Dangerous and Illegal Operations in Calculus

Do we avoid differentiating discontinuous functions because it's impossible, unwise, or simply out of ignorance and fear?

Despite the risks, many natural phenomena are best understood in terms of *generalized functions* unacceptable until relatively recently.

Fallacious arguments in false proofs of false theorems in the early nineteenth century generated too much paranoia. For example, Heaviside's mathematical innovations arising in the physics of telegraph cables (1880-7) were disregarded for 30 years.

Only in the 1930s Hadamard, Sobolev, and others made systematic use of non-classical *generalized functions*. In 1952 Laurent Schwartz won a Fields Medal for systematic treatment of these ideas.

Desiderata

We want a large enough class of *generalized functions*, or *distributions*, so that otherwise illegal operations are perfectly fine. Within this class:

- We can differentiate nearly any function as many times as we like, regardless of discontinuities.

- If $\lim_i \int u_i f$ exists for all very nice *test functions* f then the $\lim_i u_i$ exists as a generalized function.

- Any generalized function u should be *approximate-able* by a sequence of nice functions u_i in the *weak sense* that for all test functions f

$$\lim_i \int u_i f = \int u f$$

Differentiating discontinuous functions

Heaviside's function is

$$H(x) = \begin{cases} 1 & (\text{for } x \geq 0) \\ 0 & (\text{for } x < 0) \end{cases}$$

Yes, this function is discontinuous at 0, but the discontinuity is of a straightforward nature.

Let f be a continuously differentiable function. Let $a < 0$ and $b > 0$. Then

$$\int_a^b H(x)\, f'(x)\, dx = \int_0^b f'(x)\, dx$$

$$= f(b) - f(0)$$

by the fundamental theorem of calculus.

We can evaluate $\int_a^b H\, f'$ in another way. Integration by parts is

$$\int_a^b u'\, v\, dx = [uv]_a^b - \int_a^b u\, v'\, dx$$

Despite the fact that $H(x)$ is not *differentiable* at $x = 0$, and is not even *continuous* there, let's not fret about this, and just integrate by parts

$$\int_a^b H(x)\, f'(x)\, dx$$

$$= [H(x)\, f(x)]_a^b - \int_a^b H'(x)\, f(x)\, dx$$

$$= f(b) - \int_a^b H'(x)\, f(x)\, dx$$

Comparing the two expressions

$$\int_a^b H(x)\, f'(x)\, dx = f(b) - f(0)$$

$$\int_a^b H(x)\, f'(x)\, dx = f(b) - \int_a^b H'(x)\, f(x)\, dx$$

gives

$$\int_a^b H'(x)\, f(x)\, dx = f(0)$$

The alleged function $H'(x)$ is denoted

$$\delta = H'$$

δ is often called a **monopole** or **Dirac's delta function** (though it arose in the work of Heaviside decades earlier).

But there is no such function as δ *in any classical sense.*

If $\delta = H'$, since H is constant away from 0, it *looks* like $\delta(x) = 0$ for $x \neq 0$. But what happens at 0?

There is no classical δ

There is no *function* δ (in any usual sense) with the property that

$$\int_a^b \delta(x)\, f(x)\, dx = f(0)$$

for continuous f.

If $\delta(x)$ were *continuous*, then it would have to be 0 for $x \neq 0$: if $\delta(x_o) > 0$ for $x_o \neq 0$, then $\delta(y) > 0$ for y sufficiently near x_o (and away from 0). Make a continuous function f which is 0 except on a small neighborhood of x_o. Then $f(0) = 0$, but the integral of f against δ would be positive, contradiction. ///

(Proving non-existence is impeded by not having a concise description of integration of very general types of functions!)

