Re-Introduction to Structured Differentiation and Who's on First

P. Reany

June 17, 2021

Abstract

I review the foundations of Structured Differentiation (SD) and use it to solve a few partial differentiation problems I found on the Internet. Along the way, I want to review why the subject of partial differentiation is so confusing to so many students, and draw a comparison of my experience in learning mathematics to the Abbott and Costello skit 'Who's on First'.

Matters of notation play a considerable role in connection with the chain rule. Wide varieties of usage exist in mathematical writing where the chain rule is concerned.

—Taylor & Mann

1 Introduction

This paper is a re-introduction to the subject I call "Structured Differentiation" (SD), which is the straight-forward generalization of so-called "ordinary differentiation," meaning that it unashamedly employs a generalized total derivative. The subject is more commonly known as "partial differentiation," though entering that phrase into a search engine is more likely to bring up a simple applications of the chain rule from first-year calculus, and I want to address far more difficult aspects of the subject.

In this presentation I hope to prove to the reader that it's a lack of 'structure' that accounts for what Taylor-Mann note in the above quote as a wide variety of notations concerning the chain rule, I presume, to make for this lack of appropriate structure. I invented SD to give the subject enough structure to eliminate the babel of usages concerning, not only the chain rule, but also all other confusing aspects of partial differentiation, such as the change of variable problem (i.e., how a change of variable can induce a change in derivatives).

Just a quick word at this point of where the adjective 'structure' came from in Structured Differentiation. Everone know what an engineer says when one finds a system that is not broken: "Don't fix it!" But guess what the same engineer says to someone who has a system that is broken? That's right: "Fix it!"

The 'structured programming' mentality was a response to the ever-growing problem of computer program complexity. As programs got bigger and bigger, their complexity increased, and something was needed to control the programmer's tendency to produce spaghetti code that no one could read. Thus was born structured programming: Get rid of gotos, instead, use control structures (such as if-then-else), modularize code, use top-down algorithmic development. The details are not important here. What is important is that after I learned of structured programming, I soon saw that partial differentiation at the level of advanced calculus was completely unintelligible to me, and I concluded, rightly or wrongly, because the subject lacked 'structure'.

As I looked at how physicists and engineers did partial differentiation, I saw important differences comparing the realms of tensors, thermodynamics, Lagragian and Hamilonian formalisms. On the mathematics side, I saw so many differences in that domain, too. As I struggled with this broken system for months, eventually insight arrived and the pieces began to fit together smoothly. The resulting system I called Structured Differentiation by analogy with structured programming.

First, what SD is not: SD is not the generalization of introductory calculus intended to include notions of directed areas, volumes, etc, as is the case of differential forms or geometric algebra. SD is meant to be a generalization that treats variables as components of appropriate vectors and allows for the total derivative of those vectors by other vectors, employing matrices when it's convenient to do so.

2 SD: What is the 'structure' in SD?

Since I've presented SD in detail in earlier papers, I will only touch on the highlights of the system in this paper. Although SD states my preference for presenting differentiation of functions of more than one variable in a particular way, it is compatible with many of the more established systems already in wide use.[1]

Now for the details of the 'structure' that gives SD its clarity, I'll introduce a few simple variables and functions for demonstration purposes. Functions at a minimum must show their explicit dependencies on variables, but variables do not need to follow this rule. Functions which get reparameterized by a change of variables must be represented by different function names. Variables do not have to follow this rule. Consider

$$\tau = f(x, y) = x^2 y^2 \,, \tag{1}$$

$$\tau = F(x, y(x)) = x^2 y^2 \,, \quad y = e^{-x} \,, \tag{2}$$

[1]We will always consider the functions we encounter herein as differentiable.

$$\tau = G(x) = x^2 e^{-2x} \,, \tag{3}$$

and their dependency graphs in Figure 1.

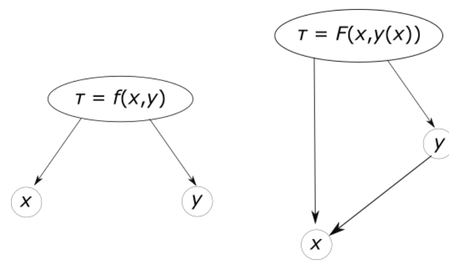

Figure 1. This figure shows how the functions representing τ are dependent on variables x and y, as represented in three flowcharts, also known as dependency graphs. Each arrow of each flowchart means 'is explicitly dependent on'. Going from left to right, τ through f is explicitly dependent on both x and y; through F, it's explicitly dependent on both x and y, and implicitly dependent on x through y; through G, it is only dependent on x.

Let

$$w = g(x) = x^2 + 3x + 9 \,, \tag{4}$$

where w is variable and g is a function of one variable x, the 'independent variable'. Also, let

$$v = f(x, y) = x^2 y + 1 \,, \tag{5}$$

where v is variable and f is a function of two independent variables x and y.

Let

$$w = h(x, y(x)) = x^2 y + 1 \,, \quad y = e^{-x} \,, \tag{6}$$

where w is variable and h is a function of one independent variable x.

Structure Rules for SD:

Rule 1) SD employs both differentials and derivatives, which should make some physicists and engineers happy.

Rule 2) In SD, variables can be treated as functions. (More on this soon.)

Rule3) In SD, the usual differential is the delta, δ. The usual derivative operator, by x, say, is $\delta/\delta x$, which is a total derivative.

Rule 4) In SD, the delta, δ, can be demoted to a d when it's clear that it is operating on a function of just one independent variable.

Rule 5) In SD, we are allowed to differentiate by both scalar and vector variables.

Rule 6) The total derivative of any scalar variable by itself is unity.

Rule 7) We can always divide one deltal differential by any other deltal differential to obtain a total derivative. For example, if δu and δx are defined, then $\delta x/\delta u$ is defined.

