微分与导数 1

微分与导数之路开启于 17 世纪, 费马（Fermat）和巴罗（Barrow）求极值与切线的方法事实上就是现代的导数. 牛顿明确提出了流数的概念, 而莱布尼茨使用了微分的说法. 但因缺乏严格的极限理论，早期的微分学一度被无穷小幽灵所笼罩. 所幸，经历了分析的严格化之后，微分和导数的逻辑基础已足够坚实. 现如今，它们以各种不同的面貌出现在世人面前，它可能被称为增长率或反应速度，也可能叫做边际成本.

> Mathematical research doesn't operate in a vacuum. New ideas have to condense around some notion of reality or someone's notion.
>
> Nigel Hitchin

1.1 微分与导数的概念

就数学而言，微分与导数至少有切线和极值两个来源. 这两个问题都与函数增长快慢有关，最核心的问题就是处理函数在一点处的增量

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

我们知道，函数 f 在 a 点连续是指它在该点的增量是无穷小量, 即

$$\Delta f_a(h) = f(a+h) - f(a) = o(1)(h \to 0).$$

但 $o(1)$ 的估计显然过于粗略，无法回答“增长快慢”这类更深入的问题. 为此，Fermat 和 Barrow 使用了魔法.[1]

- Barrow 试图计算抛物线 $y^2 = x$ 在点 $P(a, \sqrt{a})$ 处的切线. 设过 P 的切线交 x 轴于 Q，记 P 在 x 轴上的投影点为 R，则切线斜率为

  $$k = \frac{PR}{QR}.$$

  当 x 由 a 增加小量 h 时，记 $\Delta y_a(h) = v$. 只要 h 足够小，那么

  $$\frac{PR}{QR} = \frac{v}{h}$$

  应当是足够准确的. 为了求得此比例，可以将它们代入抛物线方程

  $$(\sqrt{a} + v)^2 = a + h$$

  即

  $$2\sqrt{a}v + v^2 = h.$$

  然后，魔法开始了：既然 v^2 比 v 小很多，索性删去，这样就得到 $2\sqrt{a}v = h$. 于是

  $$k = \frac{v}{h} = \frac{1}{2\sqrt{a}}.$$

1: Fermat 在他 1637 年的手稿中给出了切线和最值的算法. Barrow 在其 1669 年的著作中也给出了求切线的方法，他的方法实质上与 Fermat 是一致的；同年, Barrow 辞去了剑桥大学 Lucas 教授的职位，他的继任者是 Newton.

这个结果当然是符合预期的，我们可以通过约束直线与抛物线只有一个交点的方式求得. 那么这个魔法的合理性在哪里？事实上，如果用无穷小的语言，那么 Barrow 其实假设的是

$$\Delta y_a(h) = v = kh + o(h) \quad (h \to 0).$$

以此代入方程，有

$$\begin{aligned} 2\sqrt{a}v + v^2 &= 2\sqrt{a}(kh + o(h)) + (kh + o(h))^2 \\ &= 2\sqrt{a}kh + o(h) = h. \end{aligned}$$

由此可得 $k = 1/2\sqrt{a}$.

- Fermat 想要计算两条邻边长度之和为 A 的矩形的面积的最大值，即求函数 $f(x) = (A - x)x$ 的最大值. 假设 $x = a$ 时函数取得最大值. 考虑 x 由 a 增加一个小量 h 时 f 的变化状态. Fermat 的哲学是：既然 $x = a$ 是最大值点，那么 f 在 $x = a$ 附近几乎没有变化. 因此，当 h 充分小时，应当几乎成立

$$\Delta f_a(h) = (A - a - h)(a + h) - (A - a)a = 0.$$

简单计算表明

$$\Delta f_a(h) = (A - 2a)h - h^2.$$

既然 h^2 比 h 小很多，删去！于是 Fermat 得到了

$$\Delta f_a(h) = (A - 2a)h = 0.$$

所以 $a = A/2$. 这个结果自然无比正确，因为我们用配方法可以得到同样的结果. 那么，Fermat 的魔法到底隐藏着什么？其实，函数在最大值点附近几乎不变这个假设，用无穷小的语言可写为

$$\Delta f_a(h) = o(h) \quad (h \to 0).$$

这样就有

$$\Delta f_a(h) = (A - 2a)h - h^2 = o(h),$$

从而 $(A - 2a)h = o(h)$, 所以 $a = A/2$.

可以看到，Fermat 和 Barrow 的魔法事实上是删除高阶无穷小，并且都假设了函数的增量可写为

$$\Delta f_a(h) = Lh + o(h) \ \ (h \to 0),$$

其中 L 是一个依赖于 a 的实数（Barrow 的 L 是 $1/2\sqrt{a}$, Fermat 的 L 是零）. 显然, 这是一个比连续性更精细的假设: 它假定函数的增量 $\Delta f_a(h)$ 不仅仅是个无穷小量，事实上还可以写成**线性主部** Lh 与**高阶量** $o(h)$ 之和. 我们把这种性质称为可微性.

可微性

设 f 在 a 的某个邻域上有定义. 若存在实数 L，使得

$$\Delta f_a(h) = f(a + h) - f(a) = Lh + o(h), \quad (h \to 0)$$

则称 f 在 a 点**可微 (differentiable)**. [2]

2: 显然，若 f 在 a 点可微，则 f 在 a 点连续.

根据定义，$\Delta f_a(h) = Lh + o(h)(h \to 0)$ 意味着

$$\lim_{h\to 0} \frac{\Delta f_a(h) - Lh}{h} = 0 \text{ 即 } L = \lim_{h\to 0} \frac{\Delta f_a(h)}{h}.$$

这表明极限 $\lim\limits_{h\to 0} \frac{\Delta f_a(h)}{h}$ 存在且等于 L. 根据极限的唯一性，这也表明，若函数可微，则 L 唯一.