Approximating δ, δ' weakly

There is no *classical* function δ, but there is a sequence $u_1, u_2, \ldots$ of nice functions *weakly approximating* δ in the sense that, for any continuous function f

$$\lim_j \int_{-\infty}^{\infty} u_j(x)\, f(x)\, dx = f(0)$$

In fact, there are *many* such sequences. For example, the *taller-and-narrower-tent functions*

$$u_j(x) = \begin{cases} 0 & (\text{for } |x| \geq 1/j) \\ j(1 - j|x|)/2 & (\text{for } |x| \leq 1/j) \end{cases}$$

Generalizing the pattern used there, for u continuous on $\mathbf{R}$, with $u = 0$ off $[-1, +1]$, with $u \geq 0$, and $\int_{-1}^{+1} u = 1$, the sequence

$$u_j(x) = j \cdot u(jx)$$

approximates δ in this weak sense.

Similarly, the **dipole** on $\mathbf{R}$ is $\delta' = H''$, defined via integration by parts by

$$\int_{-\infty}^{\infty} \delta'(x)\, f(x)\, dx = -f'(0)$$

There is no classical function δ', but we can *approximate it weakly* by many sequences of nice functions. For example, with

$$u(x) = \begin{cases} \frac{3}{4}(1-x^2) & (\text{for } |x| \le 1) \\ 0 & (\text{for } |x| \ge 1) \end{cases}$$

(with the constant making the integral be 1) we have $u_j(x) = ju(jx) \to \delta(x)$. More or less taking a derivative gives

$$v(x) = \begin{cases} -\frac{3}{4}x^3 & (\text{for } |x| \le 1) \\ 0 & (\text{for } |x| \ge 1) \end{cases}$$

and then

$$v_j(x) = j^2 v(jx) \to \delta'(x)$$

δ is useful

We can systematically solve differential equations.

The **Laplacian** Δ on $\mathbf{R}^n$ is

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$$

Given a function f on $\mathbf{R}^n$, we might want to solve for u in

$$\Delta u = f$$

A **fundamental solution** E for Δ is a function such that

$$\Delta E = \delta$$

where δ is the delta on $\mathbf{R}^n$. Then solve the original equation by **convolution**

$$u(x) = (E * f)(x) = \int_{\mathbf{R}^n} E(x - y) f(y)\, dy$$

Proof?

Proof: If $\Delta E = \delta$ then

$$\Delta u(x) = \Delta \int_{\mathbf{R}^n} E(x-y) f(y)\, dy$$

$$= \int_{\mathbf{R}^n} \Delta E(x-y) f(y)\, dy$$

(bravely moving the differentiation under the integral!)

$$= \int_{\mathbf{R}^n} \delta(x-y) f(y)\, dy = f(x)$$

so we've solved the equation if we can find a fundamental solution E. ///

In $\mathbf{R}^1$ since already $H' = \delta$ one fundamental solution E for Δ would be

$$E(x) = \int_0^x H(t)\, dt = \begin{cases} 0 & (x \le 0) \\ x & (x > 0) \end{cases}$$

Another is

$$E(x) = \frac{1}{2}|x|$$

Fundamental solutions for Δ

A fundamental solution E for Δ would have $\Delta E(x) = 0$ away from 0, but something strange must happen at 0.

The *rotational symmetry* of Δ suggests that that a fundamental solution E for Δ should be rotationally invariant. Maybe

$$E(x) = \text{const} \times |x|^s$$

for some number s. Let $\rho = |x|^2$. Compute

$$\Delta |x|^s = \Delta \rho^{s/2} = \sum_i \frac{\partial}{\partial x_i} \left(s x_i \rho^{\frac{s}{2}-1} \right)$$

$$= s \sum_i \left(\rho^{\frac{s}{2}-1} + (\frac{s}{2} - 1) x_i^2 \rho^{\frac{s}{2}-2} \right)$$

$$= s(n + s - 2)\, |x|^{s-2}$$

For $n \geq 3$ we have $\Delta |x|^{2-n} = 0$ *away from* 0. *What happens at* 0*?* We claim

$$\lim_{s \to -n} (s + n)|x|^s = (\text{const}) \cdot \delta$$

Proof: This illustrates **regularization**. First, looking near 0 in $\mathbf{R}^n$,

$$\int_{|x|\le 1} |x|^s = \text{area}(S^{n-1}) \cdot \int_0^1 r^{n-1+s} dr$$

converges for $\text{Re}(s) > -n$.