Rule 8) If one scalar variable, say u, is functionally independent of another scalar variable, say x, then $\delta u/\delta x = 0$. If x is also functionally independent of u, then $\delta x/\delta u = 0$.

Before giving the rest of my 'structures', I'd like first to present some of the rules applied to the equations above. If we take the total derivative by x across (4), we can use either the deltal or the d derivative. Let's use the d derivative.

$$\frac{dw}{dx} = g'(x) = 2x + 3\,. \tag{7}$$

The d-differential taken across (4) is well defined:

$$dw = g'(x)dx = (2x+3)dx\,. \tag{8}$$

No surprises so far.

The differential taken across (5) becomes

$$\delta v = \delta f(x,y) = 2xy\delta x + x^2\delta y\,, \tag{9}$$

and now we must be careful how we convert this into a derivative. Dividing through by δx, say, we get

$$\frac{\delta v}{\delta x} = \frac{\delta}{\delta x}f(x,y) = 2xy\frac{\delta x}{\delta x} + x^2\frac{\delta y}{\delta x}\,. \tag{10}$$

Now, if we add that x and y are independent of each other, then, applying rules 6) and 7), we get

$$\frac{\delta v}{\delta x} = \frac{\delta}{\delta x}f(x,y) = 2xy\,. \tag{11}$$

Let's make things a bit more interesting by introducing the variable u, such that

$$u = \Omega\sin\theta\,, \tag{12}$$

where u, Ω, and θ are all variables but we do not know how Ω and θ are functionally related. Then, differentiating by θ, say, gives

$$\frac{\delta u}{\delta\theta} = \frac{\delta\Omega}{\delta\theta}\sin\theta + \Omega\cos\theta\,, \tag{13}$$

where we have used Rule 4), so that

$$\frac{\delta\sin\theta}{\delta\theta} = \frac{d\sin\theta}{d\theta} = \cos\theta\,. \tag{14}$$

At this point, I want to introduce the 'partial' derivative. Let's return to (5) and take the differential a different way (I'll be skipping some steps that I presume the reader will be able to understand)

$$\begin{aligned}
\delta v &= [\delta x, \delta y] \begin{bmatrix} \partial_x \\ \partial_y \end{bmatrix} f(x,y) \\
&= [\delta x\, \partial_x, \delta y\, \partial_y] f(x,y) \\
&= \ f_x(x,y)\delta x + f_y(x,y)\delta y \\
&= 2xy\delta x + x^2 \delta y \,, \qquad (15)
\end{aligned}$$

where $\partial_x = \partial/\partial x$, $f_x(x,y) = \partial_x f(x,y)$, etc. So, I've also introduced matrices and the use of subscripts to represent partial differentiation, and it's all standard stuff so far. By the way, we can rewrite the first line of (15) using the Gibbs's vector notation as well, yielding

$$\delta v = \delta \mathbf{x} \cdot \partial_{\mathbf{x}} f(x,y) \,, \qquad (16)$$

allowing us to make contact with other notations of differentiation.

But this next rule is where we leave convention by adding to our set of structure rules the following rule:

Rule 9) The partial derivative with respect to a scalar variable will always be an explicit derivative.

Our next logical step is to form the total derivative of v by x. So, dividing (15) through by δx, we get

$$\frac{\delta v}{\delta x} = 2xy + x^2 \frac{\delta y}{\delta x} \,, \qquad (17)$$

and this is as far as we can go to simplify because Equation (5) has not given us any information about y's functional dependence on x. However, if we knew that y and x were independent variables, then $\delta y/\delta x = 0$, and the last equation becomes

$$\frac{\delta v}{\delta x} = 2xy \,. \qquad (18)$$

Let's look at the situation here a bit more abstractly. Returning to the third line of (15) and dividing through by δx, we get

$$\frac{\delta v}{\delta x} = \frac{\partial}{\partial x} f(x,y) + \frac{\delta y}{\delta x}\frac{\partial}{\partial y} f(x,y) \,, \qquad (19)$$

where y may or may not be a function of x. Let's now factor out the $f(x,y)$ from the RHS:

$$\frac{\delta v}{\delta x} = \left[\frac{\partial}{\partial x} + \frac{\delta y}{\delta x}\frac{\partial}{\partial y}\right] f(x,y) \,, \qquad (20)$$

What the second derivative in the square brackets does is to ferret out the rate of change of f by x going through the variable y. Let's give this derivative its own name and symbol:

$$\frac{\not\partial}{\not\partial x} \equiv \frac{\delta y}{\delta x}\frac{\partial}{\partial y} \,, \qquad (21)$$

and we refer to the derivative $\frac{\not\partial}{\not\partial x}$ as the *copartial with respect to* x. Thus, (20) becomes

$$\frac{\delta v}{\delta x} = \left[\frac{\partial}{\partial x} + \frac{\not\partial}{\not\partial x}\right] f(x, y) \,. \tag{22}$$

We can further abstract this derivative operator by removing the variable v, to get

$$\frac{\delta}{\delta x} = \frac{\partial}{\partial x} + \frac{\not\partial}{\not\partial x} \,, \tag{23}$$

or more simply

$$\delta_x = \partial_x + \not\partial_x \,, \tag{24}$$

which, by suppressing the variable of differentiation, can be further abstracted to

$$\delta = \partial + \not\partial \,. \tag{25}$$

Thus, we have split the total derivative by x into the sum of its explicit and implict derivative parts.[2]

Now, according to Rule 6), $\delta x/\delta x = 1$, so

$$\frac{\delta x}{\delta x} = \frac{\partial x}{\partial x} + \frac{\not\partial x}{\not\partial x} = 1 \,, \tag{26}$$

which forces us to make a choice of which term should survive.