另一方面，若极限 $\lim\limits_{h\to 0} \frac{\Delta f_a(h)}{h}$ 存在，记之为 A，则

$$\frac{\Delta f_a(h)}{h} = A + o(1) \ \ (h \to 0),$$

进而

$$\Delta f_a(h) = Ah + o(1)h = Ah + o(h) \ \ (h \to 0).$$

这意味着函数 f 在 a 点可微. 由此我们引出下述定义.

导数

设 f 在 a 的某个邻域上有定义. 若极限

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

存在，则称 f 在 a 点**可导**，其极限值称为函数 f 在 a 点的**导数 (derivative)**，记作 $f'(a)$. [3]

3: “导（函）数”（derivative）的名称和符号传自 Lagrange，意为由函数“导出之（函）数”（derived function）. Newton 称之为“流数”（fluxion），他使用 $\dot{f}$ 表示关于时间的流数.

由前面的分析，**可微等价于可导**，并且在可微时成立

$$\Delta f_a(h) = f(a+h) - f(a) = f'(a)h + o(h) \ \ (h \to 0)$$

微分

若 f 在 a 点可微，称自变量增量 h 的线性函数 $f'(a)h$ 为 f 在 a 点的**微分 (differential)**，[4] 记作 df_a 或 $df|_a$，即

$$df_a(h) := f'(a)h.$$

4: “微分”的名称和符号来自 Leibniz. 只不过当时尚未澄清无穷小量的确切含义，以至于微分 df_a 成为了既是零又不是的幽灵.

考虑 $\mathbb{R}$ 上的恒等函数 id. 显然, 对于任意的 $a, h \in \mathbb{R}$ 有

$$\Delta \mathrm{id}_a(h) = h.$$

因此恒等函数 id 处处可微，且 $d\mathrm{id}_a(h) = h$. 注意到 $d\mathrm{id}_a(h)$ 与 a 无关，所以通常也简写为 $d\mathrm{id}$. 如果用 x 作为自变量，并用函数表达式代表函数本身，则可用 x 代表 id，进而可用 dx_a 代表 $d\mathrm{id}_a$, 即

$$dx_a(h) = h.$$

因此，如果 f 在 a 处可微，在以 x 记自变量的情况下，可写为

$$df_a = f'(a)dx_a.$$

于是，导数 $f'(a)$ 就是两个微分 df_a 和 dx_a 的商 $f'(a) = df_a/dx_a$. 故而导数也称为**微商**. 由于微分是无穷小量，所以导数是两个无穷小量的商，这正是 Leibniz 的原始看法. 习惯上，常把 $f'(a) = df_a/dx_a$ 简写为

$$f'(a) = \left.\frac{df}{dx}\right|_a \quad 或 \quad \left.\frac{d}{dx}\right|_a f.$$

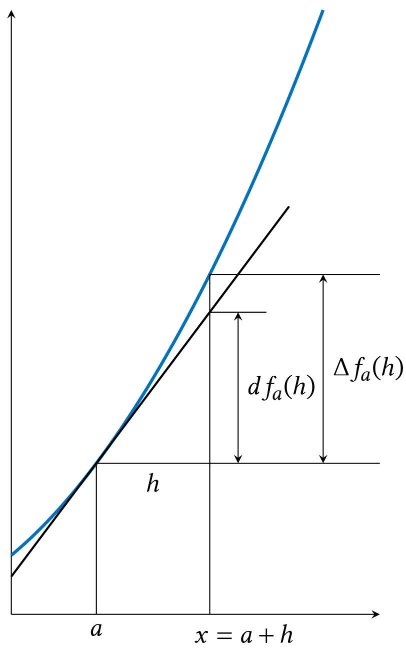
就记号 $\left.\frac{d}{dx}\right|_a f$ 而言, 我们可将 $\left.\frac{d}{dx}\right|_a$ 看作一个映射 (算子), 它将函数 f 映为 f 在 a 点的导数 $f'(a)$. Euler 使用了 D_a 来代替 $\left.\frac{d}{dx}\right|_a$, 即 $D_a f = f'(a)$.

如果 f 在开区间 I 上处处可导, 则称 f **在区间 I 上可导**, 它在 x 点的导数值 $f'(x)$ 定义了一个 I 上的函数, 称其为**导函数**（常简称为导数）, 记为 f'. 此时函数在 I 上处处可微，其微分记作 df 或 $df(x)$，通常不再添加下标，用导数表示可写为

$$df = df(x) = f'(x)dx.$$

进而导函数可写作 df/dx 或 $df(x)/dx$. 若用 Euler 的符号，导函数可写为 Df 或 $Df(x)$.

切线与线性近似



现在我们来为 Barrow 的做法正名. 若 f 在 a 点可微，则它在 a 点附近可以写为

$$f(x) = f(a) + f'(a)(x-a) + o(x-a) \quad (x \to a).$$

考虑过 $(a, f(a))$ 的直线 $y(x;k) = f(a) + k(x-a)$. 若 $k \neq f'(a)$, 则

$$f(x) - y(x;k) = O(x-a) \ \ (x \to a).$$

而当 $k = f'(a)$, 有

$$f(x) - y(x;k) = o(x-a) \ \ (x \to a).$$

因此，在 $x = a$ 附近，直线 $y(x) = f(a) + f'(a)(x-a)$ 逼近函数 $f(x)$ 的效果最佳，我们称之为函数 f 的图像在 a 处的**切线**方程. 这说明 Barrow 求切线的方法是合理的.

