高阶微分：从入门到放弃

Jul 8, 2019 | 分析代数

文章目录

本篇文字主要用来测试在 Hexo 环境下的 AMScd 包，顺便谈谈高阶微分这个东西。

微分的概念一直是微积分教学中的难点，因为学生所掌握的数学工具不够，所以无法精确地理解这个概念。高阶微分的问题则更大：有一些书定义了高阶微分，但其运算规则很不清楚，并且也没有给出什么应用，看起来就像是一个无用的概念；有一些书则因为高阶微分不具备一阶微分的那种形式不变性，所以拒绝这个概念。本文将分析这其中的问题，给出具有形式不变性的高阶微分的定义，并通过例子说明，为什么这个概念在现有的微积分/数学分析体系中其实是没有必要引入的。

高阶微分的历史与现状

莱布尼茨在 1675 年引入导数和微分的符号：

$$y_x = \frac{\mathrm{d}y}{\mathrm{d}x} \qquad \Leftrightarrow \qquad \mathrm{d}y = y_x \mathrm{d}x,$$

高阶导数和高阶微分则可类似地定义，比如：

$$y_{xx} = \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} \qquad \Leftrightarrow \qquad \mathrm{d}^2 y = y_{xx} \mathrm{d}x^2.$$

我们现在经常强调二阶导数的记号不可以看成两个东西相除，而应该把 $\frac{\mathrm{d}}{\mathrm{d}x}$ 视为一个整体的算子，二阶导数则是这个算子在 y 上作用两次的结果。但是古人完全不是这么想的。在莱布尼茨看来，y_{xx} 是一个（通常的）数，$\mathrm{d}x^2$ 是一个（无穷小）数，把这两个数相乘就会得到另一个（无穷小）数 $\mathrm{d}^2 y$，有什么不可以的呢？

拉格朗日在他 1811 年的《分析力学》中大量使用了微分和高阶微分的记号。比如我随手一翻，就能找到这样的公式：

$$(X\,\mathrm{d}x+Y\,\mathrm{d}y+Z\,\mathrm{d}z)\,\frac{\mathrm{d}m}{\mathrm{d}s}+K\,\frac{\mathrm{d}x\,\mathrm{d}^4x+\mathrm{d}y\,\mathrm{d}^4y+\mathrm{d}z\,\mathrm{d}^4z}{\mathrm{d}s^4}=\mathrm{d}\lambda,$$

或者这样的公式：

$$\begin{aligned}g&=\left(\mathrm{d}x+\mathrm{d}^2x\right)^2+\left(\mathrm{d}y+\mathrm{d}^2y\right)^2+\left(\mathrm{d}z+\mathrm{d}^2z\right)^2\\&=\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2+2\left(\mathrm{d}x\,\mathrm{d}^2x+\mathrm{d}y\,\mathrm{d}^2y+\mathrm{d}z\,\mathrm{d}^2z\right)+\mathrm{d}^2x^2+\mathrm{d}^2y^2+\mathrm{d}^2z^2\\&=\mathrm{d}s^2+2\,\mathrm{d}s\,\mathrm{d}^2s+\mathrm{d}^2x^2+\mathrm{d}^2y^2+\mathrm{d}^2z^2.\end{aligned}$$

（别问我这都是啥意思，我也不知道。）

柯西在他 1823 年的分析学教科书中也介绍了微分和高阶微分的运算法则。通过他介绍的运算法则可以看出，那个年代的人们是把高阶微分和高阶导数视为同一个东西的不同写法的：n 阶微分除以 $\mathrm{d}x^n$ 就是 n 阶导数、n 阶导数乘以 $\mathrm{d}x^n$ 就是 n 阶微分。在这种观点下做计算，一定要牢记谁是自变量 x、谁是因变量 y，因为自变量与因变量的高阶微分法则是不一样的（见下文）。古尔萨在他 1904 年的数学分析教科书中仍然沿用着这种观点。

