SIAM Review
Vol. 63, No. 4, pp. 854–864

A Precise and Reliable Multivariable Chain Rule*

Raymond Boute†

Abstract. The multivariable chain rule is often challenging to students because it is usually presented with ambiguities and other defects that hamper systematic and reliable application. A very simple formulation combines the derivation operators for functions and for expressions in a manner not found elsewhere due to common confusion between them. Some issues are rooted more deeply than others and are discussed in a broader perspective, starting with the function concept. The approach is illustrated using various applications including the transport equation, partial derivatives of a definite integral, and the distortionless (but not lossless) transmission line. This note is suitable for a lecture in any first-year course covering partial derivatives, as a complement to the other course material.

Key words. chain rule, composition, derivative, dimensional analysis, expression, function, partial derivative, transport equation, telegrapher's equation

AMS subject classifications. 00A69, 00A73, 26B05, 26B12, 35A09, 97I20, 97I40, 97I60

DOI. 10.1137/19M1296732

1. Introduction and Teaching Hints. Partial derivatives and differential equations are ubiquitous in applied mathematics. Conventions have changed little in over a century [7], yet suffer from sloppy practices. This leads to ambiguous forms of the chain rule that make its use unreliable in general and puzzling to students in particular. We present a simple solution in a wide context, assuming only basic knowledge of ordinary derivatives and an inkling (perhaps just a definition) of partial derivatives.

The material is suitable for a lecture in a first-year calculus course, complementing the textbooks used. Section 2 paves the path between both sources. Section 3 (on functions) gives a brief review of fundamentals that are assumed known, plus some addenda and warnings of pitfalls. Section 4 (on derivatives) recalls the often-forgotten but crucial distinction between operators for functions and for expressions. These sections are mainly reminders and can be presented selectively in about one third of a lecture.

Section 5 extends dimensional analysis (a blind spot in mathematics) to differential equations. Section 6 builds on the preceding principles to develop various reliable forms of the multivariable chain rule. Application examples in sections 7, 8, and 9 also show how to handle further examples provided by the instructor and the textbooks.

2. Motivation and Rationale. Typically the chain rule is stated as follows [6]: when $u = u(x, y)$, $x = x(s, t)$, and $y = y(s, t)$ (known functions of s and t),

$$\text{(2.1)} \qquad \text{(a)}\ \frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} \quad \text{and} \quad \text{(b)}\ \frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t}\,.$$

*Received by the editors October 31, 2019; accepted for publication (in revised form) January 19, 2021; published electronically November 4, 2021.
https://doi.org/10.1137/19M1296732
†INTEC, Universiteit Gent, Zwijnaarde, Ghent, Belgium (raymond.boute@pandora.be).

We usually replace expressions of the form $\frac{\partial u}{\partial x}$ by the simpler $\partial_x u$, as in [11] . This avoids unnecessary clutter and the wrong idea that ∂x in $\frac{\partial u}{\partial x}\frac{\partial x}{\partial s}$ might cancel out. Common forms of the chain rule all resemble (2.1), including its flaws.

At an early stage, students learn the distinction between a function f and its value $f(x)$ at some point x. Apparently this distinction goes out the window for partial derivatives. An example is writing $u = u(x, y)$: by the rules for equality [8], this means that u and $u(x, y)$ are interchangeable, which causes unsoundness. It also forebodes that different occurrences of u in a formula will require different interpretations.

Let, for instance, $u = \frac{1}{2}ax^2 + bxy + \frac{1}{2}cy^2$, where $x = ks + lt$ and $y = ms + nt$. Equation (2.1) requires the reader to figure out whether the partial derivatives are taken before or after the substitutions for x and y. The *intended* interpretation is as follows:

Term	Intended interpretation
$\partial_s u$	$\partial_s(\frac{1}{2}a(ks + lt)^2 + b(ks + lt)(ms + nt) + \frac{1}{2}c(ms + nt)^2)$
$\partial_x u\ \partial_s x$	$(a(ks + lt) + b(ms + nt))\ \partial_s(ks + lt)$
$\partial_y u\ \partial_s y$	$(b(ks + lt) + c(ms + nt))\ \partial_s(ms + nt)$

Moreover, the use of new variables s and t has hidden extra ambiguities that emerge when reusing variables, for instance, substituting $mx + ny$ for y and leaving x alone. Then (2.1a) becomes $\partial_x u = \partial_x u\ \partial_x x + \partial_y u\ \partial_x y$, which is trivially correct since $\partial_x x = 1$ and $\partial_x y = 0$, but uninformative and not what is intended. The intention is as follows:

Term	Intended interpretation
$\partial_x u$	$\partial_x(\frac{1}{2}ax^2 + bx(mx + ny) + \frac{1}{2}c(mx + ny)^2)$
$\partial_x u\ \partial_x x$	$(ax + b(mx + ny))\ \partial_x x$
$\partial_y u\ \partial_x y$	$(bx + c(mx + ny))\ \partial_x(mx + ny)$

Here $\partial_x u$ stands for two different expressions, in addition to $\partial_x u = ax + by$ using (2.1a), all with different tacit substitutions. No wonder the chain rule is challenging. Students looking for clarification on the web will only find utter chaos.

Our general rationale is that, using proper symbolism, the normal rules (substitution, arithmetic) must yield correct results: anything less may "undermine the students' confidence in mathematics" [9]. One might add: "or in themselves," which is worse.

Making rules reliable does not require any more formality or effort than basic good practice and the avoidance of sloppiness. In particular, making the chain rule reliable requires distinguishing between derivation operators for functions and for expressions. Most definitions in the literature respect this distinction but lapses occur later, such as writing $\frac{\partial f}{\partial x}$ for a function f, which is meaningless. In fact, $\frac{\partial f(x)}{\partial x} = f'(x)$ and $\frac{\partial f(y)}{\partial x} = 0$. Very few sources are as careful as [5], from which we quote some definitions later.

Evidently the concept of *function* is central. Potentially confusing for students are the discrepancies between calculus texts and some other texts in use today. This mostly affects function composition, which is so essential for the chain rule. Because of these ramifications, we start from a wider perspective and gradually proceed to derivatives and the chain rule, covering along the way relevant aspects not generally mentioned in textbooks. For easy reference, selected citations include page numbers.

3. Functions, Composition, and Categorization. This section clears up the discrepancies mentioned above. Some parts may be rather terse, but every item is illustrated later. Instructors can also insert material from this section "just in time" as needed.

The definition of a *function* common in many areas including set theory [4, p. 10], calculus [1, p. 53], [5, p. 4], and discrete mathematics [14, p. 167] is given as follows.