这样，微分 $df_a(x-a) = f'(a)(x-a) = y(x) - f(a)$ 事实上描述的是沿着函数图像切线的增量.

既然微分在几何上描述的是直线，且表达式又是个线性函数，因此函数在一点的**微分**也称为函数在该点的**线性化**. 换言之，线性化即是用一次函数来局部近似原来的函数

$$f(x) \approx f(a) + df_a(x-a) = f(a) + f'(a)(x-a).$$

通过这个公式可以计算一些近似值，比如 $\sin 31°$.

极值与 Fermat 定理

我们把 Fermat 求最值的方法严格化.

极值（local extrema）

若存在 $\delta > 0$ 使得在 $U^\circ(a;\delta)$ 上成立 $f(x) \geq f(a)$，则称 f 在 a 处取得**极小值** $f(a)$, 称 a 为 f 的**极小值点**. 若不等式是严格的，则称为**严格极小值**. 类似地，有**极大值**和**严格极大值**. 极小值和极大值统称为**极值**.

费马定理

若 a 是 f 的可微的极值 (内) 点, 则 $df_a = 0$.

证明. 假如 $df_a \neq 0$，则 $f'(a) \neq 0$，不妨假设 $f'(a) > 0$. 根据可微性，有

$$\lim_{h\to 0} \frac{\Delta f_a(h)}{h} = f'(a) > 0.$$

根据保号性，存在 $\delta > 0$，对任意 $h \in (-\delta, \delta)\backslash\{0\}$ 成立 $\Delta f_a(h)/h > 0$. 这意味着，当 $h \in (0, \delta)$ 时 $f(a+h) > f(a)$，当 $h \in (-\delta, 0)$ 时 $f(a+h) < f(a)$，与 a 是极值点矛盾. □

因此，对于可微函数而言，Fermat 的方法是有效的. 通常，导数为零的点一般称为**驻点/稳定点**（**stationary point**）或**临界点** (**critical point**). 因此，Fermat 定理也可叙述为“**可导的极值点必然是驻点**”. 但要注意的是，驻点未必是极值点，如 $x = 0$ 是 $f(x) = x^3$ 的驻点，但它不是极值点.

简单函数的导数

下面计算几个简单的初等函数的导函数.

常值函数 $(c)' = 0$.
幂函数 $(x^\alpha)' = \alpha x^{\alpha-1}$.
利用 $(1+x)^\alpha - 1 \sim \alpha x (x \to 0)$，有

$$\begin{aligned}\lim_{\Delta x\to 0} \frac{(x+h)^\alpha - x^\alpha}{h} &= x^\alpha \cdot \lim_{h\to 0} \frac{(1+\frac{h}{x})^\alpha - 1}{h} \\ &= x^\alpha \cdot \lim_{h\to 0} \frac{\alpha\frac{h}{x}}{\Delta x} = \alpha x^{\alpha-1}.\end{aligned}$$

指数函数 $(a^x)' = a^x \ln a$.
利用 $a^x - 1 \sim x \ln a (x \to 0)$，有

$$\lim_{h\to 0} \frac{a^{x+h} - a^x}{h} = a^x \cdot \lim_{h\to 0} \frac{e^{h\ln a} - 1}{h} = a^x \ln a.$$

对数函数 $(\log_a x)' = 1/(x \ln a)$.
利用 $(1+x)^{1/x} \to e (x \to 0)$，有

$$\lim_{h\to 0} \frac{\log_a(x+h) - \log_a x}{h} = \lim_{h\to 0} \frac{1}{x} \log_a \left(1 + \frac{h}{x}\right)^{\frac{x}{h}} = \frac{\log_a e}{x}.$$

正弦函数和余弦函数 $(\sin x)' = \cos x$，$(\cos x)' = -\sin x$.

利用 $\sin x \sim x (x \to 0)$，有

$$\lim_{h\to 0}\frac{\sin(x+h)-\sin x}{h} = \lim_{h\to 0}\frac{2\cos\left(x+\frac{1}{2}h\right)\sin\left(\frac{1}{2}h\right)}{h} = \cos x.$$

余弦函数是类似的.

很多时候我们遇到的函数是分段的，为了计算它们的导数，我们仿照函数的单侧极限引入单侧导数的概念.

单侧导数

若极限 $\lim\limits_{x\to a^+}\dfrac{f(x)-f(a)}{x-a}$ 存在，则称 f 在 a 点右可导，极限值称为右导数，记为 $f_+'(a)$. 类似的有左导数的概念，记号为 $f_-'(a)$.

下面的结论是显然的.

f 在 a 点可导 $\Longleftrightarrow$ f 在 a 点的左右导数存在且相等.

为了叙述方便，我们约定：如果函数在闭区间或半开半闭区间 I 的内部各点处可导、左（右）端点处左（右）可导，则称函数在 I 上可导.

例 1.1.1 判断 $f(x) = |x|$ 在原点的可导性.

解. 根据定义

$$f_+'(0) = \lim_{x\to 0^+}\frac{f(x)-f(0)}{x-0} = \lim_{x\to 0^+}\frac{x-0}{x} = 1,$$
$$f_-'(0) = \lim_{x\to 0^-}\frac{f(x)-f(0)}{x-0} = \lim_{x\to 0^-}\frac{-x-0}{x} = -1.$$

因此 f 在原点左右导数不相等，从而不可导. □

例 1.1.2 证明：若函数 f 在 a 点（单侧）可导，则它在该点（单侧）连续.

证明. 仅考虑右可导的情形.

$$\lim_{x\to a^+}(f(x)-f(a)) = \lim_{x\to a^+}\frac{f(x)-f(a)}{x-a}\lim_{x\to a^+}(x-a) = f_+'(a)\cdot 0 = 0.$$

这意味着 $f(x)\to f(a)(x\to a^+)$，所以函数右连续. □

例 1.1.3 分析函数 $f(x) = \begin{cases} x^2+a, & x>1 \\ kx+b, & x\le 1 \end{cases}$ 在 $x=1$ 处的可微性，其中 a,b,k 是常数.

