(ALMOST) EVERYTHING IS DIFFERENTIABLE

CHRIS JOHNSON

ABSTRACT. Mathematicians love to generalize, and this article gives a quick introduction to some important generalizations of familiar objects from calculus that were developed in the middle of the 20th century. Ultimately, our goal is to introduce a generalization of the derivative called a *distributional derivative*, which gives us a way of differentiating functions we could not differentiate in calculus. To motivate this, we spend much of the article discussing the more concrete notion of a *weak derivative*. Along the way we will also touch on the notion of a *measure*. By the end of the article we will see that (almost) every function is differentiable in a generalized sense, but curiously this generalized derivative may not actually be a function.

1. INTRODUCTION

We spend a lot of time in a first semester calculus course learning about derivatives. Ultimately we care about derivatives not for their own sake, but because they are useful tools in solving problems we find interesting or important. For example, when searching for extrema of a function $f(x)$, we first look for points where the derivative $f'(x)$ is either zero or undefined as these critical points are candidates for where the extrema may occur. More fundamentally, the derivative allows us to locally approximate complicated non-linear functions by much simpler linear functions.

Despite their usefulness in solving interesting problems, the derivative we all learn in calculus has a significant flaw: most functions are not differentiable. This might come as a surprise to students whose mathematical career is spent studying nice functions. A lot of the functions we deal with in our day-to-day lives as users of mathematics, such as polynomials, trig functions, exponential and logarithmic functions, are differentiable. In fact, they are more than just differentiable. The functions you regularly encounter in algebra and calculus are all *analytic*, which basically means that these are the nicest possible functions. Based on your experience from most undergraduate courses, it might feel like you have to go out of your way to get a function that is not differentiable, or at least not differentiable in a fundamental way.

It is not very difficult to cook up a function that has just a few isolated points where the function is not differentiable: we learn in calculus that any function whose graph has a sharp corner point is not be differentiable there. What is more striking, though, is that there are some functions which are not differentiable *anywhere*. One famous example is the Weierstrass function, seen in Figure 1, whose graph consists entirely of sharp corners.

Even though functions such as the Weierstrass function can not be differentiated in the usual sense you learn about in calculus, there is actually a way to make sense of a derivative of these functions. In fact, if we change the way we define

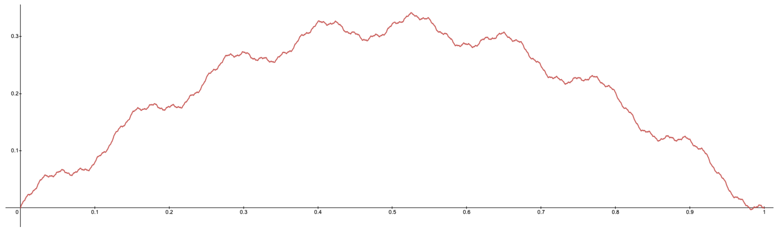

FIGURE 1. The Weierstrass function is an example of a nowhere differentiable function, given by the infinite series $\sum_{n=1}^{\infty} \frac{\sin(\pi n^2 x)}{\pi n^2}$.

the derivative a little bit, we can show that (almost) every function has a derivative! However, there is a catch: the derivative we will introduce may not itself be a function! Instead, the derivative will be an example of something called a "distribution," which generalizes our usual notion of a function.

After developing the ideas and seeing some examples, we will briefly mention two branches of mathematics where these generalizations of a derivative have proven to be particularly useful.

In what follows, we will use the words "derivative" and "differentiable" in a few different senses, so just to be clear if we use these words without any other adjectives (such as "weak" or "distributional"), then we mean the notion of a derivative you learn about in calculus. Sometimes to emphasize this we may refer to this as the *classical derivative*. When we want to discuss one of the generalizations of a derivative that we will introduce, we will explicitly refer to it as a "weak derivative" or "distributional derivative." Similarly, there is a chance the notation could get confusing, so we will adopt the convention that notation such as $\frac{d}{dx}f(x)$ or $f'(x)$ refers to the classical derivative, and we will give new (non-standard) notation for the other versions of the derivative when we introduce them.

It is also worth mentioning that there are some technical details that we will ignore in this article, if addressing them would take us far afield without giving us much more insight. That is, we are making a trade-off, sacrificing some rigor for the sake of clarity. That is not to say the technical details we are skipping over are not important or interesting, but our focus is on generalizing the notion of a derivative, and we do not want to distract from that more than necessary.

