]_y = F_x \cdot (x)_y + F_y \cdot (y)_y + F_z \cdot (f)_y = F_y + F_z f_y.$$

从而可得二元隐函数 $f(x, y)$ 的求导公式

$$\frac{\partial f}{\partial x} = -\frac{F_x}{F_z}, \quad \frac{\partial f}{\partial y} = -\frac{F_y}{F_z}.$$

也可以用微分的方式来推导. 根据一阶微分形式的不变性, 二元函数 $F(x, y, f(x, y))$ 的全微分是

$$dF(x, y, f(x, y)) = F_x dx + F_y dy + F_z df.$$

因为 $F(x, y, f(x, y)) = 0$, 所以

$$F_x dx + F_y dy + F_z df = 0, \quad df = -\frac{F_x}{F_z} dx - \frac{F_y}{F_z} dy.$$

需要注意的是, 三元函数 $F(x, y, z)$ 的全微分并不是零!

例 1.6.2 设 $2z = x^2 + y^2 + z^2$ 确定隐函数 $z = z(x, y)$, 求 z_{xx} .

解. 对方程两边连续求两次导数, 可得

$$[2z]_x = [x^2 + y^2 + z^2]_x,$$

$$2z_x = 2x + 2zz_x;$$

$$[z_x]_x = [x + zz_x]_x,$$

$$z_{xx} = 1 + z_x z_x + z z_{xx}.$$

所以

$$z_x = \frac{x}{1-z}, \quad z_{xx} = \frac{1+z_x^2}{1-z} = \frac{(1-z)^2 + x^2}{(1-z)^3}.$$

□

例 1.6.3 设 $F(u, v)$ 可微, 可微隐函数 $z = z(x, y)$ 由方程 $F(xe^z, ye^z) = 0$ 确定. 求 $xz_x + yz_y$.

解. 我们用微分的方法来做.

$$\begin{aligned} 0 &= d[F(xe^z, ye^z)] = F_u d(xe^z) + F_v d(ye^z) \\ &= e^z F_u dx + xe^z F_u dz + e^z F_v dy + ye^z F_v dz. \end{aligned}$$

由此可得

$$dz = -\frac{F_u}{xF_u + yF_v} dx - \frac{F_v}{xF_u + yF_v} dy.$$

因此 $xz_x + yz_y = -1$.

□

例 1.6.4 判断方程 $xu - yv = 0$ 和 $yu + xv = 1$ 在何处能确定隐函数 $u = u(x, y)$ 和 $v = v(x, y)$, 并求它们的偏导数.

解. 记 $F = xu - yv, G = yu + xv - 1$. 有

$$\frac{\partial(F, G)}{\partial(u, v)} = \begin{vmatrix} F_u & F_v \\ G_u & G_v \end{vmatrix} = \begin{vmatrix} x & -y \\ y & x \end{vmatrix} = x^2 + y^2.$$

所以, 在 $x^2 + y^2 \neq 0$ 的任何一个点的某个邻域上均能确定隐函数 $u = u(x, y)$ 和 $v = v(x, y)$. 对方程关于 x 求偏导, 可得

$$\begin{cases} u + xu_x - yv_x = 0, \\ yu_x + v + xv_x = 0. \end{cases}$$

解得

$$u_x = -\frac{xu + yv}{x^2 + y^2}, \quad v_x = -\frac{xv - yu}{x^2 + y^2}.$$

同理可得 $u_y = -(yu + xv)/(x^2 + y^2), v_y = -(xu + yv)/(x^2 + y^2)$.

由于隐函数定理只是一个充分性定理, 其中的条件并非必要, 所以我们需要另行分析 $x^2 + y^2 = 0$ 附近能否确定隐函数. 注意, 此时 $xu - yv = 0$ 自动成立, 而 $yu + xv = 1$ 必然不成立, 故而必然不存在隐函数. □

1.7 曲线与曲面的几何

本节我们用多元函数微分法来研究空间对象的几何性质, 主要讨论曲面的切平面和法线、曲线的切线和法平面. 我们以参数方程和一般方程这两种方式进行介绍.

参数曲线和曲面的几何

设 $\mathbf{r}(t)$ 是空间光滑曲线, 仿照平面曲线切线的定义, 我们可以把空间曲线的切线定义为割线的极限. 注意到

$$\frac{\Delta \mathbf{r}}{\Delta t} = \frac{\mathbf{r}(t_0 + \Delta t) - \mathbf{r}(t_0)}{\Delta t} = \left(\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t}, \frac{\Delta z}{\Delta t} \right)$$

就是曲线在 $t = t_0$ 处某条割线的方向向量. 所以, 当 $\Delta t \rightarrow 0$ 时, 它的极限就是曲线在 $t = t_0$ 处的切向量. 显然, 这个极限就是 \mathbf{r} 的导数 $\mathbf{r}'(t_0)$.

正则曲线的切向量

若光滑曲线 $\mathbf{r} = (\alpha(t), \beta(t), \gamma(t))$ 满足 $\mathbf{r}' \neq 0$, 则称其为**正则曲线**. 正则曲线的**切向量** $\boldsymbol{\tau} = \mathbf{r}'$, 曲线在 $t = t_0$ 点处的**切线方程**是

$$\frac{x - \alpha(t_0)}{\alpha'(t_0)} = \frac{y - \beta(t_0)}{\beta'(t_0)} = \frac{z - \gamma(t_0)}{\gamma'(t_0)}.$$

该点处与切线垂直的平面叫做曲线的**法平面**, 它的方程是

$$\alpha'(t_0) \cdot (x - \alpha(t_0)) + \beta'(t_0) \cdot (y - \beta(t_0)) + \gamma'(t_0) \cdot (z - \gamma(t_0)) = 0.$$

例 1.7.1 求曲线 $\mathbf{r} = (t, t^2, t^3)$ 在点 $(1, 1, 1)$ 处的法平面方程.

