The Derivative

Ernie Manes, University of Massachusetts
Amherst, Massachusetts, USA

FMCS 2018
May 20, 2018

0. SDG for Physicists

Forget categories. We need only add an axiom to the real numbers.

Axiom: If $\epsilon > 0$ then $\epsilon^3 = 0$.

Represent a non-abelian Lie group $G = \{g, h, \ldots\}$ by “infinitesimal matrix transformations” $I + \epsilon A$.

We have

$(I + \epsilon A)^{-1} = I - \epsilon A + \epsilon^2 A^2$

How will a group commutator $ghg^{-1}h^{-1}$ map?

$$(I+\epsilon A)(I+\epsilon B)(I-\epsilon A+\epsilon^2 A^2)(I-\epsilon B+\epsilon^2 B^2)$$

$$= \quad I + \epsilon^2 [A, B]$$

so the bracket measures "failure to commute".

1. From Lagrange to Differential Categories

Some early history of calculus

- John Wallis, 1655 (Newton was 12, Leibniz was 9). By analyzing $\int_o^\pi \sin^n x\, dx$,
$$\frac{\pi}{2} = \frac{2}{1}\frac{2}{3}\frac{4}{3}\frac{4}{5}\frac{6}{5}\frac{6}{7}\frac{8}{7}\frac{8}{9}\frac{10}{9}\frac{10}{11}\cdots$$
- Newton 1687 *Philosophi Naturalis Principia Mathematica*
- Taylor 1712. Sort of proved Taylor's theorem.

So neither Newton nor Leibniz invented calculus from scratch.

- Lagrange 1797 was Euler's student and had Fourier as a student. At Age 61, *Théorie des Fonctions Analytiques*.
  - A rigorous 600 page calculus text.
  - The "Lagrange remainder" for Taylor polynomials.
  - (With Euler) the Lagrangian in mechanics.
  - "Lagrange points" in the sun-Jupiter system predicted a concentration of asteroids which was found in 1906.
  - Lagrange's identity from vector calculus:

$$(a\times b)\,(c\times d) \;=\; (a\cdot c)(b\cdot d)-(b\cdot c)(a\cdot d)$$

A formal power series $\mathbb{R}[[t]]$ is a model of a one-variable function.

These form a commutative $\mathbb{R}$-algebra. $(\Sigma\, a_k t^k)\,(\Sigma\, b_m t^m) = \Sigma\, c_n t^n$, with

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

This algebra has a lot of structure (see Niven's paper).

Toward the end of the talk we will consider an operator-theoretic formulation of $\mathbb{R}[[t]]$ which invites categorical axiomatization.

Facts about $\mathbb{R}[[t]]$.

- $\Sigma\, a_n t^n$ is invertible $\Leftrightarrow a_0 \neq 0$. Lagrange gave a recursive formula.
- Can differentiate and integrate term-by-term.
- Big problem: cannot always define $f(g(t))$ if $g(0) = 0$.

Using $\mathbb{R}[[t]]$ to Solve Problems

$$\begin{aligned} x'(t) &= x(t) \\ x(0) &= 1 \end{aligned}$$

write

$$x = \sum_{n=0}^{\infty} a_n t^n, \quad x'(t) = \sum_{n=1}^{\infty} n a_n t^{n-1}$$

and equate coefficients to get

$$x = \sum_{n=0}^{\infty} \frac{t^n}{n!}$$

See Exercises 1, 2 for other examples.

Even if no solution in elementary functions exists, can still do this!

One of the great problems from Newton to the twentieth century was the three-body problem.

Our Colleague John Gray has a son Jeremy whose student June Barrow-Green has written a wonderful book *Poincaré and the Three Body Problem.*

It started with Newton's attempts to predict moon position. He failed badly.

In 1860 and 1867, Delaunay found a Hamiltonian for the Sun-Earth-moon system and approximated it with a series. The results were encouraging.