2. A STEPPING STONE: THE WEAK DERIVATIVE

Recall from your first semester calculus course that the product rule tells us that if $f(x)$ and $g(x)$ are differentiable, then so is the product $f(x)g(x)$, and its derivative is given by

$$\frac{d}{dx}f(x)g(x) = f'(x)g(x) + f(x)g'(x).$$

Let us modify this equation just a little bit to get another familiar equation from calculus. By rearranging the terms of the product rule just a touch, we arrive at the equation

$$f(x)g'(x) = \frac{d}{dx}f(x)g(x) - g(x)f'(x).$$

If we integrate both sides of this equation, then we will have

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int g(x)f'(x)\,dx.$$

This is the familiar *integration by parts* formula, which we should think of as something like "the product rule for integrals." Just to be clear, notice that if you take u to be $f(x)$ and v to be $g(x)$ (and so $du = f'(x)dx$ and $dv = g'(x)dx$), this becomes the mnemonic we all know and love, $\int u\,dv = uv - \int v\,du$.

If we were to integrate $f(x)g'(x)$ over the entire real line, then we would like to use the above to write

$$\int_{-\infty}^{\infty} f(x)g'(x)\,dx = f(x)g(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} g(x)f'(x)\,dx,$$

but now there are a few things we need to be careful about. In particular, just because our functions $f(x)$ and $g(x)$ are differentiable does not mean these improper integrals will converge. However, if we are willing to make a few additional assumptions about $g(x)$, we can guarantee convergence. So, we will make the following two assumptions about the function $g(x)$:

- We will suppose $g(x)$ has *compact support*. This means there is some finite interval $[a, b]$ so that $g(x) = 0$ for all x outside of $[a, b]$.
- We will also suppose $g(x)$ is not simply differentiable, but is *smooth*. This means that $g(x)$ is infinitely-differentiable. Notice that if $g(x)$ is both smooth and has compact support $[a, b]$, then $g(a) = g(b) = 0$.

Making these assumptions does a couple of things for us: it ensures that our integrals above will converge, and it actually simplifies our earlier equation a little bit. In particular, the assumption about compact support will let us drop the term $f(x)g(x)\Big|_{-\infty}^{\infty}$ since this will just be zero. Assuming the support of $g(x)$ was $[a, b]$, our earlier integral becomes just

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

Observe that when $f(x)$ is differentiable in the classical calculus sense, the derivative $f'(x)$ must satisfy the equation above for all choices of smooth, compactly supported functions $g(x)$ (though the interval $[a, b]$ will generally change if we change the function $g(x)$). This motivates the following definition of our generalization of the derivative, called the "weak derivative." We say that a function $f(x)$ is *weakly differentiable* if there exists a function we will denote $f^w(x)$ so that for every smooth, compactly supported function $g(x)$, we have

$$\int_{-\infty}^{\infty} f(x)g'(x)\,dx = -\int_{-\infty}^{\infty} g(x)f^w(x)\,dx \tag{1}$$

and we refer to $f^w(x)$ as the *weak derivative* of $f(x)$.

3. Examples of Weak Derivatives

Now that we have a definition of a weak derivative, we should see some examples. The first family of examples are simply the classically differentiable functions: if $f(x)$ is differentiable, then it is weakly differentiable and its weak derivative is the classical derivative $f'(x)$. This follows immediately from the discussion above, and shows us that weak derivatives truly are generalizations of the classical calculus derivative.

As a slightly more interesting example, consider everyone's favorite function that has a point of non-differentiability: the absolute value function $f(x) = |x|$. We will use our integral equation, Equation 1 above, to determine the weak derivative of $|x|$. This means our goal is to find an example of a function $f^w(x)$ with the property that

$$\int_{-\infty}^{\infty} |x|g'(x)\,dx = -\int_{-\infty}^{\infty} g(x)f^w(x)\,dx$$

for all smooth, compactly supported functions $g(x)$. Since this must hold for all such functions $g(x)$, in particular it holds for functions whose support is on some interval $[a, b]$ with $a > 0$. In those situations, we have

$$\int_a^b xg'(x)\,dx = -\int_a^b g(x)f^w(x)\,dx. \tag{2}$$

To evaluate the left-hand side we can use our old friend the integration by parts formula with $u = x$ and $dv = g'(x)\,dx$. This allows us to rewrite the left-hand side of the equation above as

$$\int_a^b xg'(x)\,dx = xg(x)\Big|_a^b - \int_a^b g'(x)\,dx.$$

Since $g(x)$ is smooth and has compact support $[a, b]$, $g(a) = g(b) = 0$. This allows us to rewrite Equation 2 as

$$-\int_a^b g(x)\,dx = -\int_a^b g(x)f^w(x)\,dx.$$

This last equation basically tells us that $f^w(x) = 1$ if $x > 0$, but this is actually little white lie that we will address in a moment. A similar calculation (which you should try on your own) will tell us that $f^w(x) = -1$ if $x < 0$, and again this is a little white lie.

Notice that the weak derivative is basically what you would guess it should be for $|x|$ if x is strictly positive or strictly negative. But what about the little white lies we mentioned? And what should $f^w(0)$ be? Here is where things start to get weird (or, depending on your point of view, where things start to get interesting).

