

第六册

大青花鱼

目录

| | |
|--------------------------|-----------|
| 第一章 向量函数的变化 | 5 |
| 1.1 平直空间中的向量函数 | 5 |
| 1.2 向量函数的极限 | 7 |
| 1.3 向量函数的连续性质 | 13 |
| 1.4 二元函数的变化 | 25 |
| 第二章 二元函数与曲面 | 37 |
| 2.1 曲面的切平面 | 37 |
| 2.2 曲面上一点的梯度 | 41 |
| 2.3 二元函数的高次微变 | 49 |
| 2.4 曲面的极值 | 49 |
| 第三章 离散随机变量 | 51 |
| 3.1 随机变量的性质 | 51 |
| 3.2 正态分布 | 51 |

| | |
|--------------------|------------------------|
| 4 | 目录 |
| 3.3 | 大数定律 51 |
| 3.4 | 中心极限定理 51 |
| Appendices | 53 |
| 第一章 向量函数的变化 | 55 |
| 1.1 | 二元函数的局部变化 55 |

第一章 向量函数的变化

1.1 平直空间中的向量函数

我们学习过平面、立体空间中的变换。比如，平面上关于某条直线的对称变换，把平面中任一点映射到平面中另一点。空间中关于某个旋转轴的旋转变换，把空间中任一点映射到空间中另一点。空间中关于某平面的投影变换，把空间中任一点映射到该平面上的点。研究和工作生产实践中，我们也会遇到种种和平面、空间的点和向量有关的问题。

例子 1.1.1.

数学上，我们用实系数平直空间中的映射来统一描述这些问题。给定平直空间 \mathbb{R}^m 、 \mathbb{R}^n （其中 m, n 是正整数），我们把从 \mathbb{R}^m 到 \mathbb{R}^n 的映射称为 \mathbb{R}^m 到 \mathbb{R}^n 的向量函数。 $m = n = 1$ 时的向量函数就是实函数， $m = 1, n = 2$ 时的向量函数可以是实变复函数或描述平面曲线的实变向量函数， $m = n = 2$ 时的向量函数可以是平面上的直变换和维直变换， $m = n = 3$ 时的向量函数可以是立体空间中的直变换和维直变换。我们一般把 $m = 1, n > 1$ 时的向量函数称为**实变向量函数**，把 $m > 1, n = 1$ 时的向量函数称为**多元函数**， m, n 都大于 1 时，称为**多元变换**或**多元映射**。

如何着手研究向量函数呢？首先来回顾一下，我们是怎么学习实函数的。我们研究实函数时候，首先学习最简单的函数：正比例函数。正比例函

数 $x \mapsto kx$ 把坐标轴进行放缩。另一个简单的函数是平移函数： $x \mapsto x + b$ ，它把坐标轴平移。这两个函数保持坐标轴的性质。它们结合起来，是一次函数。一次函数的图像是直线，它保持直线的性质。对于一般的函数，我们希望用这些简单函数来辅助理解它们的性质。

首先我们讨论函数的连续性。函数 f 在一点 a 连续，当且仅当 x 趋于 a 时， $f(x)$ 总趋于 $f(a)$ 。这是函数的局部性质。更进一步，我们可以把某些函数在微小局部的样貌用一次函数来表示：

$$f(x) \approx f(a) + \partial f(a) \cdot (x - a).$$

我们把这种展示称为实函数的**微直观**。

对于性质“更好”的函数，我们可以把这个表示方法拓展到多项式，乃至用无穷项的累加直接表示函数。另一方面，我们发现，这些一次函数的斜率，以某种方式累积起来，就能得到原来的函数。这就建立了微观与宏观的结合。

对于向量函数，我们也希望用类似的方法来研究。我们首先学习“最简单”的向量函数。它们的性质和正比例函数、一次函数类似，都保持坐标轴的性质，把点、直线、平面映射为点、直线、平面。具体来说，设有 \mathbb{R}^m 到 \mathbb{R}^n 的向量函数 h ，如果它满足以下条件：

- 保持平移操作： $\forall \mathbf{u}, \mathbf{v}, h(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + h(\mathbf{v})$
- 保持放缩操作： $\forall k, \mathbf{u}, h(k \cdot \mathbf{u}) = k \cdot h(\mathbf{u})$

就说 h 是**直变换**或**平直变换**。我们证明过平面和立体空间中的直变换保持坐标轴的性质，并且总把点、直线、平面映射为点、直线、平面。所有 \mathbb{R}^m 到 \mathbb{R}^n 的平直变换的集合记作 $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ 。

其他更复杂的向量函数的性质，我们希望用平直变换来辅助理解。这个研究方法与我们研究实函数时候的方法是一样的。

1.2 向量函数的极限

在第五册中，我们学习了有邻空间中函数在一点的极限概念。本节中，我们将聚焦于赋长平直空间中向量函数的极限性质，特别是经典空间 \mathbb{R}^n 中的函数。与实函数相比，向量函数的极限概念既有相似之处，也有本质区别。我们首先回顾极限的严格定义，然后探究一些特有的性质。

定义 1.2.1. 向量函数在一点的极限 给定赋长平直空间 \mathbb{R}^m 到 \mathbb{R}^n 的映射 f 。设 f 在 $\mathbf{a} \in \mathbb{R}^m$ 附近有定义，如果对任何 $r > 0$ ，都存在 $d > 0$ ，使得只要 \mathbf{x} 到 \mathbf{a} 的距离小于 d ， $f(\mathbf{x})$ 到某定点 \mathbf{l} 的距离就小于 r ，那么就说 f 在 \mathbf{a} 处有极限 \mathbf{l} 。记作：

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = \mathbf{l}.$$

向量函数的极限定义与实函数的极限定义形式上相同。区别在于，我们使用了平直空间中的距离概念。对经典空间 \mathbb{R}^n ，我们可以使用经典距离 ℓ_2 来描述“接近”关系。然而，由于 \mathbb{R}^n 中各种模是等价的，无论选择哪种模定义距离，极限的结果都相同。

定理 1.2.1. 向量函数极限的分量刻画 给定经典空间 \mathbb{R}^m 到 \mathbb{R}^n 的函数 f 以及点 $\mathbf{a} \in \mathbb{R}^m$ ，设 f 可写为：

$$\forall \mathbf{x} \in \mathbb{R}^m, \quad f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})),$$

其中 f_1, f_2, \dots, f_n 是从 \mathbb{R}^m 到 \mathbb{R} 的函数。则

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = \mathbf{l} = (l_1, l_2, \dots, l_n)$$

当且仅当

$$\forall 1 \leq i \leq n, \quad \lim_{\mathbf{x} \rightarrow \mathbf{a}} f_i(\mathbf{x}) = l_i.$$

证明： 考虑经典模 ℓ_2 , 有

$$\|f(\mathbf{x}) - \mathbf{l}\|_2 = \sqrt{\sum_{i=1}^n (f_i(\mathbf{x}) - l_i)^2}.$$

若 $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = \mathbf{l}$, 则对任意 $r > 0$, 存在 $d > 0$, 使得当 $\|\mathbf{x} - \mathbf{a}\| < d$ 时,

$$\sqrt{\sum_{i=1}^n (f_i(\mathbf{x}) - l_i)^2} < r.$$

因此, 对任意 $1 \leq i \leq n$, 当 $\|\mathbf{x} - \mathbf{a}\| < d$ 时,

$$|f_i(\mathbf{x}) - l_i| \leq \sqrt{\sum_{j=1}^n (f_j(\mathbf{x}) - l_j)^2} < r.$$

这说明 $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f_i(\mathbf{x}) = l_i$.

反之, 若对任意 $1 \leq i \leq n$, $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f_i(\mathbf{x}) = l_i$, 则对任意 $r > 0$, 对每个 i 都存在 $d_i > 0$, 使得当 $\|\mathbf{x} - \mathbf{a}\| < d_i$ 时,

$$|f_i(\mathbf{x}) - l_i| < \frac{r}{\sqrt{n}}.$$

取 $d = \min\{d_1, d_2, \dots, d_n\}$, 则当 $\|\mathbf{x} - \mathbf{a}\| < d$ 时,

$$\sqrt{\sum_{i=1}^n (f_i(\mathbf{x}) - l_i)^2} < \sqrt{n \cdot \left(\frac{r}{\sqrt{n}}\right)^2} = r.$$

这说明 $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = \mathbf{l}$. □

例子 1.2.1. 考虑二元函数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$:

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

研究 f 在 $(0,0)$ 处的极限。

我们沿着不同路径趋近 $(0,0)$ 。首先考虑沿 x 轴趋近 (即 $y=0$):

$$\lim_{x \rightarrow 0} f(x, 0) = \lim_{x \rightarrow 0} \frac{x \cdot 0}{x^2 + 0^2} = 0.$$

同样, 沿 y 轴趋近 (即 $x=0$):

$$\lim_{y \rightarrow 0} f(0, y) = \lim_{y \rightarrow 0} \frac{0 \cdot y}{0^2 + y^2} = 0.$$

然而, 沿着 $y=x$ 趋近:

$$\lim_{x \rightarrow 0} f(x, x) = \lim_{x \rightarrow 0} \frac{x \cdot x}{x^2 + x^2} = \lim_{x \rightarrow 0} \frac{x^2}{2x^2} = \frac{1}{2}.$$

由于沿不同路径极限不同, f 在 $(0,0)$ 处不存在极限。

思考 1.2.1.

1. 为什么在一元函数中, 我们只需考虑从左右两侧趋近, 而在向量函数中需要考虑从更多方向趋近? 这反映了什么本质区别?

2. 能否找到一个二元函数, 它在原点沿任何直线趋近时极限都相同, 但在原点处整体极限不存在? 尝试构造一个例子。

在研究向量函数极限时, 我们会遇到累次极限的概念。对二元函数 $f(x, y)$, 考虑先让 x 趋近 a , 再让 y 趋近 b , 或者相反次序。两种次序得到的极限一样吗?

定义 1.2.2. 累次极限 给定二元函数 $f(x, y)$ 在点 (a, b) 附近有定义。若对每个足够接近 b 的 y , $\lim_{x \rightarrow a} f(x, y)$ 存在, 且 $\lim_{y \rightarrow b} \left(\lim_{x \rightarrow a} f(x, y) \right)$ 也存在, 则称

$$\lim_{y \rightarrow b} \left(\lim_{x \rightarrow a} f(x, y) \right)$$

为函数 f 在 (a, b) 处的累次极限 (先 x 后 y)。类似定义先 y 后 x 的累次极限。

累次极限与向量极限（即向量函数在一点的极限）有重要关系：

定理 1.2.2. 向量极限与累次极限的关系 给定二元函数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, 若

$$\text{lian}_{(x,y) \rightarrow (a,b)} f(x,y) = l$$

存在, 且对每个足够接近 b 的 y , $\text{lian}_{x \rightarrow a} f(x,y)$ 存在, 且对每个足够接近 a 的 x , $\text{lian}_{y \rightarrow b} f(x,y)$ 存在, 那么两个累次极限都存在且等于 l :

$$\begin{aligned} \text{lian}_{y \rightarrow b} \left(\text{lian}_{x \rightarrow a} f(x,y) \right) &= l, \\ \text{lian}_{x \rightarrow a} \left(\text{lian}_{y \rightarrow b} f(x,y) \right) &= l. \end{aligned}$$

证明: 设 $\text{lian}_{(x,y) \rightarrow (a,b)} f(x,y) = l$. 对任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得当 $|x-a| < \delta$ 且 $|y-b| < \delta$ 时,

$$|f(x,y) - l| < \varepsilon.$$

固定 y 满足 $|y-b| < \delta$, 则当 $|x-a| < \delta$ 时, $|f(x,y) - l| < \varepsilon$. 这说明 $\text{lian}_{x \rightarrow a} f(x,y)$ 存在且等于某个值 $g(y)$, 且 $|g(y) - l| \leq \varepsilon$.