A *function* f is a functional set of (ordered) pairs x, y. *Functional* means that no two pairs have the same first member. Hence if $(x, y) \in f$, one can write y unambiguously as $f(x)$ or $f x$. Parentheses are used only for emphasis (a subjective factor) or to overrule precedence conventions. Variants are f_x, x^f, and so on, chosen by convention. The set of all first members of the pairs in f is the *domain* of f, written $\mathcal{D} f$.

A direct consequence is the *equality theorem* [1, p. 54] for functions: $f = g$ iff (shorthand for *if and only if*) (a) $\mathcal{D} f = \mathcal{D}\, g$ and (b) $f(x) = g(x)$ for each x in $\mathcal{D} f$.

This suggests an equivalent definition not based on pairs: a function f is an object *fully specified* by (a) a *domain* $\mathcal{D} f$, which is a set, and (b) a unique *value* $f(x)$ for each x in $\mathcal{D} f$. The *range* $\mathcal{R} f$ of f is then the set of all $f(x)$ for x in $\mathcal{D} f$. Tuples can now be defined as functions: if $x = a, b$, then $\mathcal{D}\, x = \{0, 1\}$ with $x_0 = a$ and $x_1 = b$.

The *composition* $g \circ f$ of *any* functions f and g is specified as follows [1, p. 140], [5, p. 11]: (a) the domain $\mathcal{D}(g \circ f)$ is the set of all x in $\mathcal{D} f$ such that $f x$ is in $\mathcal{D}\, g$, and (b) $(g \circ f) x = g(f x)$ for each x in $\mathcal{D}(g \circ f)$. A crucial property is $h \circ (g \circ f) = (h \circ g) \circ f$. The *identity function* id_S is defined for any set S by (a) $\mathcal{D}\, \mathrm{id}_S = S$ and (b) $\mathrm{id}_S\, x = x$.

Using composition and identity, the *restriction* $f \rceil S$ of a function f to a set S is defined by $f \rceil S = f \circ \mathrm{id}_S$ and the *subfunction* relation by $g \subseteq f$ iff $g = f \rceil \mathcal{D} g$. For n-fold composition, $f^0 = \mathrm{id}_{\mathcal{D} f}$ and $f^{n+1} = f \circ f^n$. This will be important in section 4.

Functions can be characterized or categorized according to their domain or range.

A *function from X (in)to Y* is a function f such that $\mathcal{D} f = X$ and $\mathcal{R} f \subseteq Y$ [4, p. 10], [1, p. 578], [14, p. 169]. The set of all such functions is written Y^X or $X \to Y$, and the common notation to introduce (or *declare*) a function of this *type* is $f : X \to Y$.

A *partial function from X to Y* is a function f such that $\mathcal{D} f \subseteq X$ and $\mathcal{R} f \subseteq Y$. The set of all such functions is written $X \nrightarrow Y$, as in $f : X \nrightarrow Y$. Any function on a real (or complex) region is partial on $\mathbb{R}$ (or $\mathbb{C}$). Simply writing $f : \mathbb{R}^n \nrightarrow \mathbb{R}^m$ obviates ad hoc notation mixtures such as "$f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$" found in some textbooks.

By common convention, parentheses are optional in expressions of the form $(f x) y$ and $X \to (Y \to Z)$. Hence they are mandatory in $f(g\, x)$ and in $(X \to Y) \to Z$.

To avoid confusion, students should be warned that some textbooks from areas outside calculus define a *function f from X to Y* as a functional subset of $X \times Y$ with domain X (so the function equality theorem holds) and casually call Y *the codomain of f*. But if $Y \subsetneq Z$, then f is also from X to Z, so *the codomain of f* is ill-defined. Some texts even say that $f : X \to Y$ and $g : X \to Z$ can be equal only if $Y = Z$, violating the equality theorem. Self-contradicting definitions should not cause self-doubt in students.

In fact, as explained in [4], a *codomain* refers to the set Y in a *triplet* $\langle f, X, Y \rangle$, which reflects a different function concept altogether. In calculus, codomains cause problems as noted in [15]. Also, for functions with codomains, $g \circ f$ is defined only if the domain of g equals the codomain of f, which would be severely limiting in calculus. It is good for instructors to bear in mind that *definitional choices must be justified* [13].

4. *Functions* versus *Expressions* in Derivative Formulas.

A derivative formula can be written in terms of either *functions* or *numbers* [1, p. 164]. This is rarely emphasized in textbooks, yet has important ramifications far beyond the chain rule.

For brevity, expressions such as f, g, $g \circ f$, $\mathrm{D} f$, which always stand for functions, will just be called *functions*. Expressions such as x, y, $f(x^2)$, $x + y$ may stand for numbers or vectors or sometimes for themselves and will just be called *expressions.*

We first apply this to ordinary derivatives, but since these are not our main topic the discussion is kept short. Consider the following definition from [5, p. 127].

Definition. Let $f : \mathbb{R} \nrightarrow \mathbb{R}$ and let x be an interior point of $\mathcal{D}f$. Then f *has a (first) derivative* at x iff $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ exists and is finite. The value of the limit is then called *the (first) derivative of* f *at* x and is denoted by $f'(x)$ or $\frac{\mathrm{d}}{\mathrm{d}x}(f(x))$. Further, the function f' whose domain is the set of interior points x of $\mathcal{D}f$ at which f has a derivative and whose value at x is $f'(x)$ is called *the (first) derivative of* f.

Analyzing the expressions $f'(x)$ and $\frac{\mathrm{d}}{\mathrm{d}x}(f(x))$ elucidates the style difference.

(i) As defined, the operator $'$ maps a function f to a function f'. An equivalent form for f' is $\mathrm{D}f$ [1, sect. 4.8], [5, p. 128]. This amounts to specifying a *derivation function* $\mathrm{D} : (\mathbb{R} \nrightarrow \mathbb{R}) \to (\mathbb{R} \nrightarrow \mathbb{R})$ such that, for any $f : \mathbb{R} \nrightarrow \mathbb{R}$, $\mathrm{D}f$ is also a function in $\mathbb{R} \nrightarrow \mathbb{R}$ with domain and values as stated in the definition. Recall that $\mathrm{D}f\,x$ stands for $(\mathrm{D}f)(x)$, not for $\mathrm{D}(f(x))$. By composition, the nth derivative of f is $\mathrm{D}^n f$.