Let us now make a quick aside about integrals in general to see what is going on. Say that we had two functions which were identical except at one point. For example, say $\phi(x) = x^2$ and $\psi(x)$ is the following piecewise function

$$\psi(x) = \begin{cases} x^2 & \text{if } x \neq 2 \\ 3 & \text{if } x = 2 \end{cases}.$$

How does changing the function $\phi(x)$ at this one point, $x = 2$, to obtain $\psi(x)$ change the values of the integrals of $\phi(x)$ and $\psi(x)$? Changing the functions in this way does not change the integral![1] The integral can only see what the function does over a range of values, and not what it does at one particular point. Once we make that observation, it is not too hard to extend this to a finite collection of values: if $\phi(x)$ equals $\psi(x)$ at all but finitely-many points, the integrals $\int_a^b \phi(x)\,dx$ and $\int_a^b \psi(x)\,dx$ will be the same. We can even go further and include infinitely-many points where the functions disagree, but we will postpone that for the moment and discuss it in the next section.

Now, what does this have to do with our little white lies earlier and our question about the value of $f^w(0)$? The little white lie is that f^w is not really a function, but more of an equivalence class of functions. That is, if we had two different candidates for the weak derivative, call them f_1^w and f_2^w, which were the same except at one point (or finitely-many points, or some other more interesting kind of set to describe later), they would both satisfy our definition of the weak derivative. So, we should not really think of the weak derivative as a particular function that maps inputs to outputs, but a family of functions that always integrate the same way. This idea feels a bit strange the first time you encounter it, but is actually quite common in certain branches of mathematics, and will be important for our second generalization of the derivative later.

Since the value of our weak derivative at an individual point does not matter, the value of $f^w(0)$ can be anything you would like. Whether we take $f^w(0)$ to be 0, $-\pi$, or $\sqrt{2}$ is immaterial: as long as our integral equation is satisfied, we have a weak derivative.

4. An Aside on Sets of Measure Zero

We mentioned in the last section that changing the value of a function at a single point does not change the value of the integral of that function over any interval. If we are willing to believe this, we should believe that we can also modify the function at finitely many points without changing the value of the integral. One easy way to think about this is to break the integral up by integrating over finitely-many intervals, each of which contains exactly one of these points with the modified value. For example, suppose $\phi(x)$ equals $\psi(x)$ at every value of x except $x = 1, 3, 5$. We can then write our integral as

$$\int_0^6 \phi(x)\,dx = \int_0^2 \phi(x)\,dx + \int_2^4 \phi(x)\,dx + \int_4^6 \phi(x)\,dx.$$

On each of these three smaller intervals, $\phi(x)$ and $\psi(x)$ agree at each point except one. But this one place where the functions disagree does not change the value of the integral, so we must have $\int_0^2 \phi(x)\,dx = \int_0^2 \psi(x)\,dx$, and similarly for the other intervals. So, the integrals of $\phi(x)$ and $\psi(x)$ must be the same!

With the previous discussion in mind, we can go a step further and easily see that $\phi(x)$ and $\psi(x)$ could differ at infinitely-many points, say all of the integer points, and their integrals still remain the same! But, wait... if we could change

[1] As an exercise, try to justify this to yourself rigorously. One way to do this is to use the definition of the integral in terms of limits of Riemann sums. Thinking about fitting rectangles under the graph of a function, the rectangle containing the modified point is the only one that differs between the two integrals, and the contribution of that one rectangle matters less and less in the limit as our rectangles get progressively skinnier.

the functions at infinitely-many points, can we not have completely different functions? Does this not mean all functions always integrate to the same thing? That can not be right... there must be something a little bit subtle going on. The issue is not *how many* points there are where the functions disagree, but the *size* of the set where the functions disagree.

In general, our notion of the integral is closely tied to how we measure the size of sets. On the real line, we usually measure the size of an interval by its length: the length of $[a, b]$ is $b-a$, and this is fundamental to the way we calculate integrals. Going back to the intuitive limit of Riemann sums description of the integral, we partition an interval we are integrating over into pieces, and multiply the length of that interval by a value of the function evaluated at a point in the interval (i.e., the widths and heights of the rectangles we try to cram under the graph). In multivariable calculus, we mimic this procedure replacing the length of an interval with the area of a small rectangle in the plane for functions of two variables, or the volume of a small rectangular prism for functions of three variables. More generally, there is an abstract notion of the size of a set called a "measure," and once we have a measure we can start integrating functions.

To be just a little bit more precise, a *measure* on a set X is a function whose inputs are subsets[1] of X, and whose outputs are non-negative numbers. If our measure is denoted μ, then to a subset E of X we associate a number $\mu(E)$ which we think of as the size of E. Your usual notions of size like length, area, and volume, give examples of measures on the real line, the plane, and 3-space. But there are a lot of other interesting examples! Probabilities are also examples of measures where the probability gives us the notion of the "size" of an event in some sample space.

