关于高阶偏导和梯度的说明

陈轶钊

2024 年 12 月 18 日

说明

这篇笔记包含了下面的一些内容：我们给梯度的方向做另一种解释，这种解释使得隐函数定理没有那么“不自然”；我们给出高阶偏导数可以交换次序的一个充分条件。大家可以自行选择内容阅读。

目录

1 高阶导数可交换次序的充分条件

我们参照伍胜健的《数学分析 III》，给出并证明高阶导数可交换次序的一个充分条件。这一条件可以简记为“二阶偏导连续推出可交换偏导次序”。

定理 1.1 设 D 为 $\mathbb{R}^2$ 中开区域（一般的 $\mathbb{R}^n$ 中是类似的），并设函数 $f: D \to \mathbb{R}$ 在 D 上处处有二阶偏导数。那么有：如果

$$\frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y \partial x}$$

中某一个连续，那么它们两者相等。

这里我们求偏导的次序和《高等数学》相同，即 $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y}(\frac{\partial f}{\partial x})$.

这一定理的证明并不是特别重要，大家能够正确使用这一定理即可。当然，这一定理可以直接使用。

证明 1 不妨设 $\frac{\partial^2 f}{\partial y\partial x}$ 连续。

对任意一点 (x,y)，我们考虑关于 $\Delta x,\Delta y$ 的函数

$$I(\Delta x,\Delta y)=\frac{1}{\Delta x\Delta y}\big((f(x+\Delta x,y+\Delta y)-f(x,y+\Delta y))-(f(x+\Delta x,y)-f(x,y))\big)$$

我们先将 Δx 视为常数，并记 $g(u)=f(x+\Delta x,u)-f(x,u)$. 那么函数 $I(\Delta x,\Delta y)$ 可以写成：

$$I(\Delta x,\Delta y)=\frac{1}{\Delta x\Delta y}\big(g(y+\Delta y)-g(y)\big).$$

固定 x，对 g 和变量 u 使用 *Lagrange* 中值定理可得：

$$I(\Delta x,\Delta y)=\frac{1}{\Delta x}g'(\eta)=\frac{\frac{\partial f}{\partial y}(x+\Delta x,\eta)-\frac{\partial f}{\partial y}(x,\eta)}{\Delta x}$$

其中 η 落在 y 和 $y+\Delta y$ 之间。对上面的式子再用一次 *Lagrange* 中值定理，就有

$$I(\Delta x,\Delta y)=\frac{\partial^2 f}{\partial y\partial x}(\xi,\eta).$$

其中 ξ 在 $x,x+\Delta x$ 之间，η 在 $y,y+\Delta y$ 之间，这样根据 $\frac{\partial^2 f}{\partial y\partial x}$ 的连续性，我们有：

$$\lim_{\Delta y\to 0}\lim_{\Delta x\to 0}I(\Delta x,\Delta y)=\lim_{\Delta y\to 0}\lim_{\Delta x\to 0}\frac{\partial^2 f}{\partial y\partial x}(\xi,\eta)=\lim_{\Delta y\to 0}\frac{\partial^2 f}{\partial y\partial x}(x,\eta)=\frac{\partial^2 f}{\partial y\partial x}(x,y). \tag{1}$$

另一方面，注意到有：

$$I(\Delta x,\Delta y)=\frac{1}{\Delta y}\left(\frac{f(x+\Delta x,y+\Delta y)-f(x,y+\Delta y)}{\Delta x}-\frac{f(x+\Delta x,y)-f(x,y)}{\Delta x}\right)$$

如果我们令 $\Delta x\to 0$，右边就会出现关于 x 的一阶偏导数，进而有

$$\lim_{\Delta y\to 0}\lim_{\Delta x\to 0}I(\Delta x,\Delta y)=\lim_{\Delta y\to 0}\frac{1}{\Delta y}\left(\frac{\partial f}{\partial x}(x,y+\Delta y)-\frac{\partial f}{\partial x}(x,y)\right)=\frac{\partial^2 f}{\partial x\partial y}(x,y)$$

将这一等式和式 *(1)* 比较就可知，定理成立。

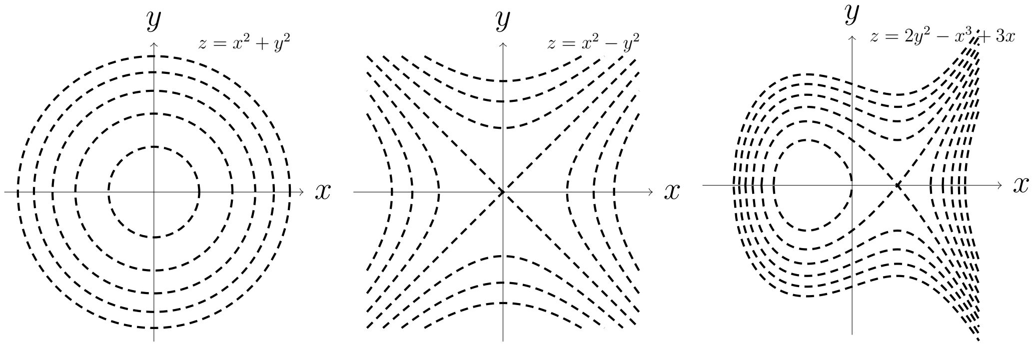
2 关于梯度的解释

教材上将梯度的方向解释为“函数的值上升最快的方向”，将它的值解释为“函数值的上升速率”。这一解释在数学上是严格的，但我认为并不直观。我在这里给出梯度的方向的另一种解释。