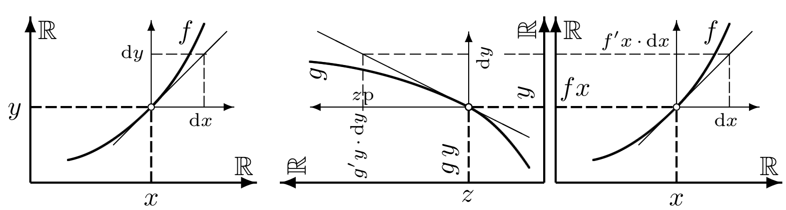
(ii) As defined, the operator $\frac{\mathrm{d}}{\mathrm{d}x}$ maps an expression $f(x)$ to an expression $\frac{\mathrm{d}}{\mathrm{d}x}(f(x))$ such that $\frac{\mathrm{d}}{\mathrm{d}x}(f(x)) = \mathrm{D}f\,x$. For instance, letting $f(x) = x^2$ yields $\frac{\mathrm{d}}{\mathrm{d}x}x^2 = \mathrm{D}f\,x = 2x$.

Since the parentheses in "$f(x)$" serve no purpose and can even be irksome, as in "$g(f(x))$," we gradually omit them. In any case, we can "disassemble" $\frac{\mathrm{d}}{\mathrm{d}x}$ by defining

$$\mathrm{d}(f(x)) = \mathrm{d}(f\,x) = (\mathrm{D}f)x \cdot \mathrm{d}x \quad \text{for all } x \text{ in the domain of } \mathrm{D}f\ . \tag{4.1}$$

As an aside, by the standards [16], a product of numbers may be written ab, $a\,b$, $a \cdot b$, or $a \times b$. The dot is often convenient for emphasis, to elucidate the structure of an expression.

Equation (4.1) allows many interpretations [2], infinitesimal and finite. One finite interpretation is that $\mathrm{d}x$ and $\mathrm{d}y$ are (new) variables and (4.1) describes the line tangent to f at x in a coordinate system with origin at (x, y), as in Figure 1(i).



(i) The d-operator (ii) The chain rule with the d-operator

Fig. 1 *Geometric interpretations.*

Note that one can properly write $\frac{\mathrm{d}}{\mathrm{d}x}(f(x))$ or $\frac{\mathrm{d}}{\mathrm{d}x}(f\,x)$ or $\frac{\mathrm{d}(f\,x)}{\mathrm{d}x}$. By contrast, $\frac{\mathrm{d}f}{\mathrm{d}x}$ is meaningless since d, just like $\frac{\mathrm{d}}{\mathrm{d}x}$ in [5], is defined only for (number) expressions. A function f is not even associated with any particular variable name such as x. For instance, $\frac{\mathrm{d}\sin}{\mathrm{d}x}$ is nonsense, whereas $\frac{\mathrm{d}(\sin x)}{\mathrm{d}x} = \cos x$ and $\frac{\mathrm{d}(\sin y)}{\mathrm{d}y} = \cos y$ are correct.

Equation (4.1) allows for fluently switching between styles during calculations *and using the normal rules for arithmetic*, as shown next for the chain rule.

(i) In terms of numbers, the chain rule can be formulated as

$$(g \circ f)'x = g'(f\,x) \cdot f'x, \text{ also written } \mathrm{D}(g \circ f)x = \mathrm{D}g\,(f\,x) \cdot \mathrm{D}f\,x, \tag{4.2}$$

for all x in $\mathcal{D}f$ such that both $g'(f\,x)$ and $f'(x)$ exist [1, p. 176], [5, p. 134].

(ii) In terms of functions, most textbooks [1, p. 175] express the chain rule as $(g \circ f)' = (g' \circ f) \cdot f'$. This does not correctly capture the existence conditions. A counterexample is $\mathsf{sqa} \circ \mathsf{abs}$, where $\mathsf{abs} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $\mathsf{abs}\, x = |x|$ and $\mathsf{sqa} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $\mathsf{sqa}\, x = x^2$. For $x = 0$, clearly $(\mathsf{sqa} \circ \mathsf{abs})'x$ is defined but $\mathsf{abs}'x$ is not. The correct expression is

$$(g \circ f)' \supseteq (g' \circ f) \cdot f', \text{ also written } \mathrm{D}(g \circ f) \supseteq (\mathrm{D}\, g \circ f) \cdot \mathrm{D} f \ , \tag{4.3}$$

where a function product $h \cdot k$ is specified by $\mathcal{D}(h \cdot k) = \mathcal{D}\, h \cap \mathcal{D}\, k$ and $(h \cdot k)\, x = hx \cdot k\, x$.

(iii) Although formulations with functions only are most elegant, in practice one often calculates with expressions and variables, so *substitution* [8] becomes central.

Letting $z := g(y)$ and $y := f(x)$ yields $z = g(y) = g(fx)$. Using (4.1), $\mathrm{d}z = \mathrm{d}(g\, y) = g'y \cdot \mathrm{d}y = g'(fx) \cdot \mathrm{d}(fx) = g'(fx) \cdot f'\, x \cdot \mathrm{d}x$ and hence the familiar expression

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\mathrm{d}z}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} \ . \tag{4.4}$$

This is consistent with (4.2). Figure 1(ii) illustrates the geometry.

This digression into ordinary derivatives was meant to illustrate three styles: (i) numbers, (ii) functions, and (iii) number expressions. In practice, all have their uses.

One last word about the improper form $\frac{\mathrm{d}f}{\mathrm{d}x}$, which gives students the impression that, in mathematics, anything goes. In physics and engineering this form is less widespread, perhaps because differential equations are usually stated in terms of expressions for physical quantities. This brings us to the following crucial topic.

5. Dimensional Analysis and Derivatives. Dimensional analysis has a direct impact on literally "everything" related to physical quantities and can be extended to "pure" real or complex numbers and vectors. It is a powerful technique for sanity checking in symbolic calculations, just like the *proof by nine* for numeric calculations.

In brief, the proof by nine checks integer multiplication $M \cdot m = p$ by verifying $M_9 \cdot m_9 = p_9$. Here n_9 stands for n modulo 9 and is calculated mentally by adding decimal digits, as in $374_9 = (3{+}7{+}4)_9 = (1{+}4)_9 = 5$ (shortcuts are evident). Division of D by d yielding quotient q and remainder r is checked using $D_9 = d_9 \cdot q_9 + r_9$.

For dimensional analysis, a complete theory is presented in [10], but here we use an elementary approach. By standard convention [16], the value of a quantity is written as a *numerical value* followed by one or more symbols expressing a *unit*. Examples are $\ell = 30\,\mathrm{mm}$, $v = 8\,\mathrm{m/s}$, and $E = 3\,\mathrm{V/m}$. The *dimension* of a quantity is expressed by symbols such as M (mass), L (length), T (time), I (electric current), and symbolic products of powers, such as LT^{-1} for velocity. The dimension of a quantity q is written $\dim q$, as in $\dim v = \mathsf{V} = \mathsf{LT}^{-1}$ and $\dim E = \mathsf{UL}^{-1}$, where $\mathsf{U} := \mathsf{ML}^2\mathsf{T}^{-3}\mathsf{I}^{-1}$.