解. 因为可微必连续，所以若 f 在 $x=1$ 可微，则必有

$$f(1+0)=\lim_{x\to 1^+}(x^2+a)=1+a=f(1)=k+b.$$

进而

$$f'_+(1)=\lim_{x\to 1^+}\frac{f(x)-f(1)}{x-1}=\lim_{x\to 1^+}\frac{(x^2+a)-(k+b)}{x-1}=\lim_{x\to 1^+}\frac{x^2-1}{x-1}=2,$$
$$f'_-(1)=\lim_{x\to 1^-}\frac{f(x)-f(1)}{x-1}=\lim_{x\to 1^-}\frac{(kx+b)-(k+b)}{x-1}=k.$$

因此，可微时必有 $k=2$ 且 $a=b+1$. 显然这一关系也是充要的. □

例 1.1.4 试辨析 $f'_+(a)$ 与 $f'(a+0)$.

解. 前者的定义是

$$f'_+(a)=\lim_{x\to a^+}\frac{f(x)-f(a)}{x-a},$$

而后者的定义是

$$f'(a+0)=\lim_{x\to a^+}f'(x)=\lim_{x\to a^+}\left(\lim_{y\to x}\frac{f(y)-f(x)}{y-x}\right).$$

可以看到，为了计算 $f'(a+0)$，前提是 f 在 a 的右侧小邻域上可导，而计算 $f'_+(a)$ 并不要求 f 在 a 附近的可导性. 另一方面，计算 $f'(a+0)$ 时完全不需要 $f(a)$，而计算 $f'_+(a)$ 则必须用到函数值 $f(a)$. 可见两者在定义上并无直接联系. □

1.2 微分与导数的运算法则

由于导数的定义就是极限，所以我们可以直接使用极限的各种性质来推导导数的性质. 在注意到可微与可导的等价性，所以我们又可以通过证明导数的运算法则得到微分的运算法则.

四则运算法则

四则运算法则[5]

线性性 $(\lambda f+\mu g)'=\lambda f'+\mu g'$, $d(\lambda f+\mu g)'=\lambda df+\mu dg$;
莱布尼茨法则 $(fg)'=f'g+fg'$, $d(fg)=gdf+fdg$;
除法法则 $(f/g)'=(f'g-fg')/g^2$, $d(f/g)=(gdf-fdg)/g^2$.

5: 与极限法则一样，此处的求导与微分法则应理解为：当右侧有意义时，那么左侧有意义，并且等式成立. 因此，左侧的可导性是结论之一.

证明. 这些性质本质上就是极限的性质. 根据

$$\Delta(\lambda f+\mu g)=\lambda\Delta f+\mu\Delta g,$$
$$\Delta(fg)=(\Delta f)g+f(\Delta g)+\Delta f\Delta g,$$
$$\Delta(1/g)=-\Delta g/(g(g+\Delta g)),$$

各式两侧除以 Δx，再取极限，容易得到导数的线性性、莱布尼茨法则以及

$$(1/g)' = -g'/g^2.$$

进而将 f/g 视为 $f \cdot 1/g$，应用莱布尼茨法则可得除法法则. □

线性性与莱布尼茨法则可以推广到多个函数的情形，后者的形式为

$$(f_1 f_2 \cdots f_n)' = \sum_{k=1}^{n} f_1 f_2 \cdots f_k' \cdots f_n,$$

$$d(f_1 f_2 \cdots f_n) = \sum_{k=1}^{n} f_1 f_2 \cdots \widehat{f_k} \cdots f_n d f_k.^6$$

6: 此处使用的符号 $\widehat{\text{term}}$ 表示删去该项.

例 1.2.1 求 $f(x) = x^2 \ln x \sin x + e^x - 3$ 的导数和微分.

解. 根据线性性和莱布尼茨法则，有

$$\begin{aligned} f'(x) &= (x^2 \ln x \sin x)' + (e^x)' - (3)' = (x^2 \ln x \sin x)' + e^x \\ &= (x^2)' \ln x \sin x + x^2 (\ln x)' \sin x + x^2 \ln x (\sin x)' + e^x \\ &= 2x \ln x \sin x + x \sin x + x^2 \ln x \cos x + e^x. \end{aligned}$$

上述过程对 $x > 0$ 均成立. 进而微分为 $df = f'(x)dx$. □

例 1.2.2 求正切函数、余切函数、正割函数、余割函数的导数.

解. 应用除法法则，有

$$\begin{aligned} (\tan x)' &= (\sin x/\cos x)' = (\cos^2 x + \sin^2 x)/\cos^2 x = \sec^2 x, \\ (\cot x)' &= (\cos x/\sin x)' = (-\sin^2 x - \cos^2 x)/\sin^2 x = -\csc^2 x, \\ (\sec x)' &= (1/\cos x)' = \sin x/\cos^2 x = \sec x \tan x, \\ (\csc x)' &= (1/\sin x)' = -\cos x/\sin^2 x = -\csc x \cot x. \end{aligned}$$

上述过程在函数相应的定义域内恒成立. □

反函数求导法

现在我们来讨论反函数的导数. 考虑函数 $y = f(x)$，记 $y_0 = f(x_0)$. 假设 $f'(x_0) = k$，那么函数图像在 (x_0, y_0) 点的切线斜率为 k. 如果我们从 y 轴上观察，反函数 $x = f^{-1}(y)$ 的图像在 (x_0, y_0) 点的切线斜率应为 $1/k$. 换言之，在适当条件下，函数与其反函数的导数互为倒数.