解. 点 $(1, 1, 1)$ 对应的参数 $t = 1$. 曲线的切向量为 $\mathbf{r}' = (1, 2t, 3t^2)$, 代入即得切向为 $(1, 2, 3)$, 进而法平面为 $(x - 1) + 2(y - 1) + 3(z - 1) = 0$. \square

下面我们来讨论参数曲面. 设 $\mathbf{r} = (\alpha(u, v), \beta(u, v), \gamma(u, v))$ 是一张光滑曲面, 它在 $\mathbf{r}(u_0, v_0)$ 点处的**法向量**是垂直于曲面上过这一点的所有曲线的非零向量. 因此, u 线 $\mathbf{r}(u, v_0)$ 在 $u = u_0$ 处的切向量 $\mathbf{r}_u(u_0, v_0)$ 垂直于法向量, 同样 v 线的切向量 $\mathbf{r}_v(u_0, v_0)$ 也应当垂直于法向量. 所以曲面的法向量可以取为 $\mathbf{n} = \mathbf{r}_u \times \mathbf{r}_v$.

正则曲面的法向量

若光滑曲面 $\mathbf{r} = (\alpha(u, v), \beta(u, v), \gamma(u, v))$ 满足 $\mathbf{r}_u \times \mathbf{r}_v \neq 0$, 则称其为**正则曲面**. 正则曲面在任一点处的**法向量**是

$$\mathbf{n} = \mathbf{r}_u \times \mathbf{r}_v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \alpha_u & \beta_u & \gamma_u \\ \alpha_v & \beta_v & \gamma_v \end{vmatrix}.$$

该点处沿着法向量的直线称为曲面的**法线**, 该点处与法线垂直的平面叫做曲面的**切平面**. 利用法向量容易写出它们的方程, 只是形式复杂, 没有罗列的必要.

例 1.7.2 求螺旋面 $\mathbf{r}(u, v) = (v \cos u, v \sin u, u)$ 的法线与 z 轴的夹角.

解. 法向量为

$$\begin{aligned}\mathbf{n} &= \mathbf{r}_u \times \mathbf{r}_v = (-v \sin u, v \cos u, 1) \times (\cos u, \sin u, 0) \\ &= (-\sin u, \cos u, -v).\end{aligned}$$

所以夹角 θ 满足

$$\cos \theta = \frac{|\mathbf{n} \cdot \mathbf{k}|}{|\mathbf{n}| \cdot |\mathbf{k}|} = \frac{|v|}{\sqrt{1+v^2}},$$

即 $\theta = \arctan 1/|v|$. □

非参数曲线和曲面的几何

我们先来讨论平面曲线的情形. 假设曲线的一般方程是 $F(x, y) = 0$, 其中 $F \in C^1$. 如果 $\nabla F \neq 0$, 则由隐函数定理可知曲线局部是函数图像, 不妨设隐函数为 $y = f(x)$. 根据求导公式, 有 $f'(x) = -F_x/F_y$, 即曲线的切向量可取为 $\boldsymbol{\tau} = (F_y, -F_x)$, 进而曲线的法向量可取 $\mathbf{n} = (F_x, F_y) = \nabla F$.

如果 $\nabla F(x, y) \neq 0$, 那么它就是曲线 $F(x, y) = 0$ 的法向量. 进而, 曲面在 $P_0(x_0, y_0)$ 处的切线方程为

$$F_x(P_0) \cdot (x - x_0) + F_y(P_0) \cdot (y - y_0) = 0.$$

该点法线的方程是

$$\frac{x - x_0}{F_x(P_0)} = \frac{y - y_0}{F_y(P_0)}.$$

下面讨论空间曲面的情形. 假设曲面的一般方程是 $F(x, y, z) = 0$, 其中 $F \in C^1$ 且 $\nabla F \neq 0$. 我们用一种新的方法来求曲面在点 $P_0(x_0, y_0, z_0)$ 处的法向量. 在曲面上任取一条过 P_0 的曲线 $\mathbf{r}(t) = (\alpha(t), \beta(t), \gamma(t))$ 使得 $\mathbf{r}(0) = (x_0, y_0, z_0)$. 因为曲线落在曲面上, 所以恒有

$$F(\alpha(t), \beta(t), \gamma(t)) = 0.$$

两端在 $t = 0$ 处求导, 可得

$$F_x(P_0)\alpha'(0) + F_y(P_0)\beta'(0) + F_z(P_0)\gamma'(0) = 0.$$

所以, F 的梯度向量 $\nabla F(P_0)$ 垂直于曲线的切向量 $\mathbf{r}'(0)$. 注意到曲线 $\mathbf{r}(t)$ 是任意取的, 而 $\nabla F(P_0)$ 是固定的, 因此当 F 的梯度非零时, 它就是曲面 $F(x, y, z) = 0$ 的法向量. 事实上, ∇F 也是 F 的等值面 $F(x, y, z) = c$ 的法向量.

如果 $\nabla F \neq 0$, 那么它就是曲面 $F(x, y, z) = 0$ 的法向量. 进而, 曲面在 $P_0(x_0, y_0, z_0)$ 处的切平面方程为

$$F_x(P_0) \cdot (x - x_0) + F_y(P_0) \cdot (y - y_0) + F_z(P_0) \cdot (z - z_0) = 0.$$

该点法线的方程是

$$\frac{x - x_0}{F_x(P_0)} = \frac{y - y_0}{F_y(P_0)} = \frac{z - z_0}{F_z(P_0)}.$$

利用空间曲面的法向量, 我们可以得到空间曲线的切向量. 假设空间曲线的一般方程为

$$\begin{cases} F(x, y, z) = 0, \\ G(x, y, z) = 0. \end{cases}$$

显然, 曲线在 P_0 处的切向量, 既垂直于曲面 $F(x, y, z) = 0$ 的法向量 $\nabla F(P_0)$, 也垂直于曲面 $G(x, y, z) = 0$ 的法向量 $\nabla G(P_0)$. 所以曲线的切向量可以取做 $\nabla F \times \nabla G$.