Second, if f is differentiable and $f(0) = 0$, then $f(x)/|x|$ is still continuous at 0, so

$$\int_{\mathbf{R}^n} f(x)\,|x|^s\,dx \quad \text{converges}$$

so for such f

$$\lim_{s\to -n} (s+n) \int f(x)|x|^s$$

$$= \int \frac{f(x)}{|x|} |x|^{1-n} \cdot \lim_{s\to -n} (s+n) = 0$$

That is, if $f(0) = 0$ then this limit is 0.

Third, taking $f(x) = e^{-|x|^2}$ in the integral

$$\int_{\mathbf{R}^n} e^{-|x|^2} \cdot |x|^s \, dx$$

$$= \text{area}(S^{n-1}) \int_0^\infty e^{-r^2} \cdot r^{n-1+s} \, dr$$

$$= \frac{\text{area}(S^{n-1})}{n+s} \int_0^\infty 2re^{-r^2} \cdot r^{n+s} \, dr$$

by integrating by parts. At $s = -n$ letting $t = r^2$ in the integral gives

$$\int_0^\infty 2re^{-r^2} \, dr = \int_0^\infty e^{-t} \, dt = 1$$

Thus

$$\lim_{s \to -n} (n+s) \int_{\mathbf{R}^n} e^{-|x|^2} |x|^s = \text{area}(S^{n-1})$$

$$= \text{area}(S^{n-1}) \cdot e^{-|0|^2}$$

Finally, for continuous f

$$f(x) = \left(f(x) - f(0) \cdot e^{-|x|^2}\right) + f(0) \cdot e^{-|x|^2}$$

The first function vanishes at 0, so

$$\lim_{s \to -n} (s+n) \int_{\mathbf{R}^n} \left(f(x) - f(0)e^{-|x|^2}\right) |x|^s = 0$$

The second is a multiple of $e^{-|x|^2}$, so

$$\lim_{s \to -n} (s+n) \int_{\mathbf{R}^n} f(0)e^{-|x|^2} |x|^s$$

$$= f(0) \lim_{s \to 2-n} \int_{\mathbf{R}^n} e^{-|x|^2} |x|^s$$

$$= f(0) \cdot \text{area}(S^{n-1})$$

That is,

$$\lim_{s \to -n} (s+n) \int_{\mathbf{R}^n} f(x)|x|^s = f(0) \cdot \text{area}(S^{n-1})$$

Thus, up to the area of an $(n-1)$-sphere, we have $\Delta |x|^{2-n} = \text{area}(S^{n-1}) \cdot \delta$ ///

Differentiating under the integral

Interchange of limits is dangerous. Differentiating under the integral with respect to a parameter is an example.

It is true that

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{e^{ixy}}{y-i}\,dy = \begin{cases} e^{-x} & (\text{for } x > 0) \\ 0 & (\text{for } x < 0) \end{cases}$$

Convergence is fragile, but ok. Differentiating has a bad effect, but do it

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{y\,e^{ixy}}{y-i}\,dy = \begin{cases} -e^{-x} & (\text{for } x > 0) \\ ? & (\text{for } x = 0) \\ 0 & (\text{for } x < 0) \end{cases}$$

Add the first to the second

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ixy}\,dy = \begin{cases} 0 & (\text{for } x > 0) \\ ? & (\text{for } x = 0) \\ 0 & (\text{for } x < 0) \end{cases}$$

In fact,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ixy}\,dy = \delta(x)$$