可导反函数定理

设 $y = f(x)$ 是 $U(x_0)$ 上严格单调的连续函数, 并且 f 在 x_0 具有非零导数，则 f 在 $U(x_0)$ 上有反函数 $x = \varphi(y)$, 且 φ 在 $y_0 = f(x_0)$ 处的导

数为 $\varphi'(y_0) = 1/f'(x_0)$.

证明. 根据连续反函数定理，φ 是严格单调的连续函数. 因此，可以使用换元法求极限，得

$$\begin{aligned}\lim_{y\to y_0}\frac{\phi(y)-\phi(y_0)}{y-y_0} &= \lim_{y\to y_0}\frac{\phi(y)-x_0}{y-f(x_0)} = \lim_{y\to y_0}\frac{\phi(y)-x_0}{f(\phi(y))-f(x_0)}\\ &\overset{x=\phi(y)}{=\!=\!=} \lim_{x\to x_0}\frac{x-x_0}{f(x)-f(x_0)} = \frac{1}{f'(x_0)}.\end{aligned}$$

上式不仅证明了 φ 在 y_0 处可导，也得到了导数公式. □

3 课时/33 课时

如果采用莱布尼茨的符号, 这一结论似乎是显然的: $dx/dy = 1/(dy/dx)$. 但以目前的逻辑，dx 和 dy 并不能独立存在，所以 dx/dy 不能看做它们的商.

应用反函数求导公式可以得到对数函数和反三角函数的导数.

对数函数 $x = \log_a y, y = a^x$.

$$x'(y_0) = \frac{1}{y'(x_0)} = \frac{1}{a^{x_0}\ln a} = \frac{1}{y_0\ln a}.$$

反正弦和反余弦 $y = \arcsin x, x = \sin y$.

$$y'(x) = \frac{1}{x'(y)} = \frac{1}{\cos y} = \frac{1}{\sqrt{1-x^2}}.$$

类似可得

$$(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}.$$

反正切和反余切 $y = \arctan x, x = \tan y$.

$$y'(x) = \frac{1}{x'(y)} = \frac{1}{\sec^2 y} = \frac{1}{1+\tan^2 y} = \frac{1}{1+x^2}.$$

类似可得

$$(\operatorname{arccot} x)' = -\frac{1}{1+x^2}.$$

至此，我们得到了所有基本初等函数的导数公式.

链式法则

链式法则是指复合函数微分法，它可以通过下述启发性推导得到.

$$\frac{d}{dx}f(u(x)) = \lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x} = \lim_{\Delta x\to 0}\frac{\Delta f}{\Delta u}\cdot\frac{\Delta u}{\Delta x} = \frac{df}{du}\frac{du}{dx}.$$

上述论证的问题在于此处 $\Delta u = u(x+\Delta x) - u(x)$ 可能等于零，从而不能做除数. 但如果在 $\Delta x \to 0$ 的过程中 Δu 可以不断等于零，那么 $f(u(x))$ 和 $u(x)$ 在该点的导数似乎也应当为零，这暗示着上述推导的结果总是成立的.

链式法则（导数）

设 $u(x)$ 在 x_0 可导, $f(u)$ 在 $u_0 = u(x_0)$ 可导. 则 $g(x) = f(u(x))$ 在 x_0 可导并且 $g'(x_0) = f'(u_0)u'(x_0)$.

证明. 问题的关键是如何处理 $u(x) = u_0$ 的情形. 事实上，这是 $\Delta f/\Delta u$ 的可去间断点，只要将 $\Delta f/\Delta u$ 连续延拓为函数

$$H(u) = \begin{cases} \dfrac{f(u) - f(u_0)}{u - u_0}, & u \neq u_0, \\ f'(u_0), & u = u_0 \end{cases}$$

就可以消除奇性. 进而

$$\frac{\Delta g}{\Delta x} = \frac{f(u(x)) - f(u(x_0))}{x - x_0} = H(u(x)) \cdot \frac{\Delta u}{\Delta x}.$$

利用 H 的连续性，上式取极限即得 $g'(x_0) = f'(u_0)u'(x_0)$. □

如果采用莱布尼茨的符号，并把复合函数简写为 $y = y(u(x))$，那么链式法则似乎也是显然的:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}.$$

事实上，我们倾向于用莱布尼茨的符号来表达链式法则，因为这套符号清楚表明了对哪个变量求导数，而这是拉格朗日的导数符号所不具备的. 顺便指出，当使用拉格朗日的符号时，$f'(u(x))$ 与 $(f(u(x)))'$ 的含义截然不同，务必注意.

利用导数的链式法则，我们可以得到微分的链式法则.

链式法则（微分）

设 u 在 x_0 点可微, f 在 $u(x_0)$ 点可微，则 $f \circ u$ 在 x_0 点可微并且

$$d(f \circ u)_{x_0} = df_{u(x_0)} \circ du_{x_0}.$$

证明. $d(f \circ u)_{x_0}(h) = (f \circ u)'(x_0)h = f'(u(x_0))u'(x_0)h = df_{u(x_0)}(du_{x_0}(h))$. □

微分的链式法则也常常表现为下述不变形.

一阶微分形式不变性

$$df(u(x)) = f'(u(x))u'(x)dx = f'(u)du.$$

这表明，无论将 f 视为最终变量 x 的函数还是中间变量 u 的函数，求微分的效果是一样的，这就是不变性. 但要注意的是，此处中间变量的微分 $du = du(x) = u'(x)dx$ 并非是 u 作为自变量的微分 Δu.

例 1.2.3 求函数 $y = \sin\ln x + x^2$ 的微分.