Rule 10) The partial derivative of a scalar variable with respect to itself is always unity, and the copartial derivative of a scalar variable with respect to itself is always zero.[3]

Definition: The variables on which a function is explicitly dependent are referred to as the *variants* of the function.[4]

Rule 11) We need to canonize what we have already been doing: In SD, when performing the chain rule, the derivative of a function by one of its scalar variants must be performed by a partial derivative.

Let's go back to Equation (20) and generalize it a bit. Let $v = f(x, y) \to v = f(x, y, z)$, then

$$\frac{\delta v}{\delta x} = \left[\frac{\partial}{\partial x} + \frac{\delta y}{\delta x}\frac{\partial}{\partial y} + \frac{\delta z}{\delta x}\frac{\partial}{\partial z}\right] f(x, y, z) \,, \tag{27}$$

or

$$\frac{\delta v}{\delta x} = f_x(x, y, z) + f_y(x, y, z)\frac{\delta y}{\delta x} + f_z(x, y, z)\frac{\delta z}{\delta x} \,, \tag{28}$$

[2] I refer to this sum as a *parametric split.*

[3] I'll keep to this convention until I find a reason to change it.

[4] This is an SD definition and is not conventional. The idea behind the definition is that a variant is a variable on which a function varies.

and, hopefully, the reader can see how to generalize to any number of variants.

I imagine that some of the readers are saying about now, "Hold on, there. I thought the chain rule goes something like this." For example's sake, let

$$v(t) = f(x(t), y(t), z(t))\,, \tag{29}$$

where t is the only independent variable. Then the chain rule goes like this

$$\frac{\delta v}{\delta t} = \frac{\partial f}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial f}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial f}{\partial z}\frac{\delta z}{\delta t}\,, \tag{30}$$

and, in conventional form is

$$\frac{dv}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\,. \tag{31}$$

To which I'd reply, "Yeah, that's the chain rule, but so is this". Let

$$F(t) = f(x(t), y(t), z(t), t)\,, \tag{32}$$

where t is the only independent variable. Then the chain rule goes like this

$$\frac{\delta F}{\delta t} = \frac{\partial f}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial f}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial f}{\partial z}\frac{\delta z}{\delta t} + \frac{\partial f}{\partial t}\,, \tag{33}$$

which can also be written as

$$\begin{aligned}
\frac{dF}{dt} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{\delta t} + \frac{\partial f}{\partial z}\frac{dz}{dt} + \frac{\partial f}{\partial t} \\
&= \frac{\partial f}{\partial x}\dot{x} + \frac{\partial f}{\partial y}\dot{y} + \frac{\partial f}{\partial z}\dot{z} + \frac{\partial f}{\partial t} \\
&= [\dot{x}, \dot{y}, \dot{z}] \begin{bmatrix} \partial_x \\ \partial_y \\ \partial_z \end{bmatrix} f + \frac{\partial f}{\partial t} \\
&= [\dot{x}, \dot{y}, \dot{z}, 1] \begin{bmatrix} \partial_x \\ \partial_y \\ \partial_z \\ \partial_t \end{bmatrix} f\,.
\end{aligned} \tag{34}$$

Now, if we let $\dot{\mathbf{x}} = [\dot{x}, \dot{y}, \dot{z}]$ and $\partial_{\mathbf{x}} = [\partial_x, \partial_y, \partial_z] = \nabla$ then

$$\begin{aligned}
\frac{dF}{dt} &= \dot{\mathbf{x}} \cdot \partial_{\mathbf{x}} f + \frac{\partial f}{\partial t} \\
&= \dot{\mathbf{x}} \cdot \nabla f + \frac{\partial f}{\partial t} \\
&= \frac{\not{\partial} f}{\not{\partial} t} + \frac{\partial f}{\partial t} \\
&= \left(\frac{\not{\partial}}{\not{\partial} t} + \frac{\partial}{\partial t}\right) f \\
&= \frac{d}{dt} f\,.
\end{aligned} \tag{35}$$

Okay, so I've belabored this to clearly show how the partial and copartial derivatives relate to the matrix notation and Gibbs's vector notation. And, to get back to the main point, total differentiation has to deal with both explicit and implicit parts.

Rule 12) In SD, the delta derivative, partial derivative, or the copartial derivative of a constant will be zero.

Now, enter the dragon of two independent variables. Say u and v are the independent variables, and

$$F(u,v) = f(x(u,v), y(u,v))\,. \tag{36}$$

What is the total derivative of $F(u,v)$ by u? First, what does SD have to say about it?

Rule 13) In SD, the delta derivative taken across a true equation yields another true equation. Again, we assume the functions in this paper are differentiable.

By contrast, the partial (explicit) derivative across a true equation is not always going to be another true equation, because it all depends on how the functions are parameterized in the first place. For example, suppose that

$$F(x,y) = 0 \tag{37}$$

is a true equation. Then, in SD, it is also true that

$$\delta F(x,y)/\delta x = 0\,, \tag{38}$$

but it is **not** true that

$$\partial F(x,y)/\partial x \equiv F_x = 0\,. \tag{39}$$

However, in conventional systems, where the partial derivative is consider to be the equivalent of the SD total derivative, this last equation becomes

$$F_x + F_y \frac{dy}{dx} = 0\,. \tag{40}$$

Thus, in SD, on taking the operator $\delta/\delta u$ across (36), we get

$$\frac{\delta F}{\delta u} = \frac{\partial f}{\partial x}\frac{\delta x}{\delta u} + \frac{\partial f}{\partial y}\frac{\delta y}{\delta u}\,, \tag{41}$$

and this is the correct form to leave it in if we do not know the functional relationship between u and v. But since we do know that these variables are functionally independent of each other, we can tease out a simpler form. Starting with Equation (41) and expanding the delta derivatives, we get

$$\frac{\partial F}{\partial u} + \frac{\partial F}{\partial v}\frac{\delta v}{\delta u} = \frac{\partial f}{\partial x}\Big(\frac{\partial x}{\partial u} + \frac{\partial x}{\partial v}\frac{\delta v}{\delta u}\Big) + \frac{\partial f}{\partial y}\Big(\frac{\partial y}{\partial u} + \frac{\partial y}{\partial v}\frac{\delta v}{\delta u}\Big)\,. \tag{42}$$

Now, using the independence condition $\delta v/\delta u = 0$, (42) simplifies down to

$$\frac{\partial F}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}\,, \tag{43}$$

where every partial derivative turns out to be explicit – but this did not occur by chance. And this equation is pretty much where convention starts when applying the chain rule. However, this result was the outcome of my choosing the functions and dependencies in (36) very carefully. In SD, one can be much less careful, yet the notation is robust enough to allow for it without ambiguity.