Another type of measure, which at first glance seems a bit silly but will reappear later, are the *Dirac delta* measures. These are measures which only care about sets that contain some given point P. For any set X and any point $P \in X$, we define the Dirac delta measure supported at P, denoted δ_P, to be the measure which gives size 1 to any set containing P, and 0 to any set not containing P:

$$\delta_P(E) = \begin{cases} 1 & \text{if } P \in E \\ 0 & \text{if } P \notin E \end{cases}.$$

This seems like a really strange way to measure size as seemingly very different sets can have the same measure. For example, on the real line the Dirac delta measure supported at 0, δ_0, assigns size 1 to the sets $(-3, 5)$, $(-\infty, \infty)$, $\mathbb{Z}$, and $\{0\}$, among many others.

We will not go into the details here, but one of the most important properties of measures on a set X is that they let us define a concept of integration for functions defined on X. This idea is extremely important in probability theory and helps to clarify ideas like random variables and their expectation. The only thing we need to know about this generalized notion of integration is that when you integrate a function over a set of measure zero, you must get zero. As a consequence, if two

[1] This is another white lie. For technical reasons we usually *do not* want the inputs to to a measure to be just any subset, but to come from a family of subsets called a "sigma algebra." Going into the details of why we need to limit ourselves to this special collection of subsets is far outside the scope of this article, but I would encourage you to view the video *How the Axiom of Choice Gives Sizeless Sets* (`https://youtu.be/hcRZadc5KpI`) from PBS' Infinite Series on YouTube to get a glimpse into these issues.

functions agree except on a set of measure zero, they must have the same integral over any set.

In the case of the real line, the measure we normally use is called the *Lebesgue measure* and it generalizes our usual notion of the length of an interval. The type of integral defined using the Lebesgue measure generalizes our normal Riemann integral.[1] We mention this to clarify our earlier comments about the integral of two functions possibly agreeing even if the functions disagree at infinitely-many points. This happens when the set of points where the functions disagree has measure zero. So, how can you determine if a subset of the real line has measure zero?

In the case of the Lebesgue measure, we can determine if a set has measure zero by seeing if we can cover it by sequences of open intervals whose total length goes to zero. Writing this down formally makes it look more complicated than it really is: all we are saying is that we can approximate our potentially complicated, weird set by nice sets, and can make those nice sets as small as we would like.

For a set with a single point, such as $\{0\}$, we can cover the set by intervals like $(-1, 1)$ which has total length 2, or $(-\frac{1}{2}, \frac{1}{2})$ which has length 1, or $(-\frac{1}{4}, \frac{1}{4})$ which has length $\frac{1}{2}$, and so on. Because we can cover $\{0\}$ with these intervals whose lengths shrink down to zero, $\{0\}$ must have size zero, as you would expect.

We can easily play the same game on any finite set, and for some infinite sets. For the set of integers $\mathbb{Z}$, we can cover the integers by a union of intervals such as

$$\begin{aligned}
&\bigcup_{n=-\infty}^{\infty} \left(n - \frac{\epsilon}{2^{|n|}}, n + \frac{\epsilon}{2^{|n|}}\right) \\
&= \cdots \cup \left(-2 - \frac{\epsilon}{4}, -2 + \frac{\epsilon}{4}\right) \cup \left(-1 - \frac{\epsilon}{2}, -1 + \frac{\epsilon}{2}\right) \cup (-\epsilon, \epsilon) \cup \\
&\quad \left(1 - \frac{\epsilon}{2}, 1 + \frac{\epsilon}{2}\right) \cup \left(2 - \frac{\epsilon}{4}, 2 + \frac{\epsilon}{4}\right) \cup \cdots
\end{aligned}$$

For $0 < \epsilon < 1$, this collection of intervals has measure 3ϵ, and so by letting ϵ go to zero, we can cover $\mathbb{Z}$ by a set of arbitrarily small total size, so it must have measure zero.

Let us mention another example of an infinite set of measure zero: the rational numbers. It is not too difficult to modify the example of $\mathbb{Z}$ above to cover the set of rational numbers, $\mathbb{Q}$, by a sequence of open intervals whose total measure can be made arbitrarily small. The key is that the rationals form a countable set, and so can be placed in some ordering of a first rational number, a second rational number, and so on (this will not agree with the usual way we order numbers, but that is okay). We then pick little intervals of some small width around each of our rational numbers. Say w_1 is the width of the interval around the first rational number, w_2 is the width of the interval around the second rational number, and so forth. We just need to pick these widths to be small enough that the series $\sum_{n=1}^{\infty} w_n$ converges to a finite number, call it w. Now repeating the process but with the intervals replaced by intervals of width ϵw_n, our total measure will be ϵw,

[1]For technical reasons, whenever we write down an integral $\int_a^b f(x)\,dx$ in this article we really want this to be the generalized type of integral associated with the Lebesgue measure. Since this generalizes the Riemann integral, for any Riemann integrable function, these two notions of the integral will agree. However, the Lebesgue integral allows us to integrate some functions that the Riemann integral does not.