由于 ε 任意, $\text{lian}_{y \rightarrow b} g(y) = l$, 即

$$\text{lian}_{y \rightarrow b} \left(\text{lian}_{x \rightarrow a} f(x,y) \right) = l.$$

同理可证另一累次极限也等于 l . □

例题 1.2.1. 考虑函数

$$f(x,y) = \frac{x-y}{x+y}$$

在点 $(0,0)$ 附近的行为。

首先, 考察重极限 $\text{lian}_{(x,y) \rightarrow (0,0)} f(x,y)$. 沿着 x 轴 ($y=0$) 趋近:

$$\text{lian}_{x \rightarrow 0} f(x,0) = \text{lian}_{x \rightarrow 0} \frac{x}{x} = 1.$$

沿着 y 轴 ($x = 0$) 趋近:

$$\lim_{y \rightarrow 0} f(0, y) = \lim_{y \rightarrow 0} \frac{-y}{y} = -1.$$

由于沿不同路径极限不同, 向量极限不存在。

其次, 考察累次极限。先计算 $\lim_{x \rightarrow 0} f(x, y)$, 对固定 $y \neq 0$:

$$\lim_{x \rightarrow 0} f(x, y) = \lim_{x \rightarrow 0} \frac{x - y}{x + y} = \frac{-y}{y} = -1.$$

然后

$$\lim_{y \rightarrow 0} \left(\lim_{x \rightarrow 0} f(x, y) \right) = \lim_{y \rightarrow 0} (-1) = -1.$$

再计算 $\lim_{y \rightarrow 0} f(x, y)$, 对固定 $x \neq 0$:

$$\lim_{y \rightarrow 0} f(x, y) = \lim_{y \rightarrow 0} \frac{x - y}{x + y} = \frac{x}{x} = 1.$$

然后

$$\lim_{x \rightarrow 0} \left(\lim_{y \rightarrow 0} f(x, y) \right) = \lim_{x \rightarrow 0} 1 = 1.$$

两个累次极限存在但不相等, 且都不等于向量极限 (向量极限不存在)。这与前面的定理并不矛盾, 因为定理的前提条件是向量极限存在。

定理 1.2.3. 极限的运算性质 设 f, g 是从 \mathbb{R}^m 到 \mathbb{R}^n 的映射, $\mathbf{a} \in \mathbb{R}^m$ 。若

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = \mathbf{l}_1, \quad \lim_{\mathbf{x} \rightarrow \mathbf{a}} g(\mathbf{x}) = \mathbf{l}_2, \quad \text{则:}$$

- $\lim_{\mathbf{x} \rightarrow \mathbf{a}} (f + g)(\mathbf{x}) = \mathbf{l}_1 + \mathbf{l}_2,$
- $\lim_{\mathbf{x} \rightarrow \mathbf{a}} (f - g)(\mathbf{x}) = \mathbf{l}_1 - \mathbf{l}_2,$
- 若 $n = 1$, $\lim_{\mathbf{x} \rightarrow \mathbf{a}} (f \cdot g)(\mathbf{x}) = \mathbf{l}_1 \cdot \mathbf{l}_2,$
- 若 $n = 1$ 且 $\mathbf{l}_2 \neq 0$, $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \left(\frac{f}{g} \right) (\mathbf{x}) = \frac{\mathbf{l}_1}{\mathbf{l}_2}.$

证明： 由向量函数极限的分量刻画，可将问题转化为实函数极限的性质。这里以加法为例证明。

设 $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x}))$, $g(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_n(\mathbf{x}))$, 则

$$(f + g)(\mathbf{x}) = (f_1(\mathbf{x}) + g_1(\mathbf{x}), f_2(\mathbf{x}) + g_2(\mathbf{x}), \dots, f_n(\mathbf{x}) + g_n(\mathbf{x})).$$

由条件，对任意 $1 \leq i \leq n$,

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} f_i(\mathbf{x}) = l_{1,i}, \quad \lim_{\mathbf{x} \rightarrow \mathbf{a}} g_i(\mathbf{x}) = l_{2,i},$$

其中 $\mathbf{l}_1 = (l_{1,1}, \dots, l_{1,n})$, $\mathbf{l}_2 = (l_{2,1}, \dots, l_{2,n})$ 。

由实函数极限的加法性质，

$$\forall 1 \leq i \leq n, \quad \lim_{\mathbf{x} \rightarrow \mathbf{a}} (f_i + g_i)(\mathbf{x}) = l_{1,i} + l_{2,i}.$$

因此，由向量函数极限的分量刻画，

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} (f + g)(\mathbf{x}) = (l_{1,1} + l_{2,1}, l_{1,2} + l_{2,2}, \dots, l_{1,n} + l_{2,n}) = \mathbf{l}_1 + \mathbf{l}_2.$$

其他性质可类似证明。 □

习题 1.2.1.

1. 研究以下函数在指定点处的极限：

$$1). \quad f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}, \quad (0, 0)$$

$$2). \quad g(x, y) = \frac{x^2 y}{x^4 + y^2}, \quad (0, 0)$$

$$3). \quad h(x, y, z) = \frac{xyz}{x^2 + y^2 + z^2}, \quad (0, 0, 0)$$

2. 研究以下累次极限：

$$1). \quad \lim_{y \rightarrow 0} \left(\lim_{x \rightarrow 0} \frac{x - y}{x^2 + y^2} \right), \quad \lim_{x \rightarrow 0} \left(\lim_{y \rightarrow 0} \frac{x - y}{x^2 + y^2} \right)$$

$$2). \quad \lim_{y \rightarrow 0} \left(\lim_{x \rightarrow 0} \frac{x^2 - y^2}{x^2 + y^2} \right), \quad \lim_{x \rightarrow 0} \left(\lim_{y \rightarrow 0} \frac{x^2 - y^2}{x^2 + y^2} \right)$$

$$3). \quad \lim_{y \rightarrow 0} \left(\lim_{x \rightarrow 0} \sin \frac{1}{x} \cdot \frac{y}{x^2 + y^2} \right), \quad \lim_{x \rightarrow 0} \left(\lim_{y \rightarrow 0} \sin \frac{1}{x} \cdot \frac{y}{x^2 + y^2} \right)$$

3. 证明：若 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ 满足

$$\forall x, y \in \mathbb{R}, \quad |f(x, y)| \leq x^2 + y^2,$$

则 $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$ 。

向量函数的极限概念是研究向量函数局部性质的基础。在下一节中，我们将基于极限概念讨论向量函数的连续性质。与实函数相比，向量函数的极限具有更丰富的直观意义和更复杂的行为模式。理解这些差异，对我们建立正确的多维直觉至关重要。

一个值得思考的问题是：在工程和科学中，我们常会遇到各种多变量系统，如何用数学语言精确描述这些系统的行为？向量函数的极限理论正是为此提供了严格框架。当我们说一个物理量依赖于多个参数时，理解这些参数如何共同影响该物理量，本质上就是理解向量函数的极限和连续性质。这种理解，将为我们打开通向更复杂数学结构和应用领域的大门。

1.3 向量函数的连续性质

给定 \mathbb{R}^m 到 \mathbb{R}^n 的向量函数 f ，对于 \mathbb{R}^m 中某点 \mathbf{a} ，我们首先定义它在 \mathbf{a} 处连续的性质。这里我们用极限来定义连续：映射在一点的极限等于该点的值。

定义 1.3.1. 向量函数在一点连续 给定赋长平直空间 \mathbb{R}^m 到 \mathbb{R}^n 的函数 f 。设 f 在点 $\mathbf{a} \in \mathbb{R}^m$ 附近有定义，如果

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a}),$$

就说 f 在 \mathbf{a} 处连续。如果 f 在定义域内某开集 U 中每一点都连续，就说 f 在 U 上连续。

向量函数在一点连续, 意味着当自变量从任意方向以任意方式趋近该点时, 函数值都会趋近于该点的函数值。这种定义与实函数的连续性定义在形式上一致, 但在多维空间中, 这意味着更复杂的可能性。

定理 1.3.1. 向量函数连续的分量刻画 给定经典空间 \mathbb{R}^m 到 \mathbb{R}^n 的函数 f , 设 f 可写为:

$$\forall \mathbf{x} \in \mathbb{R}^m, \quad f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})),$$

其中 f_1, f_2, \dots, f_n 是从 \mathbb{R}^m 到 \mathbb{R} 的函数。则 f 在 \mathbf{a} 处连续当且仅当 f_1, f_2, \dots, f_n 都在 \mathbf{a} 处连续。

证明: 由向量函数极限的分量刻画定理, 函数 f 在 \mathbf{a} 处连续当且仅当

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a}),$$

即

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x})) = (f_1(\mathbf{a}), f_2(\mathbf{a}), \dots, f_n(\mathbf{a})).$$

同样根据极限的分量刻画, 上式成立当且仅当

$$\forall 1 \leq i \leq n, \quad \lim_{\mathbf{x} \rightarrow \mathbf{a}} f_i(\mathbf{x}) = f_i(\mathbf{a}),$$

即 f_1, f_2, \dots, f_n 都在 \mathbf{a} 处连续。 □

例子 1.3.1. 考虑二元函数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$:

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

我们研究 f 在 $(0, 0)$ 处的连续性。

首先, 考虑沿 x 轴趋近 $(0, 0)$ (即 $y = 0$):

$$\lim_{x \rightarrow 0} f(x, 0) = \lim_{x \rightarrow 0} \frac{x \cdot 0}{x^2 + 0^2} = 0.$$

同样, 沿 y 轴趋近 $(0, 0)$ (即 $x = 0$):

$$\lim_{y \rightarrow 0} f(0, y) = \lim_{y \rightarrow 0} \frac{0 \cdot y}{0^2 + y^2} = 0.$$

然而, 我们需要验证以任意方式趋近时极限是否为 0。考虑使用极坐标变换: $x = r \cos \theta$, $y = r \sin \theta$, 则

$$f(x, y) = \frac{r^2 \cos \theta \sin \theta}{r^2} = \cos \theta \sin \theta.$$

当 $0 < \theta < \frac{\pi}{2}$ 时, 令 $r \rightarrow 0$, 则 $(x, y) \rightarrow (0, 0)$, 然而 $\cos \theta \sin \theta \neq 0$, 这时:

$$\lim_{r \rightarrow 0} f(x, y) = \lim_{r \rightarrow 0} \cos \theta \sin \theta = \cos \theta \sin \theta \neq 0.$$

这说明 $\lim_{(x, y) \rightarrow (0, 0)} f(x, y)$ 不一定存在, 因此 f 在 $(0, 0)$ 处不连续。

思考 1.3.1.

1. 考虑二元函数

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

研究 f 在 $(0, 0)$ 处的连续性。

2. 设想一个二元函数, 它在原点沿任何直线趋近时极限都相同, 但在原点处不连续。这样的函数存在吗? 试构造一个例子。

连续函数具有一些重要的基本性质。和实函数一样, 连续向量函数的和、差、积、商 (分母不为零) 以及复合函数仍然是连续函数。我们下面给出一些重要结论。

定理 1.3.2. 连续向量函数的运算性质 设 f, g 是 \mathbb{R}^m 到 \mathbb{R}^n 的连续映射, h 是 \mathbb{R}^n 到 \mathbb{R}^p 的连续映射。

- $f \pm g$ 在定义域上连续。
- 如果 $n = 1$, 则 fg 在定义域上连续。
- 如果 $n = 1$ 且 g 在定义域上恒不为零, 则 $\frac{f}{g}$ 在定义域上连续。
- 复合函数 $h \circ f$ 在定义域上连续。

证明： 这些结论可以从向量函数极限的运算性质及连续函数的定义直接推导。以复合函数为例：

设 f 在 \mathbf{a} 处连续, h 在 $f(\mathbf{a})$ 处连续。我们需要证明 $h \circ f$ 在 \mathbf{a} 处连续, 即

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} h(f(\mathbf{x})) = h(f(\mathbf{a})).$$

因为 f 在 \mathbf{a} 处连续, 所以 $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$ 。又因为 h 在 $f(\mathbf{a})$ 处连续, 由连续映射保持极限的性质, 有

$$\lim_{\mathbf{y} \rightarrow f(\mathbf{a})} h(\mathbf{y}) = h(f(\mathbf{a})).$$

因此

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} h(f(\mathbf{x})) = h\left(\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x})\right) = h(f(\mathbf{a})),$$

这就证明了 $h \circ f$ 在 \mathbf{a} 处连续。

其他性质可类似证明。 □

例题 1.3.1. 考察函数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$f(x, y) = \sin(x^2 + y^2).$$

证明 f 在 \mathbb{R}^2 上连续。

解答. 考虑两个函数 $g: \mathbb{R}^2 \rightarrow \mathbb{R}$, $h: \mathbb{R} \rightarrow \mathbb{R}$:

$$g(x, y) = x^2 + y^2, \quad h(t) = \sin t.$$

g 是多项式函数, 因此在 \mathbb{R}^2 上连续; h 是正弦函数, 在 \mathbb{R} 上连续。而 $f = h \circ g$, 根据复合函数的连续性质, f 在 \mathbb{R}^2 上连续。

对于实函数来说, 连续意味着很多良好的性质, 比如闭区间上的连续函数必然有界, 且达到最大值和最小值, 等等。对于向量函数来说, 连续性带来了什么呢?

连续函数把经典空间中的紧致集映射为紧致集。而 $\mathbb{R}^m, \mathbb{R}^n$ 中的紧致集就是有界闭集。因此，有界闭集经过连续映射，仍然是有界闭集。当 $n = 1$ 时，紧致集的像是 \mathbb{R} 的有界闭子集，即紧致集上的连续多元函数必达到最大值和最小值。

定理 1.3.3. 有界闭集上连续函数的性质 设 $U \subseteq \mathbb{R}^m$ 是有界闭集, $f: U \rightarrow \mathbb{R}$ 是连续函数。则:

- f 在 U 上有界。
- f 在 U 上达到最大值和最小值，即存在 $\mathbf{a}, \mathbf{b} \in U$ ，使得

$$f(\mathbf{a}) = \sup_{\mathbf{x} \in U} f(\mathbf{x}), \quad f(\mathbf{b}) = \inf_{\mathbf{x} \in U} f(\mathbf{x}).$$

- f 在 U 上一致连续。

证明: 1). 由于 U 是有界闭集 (紧致集)，且 f 连续，根据第五册中的结论， $f(U)$ 也是 \mathbb{R} 中的有界闭集。因此 f 在 U 上有界。

2). 同理， $f(U)$ 是 \mathbb{R} 中的非空有界闭集，因此它有最大值和最小值。设 $M = \max f(U)$, $m = \min f(U)$ ，则存在 $\mathbf{a}, \mathbf{b} \in U$ ，使得 $f(\mathbf{a}) = M$, $f(\mathbf{b}) = m$ 。

3). 由于 U 是紧致集，且 f 在 U 上连续，根据前面的结论， f 在 U 上一致连续。 □

例题 1.3.2. 如果多元函数 $f: \mathbb{R}^m \rightarrow \mathbb{R}$ 在自变量的模长趋于无穷大时总趋于无穷大，就说 f 是径向增函数。证明：在 \mathbb{R}^m 中连续的径向增函数总取到最小值。

解答. 首先把径向增函数的定义“翻译”成具体的语言：对任意实数 M ，总有 $d > 0$ ，使得只要 $|\mathbf{x}| > d$ ，就有 $f(\mathbf{x}) > M$ 。