For example, suppose instead of (36), we write

$$F = F(x(u,v), y(u,v), u)\,, \tag{44}$$

where F on the LHS is a variable and we leave the functional dependency between u and v unknown. Then, differentiating by u (i.e., taking the total derivative by u), we get, without ambiguities,

$$\frac{\delta F}{\delta u} = \frac{\partial F}{\partial x}\frac{\delta x}{\delta u} + \frac{\partial F}{\partial y}\frac{\delta y}{\delta u} + \frac{\partial F}{\partial u}\,. \tag{45}$$

But what do we get when we apply the conventional knee-jerk 'chain rule' to (44)? We get,

$$\frac{\partial F}{\partial u} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial F}{\partial u}\,, \tag{46}$$

and we do have problems. First, what do we do about the $\frac{\partial F}{\partial u}$ on both sides of the equation? According to the quote by Taylor-Mann on the first page, there are many ways that convention uses to deal with situations like these, none of which I found satisfactory.

On the RHS of (46), all three derivatives $\frac{\partial F}{\partial x}$, $\frac{\partial F}{\partial y}$, and $\frac{\partial F}{\partial u}$ are explicit derivatives. What then is $\frac{\partial F}{\partial u}$ on the LHS? Well, it can't also be an explicit derivative, can it? And what about $\frac{\partial x}{\partial u}$ and $\frac{\partial y}{\partial u}$? We can't conclude that they're explicit derivatives because each of x and y could vary by u through the variable v.

Let's demonstrate this by a specific example. What if we take (45) and add to it the requirement that

$$v = \phi(u)\,? \tag{47}$$

Now, if we want to expand (45), we get

$$\frac{\delta F}{\delta u} = \frac{\partial F}{\partial x}\Big(\frac{\partial x}{\partial u} + \frac{\partial x}{\partial v}\frac{\delta \phi}{\delta u}\Big) + \frac{\partial F}{\partial y}\Big(\frac{\partial y}{\partial u} + \frac{\partial y}{\partial v}\frac{\delta \phi}{\delta u}\Big) + \frac{\partial F}{\partial u}\,. \tag{48}$$

And, in consequence of (47), if we assume that u is the only ndependent variable in (44), which did not assume initially, then (48) could be rewritten as

$$\frac{dF}{du} = \frac{\partial F}{\partial x}\Big(\frac{\partial x}{\partial u} + \frac{\partial x}{\partial v}\frac{d\phi}{du}\Big) + \frac{\partial F}{\partial y}\Big(\frac{\partial y}{\partial u} + \frac{\partial y}{\partial v}\frac{d\phi}{du}\Big) + \frac{\partial F}{\partial u}\,. \tag{49}$$

So, by appending again (47) to the system, can we go from the conventional equation (46) to (49), as we just did in SD? Well, yes and no. Some annotation is needed because we can't go from $\partial x/\partial u$ in (46) to $\dfrac{\partial x}{\partial u}+\dfrac{\partial x}{\partial v}\dfrac{d\phi}{du}$ in (46) without running into the same problem as we did with $\dfrac{\partial F}{\partial u}$ in (46).

What we have here, folks, is a failure to communicate, unambiguously and concisely. But what's the reason behind this? I think this happened because long ago in the history of multivariable calculus the meme was established that: There shall be only one operator for differentiation in multivariable calculus and its symbol is the partial sign. I ask a simple question, then: Why? What is that so important about that? In 1982, I invented SD to remove ambiguity by adding a generalized total derivative with a symbol different from a partial sign. But how does convention try to remove these ambiguities?

One way to remove them is to never use unparameterized variables, such as on the LHS of (44), writing instead something like

$$F(u,v) = f(x(u,v), y(u,v), u)\,, \tag{50}$$

and appling the chain rule to get

$$\frac{\partial F}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial f}{\partial u}\,, \tag{51}$$

which is correct if u and v are independent of each other, but what if they're not? Where does that leave $\partial x/\partial u$, for example? Besides that, we've been forced to add a new function symbol to the mix, thus we now have both F and f. That may be fine for mathematicians writing in the abstract, but this is not a happy situation for the physicist, who finds that nearly all the letters in English and in Greek already have conventional meanings.

Thus, the physicist might be tempted to write (44) and remove some ambiguity when applying the chain rule by writing

$$\frac{\partial F}{\partial u} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial u} + \left.\frac{\partial F}{\partial u}\right|_{\mathrm{exp}}, \tag{52}$$

where we have annotated the right-side $\dfrac{\partial F}{\partial u}$ with a subscript to insist that that derivative is explicit only. This system seems to work, if we take as the default that the partial derivative is 1) explicit when a function is differentiated by one of its *normal* variants,[5] and 2) a total derivative unless annotated otherwise, but at the cost of a messy annotation.

If you think about it, it should be clear that this system above is just a sneaky way to introduce a new symbol into the mix, namely, $\left.\dfrac{\partial}{\partial u}\right|_{\mathrm{exp}}$. Long ago, I decide that if I must introduce a second derivative symbol, I'd prefer the pair to be $\dfrac{\delta}{\delta u}$ and $\dfrac{\partial}{\partial u}$ rather than the corresponding pair $\dfrac{\partial}{\partial u}$ and $\left.\dfrac{\partial}{\partial u}\right|_{\mathrm{exp}}$.