如果 $\nabla F \times \nabla G \neq 0$,⁹ 那么曲面 $F(x, y, z) = 0$ 与曲面 $G(x, y, z) = 0$ 的交线的切向量为

$$\boldsymbol{\tau} = \nabla F \times \nabla G = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ F_x & F_y & F_z \\ G_x & G_y & G_z \end{vmatrix}.$$

9: 此时, 曲线可以用坐标局部参数化. 比如, 当 $\frac{\partial(F, G)}{\partial(x, y)} \neq 0$ 时, 根据隐函数组定理, 可以从曲线方程

$$\begin{cases} F(x, y, z) = 0 \\ G(x, y, z) = 0 \end{cases}$$

解出 $x = x(z), y = y(z)$, 那么曲线可以用竖坐标参数化为 $\mathbf{r} = (x(z), y(z), z)$.

进一步可得到曲线的切线和法平面方程, 不再赘述.

例 1.7.3 求曲面 $x^2 + y^2 + z^2 = 6$ 和 $x + y + z = 0$ 的交线在点 $(1, -2, 1)$ 处的切线方程.

解. 两个函数的梯度为 $\nabla(x^2 + y^2 + z^2 - 6) = (2x, 2y, 2z)$ 及 $\nabla(x + y + z) = (1, 1, 1)$. 所以曲面的法向量可以取为 $\mathbf{n}_1 = (x, y, z), \mathbf{n}_2 = (1, 1, 1)$. 代入 $(x, y, z) = (1, -2, 1)$ 有 $\mathbf{n}_1 \times \mathbf{n}_2 = (1, -2, 1) \times (1, 1, 1) = (-3, 0, 3)$. 曲线的切向量可以取为 $\boldsymbol{\tau} = (-1, 0, 1)$, 进而切线方程为 $\frac{x-1}{-1} = \frac{y+2}{0} = \frac{z-1}{1}$. \square

例 1.7.4 如果曲面 $x^2 + y^2 + \frac{z^2}{a^2} = 1$ 和 $x^2 + y^2 = z^2$ 正交, 即在交线上两者的法向量互相垂直, 试求 a .

解. 两个曲面的法向量分别为 $\mathbf{n}_1 = (2x, 2y, \frac{2z}{a^2}), \mathbf{n}_2 = (2x, 2y, -2z)$. 它们垂直等价于在交线上有 $\mathbf{n}_1 \cdot \mathbf{n}_2 = 4(x^2 + y^2 - \frac{z^2}{a^2}) = 0$. 代入 $x^2 + y^2 = z^2$ 可得 $(1 - \frac{1}{a^2})z^2 = 0$. 交线处 $z \neq 0$, 所以 $a = \pm 1$. \square

函数的图像

函数 $z = f(x, y)$ 的图像既可以看做参数曲面 $\mathbf{r} = (x, y, f(x, y))$, 也可以看做一般方程 $F(x, y, z) = z - f(x, y) = 0$ 的图形. 从参数曲面角度, 它的法向量可以取为

$$\mathbf{n} = \mathbf{r}_x \times \mathbf{r}_y = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} = (-f_x, -f_y, 1). \quad (7.1)$$

从一般方程的角度, 它的法向是

$$\mathbf{n} = \nabla F = (-f_x, -f_y, -1). \quad (7.2)$$

两者相同. 注意到此法向量的第三个坐标是正的, 所以这个法向量是指向上方的. 利用这个法向量, 图形在 (x_0, y_0, z_0) 处的切平面为

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0). \quad (7.3)$$

容易看出右端就是函数 f 在 (x_0, y_0) 点的全微分, 所以切平面是函数图像的线性近似.

最后我们指出, 用函数图像理解曲面具有一般性. 至少, 曲面的每个局部我们可以用函数图像来描述它.

参数曲面 如果曲面由参数方程给出 $\mathbf{r}(u, v)$, 当它是正则参数曲面时, 有 $\mathbf{r}_u \times \mathbf{r}_v \neq 0$. 以 $\frac{\partial(\alpha, \beta)}{\partial(u, v)} \neq 0$ 为例. 根据反函数定理, $x = \alpha(u, v), y = \beta(u, v)$ 可以局部反解出 $u = u(x, y), v = v(x, y)$. 进而曲面竖坐标满足 $z = \gamma(u(x, y), v(x, y))$, 所以它是函数 $f(x, y) = \gamma(u(x, y), v(x, y))$ 的图像.

非参数曲面 如果曲面由方程 $F(x, y, z) = 0$ 确定, 并且 $\nabla F \neq 0$. 不妨设 $F_z(x, y, z) \neq 0$, 那么根据隐函数定理, 方程 $F(x, y, z) = 0$ 局部地可以确定隐函数 $z = f(x, y)$. 换言之, 曲面 $F(x, y, z) = 0$ 局部地可以看作函数 $z = f(x, y)$ 的图像.

1.8 极值与最值

本节讨论多元函数的最值问题. 与一元函数不同, 对于多元函数而言, 除了通常的极值问题之外, 还有一种有约束的条件极值问题. 我们先讨论无约束极值问题, 再介绍有约束的条件极值问题.

无约束极值

如果函数 f 在 \mathbf{x}_0 的某个邻域上成立 $f(\mathbf{x}) \geq f(\mathbf{x}_0)$, 则称 \mathbf{x}_0 为 f 的**极小值点**, 函数值 $f(\mathbf{x}_0)$ 称为极小值. 如果在去心邻域上成立严格不等式 $f(\mathbf{x}) > f(\mathbf{x}_0)$, 则称 \mathbf{x}_0 为 f 的**严格极小值点**. 把不等号反向, 就可以得到极大值点、极大值、严格极大值点的定义. 如果在 f 的整个定义域上恒有 $f(\mathbf{x}) \geq f(\mathbf{x}_0)$, 则称 \mathbf{x}_0 为 f 的**最小值点**, 称 $f(\mathbf{x}_0)$ 为 f 的最小值. 最大值的定义是类似的.