任取一点 $\mathbf{a} \in \mathbb{R}^m$ ，考虑 $M = f(\mathbf{a})$ ，存在 $d > 0$ ，使得只要 $|\mathbf{x}| > d$ ，就

有 $f(x) > M$ 。考虑以原点为中心, 半径为 d 的闭“球”:

$$\bigcirc_d = \bigcirc(\mathbf{0}, d) = \{\mathbf{x} \in \mathbb{R}^m \mid |\mathbf{x}| \leq d\}.$$

它是有界闭集, 因此是紧致集。而 f 在 \mathbb{R}^m 中连续, 因此 \bigcirc_d 经过 f 映射的像 $f(\bigcirc_d)$ 是 \mathbb{R} 的有界闭子集。这说明 $f(\bigcirc_d)$ 有最小值 $y_{\text{小}}$ 。

$$y_{\text{小}} = \min_{|\mathbf{x}| \leq d} (f(\mathbf{x})).$$

在 $f(\bigcirc_d)$ 中取一个收敛到 $y_{\text{小}}$ 的数列: $\{y_n\}_{n \in \mathbb{N}}$ 。按照像集的定义, 存在 \bigcirc_d 中元素的序列 $\{\mathbf{x}_n\}_{n \in \mathbb{N}}$, 使得 $f(\mathbf{x}_n) = y_n$ 。 \bigcirc_d 是紧致集, 因此序列 $\{\mathbf{x}_n\}$ 有收敛子列 $\{\mathbf{x}_{\varphi(n)}\}$ 。设这个子列收敛到 $x_{\text{小}}$ 。由于 $\{\mathbf{x}_{\varphi(n)}\}$ 的像序列: $\{f(\mathbf{x}_{\varphi(n)})\} = \{y_{\varphi(n)}\}$ 是 $\{y_n\}$ 的子列, 因此趋于 $y_{\text{小}}$ 。而 f 是连续函数, 所以

$$f(x_{\text{小}}) = \lim_{n \rightarrow +\infty} f(\mathbf{x}_{\varphi(n)}) = y_{\text{小}}.$$

按定义, $\forall |\mathbf{x}| > d$, $f(\mathbf{x}) > f(\mathbf{a})$, 因此 $\mathbf{a} \in \bigcirc_d$ 。于是按定义 $y_{\text{小}} \leq f(\mathbf{a})$, 因而 $|\mathbf{x}| > d$ 时 $f(\mathbf{x}) > f(\mathbf{a}) \geq y_{\text{小}}$ 。这说明 $y_{\text{小}}$ 就是 f 在 \mathbb{R}^m 上的最小值, 在 $\mathbf{x}_{\text{小}}$ 处取到。

连续的径向增函数总能取到最小值, 这个结论在讨论存在性问题时很有用, 因为它能“非构造地”保证 $x_{\text{小}}$ 和 $y_{\text{小}}$ 存在。比如, 在证明多项式函数在复数域中必然有根时, 我们可以把复数域上的多项式函数视作 $\mathbb{R}^2 \rightarrow \mathbb{R}$ 的二元函数。若多项式 p 是首一多项式, 可以证明复数域上的函数: $z \mapsto |p(z)|$ 是径向增函数。因此, 根据以上定理, 它总能取到最小值。这正是证明多项式函数在复数域中必然有根的第一步。

我们学习过一致连续的概念。

定义 1.3.2. 实函数一致连续 设有定义在区间 I 上的函数 f 。如果对任意 $r > 0$, 总有 $d > 0$, 使得区间中距离小于 d 的任意两个数 x, y , 函数值的差都小于 r , 就说 f 在 I 上一致连续。

向量函数也有一致连续的说法。

定义 1.3.3. 向量函数一致连续 设有函数 $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ 。考虑集合 $U \subseteq \mathbb{R}^m$ ，如果对任意 $r > 0$ ，总有 $d > 0$ ，使得 U 中距离小于 d 的任意两个向量 \mathbf{x}, \mathbf{y} 经过映射后，距离都小于 r ，就说 f 在 U 上**一致连续**。

我们知道，在闭区间上连续的实函数，必然一致连续。向量函数也有类似性质：

定理 1.3.4. 紧致集上的连续向量函数一致连续。

证明： 证明思路与实函数的情形相同。给定 $r > 0$ ，根据连续性质，对 U 中任一点 \mathbf{x} ，总能找到 $d_x > 0$ ，使得 $|\mathbf{y} - \mathbf{x}| \leq 2d_x$ 时，总有 $|f(\mathbf{y}) - f(\mathbf{x})| < \frac{r}{2}$ 。

为了证明一致连续，我们要找一个统一的 d ，使得 U 中任意两点只要距离小于 d ，经 f 映射后距离就小于 r 。对于一般连续函数， d_x 随着 \mathbf{x} 变化，并不统一。不过，现在 U 是紧致集，因此考虑开球集合 $\{\bigcirc(\mathbf{x}, d_x) \mid \mathbf{x} \in U\}$ ，它是 U 的开覆盖。因此根据紧致集定义，存在有限开子覆盖：

$$U \subseteq \bigcup_{i=1}^k \bigcirc(\mathbf{x}_i, d_{x_i}).$$

把这 k 个球的半径 d_{x_i} 中最小的半径记为 d_0 ，那么 d_0 就是我们要找的 d 。

对于 U 中任意两点 \mathbf{x}, \mathbf{y} ， \mathbf{x} 必然属于前述 k 个开球之一。不妨设 $\mathbf{x} \in \bigcirc(\mathbf{x}_i, d_{x_i})$ ，那么按定义 $|f(\mathbf{x}) - f(\mathbf{x}_i)| < \frac{r}{2}$ 。如果 $|\mathbf{y} - \mathbf{x}| \leq d_0$ ，那么

$$|\mathbf{y} - \mathbf{x}_i| \leq |\mathbf{y} - \mathbf{x}| + |\mathbf{x} - \mathbf{x}_i| \leq d_0 + d_{x_i} \leq 2d_{x_i}.$$

于是 $|f(\mathbf{y}) - f(\mathbf{x}_i)| < \frac{r}{2}$ 。因此：

$$|f(\mathbf{y}) - f(\mathbf{x})| \leq |f(\mathbf{y}) - f(\mathbf{x}_i)| + |f(\mathbf{x}_i) - f(\mathbf{x})| < \frac{r}{2} + \frac{r}{2} = r.$$

这就说明 f 在 U 上一致连续。 □

直接讨论向量函数的连续和一致连续性质比较麻烦。我们引入一个更方便、更直观的概念：

定义 1.3.4. 自限性质 给定向量函数 f ，如果有固定的数 k ，使得 f 定义域中任两点经过 f 映射后的距离不大于这两点距离的 k 倍，就说 f 是 (k 倍) 自限的。 k 称为 f 的自限系数。

自限性质的出发点是很自然的。连续，就是不要有局部的“跳跃”或过于剧烈的“波动”。而自限性质用自变量的变化来约束映射后的变化量，简单来说，就是约束变率的绝对值不得超过 k 。直观来说， k 倍自限就说明映射在局部“斜率”不超过 k ，不如斜率为 k 的直线或直映射那么“陡峭”。

定理 1.3.5. 自限的向量函数必然一致连续（从而连续）。

证明： 设向量函数 f 自限，系数为 k 。对任意 $r > 0$ ，任意两点 \mathbf{x}, \mathbf{y} 只要距离小于 $\frac{r}{k}$ ，那么：

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq k \cdot |\mathbf{x} - \mathbf{y}| < k \cdot \frac{r}{k} = r.$$

这说明 f 一致连续。 □

下面来看一些基础的连续向量函数。通过自限性质来证明连续性质。

例题 1.3.3. 证明：

- \mathbb{R}^m 上的模映射是连续映射。
- $\mathbb{R}^m \rightarrow \mathbb{R}^n$ 的平直变换在经典邻则下是连续映射。
- \mathbb{R}^m 上的整式映射在经典邻则下是连续映射。
- \mathbb{R}^m 上到定点的距离映射是连续映射。
- \mathbb{R}^m 上到紧致集的距离映射是连续映射。

解答. 1). 把向量映射到其模长的映射是 1 倍自限的，因此连续。

2). 首先证明：如果有常数 A 使得 $\forall \mathbf{x}, |f(\mathbf{x})| \leq A|\mathbf{x}|$, 那么平直变换 f 是连续映射。

根据平直变换的基本性质：

$$\forall \mathbf{x}, \mathbf{y}, \quad |f(\mathbf{x}) - f(\mathbf{y})| = |f(\mathbf{x} - \mathbf{y})| \leq A|\mathbf{x} - \mathbf{y}|.$$

这说明平直变换 f 是 A 倍自限的，因此连续。

再证明：对任意 $\mathbb{R}^m \rightarrow \mathbb{R}^n$ 的平直变换，这样的常数 A 总存在。

我们知道平直变换 f 的值由一组基底上的值确定。给定基底 $B = (\mathbf{e}_1, \dots, \mathbf{e}_m)$, 记 f 在基底上的值为：

$$\forall 1 \leq i \leq m, \quad f(\mathbf{e}_i) = \mathbf{v}_i \in \mathbb{R}^n.$$

那么对任意 $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_m\mathbf{e}_m \in \mathbb{R}^m$, 有：

$$\begin{aligned} |f(\mathbf{x})| &= \left| f\left(\sum_{i=1}^m x_i \mathbf{e}_i\right) \right| = \left| \sum_{i=1}^m x_i f(\mathbf{e}_i) \right| \\ &\leq \sum_{i=1}^m |x_i| |f(\mathbf{e}_i)| = \sum_{i=1}^m |x_i| \cdot |\mathbf{v}_i| \\ &\leq A \sum_{i=1}^m |x_i| = A\|\mathbf{x}\|_1 \end{aligned}$$

其中 A 是所有 $|\mathbf{v}_i|$ 中最大者， $\|\cdot\|_1$ 是出发空间 \mathbb{R}^m 上的一次模 ℓ_1 。 \mathbb{R}^m 中的经典邻则对应二次模 ℓ_2 ，而我们知道 ℓ_1 和 ℓ_2 等价。于是有常数 B 使得

$$\forall \mathbf{x}, \quad \|\mathbf{x}\|_1 \leq B\|\mathbf{x}\|_2.$$

因此，常数 AB 使得：

$$\forall \mathbf{x}, \quad |f(\mathbf{x})| \leq A\|\mathbf{x}\|_1 \leq AB\|\mathbf{x}\|_2.$$

这说明平直变换 f 在经典邻则下是连续映射。

3). \mathbb{R}^m 上的整式映射, 指给定了基底 $B = (\mathbf{e}_1, \dots, \mathbf{e}_m)$ 后, 把任意 $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_m\mathbf{e}_m \in \mathbb{R}^m$ 映射到关于分量 x_1, x_2, \dots, x_m 的整式的映射。它是一种多元函数。

要证明整式映射是连续映射, 首先看最基本的一次式情形:

$$p_i: \quad \mathbf{x} = x_1\mathbf{e}_1 + \dots + x_m\mathbf{e}_m \mapsto x_i.$$

p_i 是平直变换, 因此是连续映射。对于更复杂的多项式映射, 我们可以把问题放到局部来看。 m 元多项式 p 的形式如下:

$$p: \quad \mathbf{x} \mapsto \sum_{(i_1, i_2, \dots, i_m) \in I_m} a_{i_1, i_2, \dots, i_m} x_1^{i_1} x_2^{i_2} \cdots x_m^{i_m}.$$

其中下标集 I_m 是 \mathbb{N}^m 的有限子集, 表示多项式各项的下标。每一项的次数 s_i 是各个指数下标 i_j 的和: $s_i = \sum_{l=1}^m i_l$ 。

对给定点 \mathbf{x} , 考虑 \mathbf{x} 的有界邻域 C , 记 M 为其上界:

$$\forall \mathbf{x} \in C, \quad \|\mathbf{x}\|_2 \leq M.$$

于是, 多项式函数中的每一项 $x_1^{i_1} x_2^{i_2} \cdots x_m^{i_m}$, 都可以这样限制其大小:

$$|x_1^{i_1} x_2^{i_2} \cdots x_m^{i_m}| = |x_j| \cdot |x_j|^{i_j-1} \cdot \prod_{l \neq j} |x_l^{i_l}| \leq |x_j| \cdot M^{i_j-1} \cdot \prod_{l \neq j} M^{i_l} = M^{s_i-1} |x_j|.$$

其中 j 是所有 i_l 中最大者对应的下标, $\prod_{l \neq j}$ 表示把下标不为 j 的项乘起来。上式结果说明, 在有界邻域 C 里, 多项式函数 p 的每一项有常数 $A_{i_1, i_2, \dots, i_m} = M^{s_i-1}$ 使得:

$$|x_1^{i_1} x_2^{i_2} \cdots x_m^{i_m}| \leq A_{i_1, i_2, \dots, i_m} |x_j| \leq A_{i_1, i_2, \dots, i_m} \|\mathbf{x}\|_1$$

于是多项式函数 p 满足:

$$\forall \mathbf{x}, \quad |p(\mathbf{x})| \leq \sum_{(i_1, i_2, \dots, i_m) \in I_m} a_{i_1, i_2, \dots, i_m} A_{i_1, i_2, \dots, i_m} \|\mathbf{x}\|_1 = A \|\mathbf{x}\|_1 \leq AB \|\mathbf{x}\|_2.$$