For instance, in the usual equation for the mass-damper-spring system

$$m \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^2 x + b\frac{\mathrm{d}}{\mathrm{d}t}x + kx = F \ , \tag{5.1}$$

$\dim m = \mathsf{M}$, $\dim x = \mathsf{L}$, and $\dim t = \mathsf{T}$, and hence $\dim(m(\frac{\mathrm{d}}{\mathrm{d}t})^2 x) = \mathsf{MLT}^{-2} = \dim F = \mathsf{F}$. For consistency, $\dim(b\frac{\mathrm{d}x}{\mathrm{d}t}) = \dim(kx) = \mathsf{F}$, hence $\dim b = \mathsf{FL}^{-1}\mathsf{T}$ and $\dim k = \mathsf{FL}^{-1}$. This information can be used for sanity checking in all calculations, starting with solving the characteristic equation and ending with verifying that all elements in the solution (amplitude, frequency, decay exponent) have the right dimension (exercise).

Because of the advantages, it is recommended [16] that formulas be written as equations between quantities rather than between numerical values. In a coherent unit system, the equations look identical, which may explain (but not justify!) why equations from physics are often treated a priori as numerical in mathematics courses.

Whatever the cause, dimensional analysis is seriously underexploited in mathematics. Arguably the technique should be introduced at an early stage in high school algebra to assist with formula manipulation. This can be done gradually and in passing, for instance, in quadratic equations and by writing the equation for a circle as $x^2 + y^2 = r^2$ rather than $x^2 + y^2 = 1$. If it is true that, around the time of Viète, quantities were more common in mathematics than pure numbers, something has been lost.

For derivatives, the expression style directly supports dimensional analysis, as shown for (5.1). Even so, we have found it convenient to extend the standard notation for dimensions to *types* for quantities whose values may be real or complex numbers or vectors. This is done by simple juxtaposition, so $8\,\mathrm{m/s} \in \mathbb{R}\,\mathsf{LT}^{-1}$.

In this manner, functions between dimensioned quantities fall within the framework of section 3. An electric field is then a function $F : (\mathbb{R}\,\mathsf{L})^3 \times \mathbb{R}\,\mathsf{T} \to (\mathbb{R}\,\mathsf{UL}^{-1})^3$ or, with "genuine" vectors (not their component representation), $F : \mathbb{V}\,\mathsf{L} \times \mathbb{R}\,\mathsf{T} \to \mathbb{V}\,\mathsf{UL}^{-1}$. Given $f : \mathbb{R}\,\mathsf{D} \not\to \mathbb{R}\,\mathsf{D}'$, where D and D' are any dimensions, then the derivative $\mathrm{D}f$ is a function in $\mathbb{R}\,\mathsf{D} \not\to \mathbb{R}\,\mathsf{D}'\mathsf{D}^{-1}$, in view of the quotient used for defining $(\mathrm{D}f)x$.

The use of three different D's (derivative D, dimension D, domain $\mathcal{D}$) is unfortunate, but an alphabet of 26 letters is, like symbols in general, a scarce resource.

Anyway, D is now a *polymorphic* function since its domain consists of functions of various types depending on the dimensions. There exists a general-purpose "function toolkit" to properly handle this symbolically, but that is beyond the topic of this note.

Therefore the multivariable chain rule is explained for pure numbers, and its extension to dimensioned quantities in each case is immediate, as illustrated later.

6. Reliable Formulations of the Multivariable Chain Rule. In what follows, let $f : \mathbb{R}^n \not\to \mathbb{R}$. A running example is $\mathsf{f} : \mathbb{R}^2 \to \mathbb{R}$ with $\mathsf{f}(x, y) = \frac{1}{2}ax^2 + bxy + \frac{1}{2}cy^2$. We start from the definition in [5, p. 351], where $\mathbf{x}$ stands for any n-tuple in $\mathbb{R}^n$ and $\mathbf{e}$ for a natural basis for $\mathbb{R}^n$, for instance, if $n = 3$, then $\mathbf{e} = (1, 0, 0), (0, 1, 0), (0, 0, 1)$. Tuple arithmetic is elementwise: $(\mathbf{x} + t\mathbf{e}_j)_i = \mathbf{x}_i + t\mathbf{e}_{ji}$. We remark that in [5], $\mathbf{x}_i$ is written x_i, which demonstrates that using a boldface $\mathbf{x}$ is a waste of fonts (a scarce resource), but we leave it unchanged momentarily for easy comparison.

Definition [5, p. 351]. Let $f : \mathbb{R}^n \not\to \mathbb{R}$. Then f *has a partial derivative at* $\mathbf{x}$ *with respect to the* j*th coordinate* iff 0 is an interior point of the set $\{t : \mathbb{R} \,|\, \mathbf{x} + t\mathbf{e}_j \in \mathcal{D}f\}$ and $\lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{e}_j)-f(\mathbf{x})}{t}$ exists. This value is written $\mathrm{D}_j f \mathbf{x}$ or $\partial_{x_j}(f(\mathbf{x}))$.

As before, this introduces two operators: D_j for functions and ∂_{x_j} for expressions.

(i) D_j maps every function $f : \mathbb{R}^n \not\to \mathbb{R}$ to the function $\mathrm{D}_j f$ whose domain is the set of all $\mathbf{x}$ in $\mathcal{D}f$ where f has partial derivative with respect to the jth coordinate, and $\mathrm{D}_j f\,\mathbf{x}$, which stands for $(\mathrm{D}_j f)\mathbf{x}$, is then the value of the stated limit. For instance, $\mathrm{D}_0\mathsf{f}$ and $\mathrm{D}_1\mathsf{f}$ are functions from $\mathbb{R}^2$ to $\mathbb{R}$ such that $\mathrm{D}_0\mathsf{f}(x, y) = ax + by$ and $\mathrm{D}_1\mathsf{f}(x, y) = bx + cy$ (here $\mathbf{x}$ is x, y). In a general context, indexing from 0 has many advantages, but since it does not matter in this note we use a "neutral" index set I_n which can be $\{i : \mathbb{Z} \,|\, 0 \le i < n\}$ or $\{i : \mathbb{Z} \,|\, 0 < i \le n\}$ as the reader prefers.

(ii) ∂_{x_j} maps the expression $f(\mathbf{x})$ to $\partial_{x_j}(f(\mathbf{x}))$, where $\partial_{x_j}(f(\mathbf{x})) = (\mathrm{D}_j f)\mathbf{x}$. For instance, $\partial_x(\mathsf{f}(x, y)) = ax + by$ and $\partial_y(\mathsf{f}(x, y)) = bx + cy$.

The principle for designing a precise and reliable multivariable chain rule lies in expressing the interdependence between the variables completely yet succinctly.