Modern calculations (see Meeus' book) show it is more than a three body problem (Venus and Jupiter have big effects) and many hundred perturbations enter the algorithm.

In 1912, K. F. Sundman found an infinite series solution for a general three body problem.

Precious little qualitative information resulted (but, see Barrow-Green pages 187–192).

Sundman made extensive use of Complex variable theory.

A 1993 Precursor to Differential Categories

In his book *Real and Functional Analysis*, Serge Lang explained how calculus fits into functional analysis.

He used Banach spaces, both with smooth maps and continuous linear maps, the latter category being closed.

It seems to me a reasonable expectation of the differential categorists to *precisely* relate this example to their work.

Fundamental properties of Lang's category:

The derivative of $f : E \longrightarrow F$ has form $f' : E \longrightarrow [E, F]$.

A product rule holds for any product $E \otimes E \longrightarrow F$. This includes the usual one-variable product rule as well as the rules for differentiating a dot product or a cross product.

The usual chain rule holds for the derivative of the composition of two morphisms.

If $f' = 0$ on the convex hull $[x, y]$ of x, y, f is constant on $[x, y]$ (the proof uses Hahn-Banach).

Taylor's theorem:

$$f(x+a) = f(x) + f'(x)a + \frac{f''(x)(a \otimes a)}{2} + \\ \ldots + \frac{f^{(n-1)}(x)(a \otimes \ldots \otimes a)}{(n-1)!} + \ldots$$

Partial derivatives use $\times$, not $\otimes$.

What seems to be called for is, at least, a closed category enriched over real or complex vector spaces together with

- a cartesian structure
- a "coalgebra modality"

$$X \to X \otimes X$$

Should we ignore scalars?

> "It's not a question of the maximum generality but the *right* generality."
>
> Saunders Mac Lane

From Kelley and Namioka page 108

"This chapter, which begins our intensive use of scalar multiplication in the theory of linear topological spaces, marks the definite separation of this theory from that of topological groups."

Is the Line the Same as the Plane?

In the early twentieth century, topologists strived to show that $\mathbb{R}^m$, $\mathbb{R}^n$ are different if $m \neq n$.

This holds as real vector spaces.

But *not* as abelian groups.

2. Lagrange and Cotangents

In a paper in the Monthly on differential geometry in mechanics, Mac Lane emphasized the cotangent bundle and its use as a model for the phase space.

Let M be a manifold with tangent bundle $\pi : TM \to M$.

The fibre $\pi^{-1}r$ is $\{r\} \times M_r$ with M_r a vector space.

A (global) p-**form** is an alternating p-linear $(M_r)^p \to \mathbb{R}$ which varies over M in a C^∞ way.

Alternating means if two arguments are equal the result is 0.

The theory of the next slide depends on the Lagrangian, explaining the section title.

Conjecture: We can do the following in any tangent category if fibres are vector spaces.

$T^\star M$ is $M_r^\star$ on each fibre, the **cotangent bundle**. We have

$$T^\star M \xleftarrow{\alpha=\pi_{T^\star M}} TT^\star M \xrightarrow{\quad\beta=T\rho_M\quad} TM$$

Then $\theta x = (\alpha x)\,(\beta x)$ is a 1-form on $T^\star M$.

$d\theta$ is a closed 2-form of maximal rank, i.e. is a *symplectic* form.

See Mac Lane's paper or Bishop and Goldberg, Chapter 6 for how to use $T^\star M$ with this structure to do Hamiltonian mechanics.

3. A crisis with continuity

The Schrödinger wave equation replaces Newton's $F = ma$ with a partial differential equation (see Griffith's book). The story begins with D'Alembert in 1746 when Lagrange was 10:

D'Alembert's wave equation:

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2}, \;\; u(0,t) = 0 = u(b,t)$$

He showed that $u(x, 0)$ can be any odd twice differentiable function. See Exercises 3, 4.