这说明对任意 \mathbf{x} 多项式函数 p 在 \mathbf{x} 的某个有界邻域里连续。但连续是局部性质，因此多项式函数 p 在定义域上总连续。

4). 考虑定点 $\mathbf{p} \in \mathbb{R}^m$ ，到 \mathbf{p} 的距离映射为：

$$\delta_{\mathbf{p}} : \mathbf{x} \mapsto |\mathbf{x} - \mathbf{p}|.$$

根据距离映射的基本性质，

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m, \quad |\delta_{\mathbf{p}}(\mathbf{x}) - \delta_{\mathbf{p}}(\mathbf{y})| = ||\mathbf{x} - \mathbf{p}| - |\mathbf{p} - \mathbf{y}|| \leq |(\mathbf{x} - \mathbf{p}) + (\mathbf{p} - \mathbf{y})| = |\mathbf{x} - \mathbf{y}|.$$

这说明 $\delta_{\mathbf{p}}$ 是 1 倍自限的，因此连续。

5). 给定紧致集 $U \in \mathbb{R}^m$ ，对给定的点 \mathbf{x} ，考虑 U 中点到 \mathbf{x} 的距离映射：

$$\delta_{\mathbf{x}} : \mathbf{u} \mapsto |\mathbf{u} - \mathbf{x}|.$$

$\delta_{\mathbf{x}}$ 是连续映射，因此在紧致集 U 上的像达到最小值。即有 $\mathbf{u}_x \in U$ 使得：

$$|\mathbf{u}_x - \mathbf{x}| = \min_{\mathbf{u} \in U} |\mathbf{u} - \mathbf{x}|.$$

我们把 \mathbf{u}_x 叫做 \mathbf{x} 到 U 的垂足。以上最小值取决于 \mathbf{x} ，于是我们可以定义映射：

$$\begin{aligned} \delta_U : \mathbb{R}^m &\rightarrow \mathbb{R} \\ \mathbf{x} &\mapsto \min_{\mathbf{u} \in U} |\mathbf{u} - \mathbf{x}| \end{aligned}$$

下面证明 δ_U 连续。

考虑任意两点 \mathbf{x}, \mathbf{y} 及任意 $\mathbf{u} \in U$ ，按距离映射的定义总有：

$$|\mathbf{x} - \mathbf{u}| + |\mathbf{x} - \mathbf{y}| \geq |\mathbf{y} - \mathbf{u}|,$$

所以，对于 $\mathbf{u}_x \in U$ 也有：

$$|\mathbf{x} - \mathbf{u}_x| + |\mathbf{x} - \mathbf{y}| \geq |\mathbf{y} - \mathbf{u}_x|,$$

考虑 \mathbf{u}_y , 按定义:

$$|\mathbf{x} - \mathbf{u}_x| + |\mathbf{x} - \mathbf{y}| \geq |\mathbf{y} - \mathbf{u}_x| \geq |\mathbf{y} - \mathbf{u}_y|,$$

于是:

$$|\mathbf{x} - \mathbf{y}| \geq |\mathbf{y} - \mathbf{u}_y| - |\mathbf{x} - \mathbf{u}_x|.$$

同理有:

$$|\mathbf{y} - \mathbf{x}| \geq |\mathbf{x} - \mathbf{u}_x| - |\mathbf{y} - \mathbf{u}_y|.$$

于是:

$$|\mathbf{x} - \mathbf{y}| \geq ||\mathbf{y} - \mathbf{u}_y| - |\mathbf{x} - \mathbf{u}_x|| = |\partial_U(\mathbf{x}) - \partial_U(\mathbf{y})|.$$

这说明 ∂_U 是 1 倍自限的, 因此连续。

思考 1.3.2.

习题 1.3.1.

1. 研究以下函数在指定点处的连续性:

$$1). \quad f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}, \quad (0, 0)$$

$$2). \quad g(x, y) = \frac{x^2 y}{x^4 + y^2}, \quad (0, 0)$$

$$3). \quad h(x, y, z) = \frac{xyz}{x^2 + y^2 + z^2}, \quad (0, 0, 0)$$

2. 证明函数 $f(x, y) = \sqrt{x^2 + y^2}$ 在 \mathbb{R}^2 上连续。

3. 证明函数 $f(x, y) = x^3 \sin(y) + y^3 \cos(x)$ 在 \mathbb{R}^2 上一致连续。

4. 设 U 是 \mathbb{R}^m 的有界闭子集。

4.1 为何一般来说无法定义把 \mathbb{R}^m 的点 \mathbf{x} 映射到它到 U 的垂足的映射 p_U ? 举一个反例。

4.2. 如果 $X \subseteq \mathbb{R}^m$ 满足:

$$\forall \mathbf{u}, \mathbf{v} \in X, \quad \forall t \in [0; 1], \quad t\mathbf{u} + (1 - t)\mathbf{v} \in X,$$

就说 X 是 \mathbb{R}^m 里的**凸集**。举一个 \mathbb{R}^2 中的例子来直观理解凸集。

如果 U 不仅是有界闭子集，还是凸集，我们希望证明 p_U 存在且连续。

4.3. 给定 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ ，证明以下不等式：

$$\frac{|\mathbf{x}|^2}{2} + \frac{|\mathbf{y}|^2}{2} \geq \left| \frac{\mathbf{x} + \mathbf{y}}{2} \right|^2.$$

举一个 \mathbb{R}^2 中的例子来直观理解这个不等式。

4.4. 证明：对于经典模映射 ℓ_2 ，我们可以定义上述的映射 p_U 。

4.5. 给定 $\mathbf{x} \in \mathbb{R}^m$ ，设 $\mathbf{u}_x = p_U(\mathbf{x})$ 是它到 U 的垂足。证明，对 U 中任一点 \mathbf{p} ，对于 ℓ_2 对应的内积 $(\cdot | \cdot)$ ，有以下不等式：

$$(\mathbf{x} - \mathbf{u}_x | \mathbf{u}_x - \mathbf{p}) \geq 0.$$

给出上式的直观解释，并根据直观解释作出证明。

4.6. 根据以上不等式，证明：

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m, \quad (\mathbf{x} - \mathbf{y} | p_U(\mathbf{x}) - p_U(\mathbf{y})) \geq |p_U(\mathbf{x}) - p_U(\mathbf{y})|^2.$$

4.7. 使用内积不等式证明：

$$|p_U(\mathbf{x}) - p_U(\mathbf{y})| \leq |\mathbf{x} - \mathbf{y}|,$$

并给出最终结论。

1.4 二元函数的变化

在一点 \mathbf{a} 处连续的向量函数 f ，在 \mathbf{a} 附近如何变化？和研究实函数一样，我们也希望找到一个平直变换 L ，使得 f 在 \mathbf{a} 附近有微直观：

$$f(\mathbf{x}) \approx f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a}).$$

这个 L 显然依赖于函数 f 和 \mathbf{a} ，我们将它记为

$$L = \mathcal{D}f(\mathbf{a}).$$

或者说, 对给定的 f , 有这样的映射 $\mathcal{D}f$:

$$\begin{aligned}\mathcal{D}f: \quad \mathbb{R}^m &\rightarrow \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \\ \mathbf{a} &\mapsto L = \mathcal{D}f(\mathbf{a})\end{aligned}$$

它把某些点 \mathbf{a} 映射到我们为这点找到的平直变换 L 。我们把映射 $\mathcal{D}f$ 叫做 f 的**微变映射**, 把它在点 \mathbf{a} 处的取值^①叫做 f 在 \mathbf{a} 处的**微变换**。 f 在点 \mathbf{a} 的微变换存在, 就说 f 在 \mathbf{a} 可微。

我们将 $\mathcal{D}f(\mathbf{a})$ 平移, 得到在 \mathbf{a} 处的值为 $f(\mathbf{a})$ 的维直变换。和实函数的微变一样, 这应该是使用平直变换近似描述 f 在 \mathbf{a} 附近的行为的最好办法。也就是说, 在 \mathbf{a} 附近,

$$\forall \mathbf{x}, f(\mathbf{x}) - f(\mathbf{a}) - \mathcal{D}f(\mathbf{a})(\mathbf{x} - \mathbf{a}) = o(\mathbf{x} - \mathbf{a}).$$

对一般的 f 和 \mathbf{a} , 如何找到 $\mathcal{D}f(\mathbf{a})$ 呢? 我们从简单的情况开始研究。

考虑把平面上的点映射到实数的二元函数:

$$f: \mathbb{R}^2 \mapsto \mathbb{R}.$$

这类函数在实际问题中有着广泛的应用。比如, 地形图上的海拔高度可以看作平面上一点的函数, 温度分布可以看作平面上点的位置的函数。理解二元函数在一点附近的变化性质, 对我们分析复杂现象至关重要。

二元函数 f 把平面上的点 $\mathbf{x} = (x_1, x_2)$ 映射到实数 $f(\mathbf{x})$ 。我们也将它简单地记作:

$$(x_1, x_2) \mapsto f(x_1, x_2) \in \mathbb{R}.$$

直观上, 二元函数可以想象为三维空间中的曲面。考虑装备了直角坐标系 $Oxyz$ 的立体空间, 把自变量 (x_1, x_2) 看平面 Oxy 的点的坐标, 把函数值看作曲面在该点处的“高度”, 也就是用空间中形同的点 $(x_1, x_2, f(x_1, x_2))$

^①注意: 微变映射在一点的取值是一个平直变换, 而不是数, 也不是向量。

的集合描述二元函数。作在平面上某一点 (\mathbf{a}) 附近, 如果有平直变换 L , 使得在 \mathbf{a} 附近,

$$\forall \mathbf{x}, f(\mathbf{x}) - f(\mathbf{a}) - L(\mathbf{x} - \mathbf{a}) = o(\mathbf{x} - \mathbf{a}),$$

就说 f 在 \mathbf{a} 处可微。如何理解这里的 $o(\mathbf{x} - \mathbf{a})$ 呢? 按定义, 应该有:

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{|f(\mathbf{x}) - f(\mathbf{a}) - L(\mathbf{x} - \mathbf{a})|}{|\mathbf{x} - \mathbf{a}|} = 0.$$

如何表示 \mathbf{x} 趋近于 \mathbf{a} 呢, 我们使用平面上的经典距离 ℓ_2 来描述。上面的极限就是说 $\forall r > 0$, 都有 $d > 0$, 使得只要 $\sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2} < d$, 就有:

$$\frac{|f(\mathbf{x}) - f(\mathbf{a}) - L(\mathbf{x} - \mathbf{a})|}{\sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2}} < r.$$

其中 $\mathbf{x} = (x_1, x_2)$, $\mathbf{a} = (a_1, a_2)$ 。要注意上式中分子和分母采用的距离不同。分子的距离是 \mathbb{R} 中的绝对值函数, 分母的距离是 \mathbb{R}^2 中的经典距离 ℓ_2 。

为了方便, 我们还可以把 $\mathbf{x} - \mathbf{a}$ 作为变量来讨论, 记 $\mathbf{h} = \mathbf{x} - \mathbf{a}$, 那么 $\mathbf{x} = \mathbf{a} + \mathbf{h}$ 。我们可以说, f 在 \mathbf{a} 处可微, 当且仅当

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L(\mathbf{h})|}{|\mathbf{h}|} = 0.$$

即 $\forall r > 0$, 都有 $d > 0$, 使得只要 $|\mathbf{h}| = \sqrt{h_1^2 + h_2^2} < d$, 就有:

$$\frac{|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L(\mathbf{h})|}{\sqrt{h_1^2 + h_2^2}} < r.$$

上式中的 $L(\mathbf{h})$ 是一个实数。这个实数如何与 \mathbf{h} 相关呢? L 如何选取呢?

我们先来看一种特殊情形: \mathbf{h} 是某个特定方向上的向量, 沿着该方向趋于 \mathbf{a} 。具体来说, 给定单位长度的向量 \mathbf{v} ($|\mathbf{v}| = 1$), 令 $\mathbf{h} = t\mathbf{v}$, 其中 $t \in \mathbb{R}$ 是系数, 表示 h 的长度。固定方向 \mathbf{v} 不变而让 t 变动。根据平直变换的基

本性质, $L(\mathbf{h}) = tL(\mathbf{v})$, 而 $L(\mathbf{v}) \in \mathbb{R}$ 是固定的实数, 于是当 t 趋于 0 时:

$$\begin{aligned} \frac{|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L(\mathbf{h})|}{|\mathbf{h}|} &= \frac{|f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) - tL(\mathbf{v})|}{|t| \cdot |\mathbf{v}|} \\ &= \frac{|f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) - tL(\mathbf{v})|}{|t|} \end{aligned}$$

这就把二元函数的问题转化归结为实函数的问题。考虑实函数:

$$f_{\mathbf{v}}: t \mapsto f(\mathbf{a} + t\mathbf{v}) \in \mathbb{R}$$

$f_{\mathbf{v}}$ 在 $t = 0$ 处的微变率存在, 就是说有实数 l , 使得:

$$\lim_{t \rightarrow 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} = \lim_{t \rightarrow 0} \frac{f_{\mathbf{v}}(t) - f_{\mathbf{v}}(0)}{t} = l,$$

而对于 f 来说, 如果 f 在 \mathbf{a} 处可微, 那么:

$$\lim_{t \rightarrow 0} \left| \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} - L(\mathbf{v}) \right| = \lim_{t \rightarrow 0} \frac{|f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) - tL(\mathbf{v})|}{|t|} = 0.$$

这说明

$$\lim_{t \rightarrow 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} = L(\mathbf{v}).$$

因此 $l = L(\mathbf{v})$ 。

换句话说, 如果 f 在 \mathbf{a} 处可微, 那么对任意单位长度的向量 \mathbf{v} , 实函数 $f_{\mathbf{v}}$ 在 $t = 0$ 处都应该可微, 而且微变率应该就是 $L(\mathbf{v})$ 。也就是说, 通过求出各个方向上实函数 $t \mapsto f(\mathbf{a} + t\mathbf{v})$ 在 $t = 0$ 处的微变率, 我们就得到了平直变换 L 的全貌。