and we can let ϵ go to zero. This argument actually shows that any countable set has Lebesgue measure zero! Comparing this to the previous examples, this shows you that having measure zero is more than just having the points be "far apart" from one another, as you might have been lead to believe from the example with the integers, as every open interval contains infinitely-many rationals: they are very densely packed into the real line![2]

5. Two More Weak Derivative Examples

Our example of the weak derivative of $|x|$ may have felt a little boring or disappointing: it seems like we are just saying we can ignore this one weird point at $x = 0$ where the classical calculus derivative does not exist. So, let us consider two more examples which are fundamentally different from this. First, we consider the function which equals x^2 on all of the irrational numbers, but equals zero on the rational numbers,

$$f(x) = \begin{cases} x^2 & \text{if } x \text{ is irrational} \\ -1 & \text{if } x \text{ is rational} \end{cases}$$

This function does not just have one single sharp corner or even finitely-many sharp corners, but instead has infinitely-many jumps around each and every point! In *every* interval, no matter how small, there are points where the function looks like x^2 and points where the function looks like -1. You can not really accurately graph this kind of function on a computer, but you should imagine the graph as being something like the parabola $y = x^2$, but with infinitely-many, infinitely-close places where the graph has holes that are filled in at -1. It is a little weird.

The classical calculus derivative has no hope of dealing with any part of this function, but the weak derivative gets to "cheat" and ignore all of these weird "holes" in the graph of $y = x^2$. Because integrals get to ignore sets of measure zero, as far as the integral is concerned, this function is basically just x^2, and so its derivative should basically be $2x$. Here we should expect the weak derivative to be any function that equals $2x$, except possibly on a set of measure zero. (Remember the weak derivative is not really a single function, but more like a family of functions.) We can easily check that taking $f^w(x) = 2x$ satisfies our integral equation defining the weak derivative.

Let us modify our function above by switching the roles of rationals and irrationals, and considering the function

$$f(x) = \begin{cases} x^2 & \text{if } x \text{ is rational} \\ -1 & \text{if } x \text{ is irrational} \end{cases}.$$

What should weak derivatives of this function be? Spend a minute to think about it before you read on to the answer.

Because this function basically looks like the constant function -1 outside of a set of measure zero, its weak derivative is any function that is zero (except possibly on a set of measure zero).

[2]You might now be led to wonder if having an infinite set of measure zero requires the set to be countable, but this too is false! The standard counterexample is the Cantor set, which is an uncountable set of measure zero.

6. A NON-EXAMPLE

At this point we should be able to appreciate that the weak derivative allows us to have a notion of derivative for functions that are not at all differentiable in the sense you learn in calculus. But you may start to wonder if every function must have a weak derivative. Let us work through the details to show that a pretty innocuous-seeming function, called the Heaviside function, is not weakly differentiable.

The *Heaviside function*, $H(x)$, is a discontinuous function defined to be 1 on the non-negative reals and 0 on the negative reals. This simple function is graphed in Figure 2.

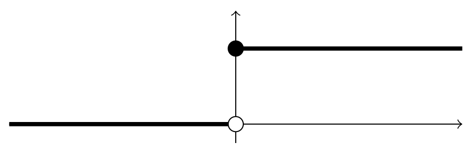

FIGURE 2. The Heaviside function.

It would seem reasonable for this to have a weak derivative; you would likely guess the weak derivative should be 0 (almost everywhere). In fact, a tiny bit of thought seems to confirm this would be the only possible candidate for $H^w(x)$. Indeed, if $g(x)$ was any compactly supported, smooth function, say whose support $[a, b]$ is contained within the positive reals, then we could write the integral defining $H^w(x)$ as

$$\int_a^b g'(x)\, dx = -\int_a^b H^w(x) g(x)\, dx.$$

Keep in mind that the left-hand integrand is really $H(x)g'(x)$, it is just that $H(x) = 1$ for positive x. Now, by the fundamental theorem of calculus the left-hand integral equals $g(b) - g(a)$, but by our assumptions on the function $g(x)$ we have $g(b) = g(a) = 0$, so the left-hand side equals zero. Thus for *every* compactly supported, smooth function $g(x)$ whose support is contained in the positive reals we must have

$$\int_a^b H^w(x) g(x)\, dx = 0$$

and this (essentially) forces $H^w(x) = 0$ for $x > 0$. The same kind of argument for a function whose support is contained in the negative reals also shows $H^w(x) = 0$ for $x < 0$, and as noted above the actual value of H^w at zero is irrelevant. So, *if* the weak derivative were to exist, it would have to be 0. But does this argument not simultaneously show us that the weak derivative of the Heaviside function exists?

