6.5 公理から考える微分

ワークシート

問題

STEP II 1992、問 4

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

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

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

? **考えてみよう**　規則 (i) は積の法則である。積の法則から、定数について何がわかるだろうか。（定数を積とみなしてみよう。）

段階 1 定数

扱う内容：定数

考えてみよう 規則 (i) は積の法則（product rule）である。それが定数（constant）について何かを教えてくれるとしたら、どのようにしてだろうか。1 を 1×1 と考えてみよ。

手を止めて、やってみよう $f = g = 1$ として規則 (i) を適用せよ。そのうえで、それ以外の定数については規則 (iv) を用いよ。

段階 2　累乗

扱う内容：累乗（powers）x^2 と x^3

考えてみよう　$x^2 = x \cdot x$、$x^3 = x^2 \cdot x$ と書き、規則 (i) と (iii) を用いよ。

手を止めて、やってみよう　$\Diamond x^n = n x^{n-1}$ を数学的帰納法（induction）で証明せよ。

段階 3 すべての多項式

扱う内容：任意の多項式（polynomial）

> **考えてみよう** 一般の多項式を $a_0 + a_1 x + \cdots + a_m x^m$ と書き、規則 (ii) と (iv) を用いよ。

背後にある数学

導分

ある代数系の上の演算 D が加法的であり、積の法則 $D(fg) = fD(g) + gD(f)$ を満たすとき、それを**導分**という。この問題が示すのは、多項式の上では、定数を 0 に写す導分は $D(x)$ によって完全に決定されるということである。実際 $D(h) = h'(x)\,D(x)$ であり、自由に選べるのは多項式 $D(x)$ だけである。規則 (iii) は $D(x) = 1$ と定めている。導分は数学のさまざまな場面に現れる。幾何学では、曲面上の**ベクトル場**（vector field）を、その曲面上の関数に対する導分として定義できる（ベクトル場の方向に関数を微分するのである）。代数学では、ある構造の導分はその「無限小の対称性」を測る。物理学では、量子力学の交換子（commutator）$[A, B] = AB - BA$ が各引数について導分になっている。

極限を使わない微積分

代数的な導関数は、極限がまったく意味をもたない状況でも機能する。素数 p を法とする算術のような有限体（finite field）の上には極限は存在しないが、多項式の形式的導関数（formal derivative）は同じ規則によって定義され、それでもなお重根（repeated root）を検出できる。多項式 f が重根をもつのは、f と f' が共通因数をもつときであり、かつそのときに限る。たとえば、$p \nmid n$ のとき $x^n - 1$ が重根をもたないことはこのようにして示される。これは、Section 6.1 の円分多項式（cyclotomic polynomial）を有限体上で定義するのに必要な事実である。同じ考え方は符号理論（coding theory）や

暗号理論（cryptography）でも用いられている。

自動微分

ここで現代的な応用を紹介しよう。$\varepsilon^2 = 0$ だが $\varepsilon \neq 0$ であるような数 ε を考え出し、**二重数**（dual numbers）$a + b\varepsilon$ で計算を行う。すると $(a + b\varepsilon)(c + d\varepsilon) = ac + (ad + bc)\varepsilon$ となり、ε の部分はまさに積の法則に従う。したがって、任意の多項式について

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

が成り立ち、二重数 $x + \varepsilon$ において f を評価すると、値と導関数の値が同時に、しかも正確に計算される。極限は不要であり、小さな差をとることによる丸め誤差も生じない（図 6.5.1）。これを**フォワードモード自動微分**（forward-mode automatic differentiation）という。その親戚であるリバースモードは、数十億のパラメータをもつニューラルネットワーク（neural networks）を訓練するのに使われるアルゴリズム（**誤差逆伝播法**（backpropagation））である。このようなシステムが学習するたびに、積の法則と連鎖律（chain rule）を導分として一歩ずつ適用しているのである。

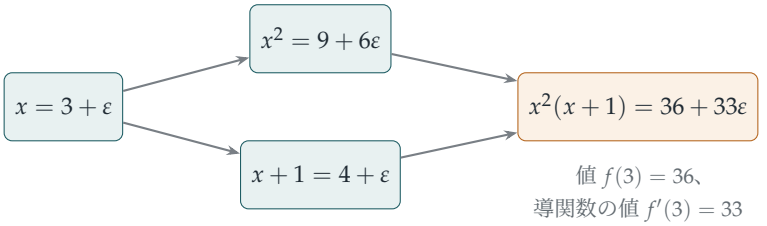

図 6.5.1　二重数（$\varepsilon^2 = 0$）を用いた、$x = 3$ における $f(x) = x^2(x+1)$ の自動微分。各段階は問題の規則に従っている。すなわち $(3+\varepsilon)^2 = 9 + 6\varepsilon$、$(9 + 6\varepsilon)(4 + \varepsilon) = 36 + (9 + 24)\varepsilon$ である。ε の係数は $f'(3) = 3 \cdot 9 + 2 \cdot 3 = 33$ であり、正確に計算されている。

次の一歩

1. **別の導分。** $\Diamond$ が規則 (i)、(ii)、(iv) を満たすが、1 の代わりに $\Diamond x = x$ であるとする。$\Diamond \mathrm{h}(x) = x\,\mathrm{h}'(x)$ を示せ。$\Diamond$ は x^n をどう写すか。

2. **商の法則。** 規則 (i) が引き続き成り立つと仮定して、$\Diamond$ を有理関数（rational functions）f/g に拡張する。$\mathrm{g} \cdot \frac{\mathrm{f}}{\mathrm{g}} = \mathrm{f}$ に規則 (i) を適用して、商の法則（quotient rule）を導け。

3. **重根。** $(x-a)^2$ が多項式 f を割り切るならば、$x - a$ は f' を割り切ることを示せ。これを用いて、$x^n - 1$ が重根をもたないことを示せ。

4 **二重数。**二重数を用いて、$f(x) = (x^2 + 1)^3$ に対する $f(2)$ と $f'(2)$ を計算し、連鎖律で検算せよ。

さらに学ぶために

- **Peter Cameron, *Introduction to Algebra***（Oxford University Press）。大学 1 年次の水準で環（rings）、多項式、体（fields）を扱い、形式的導関数も取り上げている。
- **Andreas Griewank & Andrea Walther, *Evaluating Derivatives***（SIAM）。自動微分に関する標準的な文献。上級者向け。
- **Michael Nielsen, *Neural Networks and Deep Learning***（オンラインで無料公開）。第 2 章で誤差逆伝播法がわかりやすく解説されている。

詳しい解説付きの解答は第 6.5 節にある：

https://beyondstep.keithmcnulty.org/ja/ch06/derivations.html

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com