前面已经说过, 我们可以把二元函数理解为装备了直角坐标系 $Oxyz$ 的立体空间中的曲面。把 Oxy 平面看作自变量 $\mathbf{x} = (x_1, x_2)$ 的取值平面, 用 z 轴的分量来表示函数值 $f(\mathbf{x})$ 。也就是说, 点集:

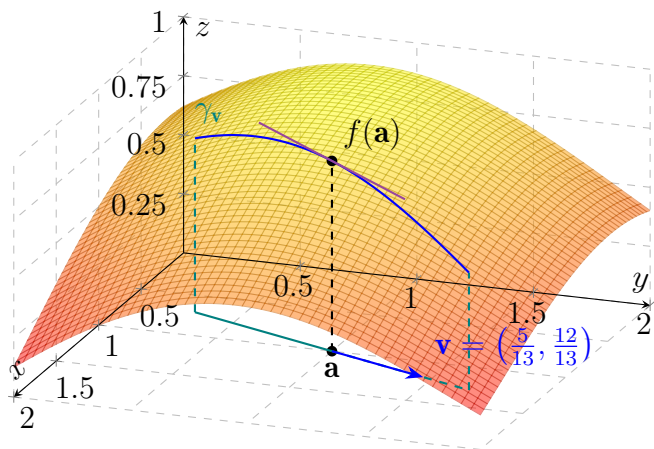
$$\{(x_1, x_2, f(x_1, x_2)) \mid x_1, x_2 \in \mathbb{R}\}$$

就是 f 在立体空间的直观表示。我们可以把这样的点集称为 f 的曲面。 f 在一点取值的大小, 直观上对应该点位置上曲面的“高度”。

考虑沿着 \mathbf{v} 方向且垂直于 Oxy 平面的竖直平面 $\gamma_{\mathbf{v}}$:

$$\gamma_{\mathbf{v}} = \{\mathbf{a} + t\mathbf{v} + s\mathbf{e}_z \mid t, s \in \mathbb{R}\}$$

其中 \mathbf{e}_z 是直角坐标系 $Oxyz$ 对应的基底中 z 轴方向的基向量。实函数 $f_{\mathbf{v}}$ 可以理解为从点 \mathbf{a} 出发, 用 $\gamma_{\mathbf{v}}$ 截取 f 的曲面得到的平面曲线。 $f_{\mathbf{v}}$ 在 $t=0$ 处的微变率就是平面曲线在 \mathbf{a} 处的切线斜率。



f 的曲面上 $\mathbf{a} = (1, 1)$ 处沿 $\mathbf{v} = (\frac{5}{13}, \frac{12}{13})$ 方向的切线

我们把 $f_{\mathbf{v}}$ 在 $t=0$ 处的微变率称为 f 沿 \mathbf{v} 方向的微变率, 简称方向微变, 记作 $\mathcal{D}_{\mathbf{v}}f(\mathbf{a})$ 。如果 \mathbf{v} 是基向量 $(1, 0)$ 或 $(0, 1)$, 就说它是 f 的首元(次元)微变率, 合称偏微变率。二元函数 f 在 \mathbf{a} 处的偏微变可以简记为 $\partial_x f(\mathbf{a}), \partial_y f(\mathbf{a})$ 或 $\partial_1 f(\mathbf{a}), \partial_2 f(\mathbf{a})$ 。

二元函数在一点可微, 则它在这点的任何方向微变和偏微变都存在。从可微二元函数在一点的偏微变出发, 我们可以求出函数在该点的微直观 L 。这是因为 L 作为平直变换, 必然由它在平面基向量上的值完全确定。如果已知:

$$L(\mathbf{e}_1) = \partial_1 f(\mathbf{a}), \quad L(\mathbf{e}_2) = \partial_2 f(\mathbf{a}),$$

那么对任意向量 \mathbf{h} , \mathbf{h} 可以唯一表示为平面基向量的直组合:

$$\mathbf{h} = h_1 \mathbf{e}_1 + h_2 \mathbf{e}_2.$$

于是:

$$L(\mathbf{h}) = L(h_1\mathbf{e}_1 + h_2\mathbf{e}_2) = h_1L(\mathbf{e}_1) + h_2L(\mathbf{e}_2) = h_1\partial_1f(\mathbf{a}) + h_2\partial_2f(\mathbf{a}).$$

也就是说, f 在 \mathbf{a} 附近可以写成:

$$\forall h_1, h_2 \in \mathbb{R}, f(\mathbf{a} + h_1\mathbf{e}_1 + h_2\mathbf{e}_2) \approx f(\mathbf{a}) + h_1\partial_1f(\mathbf{a}) + h_2\partial_2f(\mathbf{a}).$$

方向微变的定义来自实函数的微变率, 因此, 可以推出方向微变的运算法则。

定理 1.4.1. 若向量函数 $f, g: \mathbb{R}^m \rightarrow \mathbb{R}^n$ 在 $\mathbf{a} \in \mathbb{R}^m$ 处可微, 那么对任意非零向量 $\mathbf{v} \in \mathbb{R}^m$ 有:

- $\partial_{\mathbf{v}}(f \pm g)(\mathbf{a}) = \partial_{\mathbf{v}}f(\mathbf{a}) \pm \partial_{\mathbf{v}}g(\mathbf{a})$.
- 对任意常数 c , 有 $\partial_{\mathbf{v}}(cf)(\mathbf{a}) = c \cdot \partial_{\mathbf{v}}f(\mathbf{a})$.
- 若 $n = 1$, 则 $\partial_{\mathbf{v}}(f \cdot g)(\mathbf{a}) = g(\mathbf{a}) \cdot \partial_{\mathbf{v}}f(\mathbf{a}) + f(\mathbf{a}) \cdot \partial_{\mathbf{v}}g(\mathbf{a})$.
- 若 $n = 1$, 且函数 $g(\mathbf{a}) \neq 0$, 则:

$$\partial_{\mathbf{v}}\left(\frac{f}{g}\right)(\mathbf{a}) = \frac{g(\mathbf{a}) \cdot \partial_{\mathbf{v}}f(\mathbf{a}) - f(\mathbf{a}) \cdot \partial_{\mathbf{v}}g(\mathbf{a})}{g^2(\mathbf{a})}.$$

以上的运算法则显然也适用于偏微变。

例题 1.4.1. 计算以下函数在 $(0, 0)$ 处的偏微变。

- 1). $f: (x, y) \mapsto x^3 + 2x^2y + xy - y^3 - 2x - 3y + 2$, 2). $f: (x, y) \mapsto \sin(xy)$
- 3). $f: (x, y) \mapsto e^y \cos x$ 4). $f: (x, y) \mapsto \frac{x-y}{\sqrt{x^2+y^2}}$

解答.

1). 考虑非零向量 $\mathbf{v} = (1, 0)$, 函数 f 在 \mathbf{v} 方向的微变为:

$$\begin{aligned}
 \partial_{\mathbf{v}} f(x, y) &= \lim_{t \rightarrow 0} \frac{f(x+t, y) - f(x, y)}{t} \\
 &= \lim_{t \rightarrow 0} \frac{(x+t)^3 + 2(x+t)^2 y + (x+t)y - y^3 - 2(x+t) - 3y + 2 - (x^3 + 2x^2 y + xy - y^3 - 2x - 3y + 2)}{t} \\
 &= \lim_{t \rightarrow 0} \frac{(x+t)^3 - x^3}{t} + \lim_{t \rightarrow 0} 2y \frac{(x+t)^2 - x^2}{t} + \lim_{t \rightarrow 0} y \frac{(x+t) - x}{t} - \lim_{t \rightarrow 0} 2 \frac{(x+t) - x}{t} \\
 &= 3x^2 + 2y \cdot 2x + y \cdot 1 - 2 \\
 &= 3x^2 + 4xy + y - 2
 \end{aligned}$$

因此在 $(0, 0)$ 处的首元微变为 -2 。

我们注意到, 以上求偏微变的计算, 实际是把 y 看作常数, 对 $x \mapsto f(x, y)$ 求微变的结果。我们可以用这个更方便的想法计算次元偏微变。固定 x , 记 $g_x : y \mapsto f(x, y)$, 那么:

$$\begin{aligned}
 \partial_2 f(x, y) &= \partial(g_x)(y) \\
 &= \partial(-y^3 + (2x^2 + x - 3)y + (x^3 - 2x + 2)) \\
 &= \partial(-y^3) + \partial((2x^2 + x - 3)y) + \partial((x^3 - 2x + 2)) \\
 &= -3y^2 + 2x^2 + x - 3
 \end{aligned}$$

因此在 $(0, 0)$ 处的次元偏微变为 -3 。

上面的 $\partial(g_x)(y)$ 可以写为 $\partial_y(g_x)(y)$, 微变符号下标 y 表示我们在对变量 y 求微变, 以免混淆。下同。

2). 用前述方法。首先固定 y , 得到 $g_y : x \mapsto \sin(xy)$, 对 x 求微变:

$$\partial_1 f(x, y) = \partial_x(g_y)(x) = \partial_x \sin(xy) = y \cos(xy)$$

因此在 $(0, 0)$ 处的首元偏微变为 0 。

再固定 x , 得到 $g_x : y \mapsto \sin(xy)$, 对 y 求微变:

$$\partial_2 f(x, y) = \partial_y(g_x)(y) = \partial_y \sin(xy) = x \cos(xy)$$

因此在 $(0, 0)$ 处的次元偏微变为 0。

3). 首先固定 y , 得到 $g_y : x \mapsto e^y \cos x$, 对 x 求微变:

$$\partial_1 f(x, y) = \partial_x(g_y)(x) = \partial_x(e^y \cos x) = -e^y \sin x$$

因此在 $(0, 0)$ 处的首元微变为 0。

再固定 x , 得到 $g_x : y \mapsto e^y \cos x$, 对 y 求微变:

$$\partial_2 f(x, y) = \partial_y(g_x)(y) = \partial_y(e^y \cos x) = e^y \cos x$$

因此在 $(0, 0)$ 处的次元微变为 1。

4). 由于函数在 $(0, 0)$ 处无定义, 所以偏微变不存在。让我们把 $(0, 0)$ 处的值补上。记 $f(0, 0) = c$ 。其中 c 为常数。首先固定 $y = 0$, 得到 $g_y : x \mapsto \frac{x-0}{\sqrt{x^2+0^2}} = \frac{x}{x} = 1$ 。因此, 如果 $c = 1$, 那么首元微变存在且为 0, 否则不存在。同理, 固定 $x = 0$, 得到 $g_x : y \mapsto \frac{0-y}{\sqrt{0^2+y^2}} = -\frac{y}{y} = -1$, 因此, 如果 $c = -1$, 那么次元微变存在且为 0, 否则不存在。

注意: 无论 $(0, 0)$ 处的函数值是多少, 两个偏微变不能同时存在。

这些结论也可以推广到多元函数 $f : \mathbb{R}^m \rightarrow \mathbb{R}$ 。比如, 三元函数也可以定义方向微变和偏微变。我们将它对各个自变量分量的偏微变率称为**首元微变**、**次元微变**和**末元微变**。三元函数无法简单用立体空间中的曲面来直观理解, 但我们知道它在一点的微变是把 \mathbb{R}^3 中的空间向量映射到实数的直映射, 它的值可以用各个偏微变的直组合来表示。

如果多元函数在一点连续, 且任何方向的微变(包括偏微变)存在, 它是否必然在这点可微呢? 答案是否定的。我们可以找到在一点的偏微变存在, 但不可微的连续函数。下面是一个例子:

考虑函数：

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(x_1, x_2) \mapsto \begin{cases} \frac{x_1^3}{x_1^2 + x_2^2} & \text{如果 } x_1, x_2 \text{ 不全为 } 0, \\ 0 & \text{如果 } x_1 = x_2 = 0. \end{cases}$$

可以验证， f 在 $(0,0)$ 处连续，且在任何方向 $\mathbf{v} = (v_1, v_2)$ 上可微，方向微变为：

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{|f(\mathbf{0} + t\mathbf{v}) - f(\mathbf{0})|}{|t|} &= \lim_{t \rightarrow 0} \frac{|tv_1|^3}{|t|(t^2v_1^2 + t^2v_2^2)} \\ &= \lim_{t \rightarrow 0} \frac{|t|^3|v_1|^3}{|t|^3(v_1^2 + v_2^2)} \\ &= \frac{|v_1|^3}{v_1^2 + v_2^2} \end{aligned}$$

但 f 在 $(0,0)$ 处不可微。反设 f 在 $(0,0)$ 处可微，那么存在对应的平直变换 L ，而由前可知 $\forall v_1, v_2, L(v_1\mathbf{e}_1 + v_2\mathbf{e}_2) = v_1\partial_1 f(\mathbf{0}) + v_2\partial_2 f(\mathbf{0})$ 。但根据上面的计算， f 在 $(0,0)$ 处的偏微变是：

$$\partial_1 f(\mathbf{0}) = \frac{1^3}{1^2 + 0^2} = 1, \partial_2 f(\mathbf{0}) = \frac{0^3}{0^2 + 1^2} = 0.$$

于是：

$$\forall v_1, v_2, \frac{|v_1|^3}{v_1^2 + v_2^2} = L(v_1\mathbf{e}_1 + v_2\mathbf{e}_2) = v_1$$