Femart 引理

如果函数 f 在 \mathbf{x}_0 点处取得极值, 并且 f 在该点可偏导, 那么 $\nabla f(\mathbf{x}_0) = 0$.

证明. 显然 $g_j(t) = f(\mathbf{x}_0 + t\mathbf{e}_j)$ 在 $t = 0$ 处取得极值, 根据一元函数的 Femart 引理, 可知 $g'_j(0) = 0$, 即 $f_j(\mathbf{x}_0) = 0$. \square

梯度为零的点称为函数的**驻点 (稳定点、临界点)**. 因此上述定理可表述为“可微的极值点必是驻点”. 反之, 驻点不一定是极值点. 比如 $(0, 0)$ 是函数 $f(x, y) = x^2 - y^2$ 的驻点, 但沿 x 轴 $f(x, 0) = x^2 > 0 (x \neq 0)$, 沿 y 轴 $f(0, y) = -y^2 < 0 (y \neq 0)$, 所以 $(0, 0)$ 不是极值点. 像这种一个方向函数增大、一个方向函数减小的点一般叫做**鞍点**. 下面的定理给出了判定驻点是否是极值点的充分条件.

矩阵正定性

设矩阵 A 是实对称阵, A **正定**是指对任意非零向量 \mathbf{h} 总有 $\mathbf{h}^T A \mathbf{h} > 0$. 若总有 $\mathbf{h}^T A \mathbf{h} \geq 0$, 则称 A **半正定**. 类似可知 A **负定**和**半负定**的概念. 如果一个对称阵既不半正定也不半负定, 则称其为**不定**矩阵.

极值点的二阶导数判别法

设函数 f 在 x_0 的某个邻域上二阶连续可微, 并且 $\nabla f(x_0) = 0$. 记

- ▶ 若 $H_f(x_0)$ 正定, 那么 x_0 是 f 的严格极小值点.
- ▶ 若 $H_f(x_0)$ 负定, 那么 x_0 是 f 的严格极大值点.
- ▶ 若 $H_f(x_0)$ 不定, 则称 x_0 为鞍点, 它必然不是极值点.

证明概要. 不妨假设 $x_0 = 0$. 利用 Taylor 公式, 有

$$f(x) - f(0) = \frac{1}{2} x^T H_f(0) x + o(|x|^2).$$

因此影响右端正负的就是 $x^T H_f(0) x$. 根据矩阵特征值理论易知. \square

例 1.8.1 求函数 $f(x, y) = x^3 - y^3 + 3x^2 + 3y^2 - 9x$ 的极值.

解. 函数显然可微, 它的极值点一定是驻点. 所以, 我们可以在方程组

$$\begin{cases} f_x = 3x^2 + 6x - 9 = 0 \\ f_y = -3y^2 + 6y = 0 \end{cases}$$

的解集里搜索极值点. 先求出所有驻点 $A(1, 0), B(1, 2), C(-3, 0), D(-3, 2)$. 然后我们利用行列式 H 来判断各个驻点是否是极值点.

$$f_{xx} = 6(1+x), \quad H = f_{xx}f_{yy} - [f_{xy}]^2 = 36(1+x)(1-y).$$

所以

- ▶ $A(1, 0)$. 此时 $f_{xx}(A) = 12 > 0, H(A) = 72 > 0$, 所以 A 是极小值点.
- ▶ $B(1, 2)$. 此时 $f_{xx}(B) = 12 > 0, H(B) = -72 < 0$, 所以 B 不是极值点.
- ▶ $C(-3, 0)$. 此时 $f_{xx}(C) = -12 < 0, H(C) = -72 < 0$, 所以 C 不是极值点.
- ▶ $D(-3, 2)$. 此时 $f_{xx}(D) = -12 < 0, H(D) = 72 > 0$, 所以 D 是极大值点.

综上, 函数在 $A(1, 0)$ 处取得极小值 $f(1, 0) = -5$, 在 $D(-3, 2)$ 处取得极大值 $f(-3, 2) = 31$. \square

注意, 函数在极值点处未必可微. 比如 $f(x, y) = \sqrt{x^2 + y^2}$ 在原点处取得极小值, 但函数在原点不可微. 所以, 搜索极值点时, 除了驻点之外, 还要考虑不可微点.

极值点搜索指南

- ▶ 列出函数所有的驻点和不可微点. 驻点由方程 $\nabla f = 0$ 确定.
- ▶ 用 H_f 的正定、负定性判断驻点是否为极大值点或极小值点.
- ▶ 用定义判断不可微点是否为极大值点或极小值点.

进一步, 可以利用函数的极值来求函数的最值. 如果函数 f 在有界闭区域 D 上连续, 那么它在 D 上必能取到最值. 这种最值点可能在 D 的内部, 也可能在 D 的边界上. 如果在 D 的内部, 那么该最值点必是函数的极值点. 故而, 函数的最值点只要在极值点和定义域的边界上寻找.

特征值判别法

设 A 是 n 阶实对称阵, 方程 $\det[A - \lambda I_n] = 0$ 的根称为 A 的特征值.

- (1) A 正 (负) 定等价于 A 的特征值都是正 (负) 数;
- (2) A 半正 (负) 定等价于 A 的特征值都是非负 (正) 数;
- (3) A 不定等价于 A 既有正特征值也有负特征值.

Sylvester 判别法

设 $A = (a_{ij})_{i,j=1}^n$ 是 n 阶实对称阵, 则 A 正定的充要条件是

$$\det[(a_{ij})_{i,j=1}^k] > 0, \quad k = 1, 2, \dots, n.$$

最值点搜索指南

- ▶ 判断函数是否存在最值. 如果存在, 则执行以下操作.
- ▶ 求出函数所有的极值点.
- ▶ 求出函数在定义域边界上的最值.
- ▶ 比较所有极值和边界上的最值, 最大的就是最大值, 最小的就是最小值, 相应的点就是最值点.