This integral also arises in **Fourier inversion**. The **Fourier transform** $\hat{f}$ of a reasonable function f is

$$\hat{f}(\xi) = \int_{\mathbf{R}^n} f(x)\,e^{-ix\cdot\xi}\,dx$$

Fourier transform converts differentiation into multiplication by an exponential, and *vice versa*, by integration by parts

$$\widehat{x_j f}(\xi) = \int_{\mathbf{R}^n} x_j f(x)\,e^{-ix\cdot\xi}\,dx$$

$$= i\frac{\partial}{\partial \xi_j}\int_{\mathbf{R}^n} f(x)\,e^{-ix\cdot\xi}\,dx = i\frac{\partial}{\partial \xi_j}\hat{f}(\xi)$$

$$\widehat{\frac{\partial f}{\partial x_j}}(\xi) = \int_{\mathbf{R}^n} \frac{\partial}{\partial x_j} f(x)\,e^{-ix\cdot\xi}\,dx$$

$$= -\int_{\mathbf{R}^n} f(x)\,\frac{\partial}{\partial x_j}e^{-ix\cdot\xi}\,dx = i\xi_j\hat{f}(\xi)$$

Thus, to solve for u in a differential equation

$$-\Delta u + \lambda u = f$$

we take Fourier transform of both sides

$$|\xi|^2 \cdot \hat{u} + \lambda \hat{u} = \hat{f}$$

and solve

$$\hat{u} = \frac{\hat{f}}{|\xi|^2 + \lambda}$$

At least for $\lambda > 0$ if $\hat{f}$ is nice then we have a nice expression for $\hat{u}$.

How to recover u from $\hat{u}$?

Fourier inversion recovers u from $\hat{u}$ by

$$u(x) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} \hat{u}(\xi)\, e^{ix\cdot\xi}\, d\xi$$

The obvious attempt to prove Fourier inversion is

$$\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n} \hat{u}(\xi)\, e^{ix\cdot\xi}\, d\xi$$

$$= \frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} u(y)\, e^{ix\cdot\xi}\, e^{-iy\cdot\xi}\, d\xi\, dy$$

$$= \int_{\mathbf{R}^n} u(y) \left(\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n} e^{i(x-y)\cdot\xi}\, d\xi \right)\, dy$$

and if we believe that the inner expression is $\delta(x-y)$ then this is

$$= \int_{\mathbf{R}^n} u(y)\, \delta(x-y)\, dy = u(x)$$

But this begs the question.

It is necessary to express both δ and the identically-1 function $\mathbf{1}$ as *limits* of nicer functions.

The ever-popular **Gaussian** on $\mathbf{R}^n$ is

$$g(x) = e^{-|x|^2/2}$$

Its integral is $(2\pi)^{n/2}$, so

$$u_j(x) = (2\pi)^{-n/2}\, j^n g(jx) \to \delta(x)$$

in the sense that for nice f

$$\lim_j \int_{\mathbf{R}^n} u_j(x) f(x)\, dx \to f(0)$$

We can compute Fourier transforms

$$(2\pi)^{n/2} j^{-n} \hat{u_j}(\xi) = \int_{\mathbf{R}^n} e^{-j^2 x\cdot x/2}\, e^{-ix\cdot\xi} dx$$

$$= e^{-\xi\cdot\xi/2j^2} \int_{\mathbf{R}^n} e^{-(jx+i\xi/j)\cdot(jx+i\xi/j)/2} dx$$

$$= g(\xi/j) \int_{\mathbf{R}^n} e^{-jx\cdot jx/2}\, dx = (2\pi)^{n/2} \cdot \frac{g(\xi/j)}{j^n}$$

by the complex variable trick of moving the contour, so

$$\hat{u_j}(\xi) = g(\xi/j)$$

As $j \to \infty$

$$g(\xi/j) = e^{-|\xi|^2/2j^2} \to \mathbf{1}$$

since these functions flatten out more and more as $j \to \infty$. That is, for any reasonable function f

$$\int_{\mathbf{R}^n} g(\xi/j)\, f(x)\, dx \to \int_{\mathbf{R}^n} \mathbf{1} \cdot f(x)\, dx$$

though the approach to the limit certainly depends on the particular f.

Thus, if we view

$$\int e^{ix\cdot\xi} d\xi$$

as a limit

$$\lim_j \int \widehat{u_j}(\xi)\, e^{ix\cdot\xi} d\xi$$

then *by direct computation* this is

$$\int e^{ix\cdot\xi} d\xi = \lim_j u_j(x) = \delta(x)$$

Even better: duality

Yes, many tangible generalized functions are simply weak limits of more ordinary functions, and admit computations from that viewpoint.