解. 应用一阶微分的形式不变性，有

$$\begin{aligned} dy &= d(\sin\ln x) + d(x^2) = \cos\ln x d(\ln x) + 2xdx \\ &= \cos\ln x \cdot \frac{1}{x}dx + 2xdx = \left(\cos\ln x \cdot \frac{1}{x} + 2x\right)dx. \end{aligned}$$

上述过程也可以用来计算 y 的导数. □

结合四则运算法则和链式法则，我们可以解决所有初等函数的求导问题.

例 1.2.4 计算 $\sqrt{1-2x}$ 和 $\tan^2\frac{1}{x}$ 的导数.

解. 直接应用链式法则，可得

$$\begin{aligned} \left(\sqrt{1-2x}\right)' &= \frac{1}{2\sqrt{1-2x}} \cdot (1-2x)' = -\frac{1}{\sqrt{1-2x}}. \\ \left(\tan^2\frac{1}{x}\right)' &= 2\tan\frac{1}{x} \cdot \left(\tan\frac{1}{x}\right)' = 2\tan\frac{1}{x} \cdot \sec^2\frac{1}{x} \cdot \left(\frac{1}{x}\right)' \\ &= 2\tan\frac{1}{x} \cdot \sec^2\frac{1}{x} \cdot \left(-\frac{1}{x^2}\right) = -\frac{2}{x^2}\tan\frac{1}{x} \cdot \sec^2\frac{1}{x}. \end{aligned}$$

在对第二个函数的求导中，我们连续使用了多次链式法则. □

对于幂指型的初等函数，可以使用**对数法**.

例 1.2.5 计算 x^x 的导数.

解. 记 $y = x^x$，则 $\ln y = x\ln x$，两端关于 x 求导，可得

$$\frac{y'}{y} = \ln x + 1.$$

因此

$$y' = y(\ln x + 1) = x^x(\ln x + 1).$$

也可以一步到位

$$(x^x)' = \left(e^{x\ln x}\right)' = e^{x\ln x}(x\ln x)' = x^x(\ln x + 1).$$

本质上这仍然是对数法. □

对数法的用武之地不止于此，比如对函数

$$y = \frac{(x+1)^3}{\sqrt{2x-1} \cdot (4-x)^9}$$

取对数再求导，明显简便很多. 此外，对数法还暗示了一种**隐式求导**的观点：即使不知道函数的解析式，仅从函数满足的方程也可以得到其导数.

例 1.2.6 已知可导函数 $y(x)$ 满足 $e^y + x\cos y = 1 - \sin x$. 计算 $y'(0)$.

解. 等式两边对 x 求导，得

$$e^y y' + \cos y - x\sin y \cdot y' = -\cos x.$$

注意到 $y(0)=0$，代入上式可得 $y'(0)=-2$. □

例 1.2.7 证明: 不存在 $\mathbb{R}$ 上的可导函数 f 满足 $f(f(x))=-x^3+x^2+1$.

证明. 假如存在, 则 $f'(f(x))f'(x)=-3x^2+2x$. 考虑不动点，由

$$x_0=f(f(x_0))=-x_0^3+x_0^2+1$$

解得 $x_0=1$. 因此 $f\circ f$ 唯一的不动点是 $x_0=1$. 然而 $f(f(f(1)))=f(1)$, 所以 $f(1)=1$. 代入导数公式得 $f'(1)f'(1)=-1$，不可能. 因此函数必然在 $x=1$ 处不可导. □

1.3 高阶导数

如果函数 f 的导函数 f' 可以继续求导数，那么 $(f')'$ 称为 f 的二**阶导数**. 一般的，如果 f 的 $n-1$ 阶导数可以继续求导数，那么其导数就称为 f 的 n **阶导数**，记作 $f^{(n)}$. 约定 $f^{(0)}=f$. 如果用欧拉的记号，那么 n 阶导数可写为 $D^n f$. 若用莱布尼茨的符号，则可写为

$$\frac{d^n f}{dx^n},\ \frac{d^n}{dx^n}f \text{ 或 } \left(\frac{d}{dx}\right)^n f.$$

当阶数较低时，通常仍会采用拉格朗日的符号，并用撇的个数表示导数的阶数，比如 f'' 表示二阶导数、f''' 表示三阶导数. 此时用牛顿的符号也是方便的，比如 $\ddot{f}$ 和 $\dddot{f}$ 就是二阶和三阶导数. 需要注意的是，随着阶数升高，高阶导数的定义域可能会缩小.

如果函数 f 在 I 上具有直到 k 阶连续导数, 则称其为 I 上的 C^k **(光滑) 函数**，记作 $f\in C^k(I)$. **光滑函数**的概念并没有明确约定，有的文献只要求 C^1，也有文献要求 C^∞ (即任意阶可导)，本书采纳后者.

例 1.3.1 计算 $f(x)=x^m$ 的各阶导数.

解. 当 $n<m$ 时，$f^{(n)}(x)=m(m-1)\cdots(m-n+1)x^{m-n}$；当 $n=m$ 时，$f^{(n)}(x)=m!$；当 $n>m$ 时，$f^{(n)}(x)=0$. □

例 1.3.2 计算 $f(x)=e^{\alpha x}$ 的各阶导数.

解. 易见 $(e^{\alpha x})^{(n)}=\alpha^n e^{\alpha x}$. □

例 1.3.3 计算 $f(x)=\sin x$ 的各阶导数.

解. 用归纳法可以证明 $(\sin x)^{(n)}=\sin\left(x+\frac{1}{2}n\pi\right)$. □

例 1.3.4 设 $f(x)=\begin{cases} x^n\sin\frac{1}{x}, & x\neq 0\\ 0, & x=0\end{cases}$ 其中 $n\geq 2$. 计算 $f'(0), f''(0)$.