求出边界上的最值是比较困难. 对于二元函数 $f(x, y)$ 而言, 定义域 D 的边界由曲线构成, 不妨设它的方程是 $\varphi(x, y) = 0$. 那么, 边界上的最值问题其实是在条件 $\varphi(x, y) = 0$ 下求函数 $f(x, y)$ 的最值的问题, 这就是条件极值问题, 我们会在下一目中讨论.

例 1.8.2 有一边长为 24 的正方形铁板, 把它两边折起来做成断面为等腰梯形 (可以是等腰三角形或矩形) 的水槽. 问怎样折法才能使断面的面积最大?

解. 设折起来的边长为 x , 倾角为 θ . 那么等腰梯形的下底和上底分别为 $24 - 2x$ 和 $24 - 2x + 2x \cos \theta$, 高为 $x \sin \theta$. 因此梯形断面面积为

$$\begin{aligned} A &= \frac{1}{2} [(24 - 2x) + (24 - 2x + 2x \cos \theta)] \cdot x \sin \theta \\ &= (24x - 2x^2) \sin \theta + x^2 \sin \theta \cos \theta. \end{aligned}$$

显然只要在 $x \in [0, 12], \theta \in [0, \frac{\pi}{2}]$ 的范围内求最大值. 此时, 这是一个有界闭集, 连续函数 A 必然能取到最值. 显然 $x = 0$ 处和 $\theta = 0$ 处不可能取到最大值. 注意到 $A_\theta(x, \frac{\pi}{2}) = -x^2 < 0$, 所以最大值也不会出现在 $\theta = \frac{\pi}{2}$ 上. 再注意到 $A_x(12, \theta) = -24 \sin \theta (1 - \cos \theta) < 0$, 所以 $x = 12$ 上也不会有最大值点. 因此, 最大值点不在区域的边界上, 肯定在区域内部, 可以应用 Fermat 定理. 它的两个偏导数为

$$\begin{aligned} A_x &= (24 - 4x) \sin \theta + 2x \sin \theta \cos \theta, \\ A_\theta &= (24x - 2x^2) \cos \theta + x^2 (\cos^2 \theta - \sin^2 \theta). \end{aligned}$$

在 $x \in (0, 12), \theta \in (0, \frac{\pi}{2})$ 内解驻点方程 $A_x = A_\theta = 0$. 可以发现

$$0 = \cot \theta \cdot A_x - \frac{1}{x} \cdot A_\theta = (1 - 2 \cos \theta) \cdot x.$$

因此 $\cos \theta = \frac{1}{2}$, 代入 $A_x = 0$ 得 $x = 8$. 所以面积函数只有唯一的驻点 $x = 8, \theta = \frac{\pi}{3}$, 它必然是最大值点. 所以当折起长度为 8、倾角为 $\frac{\pi}{3}$ 时, 断面面积最大. \square

条件极值

前面讲的是无条件极值, 但是实际问题常常有约束条件. 比如, 在固定投资下如何将利润最大化, 为使火箭达到预定速度最少需要多少燃料, 在资源和生产力有限的情况下如何提高社会福利, 这些都是条件极值问题. 再比如, 求椭圆 $(x - 2)^2 + 2(y - 2)^2 = 1$ 上距离原点最近的点, 这也是条件极值问题, 它可以表述为:

在条件 $(x - 2)^2 + 2(y - 2)^2 = 1$ 下, 求函数 $x^2 + y^2$ 的最小值.

一般地, 二元函数的条件极值问题可以表述为:

在条件 $\varphi(x, y) = 0$ 下, 求目标函数 $f(x, y)$ 的极值 (最值) .

如果所求是目标函数的最小值, 那么通常写为

$$\min f(x, y) \quad \text{s.t.} \quad \varphi(x, y) = 0.$$

假如可以从条件 $\varphi(x, y) = 0$ 显式地解出 $y = y(x)$, 那么上述条件极值问题就等价于无条件极值问题

$$\min f(x, y(x)).$$

但在很多情形下, 显式求解并不简单, 甚至不可能. 我们另有一种直接寻求条件极值的方法, 可以不必把问题化为无条件极值, 这就是下面要介绍的 **Lagrange 乘数法**.

假定 f 和 φ 都是可微函数, 并且 f 在点 $P_0(x_0, y_0)$ 处取得条件极值. 如果方程 $\varphi(x, y) = 0$ 在 P_0 附近能确定可微的隐函数 $y = y(x)$, 那么根据费马定理, 有

$$\frac{d}{dx} f(x, y(x))|_{x=x_0} = 0.$$

应用链式求导法, 得

$$f_x(x_0, y_0) + f_y(x_0, y_0)y'(x_0) = 0.$$

另一方面, 因为 $y = y(x)$ 是 $\varphi(x, y) = 0$ 的解, 所以

$$\varphi_x(x_0, y_0) + \varphi_y(x_0, y_0)y'(x_0) = 0.$$

所以

$$f_x(P_0)\varphi_y(P_0) = f_y(P_0)\varphi_x(P_0).$$

如果 $\nabla\varphi \neq 0$, 那么一定存在常数 λ 使得

$$\nabla f(P_0) = \lambda \nabla\varphi(P_0). \quad (8.1)$$

因为极值点 P_0 处还成立 $\varphi(P_0) = 0$, 所以事实上它要满足方程组

$$\begin{cases} f_x(x_0, y_0) - \lambda\varphi_x(x_0, y_0) = 0, \\ f_y(x_0, y_0) - \lambda\varphi_y(x_0, y_0) = 0, \\ \varphi(x_0, y_0) = 0. \end{cases} \quad (8.2)$$

引入辅助函数 $L(x, y, \lambda) = f(x, y) - \lambda\varphi(x, y)$, 方程组 (8.2) 可写做

$$\nabla L(x_0, y_0, \lambda) = 0. \quad (8.3)$$

如果能解出 (x_0, y_0, λ) , 那么就可以确定极值点 $P_0(x_0, y_0)$. 通过方程组 (8.3) 求解条件极值的方法就叫做拉格朗日乘数法.