As before, we consider various styles, initially using $f : \mathbb{R}^2 \nrightarrow \mathbb{R}$, $g : \mathbb{R}^3 \nrightarrow \mathbb{R}$, and $h : \mathbb{R}^3 \nrightarrow \mathbb{R}$ together with $u := f(x, y)$, $x := g(r, s, t)$, and $y = h(r, s, t)$ for illustration. We want to make explicit all interdependencies in equations of the form $\frac{\partial u}{\partial v} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial v}$ or $\partial_v u = \partial_x u \partial_v x + \partial_y u \partial_v y$ (without "false ratios"). In these equations, v may be r, s, or t, and for other choices $\partial_v u = \partial_v x = \partial_v y = 0$.

(i) For expressions, we use a convention from [8] whereby $e[^v_d$ denotes the result of substituting the expression d for every occurrence of the variable v in the expression e. Similarly, $e[^{v,w}_{d,q}$ for tuples. Note that $(x/y)[^{x,y}_{y,x} = y/x$, but $(x/y)[^x_y[^y_x = (y/y)[^y_x = x/x$.

Now $\partial_v u = \partial_x u \partial_v x + \partial_y u \partial_v y$ can be rewritten with all terms disambiguated:

$$\begin{aligned} \partial_v f(g(r,s,t), h(r,s,t)) = {} & (\partial_x f(x,y))[^{x,y}_{g(r,s,t),h(r,s,t)} \partial_v g(r,s,t) \\ & + (\partial_y f(x,y))[^{x,y}_{g(r,s,t),h(r,s,t)} \partial_v h(r,s,t) \ . \end{aligned} \tag{6.1}$$

This is complete but not succinct. The solution is a balanced use of ∂_v and D_j. We introduce the *parameter list* $p := g(r, s, t), h(r, s, t)$, so $\partial_v f(g(r, s, t), h(r, s, t))$ becomes $\partial_v(f p)$. Similarly $\partial_v g(r, s, t) = \partial_v p_0$ and $\partial_v h(r, s, t) = \partial_v p_1$. More interestingly, $(\partial_x f(x, y))[^{x,y}_{g(r,s,t),h(r,s,t)} = (\mathrm{D}_0 f)p$ and $(\partial_y f(x, y))[^{x,y}_{g(r,s,t),h(r,s,t)} = (\mathrm{D}_1 f)p$, which effectively eliminates the troublesome variables x and y. Equation (6.1) then becomes

$$\partial_v(f p) = \mathrm{D}_0 f p \cdot \partial_v p_0 + \mathrm{D}_1 f p \cdot \partial_v p_1 \ . \tag{6.2}$$

Thus the general form for $f : \mathbb{R}^n \nrightarrow \mathbb{R}$ and any expression p with values in $\mathbb{R}^n$ is

$$\partial_v(f p) = \sum_{j:\mathrm{I}_n} \mathrm{D}_j f p \cdot \partial_v\, p_j \ \text{ or, mnemonically, } \ (f p)_v = \sum_j f_j p \cdot (p_j)_v \ . \tag{6.3}$$

Depending on the reader's preference, $\sum_{j:\mathrm{I}_n}$ can stand for either $\sum_{j=0}^{n-1}$ or $\sum_{j=1}^{n}$. Equation (6.3) contains no number variables other than those appearing in p and hence no ambiguities. It gives students a clear, safe rule that eliminates all possible confusion, second guessing, and surprises. Moreover, since the intermediary variables (x and y in the example) do not appear in the formula, even reusing them is evident. For instance, the running example f with $p := (x, mx + ny)$ and $v := x$ yields

$$\begin{aligned} \partial_x(\mathsf{f}p) &= \partial_x(\mathsf{f}(x, mx+ny)) &&= \partial_x(\tfrac{1}{2}ax^2 + bx(mx+ny) + \tfrac{1}{2}c(mx+ny)^2\,, \\ (\mathrm{D}_0\mathsf{f})p \cdot \partial_v\, p_0 &= (\mathrm{D}_0\mathsf{f})(x, mx+ny) \cdot 1 &&= (ax + b(mx+ny)) \cdot 1\,, \\ (\mathrm{D}_1\mathsf{f})p \cdot \partial_v\, p_1 &= (\mathrm{D}_1\mathsf{f})(x, mx+ny) \cdot m &&= (bx + c(mx+ny)) \cdot m \ . \end{aligned}$$

(ii) Variants with only functions avoid the issues with variables seen in section 2. We derive such variants from (6.3), exploiting the unifying power of functions to cover tuples and matrices and treating all variables (including indices) equally. For any set S, note that $S^n = \mathrm{I}_n \to S$ and hence $(S^n)^m = \mathrm{I}_m \to \mathrm{I}_n \to S$ ($m \times n$ matrices).

In the configuration with $f : \mathbb{R}^n \nrightarrow \mathbb{R}$ as in Figure 2(a) we let $p := g\, q$, where $g : \mathbb{R}^k \nrightarrow \mathbb{R}^n$ and q is a list of k variables. The idea is enabling the later removal of q so that only functions remain, yielding the configuration of Figure 2(b). For $v := q_i$, we can prepare the subexpressions in $\partial_v(f p) = \sum_{j:\mathrm{I}_n} \mathrm{D}_j f p \cdot \partial_v\, p_j$ for q-removal using

- $\partial_v(f p) = \partial_{q_i}(f(g\, q)) = \partial_{q_i}((f \circ g)\, q) = \mathrm{D}_i(f \circ g)\, q$,
- $\mathrm{D}_j f\, p = \mathrm{D}_j f\,(g\, q) = (\mathrm{D}_j f \circ g)\, q$,
- $\partial_v p_j = \partial_{q_i}(g\, q)_j = \partial_{q_i}(g^{\mathsf{T}}_j q) = (\mathrm{D}_i\, g^{\mathsf{T}}_j)\, q$, where g^{T}_j is explained next.

The operator $^{\mathsf{T}}$ is defined such that, for any sets X, Y, Z and any $F : X \nrightarrow Y \to Z$, the *transpose* F^{T} is in $Y \to X \nrightarrow Z$ and $F^{\mathsf{T}} y\, x = F x\, y$. Hence $g \in \mathbb{R}^k \nrightarrow \mathbb{R}^n$ implies $g^{\mathsf{T}} \in (\mathbb{R}^k \nrightarrow \mathbb{R})^n$ and $(g\, q)_j = (g^{\mathsf{T}})_j q$. By convention, g^{T}_j is shorthand for $(g^{\mathsf{T}})_j$.