Fourier (1822, also Euler, Bernoulli):

$$u(x,0) = \sum_{n=1}^{\infty} \beta_n \sin nax$$

One discovers this by approaching D'Alembert's equation by "separation of variables",
i.e. by assuming that

$$u(x,t) \;=\; X(x)\,T(t)$$

to get

$$u(x,t) \;=\; \sum_{n=1}^{\infty} \sin nx \,(\alpha_n \sin nat + \beta_n \cos nat)$$

For the Schrödinger equation, the terms with separated variables are eigen-states.

So we have the crisis that any twice differentiable function has infinitely many derivatives. Here's another crisis:

$$f(x) \;=\; \frac{4}{\pi}(\sin\, x+\frac{1}{3}\,\sin\, 3x+\frac{1}{5}\,\sin 5x+\cdots)$$

This converges pointwise to the Heaviside function

$$H(x) \;=\; \begin{cases} -1 & : \quad -\pi < x < 0 \\ \;\;0 & : \quad x \in \{0, \pi\} \\ \;\;1 & : \quad 0 < x \leq \pi \end{cases}$$

Thus the pointwise limit of a sequence of real-analytic functions need not be continuous.

So it's not just a need to tighten up how we use series. Continuity itself behaves badly.

This is one reason mathematicians were led to measure theory.

A topological space X is a measurable space whose measurable sets are the **Borel sets**, the σ-algebra generated by the open sets.

Continuous functions are measurable.

Theorem Unlike for continuous functions, if f_n is a sequence of bounded real-valued measurable functions, its supremum and infimum are again measurable.

Theorem (Lusin 1912) If $f : [a, b] \to \mathbb{R}$ is measurable, μ is Lebesgue measure, and $\epsilon > 0$, there exists continuous g such

$$\mu(\{x : fx \neq gx\}) < \epsilon$$

But new strangeness appears!

A space with the topology of a second countable complete metric space is a **Polish space**. A **standard Borel space** has the Borel sets of a Polish space.

We know the interval and the square are not homeomorphic. However,

Theorem (J. von Neumann, 1932) $[0,1]$ and $[0,1] \times [0,1]$ are isomorphic measure spaces. Indeed, any standard Borel space X with $\mu(X) = 1$ and $\mu(\{x\}) = 0$ is isomorphic to Lebesgue measure on $[0,1]$.

4. Derivatives via integrals

Enter functional analysis: Banach spaces.

The L^p spaces

For $1 \le p < \infty$ and measure space $(X, \mathcal{M}, \mu)$, $L^p(\mu)$ is the Banach space of almost-everywhere classes of measurable $f : X \to \mathbb{C}$ such that

$$\int_X |f|^p \, d\mu < \infty$$

with norm

$$\| f \|_p = \left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}}$$

It is a nontrivial theorem that this normed space is complete, and hence is a Banach space.

For any $f \in L^p(-\pi, \pi]$ define its **Fourier coefficients** by

$$\widehat{f}(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\, e^{-ijt} dt$$

The **Fourier series** of f is

$$\Sigma\, \widehat{f}(j)\, e^{ijt} \qquad (j \in \mathbb{Z})$$

Note: By Euler's identity

$$e^{ijt} = \cos\, jt + i\, \sin\, jt$$

so this is the same form as seen in the wave equation solutions.

Pathology for L^1

In 1923, Kolmogorov gave an example of an L^1-function whose Fourier series diverges almost everywhere.

Better Results if $p > 1$.

Theorem The Fourier series of $f \in L^p(-\pi, \pi]$ $\| \cdot \|_p$-converges to f.

Theorem If f has a continuous first derivative, its Fourier series converges uniformly to f pointwise.

Theorem (Carleson 1966, Hunt 1968, very difficult!) The Fourier series converges pointwise to f almost everywhere.

So unlike as in Taylor's theorem, we approximate f by evaluating integrals.

A function is "smooth" if its Fourier coefficients decay rapidly to 0 (Kranz *Complex Analysis*, Proposition 1.1.8).

Impediments to smoothness:

- Some $f^{(k)}$ has large L^1-norm
- Not enough derivatives exist
- Is $\sin 1000x$ smooth? Let's take a look.