拉格朗日乘数法

设 $f(x, y)$ 和 $\varphi(x, y)$ 均为可微函数, 并且 $\nabla\varphi \neq 0$. 那么可以通过以下

步骤求解 $f(x, y)$ 在条件 $\varphi(x, y) = 0$ 下的极值.

1. 设拉格朗日函数 $L(x, y, \lambda) = f(x, y) - \lambda\varphi(x, y)$.
2. 求出三元函数 L 的驻点, 即解方程 $\nabla L = 0$.
3. 上一步求出的所有 (x, y) , 包含了可能的条件极值点. 从中找出真正的极值点或者最值点.

至于第 3 步如何确定所求得的点是否为极值点, 在实际问题中往往可根据问题本身的性质来判定.

一般地, 如果 $f(x_1, \dots, x_n)$ 和 $\varphi(x_1, \dots, x_n)$ 是 n 元函数, 那么拉格朗日乘数法仍然适用, 只要将拉格朗日函数取做

$$L(x_1, \dots, x_n, \lambda) = f(x_1, \dots, x_n) - \lambda\varphi(x_1, \dots, x_n).$$

此时方程组 $\nabla L = 0$ 有 $n+1$ 个方程. 拉格朗日乘数法还能处理多条件极值问题. 比如, 为求解条件极值

$$\min f(x_1, \dots, x_n) \quad \text{s.t.} \quad \varphi(x_1, \dots, x_n) = \psi(x_1, \dots, x_n) = 0,$$

只要设拉格朗日函数为

$$L(x_1, \dots, x_n, \lambda, \mu) = f(x_1, \dots, x_n) - \lambda\varphi(x_1, \dots, x_n) - \mu\psi(x_1, \dots, x_n),$$

其它步骤不变.

例 1.8.3 试求椭圆 $(x-2)^2 + 2(y-2)^2 = 1$ 上距离原点最近的点.

解. 设 $f(x, y) = x^2 + y^2$, $\varphi(x, y) = (x-2)^2 + 2(y-2)^2 - 1$, 则问题为

$$\min f(x, y) \quad \text{s.t.} \quad \varphi(x, y) = 0.$$

作 Lagrange 函数

$$L = (x^2 + y^2) - \lambda[(x-2)^2 + 2(y-2)^2 - 1].$$

Lagrange 函数的驻点方程是

$$\begin{cases} L_x = 2x - 2\lambda(x-2) = 0, \\ L_y = 2y - 4\lambda(y-2) = 0, \\ L_\lambda = -[(x-2)^2 + 2(y-2)^2 - 1] = 0. \end{cases}$$

从前两个方程可得 $y = \frac{4x}{x+2}$, 代入第三个方程可得 $x^4 - x^2 - 36x + 44 = 0$. 可求得近似解 $x_1 = 1.2461, x_2 = 2.8645$. 进而 $y_1 = 1.5355, y_2 = 2.3554$. 它们一个是最近点, 另一个是最远点. 所以最近点是 $(1.2461, 1.5355)$. \square

例 1.8.4 求两条曲线 $x^2 - y^2 - 1 = 0$ 和 $x^2 + y^2 - 8y + 15 = 0$ 之间的最近距离.

解. 目标函数为 $F = (x-u)^2 + (y-v)^2$, 约束条件为 $x^2 - y^2 - 1 = 0$ 和 $u^2 + v^2 - 8v + 15 = 0$.

因此, 作 Lagrange 函数

$$L = (x-u)^2 + (y-v)^2 - s(x^2 - y^2 - 1) - t(u^2 + v^2 - 8v + 15).$$

临界点方程为

$$\begin{cases} L_x = 2(x-u) - 2sx = 0, \\ L_y = 2(y-v) + 2sy = 0, \\ L_u = 2(u-x) - 2tu = 0, \\ L_v = 2(v-y) - 2tv + 8t = 0, \\ L_s = -(x^2 - y^2 - 1) = 0, \\ L_t = -(u^2 + v^2 - 8v + 15) = 0. \end{cases}$$

可以解得当 $(x, y) = (\pm\sqrt{5}, 2)$, $(u, v) = (\pm\frac{\sqrt{5}}{3}, \frac{10}{3})$ 时, 取得最近距离 2. \square

1.9 含参积分的正则性

所谓**含参积分**就是多元函数关于其中一个变量的积分, 其余变量就是该积分的参量. 比如椭圆积分

$$\int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 \theta} d\theta$$

就是一个含参积分, k 就是它的参量. 一般地, 形如

$$\int_c^d f(x_1, \dots, x_n, y) dy, \quad \int_c^\infty f(x_1, \dots, x_n, y) dy$$

的表达式称为**含参积分**, 可以是正常积分, 也可以是反常积分. 本节主要以二元函数 $f(x, y)$ 为例讨论含参积分的正则性, 包括连续性、可微性.

含参正常积分

设 $Q = [a, b] \times [c, d]$, $f: Q \rightarrow \mathbb{R}$. 如果 $f(x, y)$ 关于变量 y 可积, 则可定义

$$F(x) = \int_c^d f(x, y) dy.$$

更一般地, 如果有函数 $\phi, \psi: [a, b] \rightarrow [c, d]$, 则可讨论更为复杂的含参积分

$$G(x) = \int_{\phi(x)}^{\psi(x)} f(x, y) dy.$$

为了讨论 F 和 G 的正则性, 我们先来讨论三元函数

$$\Phi(u, v, x) = \int_u^v f(x, y) dy \quad (u, v, x \in [a, b])$$

的正则性.