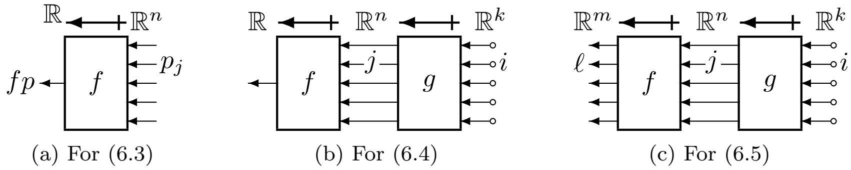


Fig. 2 *Typical configurations for the multivariable chain rule.*

Hence $\mathrm{D}_i(f\circ g)\,q = \sum_{j:\mathrm{I}_n}(\mathrm{D}_j f\circ g)\,q\cdot(\mathrm{D}_i\,g_j^{\mathsf{T}})\,q$ and, with $\cdot$ extended to functions,

$$\mathrm{D}_i(f\circ g) \supseteq \textstyle\sum_{j:\mathrm{I}_n}(\mathrm{D}_j f\circ g)\cdot \mathrm{D}_i\,g_j^{\mathsf{T}} \text{ for all } i \text{ in } \mathrm{I}_k. \tag{6.4}$$

Finally, in the configuration with $f:\mathbb{R}^n \nrightarrow \mathbb{R}^m$ as in Figure 2(c), for each $\ell:\mathrm{I}_m$ we have $\mathsf{f}_\ell^{\mathsf{T}} \in \mathbb{R}^n \nrightarrow \mathbb{R}$, so $\mathrm{D}_i(f_\ell^{\mathsf{T}}\circ g)\,q = \sum_{j:\mathrm{I}_n}(\mathrm{D}_j f_\ell^{\mathsf{T}}\circ g)\,q\cdot(\mathrm{D}_i\,g_j^{\mathsf{T}})\,q$. To prepare all variables i, ℓ, q, j for later removal we specify $\mathbf{D}$ such that, for any $f:\mathbb{R}^n \nrightarrow \mathbb{R}^m$, we have $\mathbf{D}f \in \mathbb{R}^n \nrightarrow (\mathbb{R}^n)^m$ with $\mathbf{D}f\,p\,\ell\,j = \mathrm{D}_j f_\ell^{\mathsf{T}} p$ and the domain of $\mathbf{D}f$ consists of those $p:\mathbb{R}^n$ for which all $\mathrm{D}_j f_\ell^{\mathsf{T}} p$ exist. The matrix $\mathbf{D}f p$ is the *Jacobian for f at p.*

After proving that $f_\ell^{\mathsf{T}}\circ g = (f\circ g)_\ell^{\mathsf{T}}$ and $(\mathrm{D}_j f_\ell^{\mathsf{T}}\circ g)\,q = (\mathbf{D}f\circ g)\,q\,\ell\,j$ (exercises), we find $\mathbf{D}(f\circ g)\,q\,\ell\,i = \sum_{j:\mathrm{I}_n}(\mathbf{D}f\circ g)\,q\,\ell\,j\cdot\mathbf{D}\,g\,q\,j\,i$ or, with matrix multiplication, $\mathbf{D}(f\circ g)\,q = (\mathbf{D}f\circ g)\,q\times\mathbf{D}\,g\,q$. Extending $\times$ to functions as we did for $\cdot$ yields

$$\mathbf{D}(f\circ g) \supseteq (\mathbf{D}f\circ g)\times\mathbf{D}\,g\ . \tag{6.5}$$

We take $\mathbf{D}$ to be polymorphic in n and m, but also for dimensioned quantities.

The detailed examples that follow illustrate the mechanism of (6.3) and reveal that it is more than a chain rule, but is also quite versatile for changing variables.

A useful technique is the following. If p is a tuple of variables, $\partial_{p_j}(fp) = (\mathrm{D}_j f)p$ allows decoupling the argument p in fp from the variable p_j in ∂_{p_j}. By pushing out p in every term of an equation, all these terms take the form $(\mathrm{D}_j f)p$ and p can be uniformly replaced throughout the equation by any other tuple of expressions.

7. A Simple Example: Solving the Transport Equation.

The transport equation [18, p. 20] is the simplest nontrivial example illustrating basic calculation techniques. Intuitively, it simply describes the observation, in one-dimensional space and in time, of an immutable one-dimensional "shape" moving at constant speed c.

Some students may benefit from Tisdell's YouTube lecture about this topic via the link in [18, p. 132]. Looking up the principles of slit photography is also edifying.

The transport equation for a quantity $u = f(x,t)$ (not $u = u(x,t)$) can be written

$$c\,\partial_x u + \partial_t u = g(x,t) \quad \text{with boundary condition} \quad f(x,0) = h(x)\ , \tag{7.1}$$

where c is a velocity independent of x and t, whereas g and h are given functions.

The term "quantity" recalls that all expressions are dimensioned. In (7.1), $\dim u = \mathsf{U}$ (application-dependent), $\dim x = \mathsf{L}$, $\dim t = \mathsf{T}$, $\dim c = \mathsf{LT}^{-1}$, and, formally, $\dim\partial_x = \mathsf{L}^{-1}$ and $\dim\partial_t = \mathsf{T}^{-1}$. The functions can be declared as $f:\mathbb{R}\,\mathsf{L}\times\mathbb{R}\,\mathsf{T}\to\mathbb{R}\,\mathsf{U}$, $g:\mathbb{R}\,\mathsf{L}\times\mathbb{R}\,\mathsf{T}\to\mathbb{R}\,\mathsf{UT}^{-1}$, and $h:\mathbb{R}\,\mathsf{L}\to\mathbb{R}\,\mathsf{U}$.

Using $\partial_{p_j}(fp) = (\mathrm{D}_j f)p$ with $p = (x,t)$ and $fp = u$, (7.1) becomes

$$c(\mathrm{D}_0 f)(x,t) + (\mathrm{D}_1 f)(x,t) = g(x,t)$$

for any x,t in the domain of both $\mathrm{D}_0 f$ and $\mathrm{D}_1 f$ (normally such qualifications are tacitly understood). Hence, for any p,

$$c(\mathrm{D}_0 f)p + (\mathrm{D}_1 f)p = g\,p\ . \tag{7.2}$$

If $p = (x + ct, t)$, rule (6.3) yields $\partial_t(fp) = c(\mathrm{D}_0 f)p + (\mathrm{D}_1 f)p$ and (7.2) becomes

$$\partial_t(fp) = g\,p \ .$$

The optional parentheses in $(\mathrm{D}_j f)p$ were left in place to illustrate how p was factored out of $\partial_t(fp)$, but henceforth we will often omit them.