[5]A variant of a function is said to be *normal* if and only if the function depends on the variant only explicitly.

Now, I introduced the adjective 'normal' to indicate if a variant is special in a particular useful way or not. I did this to help remove some ambiguity of the 'partial' derivative within the context of this hypothetical physicists's system of differentiation above. Looking at the functional form of F in (44) and see that the variant x is normal in F, thus $\dfrac{\partial F}{\partial x}$ is an explicit derivative. But u is not normal in F, therefore, $\dfrac{\partial F}{\partial u}$ is not explicit.

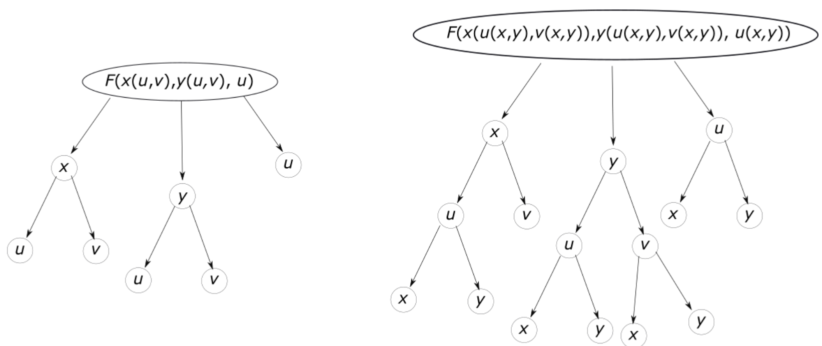

Figure 2. The dependency graph on the left shows the two variants of F, x and y, as normal. Whereas the dependency graph on the right show the convolution of x through both u and v, and hence x is no longer a normal variant of F, and similarly for y. By the way, for lack of room, I left off showing the dependency of one v-node on x and y.

But there's still room for trouble here. By the Implicit Function Theororem, under certain general conditions, the $x = x(u, v),\ y = y(u, v)$ can be turned inside out to form the pair $u = u(x, y),\ v = v(x, y)$. Then (44) becomes

$$F = F(x(u(x, y), v(x, y)), y(u(x, y), v(x, y)), u(x, y))\,, \tag{53}$$

and now none of the variants of F is normal.

Definition: A variable is said to be *convoluted* if it is dependent on itself through another variable.

For example, if $x = x(u(x))$, then x is said to be convoluted through the variable u.

This is one place where using vectors to represent variants becomes very useful. Let $\mathbf{x} = [x, y]$ and $\mathbf{u} = [u, v]$. Then (44) becomes

$$F = F(\mathbf{x}(\mathbf{u}), u)\,, \tag{54}$$

and (53) becomes

$$F = F(\mathbf{x}(\mathbf{u}(\mathbf{x})), u(\mathbf{x}))\,, \tag{55}$$

Before leaving this section, I'd like to make a comparison of the chain rule as we have applied above to the simple calculus function $F(u) = f(x(u))$. Applying the chain rule to this equation, we get

$$\frac{dF}{du} = \frac{df}{dx}\frac{dx}{du}\,, \tag{56}$$

where the RHS shows no addition signs. Is there a counterpart to this when the function is more complicated, such as in (36)? Answer: In SD, Yes. First, we rewrite (36) to this form

$$F(\mathbf{u}) = f(\mathbf{x}(\mathbf{u}))\,, \tag{57}$$

where $\mathbf{u} = [u, v]$ and $\mathbf{x} = [x, y]$. Then, by differentiating by $\mathbf{u}$, we can get derivatives by both u and v at the same time, yielding

$$\frac{\delta F}{\delta \mathbf{u}} = \frac{\partial f}{\partial \mathbf{x}}\frac{\delta \mathbf{x}}{\delta \mathbf{u}}\,, \tag{58}$$

which in matrix form is

$$\left[\frac{\delta F}{\delta u}, \frac{\delta F}{\delta v}\right] = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \begin{bmatrix} \frac{\delta x}{\delta u} & \frac{\delta x}{\delta v} \\ \frac{\delta y}{\delta u} & \frac{\delta y}{\delta v} \end{bmatrix}\,. \tag{59}$$

So, to recoup the addition signs, just perform the matrix multiplications on the RHS.

Finally, how does the chain rule look if F and f are promoted to n-component vectors, such that each component has the functional form

$$F_i(\mathbf{u}) = f_i(\mathbf{x}(\mathbf{u}))\,? \tag{60}$$

Easy:

$$\frac{\delta \mathbf{F}}{\delta \mathbf{u}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\frac{\delta \mathbf{x}}{\delta \mathbf{u}}\,. \tag{61}$$

As I stated earlier, this is all explained in detail in earlier papers I wrote on SD.

3 Who's On First?

Without doubt, at best mathematics can be hard to learn. There are subtleties and nonintuitive aspects of many areas of mathematics that act as mental barriers to rapid assimilation of mathematics by its learners. But as I look back over my last 50 years of trying so hard to learn new areas of mathematics, I have found it often to be the case that aggravating my slowness of learning new math is often what I came to recognize are bad presentation of the subjects. I know I'm being judgmental, but I find no way around it.

There's no one person or textbook to be blamed for any of this. Once a bad presentation is conventionalized as canon, it's so hard to change. Instructors themselves may have a hard time conveying the subjects they teach because they are also hampered by the same bad presentations they learned. The mess of the many systems within the subject of partial differentiation attests to this fact.

But partial differentiation is only one of the many subjects I have found wanting for a better presentation. Included in this mess is also algebra word problems, mathematical induction, stoichiometry, projective geometry, group theory, linear algebra, etc.