高阶微分和一阶微分有一个重要的区别，那就是在进行关于一阶微分的计算时，不必关心谁是自变量、谁是因变量，因为一阶微分具有所谓的“形式不变性”，高阶微分则没有这个性质。具体地说，考虑一个复合函数 $y=y(u(x))$，我们有

$$\mathrm{d}y=y_u\mathrm{d}u=y_u\left(u_x\mathrm{d}x\right)=\left(y_uu_x\right)\mathrm{d}x=y_x\mathrm{d}x,$$

即无论是用 x 做自变量还是用 u 做自变量，$\mathrm{d}y$ 都具有相同的形式。这就是所谓的形式不变性，它的几何意义是，一阶微分形式是一种与坐标系的选取无关的、整体的几何对象。需要特别指出的是，在上面这个计算中，我们用到了复合函数一阶导数的链式法则，在下文中我们会看到，在定义了正确的高阶微分之后，证明其形式不变性时也要用到复合函数的高阶导数的链式法则，即所谓的 Faà di Bruno 公式。

对于高阶微分，我们有：

$$\begin{aligned}y_{uu}\mathrm{d}u^2&=y_{uu}(u_x\mathrm{d}x)^2=\left(y_{uu}(u_x)^2\right)\mathrm{d}x^2,\\y_{xx}\mathrm{d}x^2&=\left(y_{uu}(u_x)^2+y_uu_{xx}\right)\mathrm{d}x^2.\end{aligned}$$

因为选取的变量不同，后者比前者多了一项 $y_uu_{xx}\mathrm{d}x^2$，所以我们说高阶微分不具有形式不变性，或者说高阶微分是与坐标系的选取有关的，它不是一个好的几何对象。

一种补救的办法是将二阶微分的定义修改为：

$$\mathrm{d}^2y=y_{xx}\mathrm{d}x^2+y_x\mathrm{d}^2x,$$

然后规定，当 x 为自变量时，$\mathrm{d}^2x=0$。这种新的定义具有形式不变性：

$$\begin{aligned}\mathrm{d}^2y &= y_{uu}\mathrm{d}u^2+y_u\mathrm{d}^2u\\ &= y_{uu}(u_x\mathrm{d}x)^2+y_u\left(u_{xx}\mathrm{d}x^2+u_x\mathrm{d}^2x\right)\\ &= \left(y_{uu}(u_x)^2+y_uu_{xx}\right)\mathrm{d}x^2+(y_uu_x)\,\mathrm{d}^2x\\ &= y_{xx}\mathrm{d}x^2+y_x\mathrm{d}^2x\end{aligned}$$

高木贞治在 1938 年写的教科书《解析概论》中即采用这种讲法。但是这种补救并没有解决本质的问题：首先，它仍然需要指明谁是自变量，所以不是一个整体的几何定义；其次，如果把定义式两边除以比如 $\mathrm{d}t^2$，这就是复合函数的二阶导数计算公式，所以这种定义并没有跳出高阶微分就是高阶导数的等价写法这个旧观点。

另外，在所有之前关于高阶微分的讨论中，没有人试图去说明微分和高阶微分到底是什么东西，这一点到了有了公理集合论之后的二十世纪变得尤其让人不可接受。1935 年，阿达马专门撰文批判高阶微分的记号（据说庞加莱也持类似的观点）：

> Enfin, que signifie ou que représente légalité
>
> $$\mathrm{d}^2z=r\,\mathrm{d}x^2+2\,s\,\mathrm{d}x\,\mathrm{d}y+t\,\mathrm{d}y^2?$$
>
> A mon avis, rien du tout.
>
> 最后，这个方程是什么意思或者想表达什么？在我看来，什么都没有。

在现在常见的微积分/数学分析教科书中，有一些书仍采用柯西–古尔萨或高木贞治的讲法，即给出高阶微分的定义（或修改过的定义），指出其不具有一阶微分那样的形式不变性，然后就没有下文了。在这些书中，为什么不变性不再成立、应该如何补救、有什么应用等问题全都没有涉及。在另一些比较现代的数学分析书中，作者们则选择完全不提高阶微分，这并不会影响后续的内容。