Expanding p and integrating both sides with respect to t (in any order) yields

$$f(x + cs, s)\big|_{s=0}^{s=t} = f(x + ct, t) - f(x, 0) = \int_0^t g(x + cs, s)\,\mathrm{d}s \ .$$

Since $f(x, 0) = h(x)$, substituting $x - ct$ for x yields the solution to (7.1):

$$u = f(x, t) = h(x - ct) + \int_0^t g(x + c(s - t), s)\,\mathrm{d}s \ . \tag{7.3}$$

To verify that (7.3) is indeed a solution for (7.1) one can use the following result.

8. Partial Derivatives of a Definite Integral. In view of the polymorphism explained earlier, we omit dimensions here. Consider

$$f(x, t) = \int_{a(t)}^{b(t)} g(s, t, x)\,\mathrm{d}s \ , \tag{8.1}$$

where $g : \mathbb{R}^3 \nrightarrow \mathbb{R}$, $a : \mathbb{R} \nrightarrow \mathbb{R}$, and $b : \mathbb{R} \nrightarrow \mathbb{R}$ are known functions.

This example represents a fairly general form where the integrand and the integration boundaries depend on other variables.

Letting $F(r, t, x) = \int_0^r g(s, t, x)\,\mathrm{d}s$, clearly $f(x, t) = F(b(t), t, x) - F(a(t), t, x)$. The interesting derivative is $\partial_t(f(x, t))$. Equation (6.3) yields (since $\partial_t\, x = 0$)

$$\begin{aligned}\partial_t(F(a(t), t, x)) &= (\mathrm{D}_0 F)(a(t), t, x) \cdot \partial_t(a(t)) + (\mathrm{D}_1 F)(a(t), t, x) \cdot \partial_t\, t \\ &= (\mathrm{D}_0 F)(a(t), t, x) \cdot \mathrm{D}\, a\, t + (\mathrm{D}_1 F)(a(t), t, x)\end{aligned}$$

and similarly for $\partial_t(F(b(t), t, x))$. From $F(r, t, x) = \int_0^r g(s, t, x)\,\mathrm{d}s$ it follows that $(\mathrm{D}_0 F)(r, t, x) = g(r, t, x)$ and $(\mathrm{D}_1 F)(r, t, x) = \int_0^r \partial_t(g(s, t, x))\,\mathrm{d}s$. The result, sometimes called Leibniz's rule [17, sect. 1.5], is

$$\partial_t\left(\int_{a(t)}^{b(t)} g(s, t, x)\,\mathrm{d}s\right) = g(b(t), t, x) \cdot \mathrm{D}\, b\, t - g(a(t), t, x) \cdot \mathrm{D}\, a\, t + \int_{a(t)}^{b(t)} \partial_t(g(s, t, x))\,\mathrm{d}s \ .$$

9. An Example from Practice: The Distortionless Transmission Line. The *telegrapher's equations* for voltage and current in a transmission line [12, p. 50] yield $\partial_z^2\, v(z, t) = LC\partial_t^2\, v(z, t) + (RC + GL)\partial_t\, v(z, t) + GR\, v(z, t)$ for the voltage $v(z, t)$ at position z and time t along a transmission line. Here L, C, R, G are ratios of series inductance, shunt capacitance, series resistance, and shunt conductance to length.

When calculations get messy, dimensional analysis can help. With $\dim v = \mathsf{U}$, $\dim(LC) = \mathsf{T}^2\mathsf{L}^{-2}$, $\dim(RC) = \mathsf{T}\mathsf{L}^{-2}$, $\dim(GR) = \mathsf{L}^{-2}$, and, formally, $\dim \partial_x = \mathsf{L}^{-1}$, $\dim \partial_t = \mathsf{T}^{-1}$, one can verify that each term in the equation has dimension $\mathsf{U}\mathsf{L}^{-2}$.

If the *Heaviside condition* $R/L = G/C$ is satisfied, the equation can be simplified to $\partial_z^2 v(z, t) = LC(\partial_t + R/L)^2 v(z, t)$ or, with $c = 1/\sqrt{LC}$ and $\alpha = R\sqrt{C/L}$,

$$\partial_z^2 v(z, t) = \left(\frac{1}{c}\partial_t + \alpha\right)^2 v(z, t) \ . \tag{9.1}$$

Products of operators stand for composition, so $(\frac{1}{c}\partial_t + \alpha)^2 = (\frac{1}{c}\partial_t + \alpha) \circ (\frac{1}{c}\partial_t + \alpha)$.

To solve (9.1), we write it as $(\partial_z - \frac{1}{c}\partial_t - \alpha)(\partial_z + \frac{1}{c}\partial_t + \alpha)v(z,t) = 0$. Introducing $f(z,t) = (\partial_z + \frac{1}{c}\partial_t + \alpha)v(z,t)$ transforms this into

$$\left(\partial_z - \frac{1}{c}\partial_t - \alpha\right)(f(z,t)) = 0 \quad \text{or} \quad \left(\mathrm{D}_0 f - \frac{1}{c}\mathrm{D}_1 f - \alpha f\right)(z,t) = 0 \ . \tag{9.2}$$

Unlike the leftmost form, the rightmost form allows substituting any pair for z, t. Let $p = (z, t - \frac{z}{c})$. By (6.3), $\partial_z(f p) = \mathrm{D}_0 f p - \frac{1}{c}\mathrm{D}_1 f p$, so (9.2) becomes $\partial_z(fp) - \alpha f p = 0$. The solution (as an ODE in z) is $fp = \mathrm{e}^{\alpha z}\, g\, t$ (with an arbitrary function g) or $f(z, t - \frac{z}{c}) = \mathrm{e}^{\alpha z}\, g\, t$. Replacing t by $t + \frac{z}{c}$ yields $f(z,t) = \mathrm{e}^{\alpha z}\, g(t + \frac{z}{c})$.

The second step is solving

$$\begin{aligned} &\left(\partial_z + \frac{1}{c}\partial_t + \alpha\right) v(z,t) = f(z,t) = \mathrm{e}^{\alpha z}\, g\left(t + \frac{z}{c}\right) \quad \text{or} \\ &\left(\mathrm{D}_0 v + \frac{1}{c}\mathrm{D}_1 v + \alpha v\right)(z,t) = \mathrm{e}^{\alpha z} g\left(t + \frac{z}{c}\right) \ . \end{aligned} \tag{9.3}$$

Let $p = (z, t + \frac{z}{c})$. By (6.3), $\partial_z(vp) = \mathrm{D}_0 v\, p + \frac{1}{c}\mathrm{D}_1 v\, p$, so (9.3) becomes $\partial_z(vp) + \alpha v p = \mathrm{e}^{\alpha z} g(t + \frac{2z}{c})$. The solution (as an ODE in z) is

$$vp = \mathrm{e}^{-\alpha z}\, \varphi\, t + \mathrm{e}^{-\alpha z}\int_0^z \mathrm{e}^{2\alpha x} g\left(t + \frac{2x}{c}\right)\mathrm{d}x \text{ (with an arbitrary function } \varphi) \ .$$