When I think about how much trouble I have had learning new math, I sometimes feel like Lou Costello in his skit with Bud Abbott, "Who's on First." In this skit, Lou is having a real hard time trying to learn the names of the players on a baseball team, despite Bud's best effort to convey them. Eventually, Lou learns the cue when to respond "Third base" at the correct time, but he has no idea **why** it's the right reply. With determination, Bud tries over and over to get through to Lou, but he fails, and Lou is left frustrated. It's not that Bud is a bad teacher, but rather that the subject is inherently confusing, and – let's be honest— Lou is probably not the best student Bud has ever had to teach.

But so often in my own learning of math subjects I find that the best I can do is to learn the cues of when to do this or that, without any real understanding why I am doing them, and declare, "Third base!" So, besides the fact that there's apparently a lot of Lou Costello in me (and, truth be told, probably a lot of Gilligan, too!), I also can blame part of my delayed competency in some math areas on the presentations, as well.

Definition: A function is said to be *primitive* if its variants are mutually independent of each other. (This definition is specific to SD, though I think that its use in conventional partial differentiation would be of value, too.)

Example: Suppose that $x = x(u, v)$ is primitive. Then we know that $\delta u/\delta v = 0$ and that $\delta v/\delta u = 0$. Therefore,

$$\frac{\delta x}{\delta u} = \frac{\partial x}{\partial u} \quad \text{and} \quad \frac{\delta x}{\delta v} = \frac{\partial x}{\partial v}\,. \tag{62}$$

4 Problem 1

It's now time to work a problem off the Internet, titled "Total differentials and the chain rule," which can be found at

`https://www.youtube.com/watch?v=2bF6H_xu0ao`

The instructor is David Jordan, and the problem he is solving for us is to compute $\partial z/\partial u$, given that

$$z = x^2 + y^2\,, \tag{63}$$

and

$$x = u^2 - v^2 \tag{64a}$$
$$y = uv\,. \tag{64b}$$

Now, in SD we can write

$$z(u,v) = z(x(u,v), y(u,v))\,, \tag{65}$$

where u and v are considered as independent of each other. Thus, when taking the deltal derivative on both sides of (65), we can reduce $\delta z/\delta u$ to $\partial z/\partial u$ on the LHS, to get

$$\frac{\partial z(u,v)}{\partial u} = \frac{\delta z(x(u,v), y(u,v))}{\delta u}\,. \tag{66}$$

Expanding on the RHS, we can employ the chain rule and similarly reduce deltal derivatives to partials (since both x and y are primitive in u and v), to get

$$\frac{\partial z(u,v)}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u}\,, \tag{67}$$

where we must remember that $\dfrac{\partial z}{\partial x} = \dfrac{\partial z(x,y)}{\partial x}$ and $\dfrac{\partial z}{\partial y} = \dfrac{\partial z(x,y)}{\partial y}$. Similarly, we can reduce (67) down to

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u}\,, \tag{68}$$

if we remember that $\dfrac{\partial z}{\partial u} = \dfrac{\partial z(u,v)}{\partial u}$.

Question: Is Equation (68) more virtuous than Equation (67)? Ans: Not necessarily, especially since we did not originally have z explicitly as a function of u and v, though by simple algebra we could have found $z(u,v)$ directly by substituting Equations (64a) and (64b) into (63), but then we wouldn't get to use the chain rule.

Anyway, we're now ready for the solution (noting that all the partials on the RHS are explicit):

$$\begin{aligned}\frac{\partial z}{\partial u} &= (2x)(2u) + (2y)(v)\Big|_{\substack{x=u^2-v^2\\ y=uv}} \\ &= 4(u^2 - v^2)u + 2uv^2\,. \end{aligned} \tag{69}$$

At point 5:25 in the YouTube video, David Jordan complains that the chain rule is not his favorite means to work these kinds of problem, because he finds them apparently unmotivated (calling them prescriptive, but not explanative) — a condition I at one time shared with him, and perhaps tens of millions of other calculus students also share his view. So, he chose to redo the problem with total differentials. Mr. Jordan did not say why he preferred this method, but I might venture the guess that it's because, when applying the chain rule for differentials, the partial derivatives are generally explicit and thus are easier to comprehend.

In favor of SD, I have already shown how to go from equations of total differentials to equations of total differentiations to equations of partial (explicit) derivative, under suitable revamping of variable dependencies to get rid of implicit derivatives.

5 Problem 2

This second Internet problem is found at

`https://www.physicsforums.com/threads/partial-differentiation -problem-multiple-variables-chain-rule.753651/`

If

$$z = x^2 + 2y^2 \,, \tag{70a}$$

$$x = r\cos\theta \,, \tag{70b}$$

$$y = r\sin\theta \,, \tag{70c}$$

find the partial derivative $\left(\dfrac{\partial z}{\partial \theta}\right)_x$.

Suggested answer:

$$\left(\frac{\partial z}{\partial \theta}\right)_x = 4r^2\tan\theta \,, \tag{71}$$

where convention expects us to consider $\dfrac{\partial z}{\partial \theta}$ as a total derivative[6] and $\left(\dfrac{\partial z}{\partial \theta}\right)_x$ is the total derivative of z by θ, holding x constant in the process.

Now, I must make a minor point of contention with the answer given in (71). My preference is that the only variables that should appear on the RHS should be x and θ.

Solution 1:

Before we rush into an attempted solution, it's good to take stock of what we have been given to work with. We are asked to find $\left(\dfrac{\partial z}{\partial \theta}\right)_x$ as though we know what $z = z(x, \theta)$ is, but we don't. What we actually know is $z = z(x, y)$. There are three ways to proceed at this point.[7] One way to proceed is simply to algebraically recast z accordingly

$$z = z(x, y) \longrightarrow z = z(x, \theta) \,. \tag{72}$$

[6] It's clearly a total derivative by SD nomenclature, anyway.

[7] I found three ways, but there could be more.