然而上式不可能对任意 v_1, v_2 成立，比如令 $v_1 = v_2 = 1$ 就有：

$$\frac{|v_1|^3}{v_1^2 + v_2^2} = \frac{1}{2} \neq 1 = v_1.$$

以上结果告诉我们，多元函数的微变比实函数的微变复杂得多，多元函数可微是比各向可微更强的条件。

那么，能不能根据各项可微和一些附加条件，得到多元函数可微的充分条件呢？仿照一元函数的情形，我们引入光滑函数的概念：

定义 1.4.1. 多元函数的光滑性质 如果多元函数 $f: \mathbb{R}^m \mapsto \mathbb{R}$ 在点 \mathbf{a} 的某个邻域内有定义, 且:

- f 在 \mathbf{a} 处的偏微变 $\partial_1 f, \partial_2 f, \dots, \partial_m f$ 存在;
- $\forall 1 \leq i \leq m$, 所有偏微变函数 $\partial_i f$ 在 \mathbf{a} 处连续,

就说 f 在 \mathbf{a} 处一阶光滑, 记作 C^1 。在开集 $U \subseteq \mathbb{R}^m$ 的每一点都连续可微的函数称为在 U 中一阶光滑, 所有这样的函数的集合记为 $C^1(U, \mathbb{R})$ 。

我们可以证明 (见附录):

定理 1.4.2. 可微的充分条件 如果多元函数 f 在点 \mathbf{a} 一阶光滑, 那么 f 在 \mathbf{a} 处可微。

一元函数的情形中, 一阶光滑也称为连续可微, 两者自然等价。而多元函数的情形中, 这不是显然的。

有了这个结论, 对于可以从基本函数经过运算、复合得来的向量函数, 我们可以方便地验证其偏微变函数是否连续, 从而讨论其是否可微。要注意的是: 如果偏微变函数在该点不连续, 不能说明函数在该点一定不可微。

多元函数的微变和实函数的微变有相同之处, 也有不同之处。多元函数的微变也用于构造微直观, 用相对简单的平直映射近似描述函数在一点附近的行为。但是, 实函数不区别在一点的微变率和用于构造在一点的微直观的直变换, 因为对实函数来说, 直变换就是正比例函数, 它直接由它的斜率决定。它的斜率就是函数在一点的微变率。对于多元函数来说, 直变换无法用一个数表示, 因此有必要区分微变换和 (某个方向上的) 微变率。

思考 1.4.1.

1. 可微向量函数的复合函数是否可微? 其方向微变是否有类似实函数的运算法则?

2. 在某一点有偏微变的向量函数是否在这点连续? 在某一点有任意方向微变的向量函数是否在这点连续?

习题 1.4.1.

1. 证明本节中提到的方向微变的运算法则，并说明偏微变的运算法则。
2. 计算以下函数在 $(0, 0)$ 处的偏微变，它们在 $(0, 0)$ 处是否可微？

$$1). (x, y) \mapsto x^3 - xy + xy^2 - y^3 + 3x^2 - y + 1, \quad 2). (x, y) \mapsto \sin(xy - x) + 2y$$

$$3). (x, y) \mapsto \cos(x + 2y) \sin(y - x) - e^{x - \sin y} \quad 4). (x, y) \mapsto \frac{e^{x+y} - 1}{\sqrt{x^2 + y^2}}$$

第二章 二元函数与曲面

前一节中我们学习了，如果二元函数在一点 \mathbf{a} 处可微，那么存在一个平直变换 L ，使得在 \mathbf{a} 附近，函数值可以近似表示为：

$$f(\mathbf{a} + \mathbf{h}) \approx f(\mathbf{a}) + L(\mathbf{h}).$$

其中 $\mathbf{h} = (h_1, h_2)$ 是小的向量变化量。这个平直变换 L 称为函数在 \mathbf{a} 处的微直观，记为 $\mathcal{D}f(\mathbf{a})$ 。

对于一元实函数，我们将它的行为用平面中的曲线来表示。反之，平面中满足一定条件的曲线，可以看作某个实函数对应的曲线，称为函数的图像。二元实函数也是如此，我们将它与空间中的曲面联系起来，认为它可以表示空间中满足一定条件的曲面。具体来说，就是没有重复竖直投影的曲面。换句话说，曲面上不同的两点在水平面 Oxy 上的投影必定不一样。这是由映射的定义决定的：同一自变量只映射到一个值。

2.1 曲面的切平面

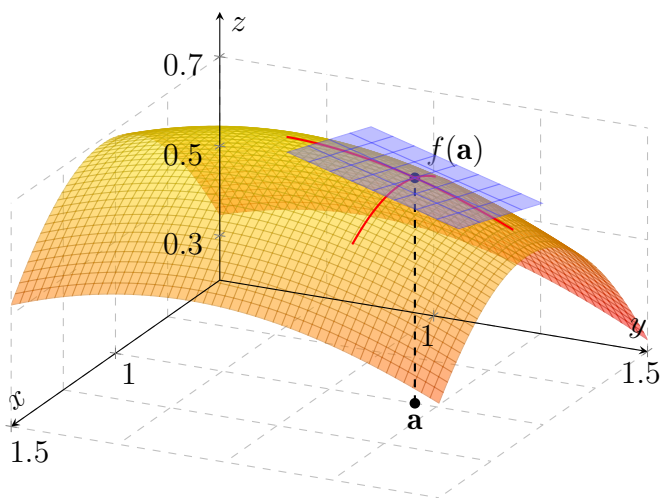
对于可微的二元函数，我们可以研究其对应的曲面的局部性质。一元的情况下，我们用函数图像在一点 a 的微变率，可以做出一条过该点的直线，与函数的曲线相切，称为切线。这条切线对应一个维直变换：

$$x \mapsto f(a) + \partial f(a)(x - a).$$

它告诉我们用直线模拟函数在该点附近行为的最佳方法。同样地，对于二元函数来说，如果它在一点 $\mathbf{a} = (a_1, a_2)$ 可微，那么我们有一个维直变换：

$$(x, y) \mapsto f(\mathbf{a}) + \mathcal{D}f(\mathbf{a})(x - a_1, y - a_2) = f(\mathbf{a}) + (x - a_1) \cdot \partial_1 f(\mathbf{a}) + (y - a_2) \cdot \partial_2 f(\mathbf{a}).$$

这个维直变换在直观上代表什么呢？它代表一个过点 $(a_1, a_2, f(\mathbf{a}))$ 的平面。我们将这个平面称为可微函数在 \mathbf{a} 处的切平面。在 \mathbf{a} 附近，它和函数 f 对应的曲面最相似。



曲面在点 $\mathbf{a} = (1, 1.2)$ 处的切平面

如果我们想知道当自变量从 \mathbf{a} 处出发，移动了 \mathbf{h} 后，二元函数的值怎么变化，可以通过在该点的切平面上移动来近似。函数值的变化量 $\Delta f(\mathbf{a})$ 可以用切平面的方程近似计算：

$$\Delta f(\mathbf{a}) = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) \approx (h_1 - a_1) \cdot \partial_1 f(\mathbf{a}) + (h_2 - a_2) \cdot \partial_2 f(\mathbf{a}).$$

与一元的情形一样，当 \mathbf{h} 接近 $\mathbf{0}$ 时，近似误差是 \mathbf{h} 长度的无穷小： $o(|\mathbf{h}|)$ ，而选择其他平面的话，就不会有这样的效果了。

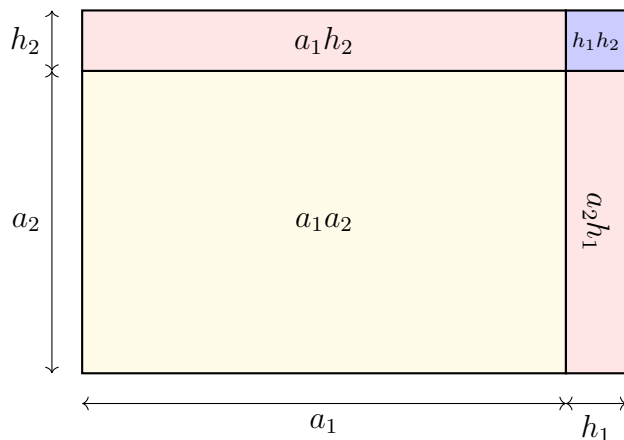
以代数的角度来看， $\Delta f(\mathbf{a})$ 的计算方式和一元函数变化量相比，可以看作在两个坐标上分别形式地计算变化分量，然后累加起来。这个特性称

为变化量的叠加性质，我们可以把多元函数的局部变化 $\Delta f(\mathbf{a})$ 视为各个分量上的变化叠加的结果，称为 f 在 \mathbf{a} 处的**全微变**。我们可以用一个简单的例子帮助理解^①：

考虑函数 $f: (x_1, x_2) \mapsto x_1 \cdot x_2$ ，它计算长和宽分别是 x_1 、 x_2 的长方形的面积。当长方形的长和宽从某个固定值 (a_1, a_2) 变化到 $(a_1 + h_1, a_2 + h_2)$ 时，长方形面积变为：

$$(a_1 + h_1)(a_2 + h_2) = a_1 a_2 + a_1 h_2 + a_2 h_1 + h_1 h_2.$$

a_2 、 a_1 分别是 f 关于自变量 x_1 、 x_2 的偏微变。它们与变化量 h_1 、 h_2 相乘，分别指出长方形面积在两个坐标方向的增长。而剩余的一项： $h_1 h_2$ ，当 h_1 、 h_2 非常小的时候，远小于前面提到的其他部分，因此可以忽略。于是我们发现，长方形面积的增长量，约等于它在各个坐标方向的增长量。



例题 2.1.1. 计算函数 $f(x, y) = \sin(xy)$ 的微直观，以及在 $(0, \frac{\pi}{2})$ 处的全微变。

解答. 首先计算偏微变：

$$\partial_1 f(x, y) = y \cos(xy), \quad \partial_2 f(x, y) = x \cos(xy).$$

^①注意：这并不是证明。

因此, 微直观为:

$$(h_1, h_2) \mapsto \partial_1 f(x, y)h_1 + \partial_2 f(x, y)h_2 = y \cos(xy)h_1 + x \cos(xy)h_2.$$

在 $(0, \frac{\pi}{2})$ 处, (h_1, h_2) 带来的变化为:

$$\begin{aligned} \Delta f(\mathbf{a}) &\approx \frac{\pi}{2} \cdot \cos\left(0 \cdot \frac{\pi}{2}\right)h_1 + 0 \cdot \cos\left(0 \cdot \frac{\pi}{2}\right)h_2 \\ &= \frac{\pi}{2}h_1 \end{aligned}$$

也就是说, 曲面在 $(0, \frac{\pi}{2})$ 处, 沿 x 轴增长 h_1 为函数值带来的变化约为 $\frac{\pi}{2}h_1$, 沿 y 轴增长 h_2 带来的变化则是 h_2 的高阶无穷小。

对于多个可微函数, 我们可以通过它们各自的全微变, 计算它们四则运算后的全微变。例如, 对于两个在 \mathbf{a} 处可微的二元函数 f 和 g , 以及常数 c , 我们有:

$$\begin{aligned} \Delta(f \pm g)(\mathbf{a}) &= \Delta f(\mathbf{a}) \pm \Delta g(\mathbf{a}), \\ \Delta(fg)(\mathbf{a}) &= f(\mathbf{a}) \cdot \Delta g(\mathbf{a}) + g(\mathbf{a}) \cdot \Delta f(\mathbf{a}), \\ \Delta(cf)(\mathbf{a}) &= c \cdot \Delta f(\mathbf{a}), \\ \Delta\left(\frac{f}{g}\right)(\mathbf{a}) &= \frac{g(\mathbf{a}) \cdot \Delta f(\mathbf{a}) - f(\mathbf{a}) \cdot \Delta g(\mathbf{a})}{g^2(\mathbf{a})} \quad (\text{当 } g(\mathbf{a}) \neq 0 \text{ 时}). \end{aligned}$$

这些性质的证明与方向微变的运算法则类似。

思考 2.1.1.

1. 如果二元函数在一点的偏微变存在, 但不可微, 这时, 依照偏微变构建的维直变换对应的平面和原曲面会有什么可能的关系? 举一些典型的例子来说明你的想法。

2. 如果 Oxy 平面上发生了基底变换, 原基底 $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ 变成了 $\mathbf{e}'_x, \mathbf{e}'_y, \mathbf{e}_z$, 这种变换如何影响可微二元函数的切平面的方程? 切平面是否改变?

习题 2.1.1.