在接下来的几节中，我们将首先回顾对（一阶）微分这个概念的现代处理方式，然后讨论高阶微分的定义，最后再通过例子说明高阶微分这个概念为什么是可以从分析学中去掉的。

一阶微分的商空间定义

设 $D\subseteq\mathbb{R}^n$ 是一个区域，$f\in C(D,\mathbb{R}^m)$，$x_0\in D$。我们说 f 在 x_0 可微，如果存在一个线性映射 $A:\mathbb{R}^n\to\mathbb{R}^m$，使得当 $x\to x_0$ 时，有

$$f(x)=f(x_0)+A(x-x_0)+o(x-x_0),$$

这个定义式可等价地写为

$$f(x)-f(x_0)+o(x-x_0)=A(x-x_0+o(x-x_0)),$$

若引入记号 $f_x(x_0) := A$，以及

$$\mathrm{d}f(x_0) := f(x) - f(x_0) + o(x - x_0), \quad \mathrm{d}x(x_0) := x - x_0 + o(x - x_0),$$

则可得一阶微分的通常定义：

$$\mathrm{d}f(x_0) = f_x(x_0)\,\mathrm{d}x(x_0).$$

小 o 记号 $o(x - x_0)$ 其实是 $C(D, \mathbb{R}^m)$ 的一个线性子空间：

$$o(x - x_0) = \left\{ h \in C(D, \mathbb{R}^m) \,\middle|\, \lim_{x \to x_0} \frac{\|h(x)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} = 0 \right\},$$

记号 $\mathrm{d}f(x_0) = f(x) - f(x_0) + o(x - x_0)$ 则表明 $\mathrm{d}f(x_0)$ 是一个商空间中的元素。定义

$$\mathfrak{m}_{x0}^k = \left\{ h \in C(D, \mathbb{R}^m) \,\middle|\, \lim_{x \to x_0} \frac{\|h(x)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}^k} = 0 \right\},$$

则有

$$\mathfrak{m}_{x0}^0 = \{ h \in C(D, \mathbb{R}^m) \mid h(x_0) = 0 \}, \quad \mathfrak{m}_{x0}^1 = o(x - x_0),$$

于是微分 $\mathrm{d}f(x_0)$ 所在的商空间就是 $\Omega_{x0} := \mathfrak{m}_{x0}^0 / \mathfrak{m}_{x0}^1$。

微分运算 $\mathrm{d} : C(D, \mathbb{R}^m) \to \Omega_{x_0}$ 可以定义为如下两个运算的复合：

$$\mathrm{d} = \pi \circ \Delta : C(D, \mathbb{R}^m) \to \mathfrak{m}_{x0}^0 \to \Omega_{x0},$$

其中 Δ 是 x_0 处的差分运算：

$$\Delta : C(D, \mathbb{R}^m) \to \mathfrak{m}_{x0}^0, \quad f(x) \mapsto f(x) - f(x_0),$$

π 则是从 $\mathfrak{m}_{x0}^0$ 到其商空间 $\Omega_{x0} = \mathfrak{m}_{x0}^0 / \mathfrak{m}_{x0}^1$ 的投影：

$$\pi : \mathfrak{m}_{x0}^0 \to \Omega_{x0}, \quad h(x) \mapsto h(x) + \mathfrak{m}_{x0}^1.$$

现在考虑形式不变性。设 $E \subseteq \mathbb{R}^l$ 是另一个区域，$\phi : D \to E$ 连续可微，定义 ϕ 的拉回映射：

$$\phi^* : C(E, \mathbb{R}^m) \to C(D, \mathbb{R}^m), \quad f(y) \mapsto f(\phi(x)).$$

设 $y_0 = \phi(x_0)$，则有：

$$\phi^*\left(\mathfrak{m}_{y0}^k\right) \subseteq \mathfrak{m}_{x0}^k, \quad k = 0, 1, 2, \ldots$$