5. Complex Analysis

For $\Omega \subset \mathbb{C}$ open, $f : \Omega \to \mathbb{C}$ is **holomorphic at** $z_o \in \Omega$ if the following limit exists:

$$f'(z_o) = \lim_{z \to z_o} \frac{f(z) - f(z_o)}{z - z_o}$$

If $f(x + iy) = u(x, y) + i\,v(x, y)$, f is holomorphic if and only if the **Cauchy-Riemann equations** (discovered by D'Alembert in 1752) hold:

$$\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

See Exercises 6,7,8,9,10.

Taylor's theorem for Holomorphic $f : U \to \mathbb{C}$

We can either integrate or differentiate:

$$
\begin{aligned}
f(z) &= \sum_{k\geq 0} a_k (z - z_o)^k \\
a_k &= \frac{1}{k!}\frac{d^k f}{dz^k}(z_o) \\
&= \frac{1}{2\pi i}\oint_\gamma \frac{f(w)}{(w - z_o)^{k+1}} dw
\end{aligned}
$$

- The radius of convergence is always > 0.
- All derivatives always exist.

For $f : \mathbb{R}^n \to \mathbb{C}$ in $L^1(\mathbb{R}^n)$, it's **Fourier transform** is the function of the same form given by

$$\widehat{f}(\xi) = \int_{\mathbb{R}^n} f(t) e^{(it)\xi} dt$$

- A similar integral computes the inverse transform.
- So long as the integral is finite, $\widehat{f}$ is uniformly continuous.

The transform of a derivative is very simple:

$$(\widehat{\frac{\partial f}{\partial x_j}})(\xi) = -i\xi_j \widehat{f}(\xi)$$

The computation time of this way of differentiating is decreased by using the *fast Fourier transform*

The **uncertainty principle** holds: f and $\widehat{f}$ cannot both have compact support.

Indeed, no nonzero f can have both itself and its Fourier transform supported on a set of finite measure.

6. Polynomials 1: approximations

Between "linear" and "differentiable" we have polynomials.

How would one put these in a differential category?

Theorem (Weierstrass 1885) A continuous function on $[a, b]$ is the uniform limit of a sequence of polynomials.

For $f : [0, 1] \to \mathbb{R}$ continuous, its nth **Bernstein polynomial** is

$$B_{f,n}(x) = \sum_{n=0}^{n} \binom{n}{k} f(\frac{k}{n})\, x^k\, (1-x)^{n-k}$$

Theorem (Bernstein 1912) On [0,1], $B_{f,n}$ converges uniformly to f.

Let's see a demonstration:

For a differentiable function, a very compelling polynomial approximation method is **Hermite interpolation**.

Given x-values $x_0 < \cdots < x_n$ on which f is defined, the desired polynomial p is the one of least degree for which $p(x_i) = f(x_i)$, $p'(x_i) = f'(x_i)$ for all i.

While developed by Hermite in 1864, there was earlier work by Chebyshev in 1859 and Laplace in 1810.

Let's see an example.

7. Polynomials 2: Operators

We have a category with $\mathbb{R}$-vector space hom-sets equipped with a special $\mathbb{R}$-algebra object "polynomials" with a differentiation endomorphism

$$\mathcal{P} \xrightarrow{\ D\ } \mathcal{P}$$

It's iterates have a "binomial property" with respect to pointwise multiplication, namely

$$D^n(pq) \;=\; \sum_{j=0}^{n} \binom{n}{j} (D^j p)(D^{n-j} q)$$

Working in $\mathcal{P}$

For $a \in k$, the **shift operator** $E^a : \mathcal{P} \to \mathcal{P}$ is defined by

$$E^a(p(x)) \;=\; p(x+a)$$

Note: if $p \in \mathbb{R}[[t]]$, the composition $p(x+a)$ may not be defined if $a \neq 0$.