含参积分的正则性 (I)

- ▶ 若 $f \in C(Q; \mathbb{R})$, 则 $\Phi \in C([a, b]^3; \mathbb{R})$.
- ▶ 若 $f \in C^1(Q; \mathbb{R})$, 则 $\Phi \in C^1([a, b]^3; \mathbb{R})$, 且

$$\Phi_u = -f(x, u), \quad \Phi_v = f(x, v), \quad \Phi_x = \int_u^v f_x(x, y) dy.$$

证明. (1) 我们需要估计

$$\begin{aligned}\Phi(u, v, x) - \Phi(u_0, v_0, x_0) \\ = [\Phi(u, v, x) - \Phi(u, v, x_0)] + [\Phi(u, v, x_0) - \Phi(u_0, v_0, x_0)].\end{aligned}$$

注意到 f 在 Q 上一致连续, 所以 $f(x, y) - f(x_0, y) = o(1)(x \rightarrow x_0)$ 关于 y 一致成立, 故而

$$\Phi(u, v, x) - \Phi(u, v, x_0) = \int_u^v [f(x, y) - f(x_0, y)] dy = o(1), \quad (x \rightarrow x_0).$$

另一方面,

$$\begin{aligned}\Phi(u, v, x_0) - \Phi(u_0, v_0, x_0) &= \int_{v_0}^v f(x_0, y) dy - \int_{u_0}^u f(x_0, y) dy \\ &= O(v - v_0) + O(u - u_0) = o(1), \quad (v \rightarrow v_0, u \rightarrow u_0).\end{aligned}$$

因此, 当 $x \rightarrow x_0, v \rightarrow v_0, u \rightarrow u_0$ 时, $\Phi(u, v, x) \rightarrow \Phi(u_0, v_0, x_0)$. 故 G 连续.

(2) 根据微积分基本定理, 显然有 $\Phi_u = -f(x, u), \Phi_v = f(x, v)$. 如果我们能证明 $\Phi_x = \int_u^v f_x(x, y) dy$, 则由 (1) 可知 Φ_x 连续, 从而 $\Phi \in C^1$. 下面计算 Φ_x . 注意到 f_x 的一致连续性, 有

$$\begin{aligned}\frac{\Phi(u, v, x) - \Phi(u, v, x_0)}{x - x_0} &= \int_u^v \frac{f(x, y) - f(x_0, y)}{x - x_0} dy = \int_u^v f_x(\xi, y) dy \\ &= \int_u^v f_x(x_0, y) dy + \int_u^v [f_x(\xi, y) - f_x(x_0, y)] dy \\ &= \int_u^v f_x(x_0, y) dy + o(1), \quad (x \rightarrow x_0).\end{aligned}$$

这意味着 Φ_x 存在且 $\Phi_x(u, v, x_0) = \int_u^v f_x(x_0, y) dy$. □

根据 Φ 的正则性, 立即可得 F 和 G 的正则性.

含参积分的正则性 (II)

► 若 $f \in C(Q; \mathbb{R})$, 则 $F, G \in C([a, b]; \mathbb{R})$, 且

$$\lim_{x \rightarrow x_0} \int_c^d f(x, y) dy = \int_c^d \lim_{x \rightarrow x_0} f(x, y) dy.$$

► 若 $f \in C^1(Q; \mathbb{R}), \phi, \psi \in C^1([a, b]; \mathbb{R})$, 则 $F, G \in C^1([a, b]; \mathbb{R})$, 且有 **Leibniz 积分求导公式**

$$\begin{aligned}\frac{d}{dx} \int_c^d f(x, y) dy &= \int_c^d f_x(x, y) dy, \\ \frac{d}{dx} \int_{\phi(x)}^{\psi(x)} f(x, y) dy &= \int_{\phi(x)}^{\psi(x)} f_x(x, y) dy \\ &\quad + \psi'(x) f(x, \psi(x)) - \phi'(x) f(x, \phi(x)).\end{aligned}$$

证明. 只要注意到 $F(x) = \Phi(c, d, x), G(x) = \Phi(\phi(x), \psi(x), x)$. □

例 1.9.1 求 $I(x) = \int_x^{x^2} \sin(xe^y)dy$ 的导数.

解. 此时 $\phi(x) = x, \psi(x) = x^2, f(x, y) = \sin(xe^y)$, 它们都连续可微, 因此适用莱布尼茨公式. 所以

$$\begin{aligned} I'(x) &= \int_x^{x^2} \cos(xe^y)e^y dy + 2x \sin(xe^{x^2}) - \sin(xe^x) \\ &= \frac{1}{x} \left(\sin(xe^{x^2}) - \sin(xe^x) \right) + 2x \sin(xe^{x^2}) - \sin(xe^x). \end{aligned}$$

需要注意的是, $x = 0$ 是上式的可去间断点, 然而第一个等式对 $x = 0$ 仍然适用, 所以 $I'(0) = 0$. 事实上, 最后一个式子的间断性是由积分法产生的, 原积分并不具有间断性. 这种情况下通常不必单独列出可去间断点处的函数值. \square

例 1.9.2 计算 $I = \int_0^1 \frac{\ln(1+x)}{1+x^2} dx$.

解. 作函数 $I(t) = \int_0^1 \frac{\ln(1+tx)}{1+x^2} dx$, 则 $I(1) = I$. 注意到 $I(0) = 0$, 所以

$$I = I(1) - I(0) = \int_0^1 I'(t) dt.$$

利用含参积分求导公式, 有

$$\begin{aligned} I'(t) &= \frac{d}{dt} \int_0^1 \frac{\ln(1+tx)}{1+x^2} dx = \int_0^1 \frac{\partial}{\partial t} \left(\frac{\ln(1+tx)}{1+x^2} \right) dx \\ &= \int_0^1 \frac{x}{(1+tx)(1+x^2)} dx = \frac{1}{1+t^2} \int_0^1 \left(\frac{-t}{1+tx} + \frac{x}{1+x^2} + \frac{t}{1+x^2} \right) dx \\ &= \frac{1}{1+t^2} \left(-\ln(1+t) + \frac{1}{2} \ln 2 + \frac{1}{4} \pi t \right). \end{aligned}$$

进而

$$I = \int_0^1 I'(t) dt = - \int_0^1 \frac{\ln(1+t)}{1+t^2} dt + \frac{\ln 2}{2} \cdot \frac{\pi}{4} + \frac{\pi}{4} \cdot \frac{\ln 2}{2} = -I + \frac{\pi \ln 2}{4}.$$

故而 $I = \pi \ln 2 / 8$. \square

上例利用 Leibniz 积分求导公式, 通过积分号内求微分的方法求得了定积分. 物理学家 Feynman 早年便习得此法, 并以此在大学中大杀四方, 以至于此方法以 Feynman 积分技巧之名流传天下.