它又可推出如下图表是交换的：

$$\begin{tikzcd}
C(E, \mathbb{R}^m) \arrow[r, "\Delta"] \arrow[d, "\phi^*"] \arrow[dr, phantom, "\circlearrowright"] & \mathfrak{m}_{y0}^0 \arrow[r, "\pi"] \arrow[d, "\phi^*"] \arrow[dr, phantom, "\circlearrowright"] & \Omega_{y0} \arrow[d, "\phi^*"] \\
C(D, \mathbb{R}^m) \arrow[r, "\Delta"] & \mathfrak{m}_{x0}^0 \arrow[r, "\pi"] & \Omega_{x0}
\end{tikzcd}.$$

另一方面，若 $f \in C^1(E, \mathbb{R}^m)$，在上图中先横走再竖走给出

$$\begin{aligned}\phi^* \mathrm{d} f(y_0) &= \phi^*\left(f_y(y_0)(y-y_0)+\mathfrak{m}_{y0}^1\right) \\ &= f_y\left(\phi(x_0)\right)\left(\phi(x)-\phi(x_0)\right)+\mathfrak{m}_{x0}^1 \\ &= f_y\left(\phi(x_0)\right)\mathrm{d}\phi(x_0),\end{aligned}$$

而先竖走再横走则给出

$$\begin{aligned}\mathrm{d}\phi^* f(y_0) &= \mathrm{d}\left(f\circ\phi(x)\right)(x_0) \\ &= (f\circ\phi)_x(x_0)(x-x_0)+\mathfrak{m}_{x0}^1 \\ &= (f\circ\phi)_x(x_0)\mathrm{d}x(x_0),\end{aligned}$$

由链式法则可知它们的确相等。这就是一阶微分的形式不变性。

高阶微分的商空间定义

高阶微分也是某个商空间的元素，用上面的语言来写就是

$$\Omega_{x0}^k = \mathfrak{m}_{x0}^{k-1}/\mathfrak{m}_{x0}^k,$$

高阶微分运算 $\mathrm{d}^k : C^{k-1}(D, \mathbb{R}^m) \to \Omega_{x0}^k$ 是如下两个映射的复合：

$$\mathrm{d}^k = \pi\circ\Delta^k : C^{k-1}(D, \mathbb{R}^m) \to \mathfrak{m}_{x0}^{k-1} \to \Omega_{x0}^k,$$

其中 Δ^k 叫 k 阶差分（k 阶差分亦可通过其它方式定义，结果是等价的）：

$$\Delta^k : C^{k-1}(D, \mathbb{R}^m) \to \mathfrak{m}_{x0}^{k-1}, \quad f(x) \mapsto f(x) - T_{k-1}(x; f, x_0),$$

$T_{k-1}(x; f, x_0)$ 表示 f 在 x_0 处的 $k-1$ 阶泰勒多项式， π 仍是投影：

$$\pi : \mathfrak{m}_{x0}^{k-1} \to \Omega_{x0}^k, \quad h(x) \mapsto h(x) + \mathfrak{m}_{x0}^k.$$

当 $f \in C^k(D, \mathbb{R}^m)$ 时，我们有

$$\mathrm{d}^k f(x_0) = \frac{1}{k!} f^{(k)}(x_0)(x-x_0)^{\otimes k} + \mathfrak{m}_{x0}^k,$$

与熟知的高阶微分运算一致（差一个阶乘因子）。

高阶微分不具有形式不变性，因为下面这个图表不交换了：

$$\begin{array}{ccccc} C^{k-1}(E, \mathbb{R}^m) & \xrightarrow{\Delta^k} & \mathfrak{m}_{y0}^{k-1} & \xrightarrow{\pi} & \Omega_{y0}^k \\ \Big\downarrow \phi^* & \times & \Big\downarrow \phi^* & \circlearrowleft & \Big\downarrow \phi^* \\ C^{k-1}(D, \mathbb{R}^m) & \xrightarrow{\Delta^k} & \mathfrak{m}_{x0}^{k-1} & \xrightarrow{\pi} & \Omega_{x0}^k \end{array}.$$

