My Defence of Structured Differentiation from 1999, 6

P. Reany

January 21, 2024

Abstract

Here I review my defence of Structured Differentiation which I had made in 1999 on sci.math.

1 Introduction

In 1999, I made a defence on sci.math of my notation in Structured Differentiation (SD), which is a notational system I invented to deal with the many confusing (and well-recognized) features that commonly arise in multi-variable calculus. A mathematician on the newsgroup thought he should counter my claims and I'll present his arguments, and my counterarguments. The reader can decide the merits of my system for him or herself.

I think it will become obvious to the reader that the reason partial derivatives is a confusing subject is simply because it employs too few symbols to chase too many concepts. All SD does is to add in a couple more symbols to better distribute the cognitive workload.

Of all the mathematics subjects I've published on in the AJNP the most controversial one is what I call *Structured Differentiation* (SD), which reorganizes and reformulates the so-called theory of "partial differentiation." "Defender" is an alias for a mathematician that defended the status quo for doing so-called partial differentiation [as it was commonly accepted at that time] against my presentation of SD (I have interjected "editorial" comments within square brackets.):

2 Defender's Reply (19 November)

Subject: Re: partial derivation
Date: 19 Nov 1999 15:57:18
From: Defender
Newsgroups: sci.math

In article <3834F1AD.A5CE458F> Patrick Reany writes:

> Defender feels that we have said everything of interest about the
> subject and, although I don't agree, I will expect no further
> replies from him in this thread.

Well here's one anyway. Like I said, I'm happy to drop it if you are. :-)

> This is supposed to be mathematics – an objective endeavor. How can
> people disagree so strongly on such a fundamental aspect of
> mathematics?

I think it's because: (1) most of us would like to see students learn the material as best they can, and probably are upset when they see students misunderstanding something; (2) we come from different perspectives/backgrounds and therefore the phrase “learn the material” means different things to us (e.g., I want students to know, among other things but probably these are the top of the list, the theorems/ideas and be able to apply them rigorously, to know when and how the hypotheses can be relaxed and the theorern/idea still be valid, and to know when it is incorrect to apply the theorem/idea; some it seems are happier when a student can express great proficiency at, computation but are comletely ignorant of the meaning of what they are doing), and I will stress that often these different meanings are valid for their purposes; (3) we all (whether we admit it or not) have some degree of “everyone should do things just like me!” feelings.

In this case, I am drawing on my own experiences (learning about differential calculus, teaching it, and applying it to some research problems), from which I have observed that an understanding of certain foundations will greatly enhance understanding and reduce conceptual errors.

> [snip]
> No other area of mathematics allows us to define our terms and then
> immediately allow us to ignore those definitions. Why should partial
> differentiation be an exception?

It’s not, provided you know and understand the definitions. (See below.)

> Here Defender gives the definition most often given for the ‘partial
> derivative. ’ Note that it is an explicit derivative w/r x_1, i.e.,
> it does not track the variation of the function, f, through any
> variable except through the explicit dependence f has on x_1.

By definition of “$f : R^n \to R$”, the only possible dependence of the value of f on x_1 is explicit. An expression such as “$f(x, y, g(x, y))$” is NOT a function $R^3 \to R$.

> So, even if f should vary by x_1 through x_2, say,

This violates the hypothesis that f is a function with domain R^n. this ‘partial’ derivative will not take account of that. There is nothing to take account of, because the hypothesis is not satisfied. If you have reason to expect that x_2 depends on x_1, then you should state this. (See the end of this post for what to do when things may depend on other things but you’re not yet sure about any specifics.)

> We need another derivative for that, but Defender doesn’t define one.

There is no need, because there is nothing to take account of.

> Neither does convention, really. All I ever seem to find are ad hoc
> attempts to get something like a total derivative out there, but it
> remains elusive.

There is a very well-defined formalism in which the “total derivative” is defined. I believe it has been most thoroughly investigated in the context of functions on jet bundles (a generalization of “$f = f(t, x(t))$” to “$f = f(t, x(t), x'(t), x"(t), ...)$”), with applications to the study of PDEs. This formalism is very heavy with projections to sub-bundles, lifts of curves to bundles, and the chain rule. It may seem cumbersome but it is powerful and rigorous.