This part is easy:

$$z(x,\theta) = x^2(1 + 2\tan^2\theta)\,, \tag{73}$$

In SD, we treat $z(x,\theta)$ as primitive in its arguments (i.e., x,θ are the new independent variables of the problem), so that the total derivative by θ taken across this last equation reduces on both sides to the partial derivative by θ, hence:

$$\frac{\partial z}{\partial \theta} = 4x^2 \tan\theta \sec^2\theta = 4r^2 \tan\theta\,, \tag{74}$$

where $\dfrac{\partial z}{\partial \theta}$ is, of course, an explicit derivative in SD, or rather $\left(\dfrac{\partial z}{\partial \theta}\right)_x$.

Definition: I have long had trouble thinking that I can change the set of independent variables on which some other variable is dependent. It's probably just a psychological thing. In any case, I prefer to refer to the ordered set of independent variables as the *fundamental*. I usually denote the old fundamental by the vector $\boldsymbol{\eta}$ and the new fundamental by $\boldsymbol{\eta}'$.

Rule 1) When one fundamental variable η_i is totally differentiated by a cofundamental variable η_j, the result is

$$\frac{\delta \eta_i}{\delta \eta_j} = \delta_{ij}\,, \tag{75}$$

where δ_{ij} is the Kronecker delta, meaning that when a fundamental variable is totally differentiated by itself, the result is unity, but when it is totally differentiated by a different cofundamental variable, the result is zero; hence, the notion of *variable independence* between different elements in a given fundamental is captured by this rule. Note: To remove ambiguity, any variable in a fundamental set is cofundamental to itself.[8]

Rule 2) In all other cases, i.e., when a new fundamental variable differentiates an old fundamental variable, or vice versa, the deltal derivative usually reduces to an explicit derivative. This is the standard thing to do in thermodynamics because, by default, we regard state variables as having no implicit dependence on other variables (at least this is my knowledge regarding them).[9]

Second Solution to Problem 2:

What we did in the last solution was to algebraically recast $z = z(x,y) \longrightarrow z = z(x,\theta)$ because it was algebraically easy to do. But what if such a transformation is algebraically clumsy or even impossible? Then what do we do? Ans:

[8] We will consider it axiomatic in SD that the total derivative of any variable with respect to itself is unity.

[9] When a state variable **does** have an implicit dependence on some variable, then we cannot just drop the implicit derivative as we have done in this case.

We proceed as if such a change of fundamental is possible, but use the chain rule instead. Let's see how this works.

The de facto fundamental of z in (70a) is

$$\boldsymbol{\eta} \equiv \begin{bmatrix} x \\ y \end{bmatrix} , \tag{76}$$

which we'll think of as the 'old fundamental'. Just as before, what we'd like is the parameterization of z as $z(x, \theta)$, but this time we're not going to make the algebraic substitution. This time, we're going to set the 'new fundamental' as

$$\boldsymbol{\eta}' \equiv \begin{bmatrix} x \\ \theta \end{bmatrix} . \tag{77}$$

From here, we can think of the problem as a 'change in fundamental' problem, given by

$$z(\boldsymbol{\eta}') = z(\boldsymbol{\eta}(\boldsymbol{\eta}')) . \tag{78}$$

We're going to assume that this change of variables is legitimate. Anyway, differentiating this by $\boldsymbol{\eta}'$, we get by the chain rule

$$\frac{\delta z}{\delta \boldsymbol{\eta}'} = \frac{\partial z}{\partial \boldsymbol{\eta}} \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'} . \tag{79}$$

Now, remembering that the z on the LHS is primitive in $\boldsymbol{\eta}'$, the total derivative can be replaced by an explicit derivative. Therefore, we have that

$$\frac{\partial z}{\partial \boldsymbol{\eta}'} = \frac{\partial z}{\partial \boldsymbol{\eta}} \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'} . \tag{80}$$

Expanding this, we get

$$\left[\frac{\partial z}{\partial x}, \frac{\partial z}{\partial \theta}\right] = \left[\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right] \begin{bmatrix} \dfrac{\delta x}{\delta x} & \dfrac{\delta x}{\delta \theta} \\ \dfrac{\delta y}{\delta x} & \dfrac{\delta y}{\delta \theta} \end{bmatrix} . \tag{81}$$

But, in the 2×2 matrix above, $\dfrac{\delta x}{\delta x} = 1$ and $\dfrac{\delta x}{\delta \theta} = 0$, so (81) becomes

$$\begin{aligned} \left[\frac{\partial z}{\partial x}, \frac{\partial z}{\partial \theta}\right] &= \left[\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right] \begin{bmatrix} 1 & 0 \\ \dfrac{\delta y}{\delta x} & \dfrac{\delta y}{\delta \theta} \end{bmatrix} \\ &= [2x \, , 4y] \begin{bmatrix} 1 & 0 \\ \tan\theta & x \sec^2\theta \end{bmatrix} . \end{aligned} \tag{82}$$

From this we get to

$$\frac{\partial z}{\partial x} = 2x + 4y \tan\theta \tag{83a}$$

$$\frac{\partial z}{\partial \theta} = 4yx \sec^2\theta . \tag{83b}$$

On eliminating y, we get

$$\frac{\partial z}{\partial x} = 2x + 4x\tan^2\theta \tag{84a}$$

$$\frac{\partial z}{\partial \theta} = 4x^2 \tan\theta \sec^2\theta\,. \tag{84b}$$

And for thosee who prefer the subscripts,

$$\left(\frac{\partial z}{\partial x}\right)_\theta = 2x + 4x\tan^2\theta \tag{85a}$$

$$\left(\frac{\partial z}{\partial \theta}\right)_x = 4x^2 \tan\theta \sec^2\theta\,. \tag{85b}$$

6 Problem 3

This third Internet problem is found at

```
https://www.physicsforums.com/threads/partial
     -derivative-problem.822012/
```

Problem: Define

$$f(x, y) = x + 2y\,, \tag{86a}$$

$$w = x + y\,. \tag{86b}$$

What is $\dfrac{\partial f}{\partial w}$?