事实上，以 $n=m=l=1$、$k=2$、$f \in C^2(E, \mathbb{R})$、$\phi \in C^2(D, E)$ 为例：

$$\begin{aligned}&\phi^* \circ \mathrm{d}^2\left(f(y_0)\right) = \phi^*\left(\frac{1}{2} f_{yy}(y_0)(y-y_0)^2 + \mathfrak{m}_{y0}^2\right)\\ =&\frac{1}{2}\left(f_{yy}\left(\phi(x_0)\right)\phi_x(x_0)^2\right)(x-x_0)^2 + \mathfrak{m}_{x0}^2\end{aligned}$$

$$\begin{aligned}&\mathrm{d}^2 \circ \phi^*\left(f(y_0)\right) = \mathrm{d}^2\left(\left(f\circ\phi\right)(x)\right)(x_0)\\ =&\frac{1}{2}\left(f_{yy}(\phi(x_0))\phi_x(x_0)^2 + f_y(\phi(x_0))\phi_{xx}(x_0)\right)(x-x_0)^2 + \mathfrak{m}_{x0}^2\end{aligned}$$

为什么第一种算法丢了 $f_y(\phi(x_0))\phi_{xx}(x_0)$ 这一项？因为二阶差分 Δ^2 减掉了一项 $f_y(y_0)(y-y_0)$，这一项经过 ϕ^* 后再做泰勒展开时会对 $(x-x_0)^2$ 贡献 $\frac{1}{2}f_y(\phi(x_0))\phi_{xx}(x_0)$。

形式不变性被破坏的原因是，不应该在 Δ^k 的定义中减掉那些在拉回映射下会对高阶项有贡献的低阶项。 为解决这个问题，我们提出以下的修改方案：首先定义“k 阶全微分”为如下空间

$$\tilde{\Omega}_{x0}^k = \mathfrak{m}_{x0}^0/\mathfrak{m}_{x0}^k,$$

然后定义相应的全微分运算 $\tilde{\mathrm{d}}^k : C(D,\mathbb{R}^m) \to \tilde{\Omega}_{x0}^k$ 为如下两个映射的复合：

$$\tilde{\mathrm{d}}^k = \Delta\circ\pi : C(D,\mathbb{R}^m) \to \mathfrak{m}_{x0}^0 \to \tilde{\Omega}_{x0}^k,$$

其中：

$$\begin{aligned}&\Delta : C(D,\mathbb{R}^m) \to \mathfrak{m}_{x0}^0, \quad f(x)\mapsto f(x)-f(x_0),\\ &\pi : \mathfrak{m}_{x0}^0 \to \tilde{\Omega}_{x0}^k, \quad h(x)\mapsto h(x)+\mathfrak{m}_{x0}^k.\end{aligned}$$

当 $f\in C^k(D,\mathbb{R}^m)$时，我们有

$$\tilde{\mathrm{d}}^k f(x_0) = \sum_{i=1}^k \frac{1}{i!} f^{(i)}(x_0)(x-x_0)^{\otimes i} + \mathfrak{m}_{x0}^k,$$

特别地，当 f 取为 D 到自身的恒等映射时（即 $f(x)=x$），有

$$\tilde{\mathrm{d}}^k x(x_0) = (x-x_0) + \mathfrak{m}_{x0}^k,$$

将 $\tilde{\mathrm{d}}^k$ 简记为 d，则有

$$\mathrm{d}f(x_0) = \sum_{i=1}^k \frac{1}{i!} f^{(i)}(x_0)(\mathrm{d}x)^{\otimes i},$$

我们以后将用这种方式来表示高阶全微分。

接下来，k 阶全微分的形式不变性即如下图表的交换性：

$$\begin{array}{ccccc} C(E,\mathbb{R}^m) & \xrightarrow{\Delta} & \mathfrak{m}_{y0}^0 & \xrightarrow{\pi} & \tilde{\Omega}_{y0}^k \\ \downarrow \phi^* & \circlearrowleft & \downarrow \phi^* & \circlearrowleft & \downarrow \phi^* \\ C(D,\mathbb{R}^m) & \xrightarrow{\Delta} & \mathfrak{m}_{x0}^0 & \xrightarrow{\pi} & \tilde{\Omega}_{x0}^k \end{array}.$$

