# 6.5 从公理出发的微分法

**工作纸**

**题目**

**STEP II 1992，第 4 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/92-s2-q4

## 提示

请先自行尝试解题。如果遇到困难，请逐一阅读提示，每读完一个便回到题目再试。标有“停一停”或“试一试”的提示是关键步骤。每个提示之后均留有空位供你演算。

> **想一想**　法则 (i) 就是积法则。积法则怎可能告诉你关于常数的任何事情？（试把常数看成一个乘积。）

### 阶段 1 常数（constant）

涵盖：常数

**想一想** 法则 (i) 是积法则（product rule）。它怎能告诉你关于常数的事情？试把 1 看成 $1 \times 1$。

**停一停，试一试** 取 $f = g = 1$，应用法则 (i)。然后用法则 (iv) 处理其他任何常数。

### 阶段 2　幂（power）

涵盖：幂 $x^2$ 和 $x^3$

**想一想**　写成 $x^2 = x \cdot x$ 和 $x^3 = x^2 \cdot x$，并运用法则 (i) 和 (iii)。

**停一停，试一试**　用数学归纳法（induction）证明 $\Diamond x^n = nx^{n-1}$。

**阶段 3 所有多项式（polynomial）**

涵盖：任何多项式

? **想一想** 把一般的多项式写成 $a_0 + a_1 x + \cdots + a_m x^m$，并运用法则 (ii) 和 (iv)。

## 背后的数学

### 导子

在一个代数系统上，若运算 $D$ 具有可加性（additive），并且满足积法则 $D(fg) = fD(g) + gD(f)$，便称为导子。这道题说明，在多项式上，一个把常数变成零的导子完全由 $D(x)$ 决定：事实上 $D(h) = h'(x)\,D(x)$，所以唯一的自由度就是多项式 $D(x)$ 的选择。法则 (iii) 规定了 $D(x) = 1$。导子遍布数学各个领域：在几何中，曲面上的向量场（vector field）可以定义为该曲面上的函数的一个导子（它沿着向量场的方向对函数求导）；在代数中，一个结构的导子量度它的“无穷小对称”（infinitesimal symmetries）；在物理中，量子力学的交换子（commutator）$[A, B] = AB - BA$ 对每个变元而言都是一个导子。

### 没有极限的微积分

在极限完全没有意义的场合，代数式的导数依然行得通。在有限域（finite field）上，例如模质数 $p$ 的算术，并没有极限可言，但多项式的形式导数（formal derivative）仍由同样的法则定义，而且仍然能够侦测重根（repeated root）：多项式 $f$ 有重根，当且仅当 $f$ 与 $f'$ 有公因式。举例说，我们就是这样证明当 $p \nmid n$ 时 $x^n - 1$ 没有重根的；要在有限域上定义 Section 6.1 的分圆多项式（cyclotomic polynomial），就需要这个事实。编码理论和密码学也用到同样的想法。

### 自动微分

以下是一个现代应用。创造一个数 $\varepsilon$，使得 $\varepsilon^2 = 0$ 但 $\varepsilon \neq 0$，然后用*对偶数*（dual number）$a + b\varepsilon$ 来做算术。这时 $(a + b\varepsilon)(c + d\varepsilon) = ac + (ad + bc)\varepsilon$：$\varepsilon$ 的部分恰好遵从积法则。因此对任何多项式，

$$f(x + \varepsilon) = f(x) + f'(x)\,\varepsilon,$$

而在对偶数 $x + \varepsilon$ 处计算 $f$，就能同时求出函数值和导数，结果完全准确，既不用极限，也没有因微小差值而产生的舍入误差（图 6.5.1）。这称为*前向模式自动微分*（forward-mode automatic differentiation）。它的近亲“反向模式”（reverse mode），就是用来训练拥有数十亿个参数的神经网络（neural network）的算法（*反向传播*，backpropagation）：这类系统每学习一次，都是在把积法则和链式法则（chain rule）当作导子，一步一步地应用。

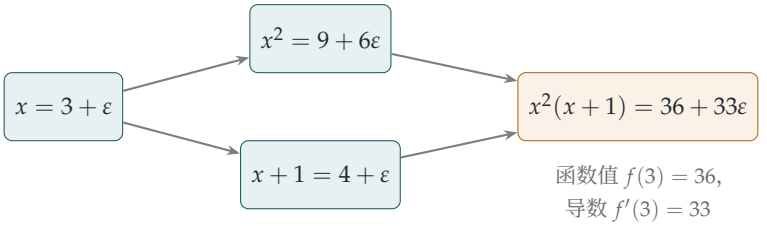


**图 6.5.1**　用对偶数（$\varepsilon^2 = 0$）对 $f(x) = x^2(x+1)$ 在 $x = 3$ 处作自动微分。每一步都遵从题目中的法则：$(3 + \varepsilon)^2 = 9 + 6\varepsilon$，以及 $(9 + 6\varepsilon)(4 + \varepsilon) = 36 + (9 + 24)\varepsilon$。$\varepsilon$ 的系数就是 $f'(3) = 3 \cdot 9 + 2 \cdot 3 = 33$，计算结果完全准确。

## 延伸练习

1 **另一个导子。**假设 $\Diamond$ 满足法则 (i)、(ii) 和 (iv)，但 $\Diamond x = x$ 而不是 1。证明 $\Diamond \mathrm{h}(x) = x\,\mathrm{h}'(x)$。$\Diamond$ 会把 $x^n$ 变成甚么？

2 **商法则（quotient rule）。**把 $\Diamond$ 推广到有理函数（rational function）$\mathrm{f}/\mathrm{g}$，并假设法则 (i) 仍然成立。把法则 (i) 应用于 $\mathrm{g} \cdot \frac{\mathrm{f}}{\mathrm{g}} = \mathrm{f}$，从而推导出商法则。

3 **重根。**证明若 $(x - a)^2$ 整除多项式 f，则 $x - a$ 整除 f′。利用这个结果证明 $x^n - 1$ 没有重根。

4 **对偶数。**设 $f(x) = (x^2 + 1)^3$，用对偶数计算 $f(2)$ 和 $f'(2)$，并用链式法则验算。

## 延伸阅读

- **Peter Cameron, *Introduction to Algebra***（Oxford University Press）。大学一年级程度的环、多项式和域，并包括形式导数。
- **Andreas Griewank & Andrea Walther, *Evaluating Derivatives***（SIAM）。自动微分的标准参考书。程度较深。
- **Michael Nielsen, *Neural Networks and Deep Learning***（网上免费阅读）。第 2 章清楚讲解反向传播。

完整的引导式解答见第 6.5 节：

https://beyondstep.keithmcnulty.org/zh-hans/ch06/derivations.html

每星期一道 STEP 题目，完整讲解：The STEP Up，以电子邮件免费发送，订阅网址：keithmcnulty.substack.com