Remember our integral equation has to hold for all choices of smooth, compactly supported $g(x)$. Let us see if we can construct a $g(x)$ which will cause us some trouble when we consider the integral definition of the weak derivative of $H(x)$. Consider a function $g(x)$ whose derivative has the graph indicated in Figure 3. As long as we are careful that the integrals of $g'(x)$ over $[-2, -1]$ and

$[-1, 1]$ cancel out, we can find an antiderivative $g(x)$ with compact support. Notice, though, that when we consider the integral

$$\int_{-\infty}^{\infty} H(x)g'(x)\,dx,$$

this integral will equal $\int_0^\infty g'(x)\,dx$. But this integral will now yield a positive number since the area under the graph of $y = g'(x)$ over $[0, \infty)$ never dips below the x-axis again. However, $H^w(x)$ needs to equal zero for (almost all) x, and so the right-hand side of the defining integral, $\int_{-\infty}^{\infty} H^w(x)g(x)\,dx$, will equal zero. Now we have a problem. On the one hand, we have argued that if H^w exists it must equal zero almost everywhere. On the other hand, we have seen that if $H^w(x)$ were to equal zero almost everywhere then we can not satisfy the integral equation for all smooth, compactly supported $g(x)$. Thus we must conclude that $H(x)$ is not weakly differentiable!

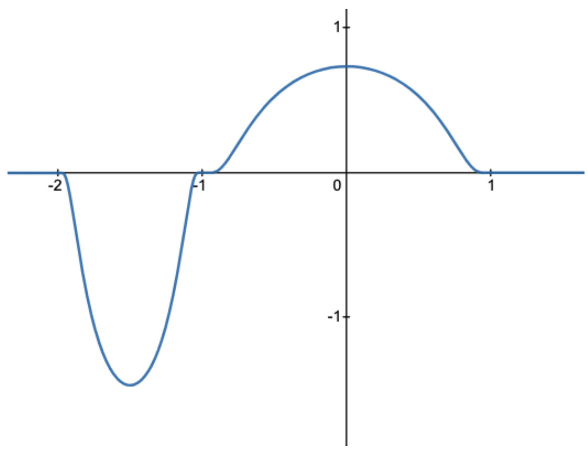

FIGURE 3. The derivative $g'(x)$ of a function $g(x)$ which is smooth and compactly supported, but which causes some trouble for our supposed weak derivative of $H(x)$.

This example highlights an important point about the weak derivative: being weakly differentiable is distinct from being differentiable almost everywhere. The Heaviside function seems like it ought to have a weak derivative because it is basically just a constant, and there is this one point of discontinuity that seems to mess everything up. However, this one minor change fundamentally breaks the integral equation a weak derivative is required to satisfy.

We can also show that some continuous functions like the Weierstrass function mentioned earlier do not have a weak derivative, but that would require us to build up some more theory that we will not take the time to do in this article.

7. Distributions and Distributional Derivatives

We mentioned that weak derivatives should not really be thought of as functions in the normal sense, but as an equivalence class of functions. This was because the weak derivative is defined in terms of integrals, and integrals can not see what functions do on a set of measure zero. In particular, we should not think of weak derivatives as acting on points the way that functions do.

However, from our definition in terms of integrals, we see there is a way to think about weak derivatives as acting on functions. Even though two choices for a weak derivative f^w of f may disagree at any given point, there is no ambiguity if we think of a weak derivative as acting on the set of smooth, compactly supported functions by mapping such a function ϕ to $\int_{\mathbb{R}} \phi(x) f^w(x)\, dx$. This motivates the following notion of a "distribution," sometimes also referred to as a "generalized function."

Instead of thinking of a function as acting on points (i.e., assigning a value $f(x)$ to each point x in the domain of the function), we will think of generalized functions as acting on functions, assigning a value to functions that live on our set. That is, we would like to define a distribution T as assigning a value $T(\phi)$ to each function ϕ, but we need to be a little bit careful. If we want weak derivatives to define distributions via the assignment $\phi \mapsto \int \phi(x) f^w(x)\, dx$ above, we can not literally consider *all* functions ϕ as the integral may not be defined. As mentioned earlier, though, the integral will be defined if we restrict ourselves to compactly supported, smooth functions. Notice also that properties of integrals imply our action of f^w on these functions is linear: for every pair of smooth, compactly supported functions ϕ and ψ and every real number λ we have

$$\int_{-\infty}^{\infty} (\phi(x) + \lambda\psi(x)) f^w(x)\, dx = \int_{-\infty}^{\infty} \phi(x) f^w(x)\, dx + \lambda \int_{-\infty}^{\infty} \psi(x) f^w(x)\, dx.$$

With all of this in mind, we can finally define distributions, which we think of as being a sort of generalized function.

A *distribution* T is a map which assigns to each smooth, compactly supported function ϕ on $\mathbb{R}$ a number $T(\phi)$ such that for every pair of such functions ϕ and ψ and every real number λ we have

$$T(\phi + \lambda\psi) = T(\phi) + \lambda T(\psi).$$

The collection of smooth, compactly supported functions on $\mathbb{R}$ is denoted $C_0^\infty(\mathbb{R})$, and these are sometimes referred to as *test functions* for our notion of distribution.