1. 证明以下函数在指定点 \mathbf{a} 处可微, 并计算其全微变:

$$1). \quad f(x, y) = x^3 + y^3 - 3xy, \quad \mathbf{a} = (1, 1)$$

$$2). \quad g(x, y) = e^{x+y} \cos(x - y), \quad \mathbf{a} = (0, 0)$$

$$3). \quad h(x, y) = \ln(x^2 + y^2 + 1), \quad \mathbf{a} = (1, 0)$$

2. 证明以下函数在 $(0, 0)$ 处偏微变存在, 但不可微:

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

3. 设 $f(x, y) = x^2y + 2xy^2$, 利用全微变计算 $f(1.01, 1.98)$ 的近似值。

4. 证明本节最后, 函数的四则运算对应的全微变公式。

2.2 曲面上一点的梯度

我们已经看到, 如果二元函数 f 在点 \mathbf{a} 处可微, 那么有平直变换 $\mathcal{D}f(\mathbf{a})$, 使得当 \mathbf{h} 很小时,

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = \mathcal{D}f(\mathbf{a})(\mathbf{h}) + o(|\mathbf{h}|),$$

其中 $\mathcal{D}f(\mathbf{a})$ 由 f 在 \mathbf{a} 处的偏微变唯一确定:

$$\mathcal{D}f(\mathbf{a})(\mathbf{h}) = h_1 \partial_1 f(\mathbf{a}) + h_2 \partial_2 f(\mathbf{a}).$$

多元函数的微变是平直变换, 是不是就没办法用一个量来代表它了呢? 也不尽然。考虑平面经典内积的话, 可以把 $\mathcal{D}f(\mathbf{a})$ 表示为:

$$\forall \mathbf{h} = (h_1, h_2) \in \mathbb{R}^2, \quad \mathcal{D}f(\mathbf{a})(\mathbf{h}) = (\mathbf{h} | \nabla f(\mathbf{a})),$$

其中 $\nabla f(\mathbf{a}) = (\partial_1 f(\mathbf{a}), \partial_2 f(\mathbf{a}))$ 。我们把这个向量 $\nabla f(\mathbf{a})$ 叫做函数 f 在点 \mathbf{a} 处的梯度。

梯度有什么直观意义呢？考虑从点 \mathbf{a} 出发，沿单位向量 \mathbf{v} 方向移动微小距离 t 。函数值的变化为：

$$f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) = t \mathcal{D}f(\mathbf{a})(\mathbf{v}) + o(t) = t \cdot (\mathbf{v} | \nabla f(\mathbf{a})) + o(t).$$

根据内积不等式，

$$-|\mathbf{v}| \cdot |\nabla f(\mathbf{a})| \leq (\mathbf{v} | \nabla f(\mathbf{a})) \leq |\mathbf{v}| \cdot |\nabla f(\mathbf{a})|.$$

由于 \mathbf{v} 是单位向量， $|\mathbf{v}| = 1$ ，因此当 \mathbf{v} 与 $\nabla f(\mathbf{a})$ 同向时， $(\mathbf{v} | \nabla f(\mathbf{a}))$ 达到最大值 $|\nabla f(\mathbf{a})|$ ；当 \mathbf{v} 与 $\nabla f(\mathbf{a})$ 反向时，达到最小值 $-|\nabla f(\mathbf{a})|$ 。

梯度向量与方向微变有密切关系。 f 沿单位向量 \mathbf{v} 方向的微变为：

$$\begin{aligned} \partial_{\mathbf{v}} f(\mathbf{a}) &= (\mathbf{v} | \nabla f(\mathbf{a})) \\ &= |\mathbf{v}| \cdot |\nabla f(\mathbf{a})| \cdot \cos \theta \\ &= |\nabla f(\mathbf{a})| \cdot \cos \theta, \end{aligned}$$

其中 θ 是 \mathbf{v} 与 $\nabla f(\mathbf{a})$ 的夹角。这说明方向微变是梯度大小乘以夹角的余弦。方向微变表示沿此方向变动时，函数值的变化。因此，函数值的变化在梯度方向最快，而在垂直梯度方向最不明显。

为了帮助理解，可以把函数值变化量写成这样：

$$f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) = t \cdot |\mathbf{v}| \cdot |\nabla f(\mathbf{a})| \cdot \frac{(\mathbf{v} | \nabla f(\mathbf{a}))}{|\mathbf{h}| \cdot |\nabla f(\mathbf{a})|} + o(t) = t \cdot |\nabla f(\mathbf{a})| \cdot \cos \theta + o(t).$$

其中 θ 是向量 \mathbf{v} 和梯度向量 $\nabla f(\mathbf{a})$ 的夹角。直观来说， t 很小的时候，往任意方向移动 t 长度，函数值的变化量就和移动向量在梯度向量上的投影长度成正比。投影长度小于等于移动向量的原长度。如果移动的方向与梯度向量相同或相反，那么投影长度就是原长度，于是函数值变化得最快。可以理解为，在梯度方向上的微小移动，对函数值变化的“效率”是最高的。从点 \mathbf{a} 出发，沿着梯度方向 $\nabla f(\mathbf{a})$ 移动，函数值增长最快；沿着梯度的反方向移动，函数值下降最快。梯度的大小 $|\nabla f(\mathbf{a})|$ 表示函数在该点的最大增长率，梯度的方向是函数值局部增长最快的方向。

例题 2.2.1. 计算函数 $f(x, y) = x^2 + y^2$ 在点 $P = (1, 2)$ 处的梯度，解释其直观意义。

解答. 首先计算偏微变：

$$\partial_1 f(x, y) = 2x, \quad \partial_2 f(x, y) = 2y.$$

在点 $(1, 2)$ 处：

$$\nabla f(1, 2) = (\partial_1 f(1, 2), \partial_2 f(1, 2)) = (2, 4).$$

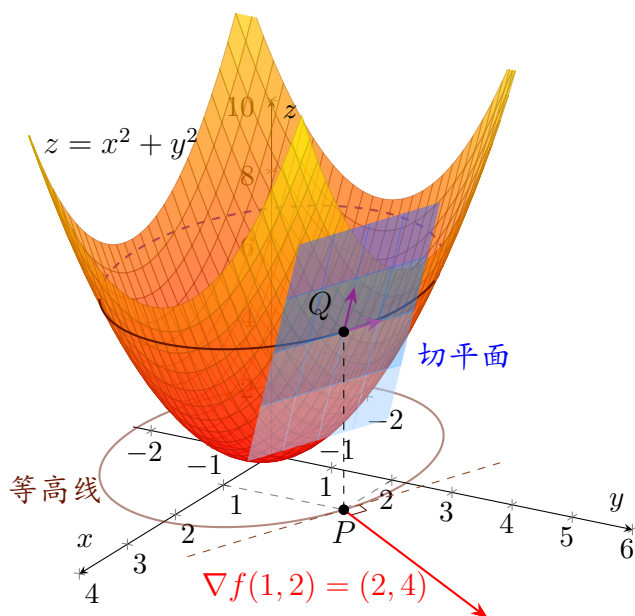
这说明在 $(1, 2)$ 处，函数 f 的增长最快方向是向量 $(2, 4)$ 的方向。而函数值在该方向的增长率为 $\sqrt{2^2 + 4^2} = 2\sqrt{5}$ 。直观上看，函数的曲面是一个旋转抛物面，方程为 $z = x^2 + y^2$ 。 $f(1, 2) = 5$ ，对应曲面上点 $Q = (1, 2, 5)$ ，切平面方程为

$$\gamma_{(1,2)} : z = 2(x - 1) + 4(y - 2) + 5 = 2x + 4y - 5.$$

梯度向量 $(2, 4)$ 是一个平面向量，它说明自变量 (Oxy 平面上的点) (x, y) 沿着 $(2, 4)$ 方向移动时， f 的值变化最快。我们可以作过 (x, y) 垂直于 Oxy 的竖直直线，研究它自 $(1, 2)$ 略微移动时，直线和曲面、切平面的交点的变化，来查看两者的增长方式。具体来说，比如从 $(1, 2)$ 移动到 $(1.1, 2.2)$ 时，直线和曲面的交点高度从 5 变为 6.05，而直线和切平面的交点高度从 5 变为 6。所以在切平面，梯度方向的向量为 $(1, 2, 10)$ 。而如果我们的移动垂直于 $(2, 4)$ ，比方说从 $(1, 2)$ 移动到 $(1.2, 1.9)$ (移动量为 $(0.2, -0.1)$)，则直线与切平面交点高度还是 5，没有变化 (直线和曲面交点高度变为 5.05)。我们可以把切平面写作经过 $\mathbf{u} = (1, 2, 5)$ ，以 $\mathbf{e}_1 = (1, 2, 10)$ 和 $\mathbf{e}_2 = (2, -1, 0)$ 为基底构成的平面：

$$\gamma_{(1,2)} = \{\mathbf{u} + s\mathbf{e}_1 + t\mathbf{e}_2 \mid s, t \in \mathbf{R}\}$$

沿 \mathbf{e}_2 方向，函数 f 变化最快，沿 \mathbf{e}_1 方向， f 变化最不明显 (实际上是等高线的切线)。



定理 2.2.1. 梯度的运算性质 设 \mathbb{R}^2 上的函数 f, g 在 \mathbf{a} 处可微, 则:

- $\nabla(f \pm g)(\mathbf{a}) = \nabla f(\mathbf{a}) \pm \nabla g(\mathbf{a})$
- 设 c 为常数, 则 $\nabla(cf)(\mathbf{a}) = c \cdot \nabla f(\mathbf{a})$
- $\nabla(f \cdot g)(\mathbf{a}) = g(\mathbf{a}) \cdot \nabla f(\mathbf{a}) + f(\mathbf{a}) \cdot \nabla g(\mathbf{a})$
- 若 $g(\mathbf{a}) \neq 0$, 则 $\nabla\left(\frac{f}{g}\right)(\mathbf{a}) = \frac{g(\mathbf{a}) \cdot \nabla f(\mathbf{a}) - f(\mathbf{a}) \cdot \nabla g(\mathbf{a})}{g^2(\mathbf{a})}$

证明: 这些性质可以从偏微变的运算性质直接推导。以乘积法则为例:

$$\begin{aligned}
 \nabla(fg)(\mathbf{a}) &= (\partial_1(fg)(\mathbf{a}), \partial_2(fg)(\mathbf{a})) \\
 &= (g(\mathbf{a}) \cdot \partial_1 f(\mathbf{a}) + f(\mathbf{a}) \cdot \partial_1 g(\mathbf{a}), g(\mathbf{a}) \cdot \partial_2 f(\mathbf{a}) + f(\mathbf{a}) \cdot \partial_2 g(\mathbf{a})) \\
 &= g(\mathbf{a}) \cdot (\partial_1 f(\mathbf{a}), \partial_2 f(\mathbf{a})) + f(\mathbf{a}) \cdot (\partial_1 g(\mathbf{a}), \partial_2 g(\mathbf{a})) \\
 &= g(\mathbf{a}) \cdot \nabla f(\mathbf{a}) + f(\mathbf{a}) \cdot \nabla g(\mathbf{a}).
 \end{aligned}$$

其他性质可类似证明。 □

例题 2.2.2. 设 $z = f(x, y)$ 由方程 $x^2 + y^2 + z^2 = 1$ 隐式定义, 计算梯度 $\nabla z(x, y)$ 。

解答. 将方程视为隐函数 $F(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$ 。对 x 和 y 求偏微变:

$$\partial_1 F + \partial_3 F \cdot \partial_1 z = 0$$

$$\partial_2 F + \partial_3 F \cdot \partial_2 z = 0$$

代入 $\partial_1 F = 2x$, $\partial_2 F = 2y$, $\partial_3 F = 2z$, 得:

$$\begin{aligned} 2x + 2z \cdot \partial_1 z &= 0 \quad \Rightarrow \quad \partial_1 z = -\frac{x}{z} \\ 2y + 2z \cdot \partial_2 z &= 0 \quad \Rightarrow \quad \partial_2 z = -\frac{y}{z} \end{aligned}$$

因此,

$$\nabla z(x, y) = \left(-\frac{x}{z}, -\frac{y}{z} \right) = -\frac{1}{z}(x, y).$$

从直观上看, 这个结果表示球面 $x^2 + y^2 + z^2 = 1$ 上任一点 (x, y, z) 处, 函数 $z = f(x, y)$ 的梯度指向原点在 Oxy 平面上的投影, 这与球面的直观性质一致。

需要注意的是, 梯度向量的存在依赖于函数在该点的可微性。我们已知函数在某点的偏微变存在, 并不代表函数在该点可微。如果函数在点 \mathbf{a} 处不可微, 那么即使能构造出形式上的梯度向量, 它也不能正确描述函数在该点附近的变化。例如, 考虑函数:

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

虽然 f 在 $(0, 0)$ 处的偏微变存在且均为 0, 但 f 在该点不可微。此时, 形式上的梯度向量 $\nabla f(0, 0) = (0, 0)$ 不能描述函数在原点附近的真实变化。比如, 考虑 $(1, k)$ 方向上函数的变化。令 $x = t, y = kt$ 。代入表达式得到

$$f(x, y) = \frac{t \cdot kt}{\sqrt{t^2 + k^2 t^2}} = \frac{k}{\sqrt{k^2 + 1}} \cdot t.$$

t 增大时, f 的值成比例增加。偏微分构成的 $(0, 0)$ 向量无法描述 f 在 $(0, 0)$ 附近的变化了。因此, 在使用梯度之前, 必须先验证函数在该点的可微性。

学习实函数时, 我们证明了中值定理。中值定理为我们搭建了函数局部性质与整体行为之间的桥梁。那么, 多元函数是否有类似的结论呢? 回顾实函数的中值定理, 我们看到, 中值定理描述函数在线段 $[a, b]$ 上的性质, 而二元函数的定义域则是平面的点集。对于二元函数来说, 连接定义域内 \mathbf{a}, \mathbf{b} 两点, 函数值从 \mathbf{a} 到 \mathbf{b} 的变化趋势, 是否也能用两点连线上的一点的微变性质表示呢?