But there is another completely different definition which offers new insight.

Observe that for a nice function u the function-on-functions

$$\lambda_u(f) = \int_{\mathbf{R}^n} u(x)\, f(x)\, dx$$

has the properties that

$$\lambda_u(\alpha f + \beta g) = \alpha\lambda_u(f) + \beta\lambda_u(g)$$

for complex α, β and functions f, g. Further, with various mild hypotheses on u, if $f_i \to f$ *uniformly* then

$$\lambda_u(f_i) \to \lambda_u(f)$$

This may suggest defining a generalized function similarly.

The **test functions** $\mathcal{D}$ on $\mathbf{R}^n$ are infinitely differentiable functions which are identically 0 outside some sufficiently large ball. This may seem tricky or even *grotesque* (in the words of H. Cartan), but there are many such functions. For example,

$$f(x) = \begin{cases} e^{-1/(1-x^2)} & (\text{for } |x| < 1) \\ 0 & (\text{for } |x| \geq 1) \end{cases}$$

Test functions f_i *approach* a test function f if there is a common ball outside which they all vanish, and if the f_i *and all their derivatives* approach f and its derivatives *uniformly.*

That is, for each j, given $\varepsilon > 0$ there is N sufficiently large such that for $i \geq N$ and for all x

$$|f_i^{(j)}(x) - f^{(j)}(x)| < \varepsilon$$

The advantage of including derivatives in the definition of convergence of a sequence is that it makes all differentiation operators

$$f \to \frac{\partial}{\partial x_j} f$$

into *continuous* linear maps

$$\frac{\partial}{\partial x_j} : \mathcal{D} \to \mathcal{D}$$

from the space of test functions to itself.

The **distributions** $\mathcal{D}'$ or **generalized functions** on $\mathbf{R}^n$ are functions $\lambda : \mathcal{D} \to \mathbf{C}$ with the **linearity** property

$$\lambda(\alpha f + \beta g) = \alpha\lambda(f) + \beta\lambda(g)$$

for complex α, β and functions f, g, and with the **sequential continuity** property that if a sequence of test function f_i approaches f in the sense above, then

$$\lambda(f_i) \to \lambda(f)$$

Certainly *continuous functions* or reasonably *integrable functions* u give such continuous functionals by

$$\lambda_u(f) = \int_{\mathbf{R}^n} u(x)\, f(x)\, dx$$

Differentiation of distributions is allowed without restriction, defined exactly to extend integration by parts to weak limits:

$$\left(\frac{\partial}{\partial x_j}\lambda\right)(f) = -\lambda\left(\frac{\partial}{\partial x_j} f\right)$$

The continuity of differentiations as maps $\mathcal{D} \to \mathcal{D}$ assures that derivatives of distributions are again *continuous* linear functionals, hence are still inside the class of distributions.

The **weak limit topology** on distributions says that $\lambda_i \to \lambda$ in $\mathcal{D}'$ if for every test function f the *numbers* $\lambda_i(f)$ approach $\lambda(f)$.

Theorem: Weak limits of sequences of distributions are distributions.

Smoothing

But what have we accidentally included by this inclusive rather than constructive definition? Any unnecessary crazy things?

The theorem below says we can *smooth* a distribution into a *nearby* distribution which is literally a differentiable function. Thus the definition of distribution does not include unnecessary things.

Let T_x be translation $T_x f(y) = f(x+y)$. This is a continuous linear map of test functions to themselves.

Theorem: Let f_i be test functions weakly approaching δ. Let u be any distribution. Then the **smoothings**

$$T_{f_i} u(x) = u(T_x f) \sim \int_{\mathbf{R}^n} f(x+y)\, u(y)\, dy$$

are infinitely differentiable functions and

$$T_{f_i} u \to u \ \ \text{(weakly)}$$

Fourier transforms revisited

Fourier transforms of test functions are *never* test functions. They are *holomorphic* functions defined on open sets containing the real line.

The class $\mathcal{S}$ of **Schwartz functions** on $\mathbf{R}^n$ is slightly larger than the test functions $\mathcal{D}$.