Observe that each integrable function f determines a distribution, denoted T_f, by

$$T_f(\phi) = \int_{-\infty}^{\infty} \phi(x) f(x)\, dx.$$

Though we have not defined the notion of an integral determined by a general measure, it is worth pointing out that measures μ also give rise to distributions T_μ via

$$T_\mu(\phi) = \int_{-\infty}^{\infty} \phi(x)\, \mu(dx).$$

In the case of a Dirac delta measure δ_P, this will simply recover the value of the test function at P:

$$T_{\delta_P}(\phi) = \int_{-\infty}^{\infty} \phi(x)\, \delta_P(dx) = \phi(P).$$

As another example whose details are outside of the scope of this article but still interesting to mention, is when μ is a probability measure on $\mathbb{R}$, the associated distribution simply assigns the expectation $\mathbb{E}[X]$ to each random variable X!

Now that we have defined these "generalized functions," how should we go about differentiating them? The usual calculus definition makes no sense here since distributions do not act on points. But the definition of a weak derivative from earlier did not need to know what a function did on individual points since it was defined in terms of integrals. So, how about we use the defining property of the weak derivative as our guide to define the derivative of a distribution? That is, for any distribution, T, we consider its *distributional derivative* to be the distribution, which we will denote T^d, satisfying the equation

$$T(\phi') = -T^d(\phi) \tag{3}$$

for every test function ϕ. The distribution corresponding to the weak derivative from earlier will satisfy this definition:

$$T_f(\phi') = \int_{-\infty}^{\infty} f(x)\phi'(x)\,dx = -\int_{-\infty}^{\infty} \phi(x)f^w(x)\,dx = -T_{f^w}(\phi),$$

so this seems like a reasonable way to generalize the weak derivative. Moreover, notice that Equation 3 does not simply tell us "we need to find the distribution which satisfies this equation," it actually tells us what the distribution is! The distributional derivative of a distribution T is the distribution defined by $T^d(\phi) = -T(\phi')$! And now this is the amazing thing: this quantity *always* exists! That is, every distribution necessarily has a distributional derivative. This is just because the derivative of a test function is itself a test function, and classical derivatives and distributions are both linear, which makes it easy to see that T^d will also be linear:

$$\begin{aligned}
T^d(\phi + \lambda\psi) &= -T(\frac{d}{dx}(\phi + \lambda\psi)) \\
&= -T(\phi' + \lambda\psi') \\
&= -T(\phi') - \lambda T(\psi') \\
&= T^d(\phi) + \lambda T^d(\psi)
\end{aligned}$$

8. Examples of Distributional Derivatives

At this point we have defined distributional derivatives and we have seen that the distribution corresponding to a weak derivative corresponds to the distributional derivative of a weakly differentiable function. What about something that is not weakly differentiable, like our Heaviside function from earlier?

If we let T_H denote the distribution associated to the Heaviside function,

$$T_H(\phi) = \int_{-\infty}^{\infty} H(x)\phi(x)\,dx = \int_0^{\infty} \phi(x)\,dx,$$

then its distributional derivative T_H^d is given by

$$T_H^d(\phi) = -T_H(\phi') = \int_0^{\infty} \phi'(x)\,dx.$$

By the fundamental theorem of calculus and the assumption ϕ has compact support, this will equal

$$-\left(\lim_{b\to\infty} \phi(b) - \phi(0)\right) = \phi(0).$$

So, our distribution derivative of the Heaviside function just spits out the value of a function at $x = 0$. But we just saw another distribution that did the same thing: the distribution associated with the Dirac measure δ_0. In this sense, we see that a function can have a (distributional) derivative which is actually a measure!

In the case of the Heaviside function, we kind of lucked out a little bit because its distributional derivative had a nice, easy-to-describe form. This does not need to happen in general. For a function such as the Weierstrass function, the distributional derivative has a nice expression in terms of an integral:

$$T_W^d(\phi) = -T_W(\phi') = -\int_{-\infty}^{\infty} W(x)\phi'(x)\,dx$$

On the other hand, you can not really write the distributional derivative as anything simpler than an expression like this. This is analogous to the situation of antiderivatives in calculus where every continuous function has an antiderivative, but sometimes the best we can hope to do is express that antiderivative in terms of an integral. This happens, for example, with $f(x) = e^{-x^2}$. This is a perfectly nice, continuous function, and so it has an antiderivative. However, the only way we can express that antiderivative is as

$$F(x) = \int_{-\infty}^{x} e^{-t^2}\,dt + C.$$

(In fact, this is not simply analogous, this really is the exact same issue.)

9. Some History and Applications

A totally reasonable question to ask at this point is: who cares? Why does anyone care about these generalized notions of derivatives? Were they invented just so mathematicians have something else to think about, or are they actually useful in solving problems?