An operator $T : \mathcal{P} \to \mathcal{P}$ is **shift-invariant** if $E^a \circ T = T \circ E^a$ for all a, that is,

$$(Tp)(x+a) \;=\; T(p(x+a))$$

Shift-invariant operators form an $\mathbb{R}$-algebra where multiplication is composition.

Theorem (Rota, Kahaner and Odlyzko)
If T is shift-invariant then there exist unique a_n with

$$T = \sum_{n \geq 0} \frac{a_n}{n!} D^n$$

- $a_n = (Tx^n)(0)$
- There is an $\mathbb{R}$-algebra isomorphism between shift-invariant operators and formal power series:

$$\sum \frac{a_n}{n!} D^n \mapsto \sum \frac{a_n}{n!} t^n$$

- Hence shift-invariant operators commute.
- T is invertible if and only if $T1 \neq 0$.

Examples of shift-invariant operators

- $E^a = \Sigma \frac{a^n}{n!} D^n$ with inverse E^{-a}.
- The **Bernoulli operator**

$$Tp = \int_x^{x+1} p(t)dt$$

$$T = \Sigma \frac{1}{(n+1)!} D^n \text{ is invertible.}$$

- The **Euler mean operator**

$$Tp \;=\; \frac{1}{2}(p(x) + p(x+1))$$

$$T = \frac{1}{2}(I{+}E^1) \;=\; I + \Sigma\,(\frac{1}{2n!}D^n : n \geq 1)$$

$$T^{-1} \longleftrightarrow \tfrac{2}{e^t - 1}$$

- The **difference operator**

$$\Delta = E^1 - I$$

$$Tp \;=\; p(x+1) - p(x)$$

$$T = \Sigma\,(\frac{1}{n!}D^n : n \geq 1)$$

$T \longleftrightarrow e^t - 1$ is not invertible.

Differentiating T

Let $\widehat{x} : \mathcal{P} \longrightarrow \mathcal{P}$ be the linear operator $p(x) \mapsto x\,p(x)$.

See Exercise 11.

For linear $T : \mathcal{P} \longrightarrow \mathcal{P}$ define its **Pincherle derivative** as

$$T' = T\widehat{x} - \widehat{x}T$$

Theorem

- If T is shift-invariant, so is T'.
- If shift-invariant T corresponds to the series $f(t)$, T' corresponds to the term-by-term derivative $f'(t)$.

8. Variation of parameters

In an undergraduate ODE course we explain how to solve

$$Ty \;=\; y'' - 2y' + 3y \;=\; t^3 - t^2 + 1$$

- Find an eigenvector basis z_1, z_2 of $Ty = 0$.
- Find one solution y_p by "variation of parameters".
- The general solution is

$$a\, z_1(t) + b\, z_2(t) + y_p(t)$$

.

Well let's see.

- $T = D^2 - 2D + 3I$ is shift-invariant.
- The problem is $Ty = t^3 - t^2 + 1$.
- $T1 = 0 - 0 + 3 \neq 0$ so T is invertible.

So a (polynomial!) solution is

$$T^{-1}(t^3 - t^2 + 1)$$

As Lagrange well knew, if

$$\left(\Sigma\, a_n t^n\right)\left(\Sigma\, b_n t^n\right) = 1$$

then

$$b_0 = \frac{1}{a_0}$$

$$b_n = -\frac{1}{a_0} \sum_{i=1}^{n} a_i b_{n-i}$$

Inverting T this way gives

$$T^{-1} = \frac{1}{3}I + \frac{2}{9}D + \frac{1}{27}D^2 - \frac{4}{81}D^3 + \cdots$$

so the desired polynomial solution $y_p(t)$ is

$$T^{-1}(t^3 - t^2 + 1) = -\frac{1}{27} - \frac{2}{9}t + \frac{1}{3}t^2 + \frac{1}{3}t^3$$

At this juncture, Robin's graduate students can take over.

At least those who are still awake.

To all who are still awake, congratulations for surviving another Manes tutorial!