它的证明是显然的。

当 $f \in C^k(D,\mathbb{R}^m)$ 时，k 阶全微分的形式不变性也可通过 Faà di Bruno 公式证明，它是链式法则的高阶推广。因为这个公式的多分量版本写起来过于啰嗦，此处就省略了。它的单分量版本即《形式幂级数与 Lagrange 反演公式》的定理 2.5。

例：隐函数的高阶导数

下面来看一个例子。设函数 $y(x)$ 由隐函数方程 $f(x,y(x))=0$ 确定，其中 f 二阶连续可导，求 $y''(x)$。

常规做法是：

$$\begin{aligned} y''(x) =& \frac{\mathrm{d}}{\mathrm{d}x}\left(-\frac{f_x(x,y(x))}{f_y(x,y(x))}\right) \\ =& -\frac{1}{f_y}\left(f_{xx}+f_{xy}\left(-\frac{f_x}{f_y}\right)\right)+\frac{f_x}{f_y^2}\left(f_{yx}+f_{yy}\left(-\frac{f_x}{f_y}\right)\right) \\ =& -f_{xx}\frac{1}{f_y}+2f_{xy}\frac{f_x}{f_y^2}-f_{yy}\frac{f_x^2}{f_y^3}. \end{aligned}$$

这种计算需要对隐函数求导法和链式法则有很好的理解才能算对。

利用高阶全微分则可以这样做：先写出 $f(x,y)=0$ 的二阶全微分：

$$\left(f_x\mathrm{d}x+f_y\mathrm{d}y\right)+\frac{1}{2}\left(f_{xx}(\mathrm{d}x)^2+2f_{xy}\left(\mathrm{d}x\right)\left(\mathrm{d}y\right)+f_{yy}(\mathrm{d}y)^2\right)=0,$$

然后将 $y(x)$ 的二阶全微分代入：

$$dy=y_x\mathrm{d}x+\frac{1}{2}y_{xx}(\mathrm{d}x)^2,$$

扔掉 $\mathrm{d}x$ 的二阶以上的项，然后通过对比系数可直接解得：

$$y_x=-\frac{f_x}{f_y},\quad y_{xx}=-f_{xx}\frac{1}{f_y}+2f_{xy}\frac{f_x}{f_y^2}-f_{yy}\frac{f_x^2}{f_y^3}.$$

这种算法不需要知道什么求导法则，完全就是待定系数法。

所以，高阶微分的问题解决了吗？看起来我们讲清楚了到底什么是 $\mathrm{d}^kf(x)$，定义了具有形式不变性的高阶全微分这一概念，并且它跟高阶导数的确是不一样的东西，它还能帮助我们计算。但是从上面

这个计算我们马上会发现这种定义的新问题：新定义的高阶全微分跟带有皮亚诺余项的泰勒公式有什么区别？答案是：没有任何区别，我们绕了一圈，只是重新发现了泰勒公式而已，而这是牛顿在 1669 年的小册子里就已经写得很清楚的东西。

对今后教学工作的建议

基于上面的讨论，我对今后微积分/数学分析的教学工作有如下的建议：

- 在数学分析课程中用商空间方式定义一阶微分形式并证明其形式不变性；
- 不要引入“高阶微分”的概念，更不要去强调其不具备形式不变性；
- 在带有皮亚诺余项的泰勒公式部分增加用待定系数法求反函数高阶导数的例子；
- 在数学分析课程中介绍 Faà di Bruno 公式，并利用数学归纳法或泰勒公式证明之。

📎 Newton 📎 Leibniz 📎 Taylor 📎 Faà di Bruno

形式幂级数与 Lagrange 反演公式 ▸

撰写评论

发布

账号（邮件地址）

评论 3

时间正序 时间倒序 同感正序

Scriabin 2020.08.07 12:33

厉害了呀

回复 0 0

Dreamerxic 2019.11.21 11:47

请问新定义的高阶微分不能再看作微分的复合了吗

回复 0 0

taofuan 2019.09.25 07:25

您好，请问那个一阶微分的等价表达式是怎么推导的?

回复 0 0