# My Defence of Structured Differentiation from 1999, 8

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 (22 November)

Subject: Re: partial derivation
From: Defender
Newsgroups: sci.math
Date: 22 Nov 199912:11:48

In article <3836CO70.l8ABFAC3> Patrick Reany writes:

>> By definition of "$f : R^n \to R$", the only possible dependence of the value
>> of f on xl 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 $R^n$.
>
> I assume you're saying that all the variants (arguments) of fare
> mutually independent of each other, right? (SD refers to such a

> function as 'primitive.')

For the purposes of making a precise definition, yes. (Of course there are times when one will write down, for example, $f(x, y, g(x, y))$, a point which you raise below.)

> I went back to Buck (**Advanced Calculus**, 3rd ed, p. 23) and he
> defines a function $F$ from $R^n$ to $R^m$ by $y = F(x)$, where $x =$
> $(x_1, \dots, x_n)$, $y = (y_1, \dots, y_m)$ , and $y_i = f_i\ (x_1, \dots, x_n)$ where
> $x_1, \dots, x_n$ are $n$ real variables. Buck doesn't explicitly say that
> these $n$ variables are mutually independent, so if they are, I should
> be forgiven this transgression. It's hard to second guess a
> definition.

I have to admit, I had never stopped to think about this, but yes, when one defines a function $f(x_1, \dots, x_n)$, the $n$ variables are independent. The reason for this is that (as Buck says) $f$ has domain $R^n$, which (as a set) is the set of *all* possible $(x_1, \dots, x_n)$. You can vary any of them independently from the rest.

> Defender, you said that the real confusion lies in this functional
> dependence stuff, so here's an opportunity to help a confused soul
> out. I'll come back to this shortly.

If you start with a function $f : R^3 \to R$, $f(x, y, z)$, then you have three independent variables. If you then write down something like $f(x, y, g(x, y))$, you have used this function, together with another function $g : R^2 \to R$, to define a third function, $h(x, y) = f(x, y, g(x, y))$ (although you have not actually given this third function a name).

>> if [the derivative of $f$] exists at $x$, it is a
> linear operator $Df(x)$ such that $\lim_{h\to 0} |f(x + h) - f(x) - Df(x)h|/|h| = 0$.
>> (I may have put an extra $|.|$ in the definition I gave before.)

> I take it that in $Df(x)$, the variants of $f$ are also mutually
> independent of each other, right? I.e., f is primitive?

Strictly speaking, there is only one independent variable in $f(x)$ and $Df(x)$, namely, $x$; it just so happens that $x$ is a point in $R^n$. Being a point in $R^n$, the $n$ coordinates $(x_1, \dots, x_n)$ are independent.

>> 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.
>
> This is the kind of stuff that puts your '$D$' in relation to the ordinary
> derivative. It is useful to students. It is useful for computation too.

The "freshman definition" $(\lim_{h\to 0}(f(x + h) - f(x))/h)$ is very nice in the sense that it is (1) explicit, and (2) shows the intuitive motivation behind the idea of the derivative (rate of change). This probably makes it most appropriate to most first-year calculus courses. Unfortunately it doesn't work when $f$ has domain $R^n (n > 1)$ because then you can't divide by $h$, so you really need to come up with a better definition; the linear approximation of the function as $f(x + h) =$ approx $f(x) + Df(x)h$ is probably one of the most important uses of the derivative, so we might as well turn it into a definition, (which fortunately is equivalent to the "freshman definition" when the domain is $R$).

>> If by "explicit derivative" you mean the definition of partial
>> derivative which I gave, $Df(x)$ is represented by $[D_i f_j(x)]$.
>
> Where $f$ and $f_j$ are primitive??

Yes.

> OK, I need to rephrase my question. How committed are you to
> maintaining that a partial derivative acts only on primitive
> functions? (Be patient with me. Remember, I'm just a confused
> soul.)

Well, I often will write down expressions such as

$$\partial_x f(x, y, g(x, y)) \tag{1}$$

**so probably not very committed in the strict sense**.

[How is a confused student, learning so-called "partial differentiation" for the first time, ever supposed to understand the subject when all the examples he or she encounters break the very definitions that are fundamental to the subject? In SD one is **not** allowed to break a definition, just as is true in every other discipline in mathematics. This is partly why SD is said to be "structured."]

However, I use the interpretation that what the above really means is $D_1 h(x, y)$ where $h(x, y) = f(x, y, g(x, y))$, and if I think there is a large chance of misinterpreting the above as $D_1 f(x, y, g(x, y))$, then I will probably be more explicit. In this latter sense, I suppose I am very committed.

> OK, I am really confused here. First you say that you're applying
> the 'partial derivative' to $F(x, y, g(x, y)) = 0$, then you say it's THE
> DERIVATIVE you applied.

To make everything explicit (I hope): Define $H(x, y) = F(x, y, g(x, y))$. By construction of $g$, $H(x, y)$ is identically zero; therefore, so is its derivative, $DH(x, y) = 0$. The partial derivatives, being "components" of the derivative (more accurately, they are used to represent the derivative), must also be identically zero, in particular, $D_1 H(x, y) = 0$. Appealing to the coordinatized version of the chain rule (see below),

$$D_1 H(x, y) = D_1 F(x, y, g(x, y)) + D_3 F(x, y, g(x, y)) D_1 g(x, y) \tag{2}$$

(we've aleady been through how to get this expression so I won't do that gain).