解. 根据定义

$$f'(0)=\lim_{x\to 0}\frac{f(x)-f(0)}{x-0}=\lim_{x\to 0}x^{n-1}\sin\frac{1}{x}=0.$$

为了求 $f''(0)$，需要计算原点附近的一阶导数 $f'(x)$. 而在 $x\neq 0$ 处，f 是初等函数，可以直接应用导数公式.

$$f'(x)=nx^{n-1}\sin\frac{1}{x}+x^{n-2}\cos\frac{1}{x}.$$

显然，当 $n=2$ 时，$f'(0\pm 0)$ 不存在，所以 f' 在原点不连续，进而 $f''(0)$ 不存在. 当 $n\geq 3$ 时考虑极限

$$\lim_{x\to 0}\frac{f'(x)-f'(0)}{x-0}=\lim_{x\to 0}\left(nx^{n-2}\sin\frac{1}{x}+x^{n-3}\cos\frac{1}{x}\right).$$

当 $n=3$ 时，上述极限不存在，因而 $f''(0)$ 不存在. 当 $n\geq 4$ 时，极限为零，即 $f''(0)=0$. □

高阶导数的线性法则自不必说，对于乘积型函数的高阶导数，有下述莱布尼茨公式.

莱布尼茨公式

$$(uv)^{(n)}=\sum_{k=0}^{n}C_n^k u^{(k)}v^{(n-k)}.$$

此公式可用数学归纳法证明，留给读者. 下面来看两个应用.

例 1.3.5 设 $f(x)=(x^2+2x+3)e^{2x}$，求 $f^{(n)}$.

解. 应用莱布尼茨公式，当 $n\geq 2$ 时成立

$$\begin{aligned} f^{(n)}(x)&=\sum_{k=0}^{n}C_n^k(x^2+2x+3)^{(k)}(e^{2x})^{(n-k)}\\ &=\sum_{k=0}^{2}C_n^k(x^2+2x+3)^{(k)}\cdot 2^{n-k}e^{2x}\\ &=(x^2+2x+3)\cdot 2^ne^{2x}+n(2x+2)\cdot 2^{n-1}e^{2x}+n(n-1)\cdot 2^{n-2}e^{2x}\\ &=\Big(4x^2+4(n+2)x+12+4n+n(n-1)\Big)2^{n-2}e^{2x}.\end{aligned}$$

可以验证，当 $n=0,1$ 时，上式仍成立. □

例 1.3.6 设 $y=\arctan x$，求 $y^{(n)}(0)$.

解. 注意到 $(1+x^2)y' = 1$，对其两端求 n 阶导数，可得

$$y^{(n+1)}(0) = -n(n-1)y^{(n-1)}(0).$$

由于 $y(0) = 0, y'(0) = 1$，所以 $y^{(2k)}(0) = 0, y^{(2k+1)}(0) = (-1)^k(2k)!$，其中 $k \geq 0$. □

1.4 向量值函数的微分与导数

向量值函数

如果把平面曲线想象成点的轨迹，那么用参数方程表示曲线是自然的. 通常将**平面参数曲线**写为

$$C : \begin{cases} x = \phi(t), \\ y = \psi(t), \end{cases} \quad t \in I,$$

其中 I 是区间. 它的向量形式是

$$\boldsymbol{r}(t) = (\phi(t), \psi(t)), \qquad t \in I.$$

因此平面参数曲线事实上是一个映射

$$\boldsymbol{r} : I \to \mathbb{R}^2, \quad t \mapsto (\phi(t), \psi(t)).$$

由于上述映射 $\boldsymbol{r}$ 把实数 t 变成一个向量 $(\phi(t), \psi(t))$，所以也称其为**向量值函数**. 习惯上，也把曲线 C 称为向量值函数的终端曲线.

向量值函数的光滑性

设向量值函数 $\boldsymbol{r}(t) = (\phi(t), \psi(t)), t \in I$.

- 若 ϕ, ψ 在 t_0 处连续，则称 $\boldsymbol{r}$ 在 t_0 处连续，
- 若 ϕ, ψ 在 t_0 处可微，则称 $\boldsymbol{r}$ 在 t_0 处可微. 此时，$\boldsymbol{r}$ 在 t_0 点的导数和微分分别定义为

$$\boldsymbol{r}'(t_0) := (\phi'(t_0), \psi'(t_0)), \quad d\boldsymbol{r}_{t_0} := (d\phi_{t_0}, d\psi_{t_0}) = \boldsymbol{r}'(t_0)dt.$$

- 若 $\phi, \psi \in C^k(I)$, 则称 $\boldsymbol{r}$ 在 I 上 k 阶连续可微, 记作 $\boldsymbol{r} \in C^k(I, \mathbb{R}^2)$.

若视 $\boldsymbol{r}(t) = (\phi(t), \psi(t))$ 为质点在 t 时刻的位置，则 $\phi'(t)$ 就是质点的水平速度，而 $\psi'(t)$ 就是垂直速度，从而向量函数的导数就是质点 $\boldsymbol{r}'(t)$ 的**速度向量**，进一步向量函数的二阶导数 $\boldsymbol{r}''(t)$ 就是**加速度向量**.

平面正则曲线

通常我们所说的曲线 (段) 是连续的. 换言之，若 $\boldsymbol{r}(t) : I \to \mathbb{R}^2$ 连续，则称其为定义在参数区间 I 上的**连续 (参数) 曲线**. 更一般地，若 $\boldsymbol{r} \in C^k(I, \mathbb{R}^2)$，则称该参数曲线为 C^k **（参数）曲线**或者 C^k **光滑（参数）曲线**.