[Partial differentiation is ALL about taking TOTAL derivatives! Defender has frequently gotten confused about what I was talking about because he simply refused to except my stipulated definitions.]

>> “The Operator” which I applied was the derivative. I then expressed
>> the various derivatives in terms of the matrices as represented in the

>> standard coordinate systems on R^2 and R^3.
> > What Defender is saying is that conventional partial differentiation
> has some invisible operator, called 'the derivative,' which has been
> applied, out of sight of the beginner, and then we rely on the
> result of a theorem to write the derivative.

No, what I am saying is that "the derivative" can be (and is) defined without reference to a choice of coordinate system. I seriously question the quality of any vector-calculus type of course which does otherwise (except that it's not necessary to actually talk about "R^n without a coordinate system"). Of course a lot of things are most I easily computed once a coordinate system is chosen; by choosing I coordinates (or using the standard ones) an almost trivial theorem tells us how to write down the derivative.

> Oh, there are so many questions I'd still like to ask Defender like:
> What symbol represents this invisible 'derivative'?

Usually the derivative of f is represented by Df.

> And what is its definition?

I already gave it, but here it is again: if it exists at x, it is a linear operator $Df(x)$ such that

$$\lim_{h \to 0} \frac{|(x+h) - f(x) - Df(x)h|}{|h|} = 0 \,. \tag{1}$$

(I may have put an extra $|.|$ in the definition I gave before.)

This is an implicit definition, but an almost trivial theorem tells you it is unique, and (as above) an almost trivial theorem tells you how to write it down.

> And how does it relate to the total derivative of ordinary
> differential calculus when there is only one independent variable?

It is the same thing: let $f : R \to R$, and suppose $Df(x)$ satisfies the above definition. Then
$\lim_{h \to 0} |(f(x+h) - f(x) - Df(x)h)/h| = 0$
$\lim_{h \to 0} (f(x+h) - f(x) - Df(x)h)/h = 0$
$\lim_{h \to 0} [(f(x+h) - f(x)/h - Df(x)] = 0$
$\lim_{h \to 0} (f(x+h) - f(x))/h = Df(x)$
the converse is equally easy.

The definition of derivative which I gave has two (very large) advantages: it generalizes to $f : R^n \to R^m$ (or more generally, $f : N \to M$ for manifolds), and it makes clear that the derivative gives the best linear approximation to the function: $f(x+h)$ is approx. $f(x) + Df(x)h$. The disadvantage, that it is not an explicit definition ($Df(x) = \ldots$), is easily rectified by the application of a simple theorem.

> How does it relate to our explicit derivative defined above?

Since I claim that the "explicit derivative" which you use is unnecessary, this last question is irrelevant if not meaningless. If by "explicit derivative" you mean the definition of partial derivative which I gave, $Df(x)$ is represented by $[D_i f_j(x)]$.

> SD has good answers to all these questions.

The answers I gave are not only good, they are rigorous. [But this is not an argument between notation vs rigor. That fact that I did not consider this thread to be the place to deal with rigor does not mean that I don't care about rigor.]

> I have written a great deal more on total differentiation than I'll
> offer as a post. My immediate goal was to warn against the
> confusions and mathematically incorrect procedures used in partial
> differentiation.

We are, I think, in strong agreement on this goal. The cause of the potential confusions is where we disagree.

> [snip]
>> its value depends on the coordinate system.
>
> Yes, it is true of the partial (explicit) derivative. But how
> committed are you to maintaining that a partial derivative is an
> explicit derivative?

Since I have no need for what you call an "explicit derivative", I am very happy to discard it entirely and work only with what I call "the derivative" and (given a coordinate system) "a partial derivative". (Though to be truthful, I often work with differentials, as in differential forms on manifolds; these can be interpreted as a generalization of "the derivative".)

[Except that Defender uses the explicit derivative all over the place. He just doesn't want to admit to it! In fact, the definition of a partial derivative is trvially explicit.]