含参反常积分

此处以无穷区间反常积分为例. 设 $Q = [a, b] \times [c, +\infty)$, $f(x, y) : Q \rightarrow \mathbb{R}$ 关于 y 可积, 则可定义

$$F(x) = \int_c^\infty f(x, y) dy.$$

为了讨论 F 的正则性, 我们需要引入含参反常积分一致收敛的概念. 引入一族 x 的函数

$$F_t(x) = \int_c^t f(x, y) dy.$$

若

$$\lim_{t \rightarrow +\infty} \|F_t - F\| = 0,$$

则称 $\int_c^\infty f(x, y) dy$ 一致收敛.

含参积分的正则性 (III)

- ▶ 若 $f \in C(Q; \mathbb{R})$, 且 $\int_c^\infty f(x, y) dy$ 一致收敛, 则 $F \in C([a, b]; \mathbb{R})$, 且

$$\lim_{x \rightarrow x_0} \int_c^\infty f(x, y) dy = \int_c^\infty \lim_{x \rightarrow x_0} f(x, y) dy.$$

- ▶ 若 $f \in C^1(Q; \mathbb{R})$, 且 $\int_c^\infty f(x, y) dy$ 和 $\int_c^\infty f_x(x, y) dy$ 均一致收敛, 则 $F \in C^1([a, b]; \mathbb{R})$, 且

$$\frac{d}{dx} \int_c^\infty f(x, y) dy = \int_c^\infty \frac{\partial}{\partial x} f(x, y) dy.$$

证明. 取 $t = n > c$. 根据含参积分正则性 (II), 可知在对应条件下有 F_n 连续或连续可微. 根据条件, $F_n \Rightarrow F$, 所以 F 连续. 若 F'_n 也一致收敛, 则 F 可导, 且导数与极限可交换. \square

判断含参反常积分最方便的方法是 Weierstrass 的优函数法.

Weierstrass M 判别法

若 $|f(x, y)| \leq M(y)$, 且 $\int_c^\infty M(y) dy$ 收敛, 则 $\int_c^\infty f(x, y) dy$ 一致收敛.

证明. 利用 $\|F_t - F\| \leq \int_t^\infty M(y) dy$ 即可. \square

例 1.9.3 证明: Gamma 函数 $\Gamma(s) = \int_0^{+\infty} x^{s-1} e^{-x} dx$ 在 $(0, +\infty)$ 上任意阶可微.

证明. 将积分分为两个

$$\Gamma_0(s) = \int_0^1 x^{s-1} e^{-x} dx, \quad \Gamma_\infty(s) = \int_1^{+\infty} x^{s-1} e^{-x} dx.$$

考虑积分号内对 s 求 k 导数的积分

$$\Gamma_0^k(s) = \int_0^1 (\ln x)^k x^{s-1} e^{-x} dx, \quad \Gamma_\infty^k(s) = \int_1^{+\infty} (\ln x)^k x^{s-1} e^{-x} dx.$$

只要说明它们都在 $(0, +\infty)$ 上内闭一致收敛即可. 设 $b > a > 0$, 则对 $s \in [a, b]$, 有

$$\begin{aligned} x \in (0, 1] : |(\ln x)^k x^{s-1} e^{-x}| &\leq |\ln x|^k x^{a-1} e^{-x}; \\ x \in [1, +\infty) : |(\ln x)^k x^{s-1} e^{-x}| &\leq |\ln x|^k x^{b-1} e^{-x}. \end{aligned}$$

根据比较判别法易知

$$\int_0^1 |\ln x|^k x^{a-1} e^{-x} dx, \int_1^\infty |\ln x|^k x^{b-1} e^{-x} dx$$

都收敛. 所以由 Weierstrass M 判别法知, 各阶求导积分均内闭一致收敛. \square

也可以用 Leibniz 积分求导法计算反常积分.

例 1.9.4 计算 Gauss 分布的各阶矩 $I_n = \int_{-\infty}^{+\infty} x^n e^{-x^2} dx$.

解. 根据奇偶性, 当 n 是奇数时 $I_n = 0$. 下面考虑 $n = 2k$ 是偶数的情况. 引入含参变量积分 $G(t) = \int_{-\infty}^{+\infty} e^{-x^2 t} dx (t > 0)$. 容易发现, 它关于 t 的各阶导数积分

$$G^{(k)}(t) = \int_{-\infty}^{+\infty} \frac{\partial^k}{\partial t^k} (e^{-x^2 t}) dx = (-1)^k \int_{-\infty}^{+\infty} x^{2k} e^{-x^2 t} dx$$

均内闭一致收敛, 因此 $G^{(k)}(t) = G^{(k)}(t)$. 另一方面, 根据 Gauss 积分可得

$$G(t) = \frac{1}{\sqrt{t}} \int_{-\infty}^{+\infty} e^{-u^2} du = \sqrt{\pi} \cdot t^{-\frac{1}{2}}.$$

从而

$$G^{(k)}(t) = \sqrt{\pi} \cdot \frac{(-1)^k \cdot (2k-1)!!}{2^k} \cdot t^{-\frac{2k+1}{2}}.$$

所以, $I_{2k} = G^{(k)}(1) = 2^{-k} \sqrt{\pi} \cdot (2k-1)!!$. \square