我们使用“等高线”来描述函数。一个二元函数 $f(x,y)$ 的“等高线”是下面的一系列集合[1]：

$$f^{-1}(c) = \{(x,y) \mid f(x,y) = c\}, \quad \forall c \in \mathbb{R}.$$

下面的图形展示了一些函数的等高线。



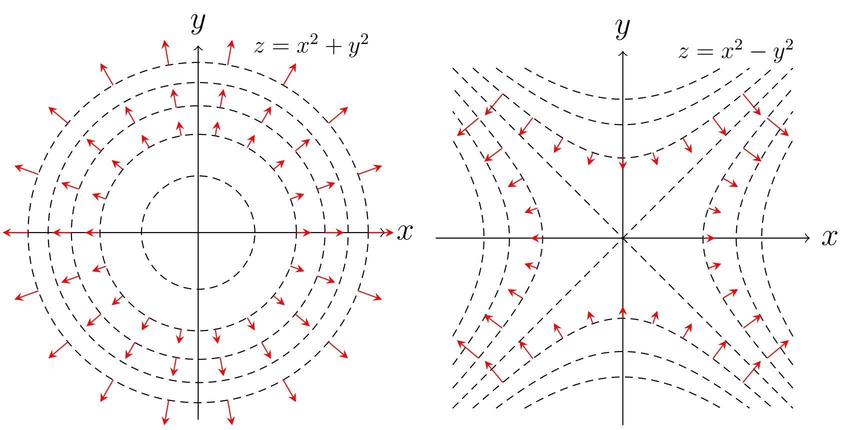
利用函数的等高线，我们可以给出函数梯度的另一种描述方式。

命题 2.1 对于一个连续可微的函数 $f(x,y)$，它在点 (x_0,y_0) 处的梯度的方向是等高线的法向（或者说垂直于等高线的切线）。

对于一些简单的情况，我们可以依靠计算直接验证这件事。例如对一个二次函数 $f(x,y) = ax^2 + by^2$，它的等高线是一条二次曲线 $ax^2 + by^2 = c$. 经过上面一点 (x_0,y_0) 的切线是

$$\ell : ax_0x + by_0y = c$$

它的法向量是 (ax_0, by_0)，和 f 的梯度 $\operatorname{grad} f(x_0,y_0) = (2ax_0, 2by_0)$ 正好平行。下面两张图分别展示了在 $x^2 + y^2$ 和 $x^2 - y^2$ 的几条等高线上的梯度，注意等高线越密集的地方梯度越大（这也是为什么我们不画出所有等高线上的梯度）。



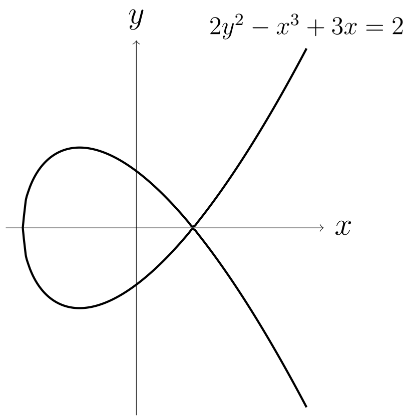
[1]对一般的多元函数 $f : D \to \mathbb{R}$，一般将 $f^{-1}(c)$ 称为水平集 (level set)。

尽管这一定理的陈述相当直观，但证明起来却有些麻烦——严格的数学证明需要给出等高线 $f^{-1}(c)$ 的一个参数方程并计算它的切向量，而这会用到隐函数定理。所以我们这里不会给出详细的证明过程。

2.1 等高“线”

严格来讲，上面的讨论有一个疏漏：我们过早地假设了 $f^{-1}(c)$ 真的就是一条“线”。实际上我们的水平集 $f^{-1}(c)$ 未必为一条曲线。

在前面给出的等高线的例子里就有反例，如对 $f(x,y)=x^2-y^2$，它经过原点的等高线实际上是两条直线，这条等高线在 $(0,0)$ 处没有办法定义法向量（不过好消息是，原点处的梯度等于 $\overrightarrow{0}.$）。又比如对于 $f(x,y)=2y^2-x^3+3x$



它的水平集 $f^{-1}(2)$ 也不完全是一条曲线。它在点 $(1,0)$ 处没有办法定义法向量。

这些反例的存在使得一个形如下面的定理变得非常有用：

当连续可微函数 $f(x,y)$ 在一点 (x_0,y_0) 处满足某某性质时，f 经过这点的等高线 $f^{-1}(c)$ 在这一点的某个邻域上是一条“曲线”。

这里我们省略了一些技术细节，比如“曲线”的严格数学定义。而隐函数定理正好是这样一个形状的定理，它告诉了我们：如果函数 $f(x,y)$ 在点 P 处的梯度不为 0 的话，这个函数在 P 处的等高线的确是一条“曲线”。事实上，隐函数定理还告诉了我们更多的信息（所以隐函数定理的陈述有些复杂）：这条等高线可以被看作某个函数的图像，并且我们可以计算等高线的切向量。