作为讨论的前提, 我们希望函数在两点的连线上总有定义。为了方便说话, 我们引入凸集的概念:

定义 2.2.1. 凸集 如果集合 $D \subseteq \mathbb{R}^2$ 中任意两点 \mathbf{a}, \mathbf{b} 的连线段全部位于 D 内, 即对任意 $t \in [0, 1]$, 点 $t\mathbf{a} + (1-t)\mathbf{b}$ 都属于 D , 就说 D 是凸集。

凸集的直观意义是: 集合中任意两点之间的连线完全包含在集合内。常见的凸集包括圆盘、矩形区域等。假设定义域为凸集, 或者在定义域中划出一个凸集作为邻域, 我们就可以讨论其中两点的连线, 以及中值定理了。中值定理的整体想法是: 函数在两点间的变化量, 等于函数在两点连线上某点的梯度与位移向量的内积, 换句话说, 我们总能找到某点的梯度来代表整体的趋势。这与一元函数中值定理的思想一脉相承, 只是将标量乘法扩展为向量内积。

定理 2.2.2. 设二元函数 f 在凸集 $D \subseteq \mathbb{R}^2$ 上可微, $\mathbf{a}, \mathbf{b} \in D$ 。则存在点 \mathbf{c} 在线段 \mathbf{ab} 上, 使得:

$$f(\mathbf{b}) - f(\mathbf{a}) = (\nabla f(\mathbf{c}) | \mathbf{b} - \mathbf{a}).$$

证明: 构造辅助函数 $g: [0; 1] \rightarrow \mathbb{R}$:

$$t \mapsto f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})).$$

由于 D 是凸集, g 在 $[0; 1]$ 上有定义。由于 f 在 D 上可微, 因此在 D 上连续。于是 g 在 $[0; 1]$ 上连续。计算 g 的微变:

$$\begin{aligned}\partial g(t) &= \lim_{h \rightarrow 0} \frac{g(t+h) - g(t)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + (t+h)(\mathbf{b} - \mathbf{a})) - f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))}{h}\end{aligned}$$

$\forall t \in (0; 1)$, f 在 $\mathbf{a} + t(\mathbf{b} - \mathbf{a})$ 处可微, 因此对足够小的 h :

$$f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}) + h(\mathbf{b} - \mathbf{a})) - f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) = (\nabla f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) | h(\mathbf{b} - \mathbf{a})) + o(h).$$

因此 g 在 $(0; 1)$ 上可微, 且其微变可以写为:

$$\partial g(t) = (\nabla f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) | \mathbf{b} - \mathbf{a}).$$

由于 g 在 $[0; 1]$ 上连续, 在 $(0; 1)$ 内可微, 根据一元中值定理, 存在 $\theta \in (0; 1)$, 使得:

$$g(1) - g(0) = \partial g(\theta) \cdot (1 - 0) = \partial g(\theta).$$

即:

$$f(\mathbf{b}) - f(\mathbf{a}) = (\nabla f(\mathbf{a} + \theta(\mathbf{b} - \mathbf{a})) | \mathbf{b} - \mathbf{a}).$$

令 $\mathbf{c} = \mathbf{a} + \theta(\mathbf{b} - \mathbf{a})$, 则 \mathbf{c} 在线段 \mathbf{ab} 上, 且:

$$f(\mathbf{b}) - f(\mathbf{a}) = (\nabla f(\mathbf{c}) | \mathbf{b} - \mathbf{a}).$$

□

直观上, 我们可以这样理解: 用 \mathbf{ab} 所在的竖直平面截二元函数 f 的曲面, 得到交线。这个交线可以看作一元函数的图像。具体来说, 它对应的函数就是:

$$t \mapsto f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})).$$

因此, 这个一阶光滑函数满足的中值定理就是我们所需要的二元函数中值定理。

中值定理揭示了曲面在两点间的变化与梯度的关系。若将函数 f 视为描述地形高度的函数, 则 $f(\mathbf{b}) - f(\mathbf{a})$ 表示从点 \mathbf{a} 到 \mathbf{b} 的高度变化。中值定理告诉我们, 这一变化量等于在路径上某点 \mathbf{c} 处, 梯度 $\nabla f(\mathbf{c})$ 与位移 $\mathbf{b} - \mathbf{a}$ 的内积, 即梯度在位移方向上的投影乘以位移的长度。

思考 2.2.1.

1. 考虑二元函数 $f(x, y) = \sqrt{x^2 + y^2}$, 研究它在 $(0, 0)$ 处的梯度。
2. 梯度与我们前面定义的切平面有何关系? 从直观角度解释梯度向量如何与切平面相互作用。

习题 2.2.1.

1. 计算以下函数在指定点处的梯度:

- 1). $(x, y) \mapsto x^3 - 3xy^2, \quad (1, 1)$
- 2). $(x, y) \mapsto \ln(x^2 + y^2), \quad (1, 0)$
- 3). $(x, y) \mapsto e^{x-y} \cos(x + 2y), \quad (0, 0)$

2. 证明: 设 f 是径向函数, 即存在实函数 $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ 使得 $f(x, y) = \varphi(\sqrt{x^2 + y^2})$ 。若 f 在原点外可微, 证明梯度 $\nabla f(x, y)$ 与 (x, y) 共轴。

3. 设 $f(x, y) = x^2 + 2y^2 + 2xy + 3x + 4y + 5$ 。找出点 \mathbf{a} 使得 f 在 \mathbf{a} 处的梯度为零向量。这样的点有什么直观意义?

4. 考察函数 $f(x, y) = (|x| + |y|)^2$ 。

4.1. f 在 $(1, 0)$ 处是否连续? 是否可微? 是否有偏微变?

4.2. 考察 f 在 $(0, 1)$ 处的情况。

4.3. 找出使 f 不可微的点的集合。把 f 可微的点合理分成几个区域。

4.4. f 的图像是怎样的曲面?

2.3 二元函数的高次微变

2.4 曲面的极值

第三章 离散随机变量

3.1 随机变量的性质

3.2 正态分布

3.3 大数定律

3.4 中心极限定理

Appendices

第一章 向量函数的变化

1.1 向量函数的极限与连续性质

从实函数的情形，我们知道，函数有极限，其复合函数不一定有极限。向量函数同样如此。不过，附加了连续性后，就可以保证复合函数的极限存在了。

定理 1.1.1. 复合函数的极限 设有向量函数 $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^p$, f 在点 $\mathbf{a} \in \mathbb{R}^m$ 处有极限 $\mathbf{b} \in \mathbb{R}^n$, g 在 \mathbf{b} 处连续，有极限 $\mathbf{c} = g(\mathbf{b}) \in \mathbb{R}^p$, 则复合函数 $g \circ f$ 在 \mathbf{a} 处有极限 \mathbf{c} 。

证明： 根据极限的定义，我们要证明，对任意 $r > 0$ ，都存在 $d > 0$ ，使得只要 $\mathbf{x} \in \mathbb{R}^m$ 到 \mathbf{a} 的距离小于 d ， $g(f(\mathbf{x}))$ 到 \mathbf{c} 的距离就小于 r 。

基于经典空间距离的等价性，我们只证明经典距离 ℓ_2 的情况。首先，由于 g 在 \mathbf{b} 处连续，按照定义，存在 $r_1 > 0$ ，使得对 $\|\mathbf{y} - \mathbf{b}\| < r_1$ 的 $\mathbf{y} \in \mathbb{R}^n$ ，总有 $\|g(\mathbf{y}) - g(\mathbf{b})\| < r$ 。而由于 f 在 \mathbf{a} 处有极限，按照定义，存在 $d_1 > 0$ ，使得只要 $\|\mathbf{x} - \mathbf{a}\| < d_1$ ，就有 $\|f(\mathbf{x}) - \mathbf{b}\| < r_1$ 。这说明

$$\|g(f(\mathbf{x})) - \mathbf{c}\| = \|g(f(\mathbf{x})) - g(\mathbf{b})\| < r.$$

□ 注意：

- 证明中使用 $\|\cdot\|$ 标记不同空间中的 ℓ_2 长度，要注意分辨。

- g 在 \mathbf{b} 处的连续性是确保不出现唯一的反例： g 在 \mathbf{b} 处的值和 g 在 \mathbf{b} 处的极限不相等，但 f 在任意接近 \mathbf{a} 时总取值为 \mathbf{b} 。实际上，我们可以用更直接的条件避免这个例外情况出现，从而保证复合函数极限存在。具体参见习题 2。

习题 1.1.1.

2. 设有向量函数 $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^p$, f 在点 $\mathbf{a} \in \mathbb{R}^m$ 处有极限 $\mathbf{b} \in \mathbb{R}^n$, g 在 \mathbf{b} 处有极限 $\mathbf{c} \in \mathbb{R}^p$ 。证明：若存在 \mathbf{a} 的去心邻域 U , 使得 f 在 U 中的值总不是 \mathbf{b} , 则复合函数 $g \circ f$ 在 \mathbf{a} 处有极限 \mathbf{c} 。

2. 根据极限的定义，我们要证明，对任意 $r > 0$, 都存在 $d > 0$, 使得只要 $\mathbf{x} \in \mathbb{R}^p$ 到 \mathbf{a} 的距离小于 d , $g(f(\mathbf{x}))$ 到 \mathbf{c} 的距离就小于 r 。

基于经典空间距离的等价性，我们只证明经典距离 ℓ_2 的情况。首先，由于 g 在 \mathbf{b} 处有极限，按照定义，存在 $r_1 > 0$, 使得对 $\|\mathbf{y} - \mathbf{b}\| < r_1$ 的 $\mathbf{y} \in \mathbb{R}^n$, 总有 $\|g(\mathbf{y}) - \mathbf{c}\| < r$ 。而由于 f 在 \mathbf{a} 处有极限，按照定义，存在 $d_1 > 0$, 使得只要 $\|\mathbf{x} - \mathbf{a}\| < d_1$, 就有 $\|f(\mathbf{x}) - \mathbf{b}\| < r_1$ 。这说明

$$\|g(f(\mathbf{x})) - \mathbf{c}\| = \|g(f(\mathbf{x})) - g(\mathbf{b})\| < r.$$

1.2 二元函数的局部变化

定理 1.2.1. 可微的充分条件 如果二元函数 f 在点 \mathbf{a} 的某个邻域内有定义，且：

- f 在 \mathbf{a} 处的偏微变 $\partial_1 f$ 和 $\partial_2 f$ 存在；
- 偏微变函数 $\partial_1 f$ 和 $\partial_2 f$ 在 \mathbf{a} 处连续，

那么 f 在 \mathbf{a} 处可微。

证明： 我们只需证明存在平直变换 L ，使得：

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L(\mathbf{h})}{|\mathbf{h}|} = 0.$$

令 $L(\mathbf{h}) = h_1 \partial_1 f(\mathbf{a}) + h_2 \partial_2 f(\mathbf{a})$ 。考虑：

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2).$$

我们可以把这个差分分为两部分：

$$[f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2)] + [f(a_1, a_2 + h_2) - f(a_1, a_2)].$$

对第一部分，固定 $a_2 + h_2$ ，把 f 看作 a_1 的函数，根据一元函数的中值定理，存在 $\theta_1 \in (0, 1)$ ，使得：

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2) = h_1 \partial_1 f(a_1 + \theta_1 h_1, a_2 + h_2).$$

对第二部分，固定 a_1 ，把 f 看作 a_2 的函数，同样根据中值定理，存在 $\theta_2 \in (0, 1)$ ，使得：

$$f(a_1, a_2 + h_2) - f(a_1, a_2) = h_2 \partial_2 f(a_1, a_2 + \theta_2 h_2).$$

因此：

$$\begin{aligned} f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L(\mathbf{h}) &= h_1 [\partial_1 f(a_1 + \theta_1 h_1, a_2 + h_2) - \partial_1 f(a_1, a_2)] \\ &\quad + h_2 [\partial_2 f(a_1, a_2 + \theta_2 h_2) - \partial_2 f(a_1, a_2)]. \end{aligned}$$

于是：

$$\begin{aligned} &\frac{|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L(\mathbf{h})|}{|\mathbf{h}|} \\ &\leq \frac{|h_1|}{|\mathbf{h}|} \cdot |\partial_1 f(a_1 + \theta_1 h_1, a_2 + h_2) - \partial_1 f(a_1, a_2)| \\ &\quad + \frac{|h_2|}{|\mathbf{h}|} \cdot |\partial_2 f(a_1, a_2 + \theta_2 h_2) - \partial_2 f(a_1, a_2)| \\ &\leq |\partial_1 f(a_1 + \theta_1 h_1, a_2 + h_2) - \partial_1 f(a_1, a_2)| \\ &\quad + |\partial_2 f(a_1, a_2 + \theta_2 h_2) - \partial_2 f(a_1, a_2)|. \end{aligned}$$

当 $\mathbf{h} \rightarrow \mathbf{0}$ 时, $(a_1 + \theta_1 h_1, a_2 + h_2) \rightarrow (a_1, a_2)$, $(a_1, a_2 + \theta_2 h_2) \rightarrow (a_1, a_2)$ 。由于 $\partial_1 f$ 和 $\partial_2 f$ 在 \mathbf{a} 处连续, 上式右边趋于 0。因此:

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L(\mathbf{h})}{|\mathbf{h}|} = 0.$$

这证明 f 在 \mathbf{a} 处可微。 □

例题 1.2.1. 考察函数 $f(x, y) = x^2 y + xy^2$ 在点 $(1, 2)$ 处的可微性。

解答. 首先计算偏微变:

$$\partial_1 f(x, y) = 2xy + y^2, \quad \partial_2 f(x, y) = x^2 + 2xy.$$

在点 $(1, 2)$ 处:

$$\partial_1 f(1, 2) = 2 \cdot 1 \cdot 2 + 2^2 = 8, \quad \partial_2 f(1, 2) = 1^2 + 2 \cdot 1 \cdot 2 = 5.$$

偏微变函数 $\partial_1 f$ 和 $\partial_2 f$ 都是多项式函数, 因此连续。根据可微的充分条件, f 在 $(1, 2)$ 处可微。

f 在 $(1, 2)$ 处的微变换 L 满足:

$$L(h_1, h_2) = h_1 \partial_1 f(1, 2) + h_2 \partial_2 f(1, 2) = 8h_1 + 5h_2.$$

因此, f 在 $(1, 2)$ 附近的线性近似为:

$$f(1 + h_1, 2 + h_2) \approx f(1, 2) + 8h_1 + 5h_2 = 6 + 8h_1 + 5h_2.$$

定理 1.2.2. 复合向量函数的微变