The chain rule says if $f$ and $g$ are differentiable, and their composition $h(x) = f(g(x))$ is defined (in an open set), then $h$ is differentiable and $Dh(x) = Df(g(x))Dg(x)$. This is an equation which relates linear operators: $Dh(x)$ on the left is the composition of linear operators $Df(g(x))$ and $Dg(x)$ on the right. Linear operators (finite-dimensional) can be represented as matrices by choosing a basis, and (from linear algebra) the composition of linear operators is represented by the product of the corresponding matrices. When we represent the derivatives by their matrices constructed out of partial derivatives, the equation $Dh(x) = Df(g(x))Dg(x)$ becomes a matrix equation, from which we can read off whichever partial derivatives we are interested in. This is the "coordinatized version of the chain rule" to which I previously referred.

> In SD I'd write this operator as $\delta_x$ (or

> $\delta_1$ if you prefer), but I don't see anyway to write this in your
> notation. If I use $D_1$, I get into trouble because you already have
> $D_1F(x, y, g(x, y))$ in the expansion.

**OK, I see the problem here; either we need to adopt a new notation** (e.g.,

$$D_x[F(x, y, g(x, y))] = D_1F(x, y, g(x, y)) + D_3F(x, y, g(x, y))D_1g(x, y)) \tag{3}$$

**which, apart from using different symbols, is more or less half of your notation**;

or we need to define a new function $H(x, y)$ as I did above, thereby making explicit all of our functional dependencies; or we can make statements like "differentiate with respect to $x$"

[Yes, I suppose it's possible to reduce non-primitive functions down to primitive functions, but I don't think that physicists would always want to. For examples, how time enters the Lagrangian and the Hamiltonian explicitly vs implicitly makes a big difference in the physical interpretation of what's going on. This distinction for them is not purely formal. That's why it's the physicists who are to most inclined to impliment some form of generalized total derivative with a parametric split and ignore the automatic reductions to primitive forms.]

and assume that the reader will know what to do.

I'd recommend the beginning student make things explicit (the $H(x, y)$ option) until they are comfortable with the subject, but if SD helps them get comfortable with it as well, that's fine; my main concern is that they might not understand the underlying theory (in this case, that there is a function $g(x, y)$ whose graph $z = g(x, y)$ *is* locally the solution set).

[I have only claimed that SD is a superior notation to those that are commonly used. I never said or suggested that it be taught to students apart from the correct underlying theory.]

> My second confusion comes from your insistence that the 'derivative'
> and/or 'partial derivative' must ONLY act on primitive functions
> (because those are the only objects you have defined them to act
> on).

I suppose I'm coming from the perspective where a function is defined with a domain; by definition of the domain of the function, I can choose any point in the domain and evaluate the function. If the domain is $R^n$, then I can independently vary (say) $x_1$. On the other hand, if the domain is the unit sphere $S^2$ defined by $x^2 + y^2 + z^2 - 1 = 0$ in $R^3$, then I cannot vary $x$ independently; however, if I parametrize the sphere by (say) $x = \cos(s) * \cos(t)$, $y = cos(s) * sin(t)$, $z = \sin(s)$, then I will induce a new function of $(s, t)$, and I can vary $s$ independently of $t$, and I can talk about the partial derivative with respect to $s$. If I don't parametrize the sphere, then its defining equation gives a relation between the infinitessimal variations" $(dx, dy, dz)$ I can make (namely, $2x * dx + 2y * dy + 2z * dz = 0$). In this latter approach, I can still define "the derivative", but I cannot talk about the partial derivative with respect to $x$ (because I do NOT have a function defined on $R^3$). At a point $p$ on the sphere, the derivative of $f$ at $p$, $Df(p)$, is a linear map from the "tangent space" (tangent plane) to the sphere at $p$, to the tangent space of whatever is the range of the function. (But the definition is a bit more complicated; just as we had to modify the definition to allow a domain $R^n$ because we couldn't divide by $h$, now we have to modify the definition to allow a domain which is not a vector space, because we can't do addition $x + h$ any more.)

> I find the following confusing for two reasons:
>>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 $R^3$.
>

$>$ Obviously in '$D_1F(x, y, g(x, y))$' $F$ is not primitive, nevertheless has
$>$ $D_1$ (a partial derivative) acting on it, but you defined the partial
$>$ derivative ONLY to act on primitive functions, right?

$F$, being a function with domain $R^3$, has a corresponding $D_1F$. $D_1F(x, y, g(x, y))$ is this $D_1F$ evaluated at the point $(x, y, g(x, y))$ in $R^3$. I could just as easily evaluate at $(x, y, 1 + g(x, y))$ (though for what purpose I don't know).

$>$ Furthermore, you said previously that this $F$ is not a function
$>$ $F : R^3 \to R$, but in this last paragraph you say that it is. I'm lost
$>$ here.

Sorry, I could have been more clear. An expression such as $F(x, y, g(x, y))$ has only two "independent variables" $(x, y)$ in it; this expression *uses* a function $F : R^3 \to R$ in order to define a function with domain $R^2$, and it is this latter function to which I was referring when I said that it is not a function $R^3 \to R$.

$>$ P.S. My posting to this thread is over.

Thank you for the interesting discussion. Defender.