These ideas originated with the Soviet mathematician Sergei Sobolev in the 1930's in his work on partial differential equations, and were later extended and popularized by Laurent Schwartz in the 1950's (Schwartz in fact won a Fields medal for his work on distributions). Today, the ideas of weak and distributional derivatives are widely employed in the theory of partial differential equations. The idea is essentially to represent a PDE by a linear operator between certain special types of vector spaces of functions called *Sobolev spaces*. Once this can be done, studying the differential equation and its solutions more-or-less becomes an exercise in linear algebra. (This is not to say it is an easy exercise in linear algebra, however.) There are quite a lot of technical details required to make all of this precise, but the key ideas are that these generalizations of a derivative are required to define these spaces.

Another place where weak and distributional derivatives arise is in a subfield of complex analysis called *Teichmüller theory*. One of the miracles of complex analysis is that the natural functions to consider, called *holomorphic functions*, can be interpreted in many different ways. One geometric way of thinking of holomorphic functions is that they are *conformal*, which just means these are functions which preserve angles between tangent vectors. The natural place to do complex analysis, besides the complex plane, is on a *Riemann surface* and one interesting problem is to determine how different two Riemann surfaces are. We can actually measure the distance between two distinct Riemann surfaces not by looking for conformal maps between them (which will not exist if the surfaces are truly distinct), but by looking for *quasiconformal maps* which essentially distort angles, but in a controlled kind of way. Defining quasiconformal maps precisely requires the weak derivative, and this was where I first encountered these ideas.

10. Where to Learn More?

There is quite a lot we have not mentioned about distributions, distributional derivatives, weak derivatives, or measures. Several obvious questions may have crossed your mind while reading, such as "is there a notion of a weak partial derivative?" or "what about higher-order derivatives?" These are questions that have well-developed answers, but there are still lots of other things we can say about these objects and lots of places where they are all used to solve various types of problems.

Most people first encounter weak derivatives and distributions in a graduate-level course on partial differential equations, and most textbooks for such a course will introduce these ideas. One standard reference is the book of Evans [3]. We only very briefly alluded to how weak derivatives arise in PDE's. To be just slightly more precise, once the PDE is expressed in terms of a linear operator, it is not simply linear algebra that becomes a useful tool, but a branch of mathematics called *functional analysis* which studies vector spaces (often infinite dimensional) that have some additional structure. Most students will get at least a basic introduction to functional analysis in graduate school, and some standard textbooks are Conway [2] and Stein & Shakarchi [9].

Measures are another topic that are typically first learned in graduate school, but there are some undergraduate-level textbooks available. The book of Capinski and Kopp [1] takes a very probabilistic view of measure theory and could be read by undergraduates with a good background in basic probability theory. Another nice and very readable book is by Silva [8], and is described as an undergraduate book in ergodic theory (which is essentially the study of statistical properties of deterministic systems from a measure-theoretic point of view). However, since the book assumes no background in measure theory, the first half of the book doubles as a very nice introduction to the subject.

Complex analysis is certainly taught at an undergraduate level, but the topics of Riemann surfaces, Teichmüller space, and quasiconformal maps often are not. There are several books about these topics, but the books of Miranda [6] and Kirwan [5] are two that students comfortable with complex analysis could read to begin their study of Riemann surfaces. After Riemann surfaces are understood, one of the more digestible books on Teichmüller theory is the book by Hubbard [4].

Finally, there are books dedicated primarily to the study of distributions. The book of Laurent Schwartz [7] is still highly regarded as a reference in the subject, although it is only available in French. The more recent textbook by Strichartz [10] is very readable, contains several examples, and also explains in more detail how distributions are used in the theory of PDE's.

References

[1] Marek Capiński and Ekkehard Kopp. *Measure, integral and probability*. Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, second edition, 2004.

[2] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.
[3] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.
[4] John Hamal Hubbard. *Teichmüller theory and applications to geometry, topology, and dynamics. Vol. 1*. Matrix Editions, Ithaca, NY, 2006. Teichmüller theory, With contributions by Adrien Douady, William Dunbar, Roland Roeder, Sylvain Bonnot, David Brown, Allen Hatcher, Chris Hruska and Sudeb Mitra, With forewords by William Thurston and Clifford Earle.
[5] Frances Kirwan. *Complex algebraic curves*, volume 23 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1992.
[6] Rick Miranda. *Algebraic curves and Riemann surfaces*, volume 5 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1995.
[7] Laurent Schwartz. *Théorie des distributions. Tome II*, volume 10 of *Publ. Inst. Math. Univ. Strasbourg*. Hermann & Cie, Paris, 1951.
[8] C. E. Silva. *Invitation to ergodic theory*, volume 42 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2008.
[9] Elias M. Stein and Rami Shakarchi. *Functional analysis*, volume 4 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2011. Introduction to further topics in analysis.
[10] Robert S. Strichartz. *A guide to distribution theory and Fourier transforms*. World Scientific Publishing Co., Inc., River Edge, NJ, 2003. Reprint of the 1994 original [CRC, Boca Raton; MR1276724 (95f:42001)].

WESTERN CAROLINA UNIVERSITY
Email address: cjohnson@wcu.edu