注意，参数曲线的光滑性并不意味着它的像 $\boldsymbol{r}(I)$ 在几何上是光滑的. 比如 $\boldsymbol{r}(t)=(t^3,t^2)$ 看似光滑, 实则有尖点. 为此我们引入一种更好的参数曲线.

正则参数曲线

设参数曲线 $\boldsymbol{r}(t)=(\phi(t),\psi(t))\in C^k(I,\mathbb{R}^2)$, 其中 $k\geq 1$. 若 $(\phi')^2+(\psi')^2\neq 0 (t\in I)$, 则称 $\boldsymbol{r}(t)$ 为**正则参数曲线**，简称正则曲线.

正则曲线在任意参数 t_0 附近是几何上光滑的. 不妨设 $\phi'(t_0)\neq 0$, 由于 ϕ 至少 C^1，则存在邻域 $U(t_0)$, 使得 ϕ' 在 $U(t_0)$ 上恒不为零. 根据可导的反函数定理，$x=\phi(t)(t\in U(t_0))$ 存在 C^k 的反函数，记作 $t=t(x)$. 进而，当 $t\in U(t_0)$ 时，参数曲线也可写为

$$\boldsymbol{r}=(\phi(t),\psi(t))=(x,\psi(t(x))).$$

换言之，这一段曲线是函数 $y=\psi(t(x))$ 的图像. 根据链式法则和反函数导数公式，曲线在 t_0 处的切线的斜率为

$$k(t_0)=\frac{dy}{dx}|_{x_0}=\psi'(t_0)t'(x_0)=\psi'(t_0)\cdot\frac{1}{\phi'(t_0)}.$$

这一点也可以用 $\Delta\boldsymbol{r}_{t_0}=\boldsymbol{r}(t)-\boldsymbol{r}(t_0)$ 来逼近得到，即

$$k(t_0)=\lim_{t\to t_0}\frac{\psi(t)-\psi(t_0)}{\phi(t)-\phi(t_0)}=\lim_{t\to t_0}\frac{\frac{\psi(t)-\psi(t_0)}{t-t_0}}{\frac{\phi(t)-\phi(t_0)}{t-t_0}}=\frac{\psi'(t_0)}{\phi'(t_0)}.$$

因此，曲线在 $\boldsymbol{r}(t_0)$ 处的**切向量**可取为 $\boldsymbol{r}'(t_0)$. 换言之，终端曲线的切向量就是向量值函数的导数.

方便起见，通常也把参数曲线简记为

$$\boldsymbol{r}(t)=(x(t),y(t)).$$

那么正则曲线的切向量就是

$$\boldsymbol{r}'(t)=(x'(t),y'(t)).$$

作为一个副产品，我们得到了**参数方程求导公式**.

参数方程求导

设 $y=f(x)$ 的图像是正则参数曲线 $(x(t),y(t))$ 的像，其中 $x'\neq 0$，则

$$\frac{dy}{dx}=\frac{dy/dt}{dx/dt}=\frac{y'(t)}{x'(t)}.$$

需要指出的是，参数方程

$$\begin{cases} x=\phi(t),\\ y=\psi(t)\end{cases}$$

的二阶导数计算并不容易. 事实上

$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}\left(\frac{\psi'}{\phi'}\right) = \frac{d}{dt}\left(\frac{\psi'}{\phi'}\right)\cdot\frac{dt}{dx} \\ &= \left(\frac{\psi'}{\phi'}\right)'\cdot\frac{1}{\phi'} = \frac{\psi''\phi'-\psi'\phi''}{(\phi')^3}.\end{aligned}$$

下面是一种常见的错误

$$y'' = \left(\frac{\psi'}{\phi'}\right)' = \frac{\psi''\phi'-\psi'\phi''}{(\phi')^2},$$

其中的问题在于混淆了撇的含义：最左端的撇是关于 x 求导数，而之后的撇是关于 t 求导数. 因此当变量较多时，建议使用莱布尼茨的符号.

例 1.4.1 极坐标方程 $r = a^\theta$ 表示的曲线称为**等角螺线**.[7] 证明: 等角螺线在各点处的切线与径向夹角是常量.

7: 等角螺线也叫对数螺线，它有很多有趣的性质，本例所证等角性便是其一. 由于飞蛾以路线与平行光的夹角导航，因此当光源较近时会导致飞蛾扑火的现象. 此外，它还有自相似性，将其放缩 k 倍之后的方程是 $r = ka^\theta$ 恰与原曲线旋转一定角度的方程 $r = a^{\theta+\log_a k}$ 一致. 这一性质具有生物学上的意义: 生物可以不改变形状而长大. 比如鹦鹉螺的贝壳、向日葵的种子均具有对数螺线结构. Jacob Bernoulli 对等角螺线有深入研究，对其性质大为惊叹，以致立下遗嘱要求在其墓碑上刻上等角螺线并附词“纵使改变, 依然故我”. 故事的结局是部反转剧: 雕刻师将等角螺线误刻为等速螺线(阿基米德螺线 $\rho = a\theta$).

解. 极坐标方程可以化为参数方程

$$\boldsymbol{r} = (r\cos\theta, r\sin\theta) = (a^\theta\cos\theta, a^\theta\sin\theta).$$

它的切向为

$$\begin{aligned}\boldsymbol{r}'(\theta) &= (a^\theta\ln a\cos\theta - a^\theta\sin\theta, a^\theta\ln a\sin\theta + a^\theta\sin\theta) \\ &= a^\theta\sqrt{(\ln a)^2+1}\cdot(\cos(\theta+\alpha), \sin(\theta+\alpha)),\end{aligned}$$

其中 $\tan\alpha = 1/\ln a$. 显然切向与径向的夹角为常量 α. □