A function φ on $\mathbf{R}^n$ is of **rapid decay** if, for every N,

$$\sup_{x \in \mathbf{R}^n} (1 + |x|^2)^N \cdot |\varphi(x)| < \infty$$

The Schwartz functions $\mathcal{S}$ are infinitely differentiable functions f such that f and all its derivatives are of *rapid decay.*

Lemma: Fourier transform maps $\mathcal{S}$ to itself, and is a bijection.

Proof: Fourier transform interchanges multiplication to differentiation. ///

For a multi-index $m = (m_1, \ldots, m_n)$ in $\mathbf{Z}^n$, use the standard abbreviation

$$\partial^m = \left(\frac{\partial}{\partial x_1}\right)^{m_1} \ldots \left(\frac{\partial}{\partial x_n}\right)^{m_n}$$

A sequence f_i of Schwartz functions converges to f if for all N and for all m

$$\sup_{x \in \mathbf{R}^n} (1 + |x|^2)^N \cdot |\partial^m (f - f_i)(x)| \to 0$$

Lemma: Fourier transform $\mathcal{S} \to \mathcal{S}$ is *continuous.*

Proof: Fourier transform interchanges differentiation and multiplication. ///

In fact, $\mathcal{S}$ is **sequentially complete**. And $\mathcal{D}$ is sequentially complete with the corresponding notion of convergence, and the inclusion $\mathcal{D} \to \mathcal{S}$ is continuous.

More duality

The space $\mathcal{S}' \subset \mathcal{D}'$ of **tempered distributions** consists of distributions u which extend to a continuous linear functionals on $\mathcal{S}$ (not merely on $\mathcal{D}$).

Not every distribution is tempered.

Fourier transforms $\hat{u}$ of tempered distributions u are defined via the Plancherel formula as

$$\hat{u}(f) = u(\hat{f})$$

The continuity of Fourier transform on $\mathcal{S}$ assures that $\hat{u}$ is again a tempered distribution, and $u \to \hat{u}$ is continuous in the weak topology on $\mathcal{S}'$.

As with distributions, every tempered distribution can be weakly approximated by infinitely differentiable functions.

A function u is of **moderate growth** if for some N

$$\sup_{x \in \mathbf{R}^n} (1 + |x|^2)^{-N} \cdot |u(x)| < \infty$$

A continuous function u of moderate growth gives a tempered distribution λ_u by

$$\lambda_u(f) = \int_{\mathbf{R}^n} u(x)\, f(x)\, dx$$

We have already insinuated that the Fourier transform of $\mathbf{1}$ is δ. Similarly

Proposition: the Fourier transform of the tempered distribution $x_1^{m_1} \ldots x_n^{m_n}$ is $i^{|m|}$ times $\partial^m \delta$. ///

$u_s(x) = |x|^s$ is a tempered distribution for $\mathrm{Re}(s) \geq -n$. Compute its Fourier transform?

It is rotationally invariant, so its Fourier transform will be also. It is positive-homogeneous of degree s, and the change-of-variables property

$$f(c \cdot x)\hat{}(\xi) = c^{-n} \hat{f}(\xi/c)$$

suggests that

$$c^s \, \hat{u}_s(\xi) = c^{-n} \hat{u}_s(\xi/c)$$

or replacing c by $1/c$

$$c^{-(s+n)} \, \hat{u}_s(\xi) = \hat{u}_s(c\,\xi)$$

Thus, possibly

$$\hat{u}_s = \text{const}_s \cdot u_{-(s+n)}$$

However, if $\text{Re}(s) > -n$ to have convergence of the integral for u_s, then $\text{Re}(-(s+n)) < 0$. We need $\text{Re}(-(s+n)) > -n$ for convergence of the integral for $u_{-(s+n)}$. The range in which both inequalities are met is

$$-n < \text{Re}(s) < 0$$

But for $\text{Re}(s) >> 0$ *the Fourier transform* $u_{-(s+n)}$ *of* u_s *seems to be given by a horribly divergent integral...*