> My guess is that your notion of "invariantly defined" and mine of
> "complete equational operator" is the same thing. An operator is
> said to be "complete equational" if, when applied to any equation,
> it *always* maintains the equality.

An operator which is invariantly defined is (loosely) one which has the same value independent of choice of coordinates. (To put it another way, it can be defined without reference to a coordinate system.) In the case of the derivative, its value at a point x (note that "the point x" is very different from "the point defined by coordinates $(x_1, \ldots, x_n)$") is a linear operator (*not* a matrix) . As to maintaining equality, I have yet to encounter an operator which, when (correctly) applied to both sides of a true equation, results in a false equation. (I inserted the "correctly" in there to rule out all the "proofs" that "$1 = 0$" and similar nonsense.)

> As I said many, many times before, the total derivative has this
> property, the explicit and implicit derivatives do not.

As I (more or less) said a few times already, I agree that "∂_x" is not well-defined without first identifying the function in question, and choosing a coordinate system on the domain of the function. Having chosen a coordinate system, "∂_x" is not invariantly defined.

>> certain values (x, y, z). Generically, the equation will be false, and
>> if you want to apply the "partial derivative operator" ∂_x to
>> the equation, you will get something which is generically false.
>
> That was my point too.

Then you misunderstand my point. (An equation can be thought of as a sentence, and as such, has a truth value. Simply writing down an equation does not make it true. Perhaps this subtlety is lost on a lot of students, I don't know.)

> The 'del' [∂] (explicit) derivative is an incomplete equational
> operator. Although you can always apply it to an expression, you

> can't always apply it both sides of an equation and get back an

Yes I can: $x^2 + y^2 - 1 = 0 \to 2x = 0$.
I applied "∂_x" to both sides of an equation, and got back an
equation. It just so happens that both equations are generically false.
However, it is easy to prove that if an equation $F(x, y) = 0$ is true for all (x, y) in an open set, then anywhere in that open set, $D_1F(x, y) = 0$ will also be true.

>> What you can do, though, is appeal to the implicit function theorem,
>> which says (in part, and given additional hypotheses) that the local
>> solution of the equation $F(x, y, z) = 0$ is given by $z = g(x, y)$. Then, by
>> construction, $F(x, y, g(x, y)) = 0$ will be true for all x, y (in a
>> neighbourhood) , and then applying the "partial derivative operator"
>> ∂_x to this always-true equation will give another true
>> equation: $D_1F(x, y, g(x, y)) + D_3F(x, y, g(x, y))D_1g(x, y) = 0$.
>

[Just as I said: Defender uses the explicit derivative, in this case "$D_1F(x, y, g(x, y))$". SD prefers the notation $\partial_x F(x, y, g(x, y))$. By the way, as Defender has defined the partial derivative, this usage is illegal, since F is NOT a proper function of R^3 as he alleges it must be!]

> No. This is the same old error done here that convention
> practically forces us to make and that most people make, I
> presume. We have already defined above the "partial derivative
> operator" as an explicit derivative and we cannot apply it across an
> equal sign and guarantee that the equality will hold. In this case
> it certainly doesn't hold. Aren't mathematicians supposed to hold
> true to their definitions?

Excuse me? Just to make this clear, please show me the error in the following statement, and which equality does not hold: Suppose $F(x, y, z)$ is smooth, $F(x_0, y_0, z_0) = 0$, and $D_3F(x_0, y_0, z_0) \neq 0$. Then there exists a neighbourhood of (x_0, y_0) in R^2, and a smooth function g defined on that neighbourhood, such that, in a neighbourhood of (x_0, y_0, z_0), the solution set of the equation $F(x_0, y_0, z_0) = 0$ is given by $z = g(x, y)$. Moreover, in this (first) neighbourhood, $D_1F(x, y, g(x, y)) + D_3F(x, y, g(x, y))D_1g(x, y) = 0$.

> In SD this ghostly 'derivative' that Defender refers to comes out of
> the shadows

This "ghostly derivative" is THE DERIVATIVE. It is the fundamental object in all of differential calculus.