Symmetry suggests $\int_0^z \mathrm{e}^{2\alpha x} g(t + \frac{2x}{c})\mathrm{d}x = \mathrm{e}^{2\alpha z}\psi(t + \frac{2z}{c})$ for suitable ψ. By differentiation, $\mathrm{e}^{2\alpha z} g(t + \frac{2z}{c}) = \frac{2}{c}\mathrm{e}^{2\alpha z}\psi'(t + \frac{2z}{c}) + 2\alpha \mathrm{e}^{2\alpha z}\psi(t + \frac{2z}{c})$ and hence $g\, t = \frac{2}{c}\psi' t + 2\alpha\, \psi\, t$. Thus one finds g for given ψ or, conversely, ψ for given g. Finally,

$$\begin{aligned} &v\left(z, t + \frac{z}{c}\right) = \mathrm{e}^{-\alpha z}\varphi(t) + \mathrm{e}^{\alpha z}\psi\left(t + \frac{2z}{c}\right) \text{ or, replacing } t \text{ by } t - \frac{z}{c}, \\ &v(z,t) = \mathrm{e}^{-\alpha z}\varphi\left(t - \frac{z}{c}\right) + \mathrm{e}^{\alpha z}\psi\left(t + \frac{z}{c}\right) \ , \end{aligned} \tag{9.4}$$

which solves (9.1). For $\alpha = 0$ this yields d'Alembert's solution [18, p. 40].

Equation (9.4) describes two waves moving in opposite directions with just delay $(-z/c)$ and attenuation $(\mathrm{e}^{-\alpha z})$, so the line is *distortionless* [12, p. 80]. This requires satisfying the Heaviside condition (in practice by inserting coils, since G is small).

10. Conclusion. The stated objective of making the chain rule precise and reliable has been met by combining derivation operators for functions and for expressions in a balanced manner, as in (6.3). Examples demonstrate how this rule facilitates changing variables without errors and how it conveys structure to the calculations.

Since students may encounter both kinds of operators in various forms in courses, on the web, and in their later work, learning how to use them properly with clear, safe calculation rules eliminates confusion and enhances understanding. Also, exploiting the functional style offers many opportunities [3] in other areas of mathematics.

This note is also an object lesson in the rewards of adhering to good practices. Much can be said about this, but some thoughts are summarized in Appendix A.

An additional reason for being meticulous is that, for many decades, notational conventions are increasingly being "cast in concrete" through mathematical software. It would be unfortunate if poor design decisions were perpetuated in this manner.

The teaching suggestions from section 1 can be adapted by an instructor to the specific scenario for their course and students. If a two-hour lecture is preferable, a good place for a break would be between items (i) and (ii) in section 6, since the conclusion of (i) wraps back around to the issues in section 2 and (ii) initiates the part with applications, some of which may also require some extra outlining of the context.

Appendix A. On Teaching Good Mathematical Practices. For (2.1) Humpty Dumpty might well say, "Every term means just what I choose it to mean," but if students during an exam invented a formula where, for "succinctness," every term required its own interpretation, they would be dismissed on the spot. Yet such is the power of tradition that students are supposed to accept such practices from people in authority. This double standard is very damaging in education.

Adhering to better practices than in the textbook can be problematic for instructors. However, even when some flawed notation is considered "common," it is in the best interest of the students to provide a better alternative. To prepare for other conventions and possible flaws, a simple warning suffices: students are bound to remember.

Acknowledgment. The author wishes to thank the referees and the editors for their useful comments.

REFERENCES

[1] T. M. Apostol, *Calculus*, Vol. I, 2nd ed., John Wiley, New York, 1967. (Cited on pp. 856, 857, 858)

[2] J. Bair et al., *Is mathematical history written by the victors?*, Notices Amer. Math. Soc., 60 (2013), pp. 886–904. (Cited on p. 857)

[3] R. Boute, *Pointfree expression and calculation: From quantification to temporal logic*, Form. Meth. Syst. Des., 37 (2010), pp. 95–140. (Cited on p. 863)

[4] A. Dasgupta, *Set Theory*, Birkhäuser, Berlin, 2014. (Cited on p. 856)

[5] T. M. Flett, *Mathematical Analysis*, McGraw-Hill, New York, 1966. (Cited on pp. 855, 856, 857, 859)

[6] J. Gibbon, *The Chain Rule in Partial Differentiation*, http://wwwf.imperial.ac.uk/~jdg/AECHAIN.PDF. (Cited on p. 854)

[7] É. Goursat, *A Course in Mathematical Analysis*, Dover, New York, 2005 (reprint of the 1904 edition). (Cited on p. 854)

[8] D. Gries and F. B. Schneider, *A Logical Approach to Discrete Math*, Springer, Berlin, 2010. (Cited on pp. 855, 858, 860)

[9] E. A. Lee and P. Varaiya, *Introducing signals and systems, the Berkeley approach*, in First Signal Processing Education Workshop, Hunt, TX, 2000. (Cited on p. 855)

[10] G. Meinsma, *Dimensional and scaling analysis*, SIAM Rev., 61 (2019), pp. 159–184, https://doi.org/10.1137/16M1107127. (Cited on p. 858)

[11] P. J. Olver, *Introduction to Partial Differential Equations*, Springer, 2014. (Cited on p. 855)

[12] D. M. Pozar, *Microwave Engineering*, 4th ed., Wiley, 2012. (Cited on pp. 862, 863)

[13] P. Rogaway, *On the role of definitions in and beyond cryptography*, in Advances in Computer Science, Lecture Notes in Comput. Sci. 3321, M. J. Maher, ed., Springer, Berlin, 2004, pp. 13–32. (Cited on p. 856)

[14] E. R. Scheinerman, *Mathematics—A Discrete Introduction*, 3rd ed., Cengage Learning, Boston, 2012. (Cited on p. 856)

[15] H. S. Shuard, *Does it matter?*, Math. Gaz., 59 (1975), pp. 7–15. (Cited on p. 856)

[16] A. Thompson and B. N. Taylor, *Guide for the Use of the International System of Units (SI)*, NIST Special Publication 811, National Institute of Standards and Technology, Gaithersburg, MD, 2008, https://www.nist.gov/pml/special-publication-811. (Cited on pp. 857, 858, 859)

[17] C. C. Tisdell, *Engineering Mathematics: YouTube Workbook*, bookboon.com, 2013. (Cited on p. 862)

[18] C. C. Tisdell, *Partial Differential Equations: YouTube Workbook*, bookboon.com, 2018. (Cited on pp. 861, 863)