Solution 1:

In this and the following attempts at a solution, we will adopt the rule that the old fundamental is $\boldsymbol{\eta} = (x, y)$ and that the new fundamental is $\boldsymbol{\eta}' = (w, y)$.

So, let's first try to use algebra to get

$$f = f(w, y) = w + y\,. \tag{87}$$

Then

$$\frac{\delta f}{\delta w} = \frac{\delta}{\delta w}\big(w + y\big) = 1\,, \tag{88}$$

since $\delta y/\delta w = 0$. And we really didn't use the chain rule.

Solution 2:

In this solution, we will use the chain rule. Begin with

$$f = f(x(w, y), y(w, y))\,, \tag{89}$$

then

$$\frac{\delta f}{\delta w} = \frac{\partial f}{\partial x}\frac{\delta x}{\delta w} + \frac{\partial f}{\partial y}\frac{\delta y}{\delta w}\,, \tag{90a}$$

$$= (1)(1) + (2)(0) = 1\,. \tag{90b}$$

Solution 3:

In this solution, we employ the full machinery of SD.

The de facto fundamental of f in (86a) is

$$\boldsymbol{\eta} \equiv \begin{bmatrix} x \\ y \end{bmatrix}, \tag{91}$$

which we'll think of as the 'old fundamental'. The 'new fundamental' is

$$\boldsymbol{\eta}' \equiv \begin{bmatrix} w \\ y \end{bmatrix}. \tag{92}$$

Again, we think of the problem as a 'change in fundamental' problem, given by

$$f(\boldsymbol{\eta}') = f(\boldsymbol{\eta}(\boldsymbol{\eta}')). \tag{93}$$

Differentiating this by $\boldsymbol{\eta}'$, we get

$$\frac{\delta f}{\delta \boldsymbol{\eta}'} = \frac{\partial f}{\partial \boldsymbol{\eta}} \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'}. \tag{94}$$

Expanding this, we get

$$\left[\frac{\delta f}{\delta w}, \frac{\delta f}{\delta y}\right] = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \begin{bmatrix} \frac{\delta x}{\delta w} & \frac{\delta x}{\delta y} \\ \frac{\delta y}{\delta w} & \frac{\delta y}{\delta y} \end{bmatrix}. \tag{95}$$

But $\frac{\delta y}{\delta y} = 1$ and $\frac{\delta y}{\delta w} = 0$, etc, so (95) becomes

$$\left[\frac{\delta f}{\delta w}, \frac{\delta f}{\delta y}\right] = [1, 2] \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}. \tag{96}$$

Therefore,

$$\frac{\delta f}{\delta w} = 1, \tag{97a}$$

$$\frac{\delta f}{\delta y} = 1. \tag{97b}$$

And that's the SD version of the answer. Of course, the conventional form of the answer would be

$$\frac{\partial f}{\partial w} = 1, \tag{98a}$$

$$\frac{\partial f}{\partial y} = 1. \tag{98b}$$

7 Problem 4

This fourth Internet problem is found at

```
https://www.youtube.com/watch?v=u9u9a1NTU-Q
```

Problem: Given that

$$x = u + v + w \tag{99a}$$
$$y = u^2 + v^2 + w^2 \tag{99b}$$
$$z = u^3 + v^3 + w^3 \,, \tag{99c}$$

show that

$$\frac{\partial u}{\partial x} = \frac{vw}{(u-v)(u-w)} \,. \tag{100}$$

Solution: For starters, there's no way we want to try to algebraically invert these equations to cast u, v, w in terms of x, y, z.

For the stout-hearted, you can enter

```
solve for [u,v,w] if  x = u+v+w, y= u^2+v^2+w^2,  z =  u^3+v^3+w^3
```

into WolframAlpha.com to see what it returns.

Anyway, let $\mathbf{x} = [x, y, z]$ and $\mathbf{u} = [u, v, w]$.[10] What can't be easily done with algebra can often be easily done with convolution and differentiation. Let

$$\mathbf{x} = \mathbf{x}(\mathbf{u}(\mathbf{x})) \,, \tag{101}$$

and differentiate across by $\mathbf{x}$, to get

$$\frac{\delta \mathbf{x}}{\delta \mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \mathbf{u}} \frac{\delta \mathbf{u}}{\delta \mathbf{x}} \,. \tag{102}$$

Next, two simplifications. First, because the components of $\mathbf{x}$ are considered independent among themselves, $\dfrac{\delta \mathbf{x}}{\delta \mathbf{x}} = \mathbf{I}$, the 3×3 identity matrix, and second, because $\mathbf{u}$ is considered primitive in $\mathbf{x}$, $\dfrac{\delta \mathbf{u}}{\delta \mathbf{x}} \to \dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$, yielding

$$\mathbf{I} = \frac{\partial \mathbf{x}}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \,. \tag{103}$$

Now, the partial derivative we seek is the entry in the first row, first column of matrix $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$. From the last equation we get:

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{u}} \right)^{-1} \,. \tag{104}$$

[10] I wrote these vectors as row type but they're really column type.

Then

$$\left(\frac{\partial \mathbf{x}}{\partial \mathbf{u}}\right)^{-1} = \frac{\begin{bmatrix} 6vw^2 - 6v^2w & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 2u & 2v & 2w \\ 3u^2 & 3v^2 & 3w^2 \end{vmatrix}} . \tag{105}$$

If we multiply out the determinant as it now is, we get a mess. Better is to use the YouTube presenter's method of first performing a couple column simplifications: c2→ c2 − c1, and c3→ c3 − c1, and then take the determinant to get $6(v-u)(w-u)(w-v)$. Then, extracting the component in row 1, column 1, we get, with some algebraic simplification, Equation (100).