[What Defender has NOT made clear to us is how different his presentation of so-called partial differentiation is compared to that of standard textbooks.]

> The term $D_1F(x, y, g(x, y))$ is an explicit derivative – a partial
> derivative by our previous definition above.

The term $D_1F(x, y, g(x, y))$ is the *value* of D_1F (where $F : R^3 \to R$) evaluated at the point $(x, y, g(x, y))$ in R^3.

> So the equation
> $D_1F(x, y, g(x, y)) + D_3F(x, y, g(x, y))D_1g(x, y) = 0$
> is the result of applying the (total) deltal derivative to the equation
> $F(x, y, g(x, y)) = 0$. However, SD prefers the derivatives to take the following

> forms:
> $\delta_x F(x, y, z(x, y)) = \delta_x 0 = 0$
> which expands to
> $\partial_x F(x, y, z(x, y)) + \partial_y F(x, y, z(x, y))\delta_x y$
> $+\partial_z F(x, y, z(x, y))\delta_x z(x, y) = 0$
> which reduces to

$$\partial_x F(x, y, z(x, y)) + \partial_z F(x, y, z(x, y))\partial_x z(x, y) = 0 \tag{2}$$

> on assumption that x and y are fundamental (independent) variables.

There is a point here: we may not know whether x and y are independent. A more sophisticated formalism (that of differential forms and the exterior derivative - see Spivak, for example) deals with this, but we won't need the full machinery here.

Let's suppose that we are using an equation $F(x, y, z) = 0$ to define a set of points (namely, those points at which $F(x, y, z) = 0$). Let's further suppose that all we know is that F is smooth. We can apply the exterior derivative to both sides of the equation:

$$dF(x, y, z) = d(0) = 0\,. \tag{3}$$

(The "0" on the right-hand-side is the 0 in the vector space of 1-forms, not the real number zero.) Writing this out,

$$D_1F(x, y, z)dx + D_2F(x, y, z)dy + D_3F(x, y, z)dz = 0dx + 0dy + 0dz\,. \tag{4}$$

In other words, a certain linear combination of (dx, dy, dz) vanishes; so (dx, dy, dz) are (unless $\nabla(F)(x, y, z) = 0$) linearly dependent (and hence (x, y, z) are not independent).

In brief, then, if F is not identically 0, the equation $F(x, y, z) = 0$ implies that (x, y, z) are not independent, and the expression above gives an explicit linear relation among (dx, dy, dz). As you will recall, the truth value of "A implies B", when A is false, is true; it is only when A is true that we can conclude that B is true. In this context, this says that the truth of $F = 0$ (i.e., if we impose that restriction on (x, y, z) implies a relation among (dx, dy, dz) and hence that (x, y, z) are not independent.

What is really happening here (more or less) is there is a pair of spaces: $M = R^3$ (the domain of F) and $T_p = R^3$ (the formal span of (dx, dy, dz); for each $p = (x, y, z)$, there is a different copy of T_p. The equation $F(x, y, z) = 0$ (generically) defines a 2-dimensional surface in M, and at each point p on this surface, you can look at the corresponding T_p. The induced relation on (dx, dy, dz) defines a 2-dimensional vector subspace of T_p. This subspace will vary as p varies. (To be precise, I've assumed that F is smooth, that $\nabla(F)$ is never zero when $F = 0$; and T_p is really the span of the dual to (dx, dy, dz).)

This differential forms approach is pretty much bulletproof. It does have some superficial similarities to what you are presenting (in the sense that it does not require an a priori specification of what depends on what), but it has a firm foundation (though I didn't present it) in the standard notation and definitions as I have presented them. (And it generalizes to higher-degree forms, and has applications from electromagnetics to thermodynamics to topology to integration theory to differential equations to differential geometry to. ..)

[What Defender misses here is that I haven't done anything in SD that is not commonly found in every standard text on advanced calculus, whether it uses differential forms or not! All SD is and has ever claimed to be is a improvement in the notation and terminology used to perform differentiations in these texts.]