FORMAL DIFFERENTIAL VARIABLES AND AN ABSTRACT CHAIN RULE

SAMUEL ALEXANDER

Abstract. One shortcoming of the chain rule is that it does not iterate: it gives the derivative of $f(g(x))$, but not (directly) the second or higher-order derivatives. We present iterated differentials and a version of the multivariable chain rule which iterates to any desired level of derivative. We first present this material informally, and later discuss how to make it rigorous (a discussion which touches on formal foundations of calculus). We also suggest a finite calculus chain rule (contrary to Graham, Knuth and Patashnik's claim that "there's no corresponding chain rule of finite calculus").

1. Introduction

Consider the following statement, uncontroversial in an elementary calculus context $(*)$: "For all variables u and v, $d(uv) = v\,du + u\,dv$." In his popular calculus textbook [11], Stewart says:

> ...the **differential** dx is an independent variable...

So if $*$ really does hold for *all* variables u and v, and if x is a variable, and if (as Stewart says) dx is also a variable, then, by letting $u = x$ and $v = dx$, we get $d(x\,dx) = dx\,dx + x\,ddx$. We do not know whether Stewart intended us to make such an unfamiliar-looking conclusion from his innocent-looking statement, but let's continue along these lines and see where it leads us. We will formalize this kind of computation using machinery from first-order logic, and show that it leads to an elegant higher-order multivariable chain rule.

A weakness of the familiar chain rule is that it does not iterate: it tells us how to find the *first* derivative of $f(g(x))$, but it does not tell us how to find second- or higher-order derivatives of the same (at least not directly). Our abstract chain rule will iterate: the exact same rule which tells us $df(g(x))$ will also tell us $d^k f(g(x))$ for any integer $k > 1$.

Our d operator has some similarities with the Δ operator of Huang et al [7]. Our work improves on theirs in that we explicitly distinguish differential variables from others, so that the operator we develop better reveals the connection to differentials. For example, in Huang et al, one has $\Delta_1 e^{x_0} = e^{x_0} x_1$ and $\Delta_2 e^{x_0} = e^{x_0}(x_1^2 + x_2)$, which is equivalent to our $de^{x_0} = e^{x_0}\,dx_0$ and $d^2 e^{x_0} = e^{x_0}(dx_0\,dx_0 + ddx_0)$. Besides better emphasizing the connection to differentials, the latter version should also be more familiar, since we already routinely write things like $de^x = e^x\,dx$ in elementary calculus classes.

2. Computing iterated partial derivatives: informal examples

In this section, we will informally describe a way to compute iterated partial derivatives of a multivariable function. We will make the method formal in subsequent sections.

Example 2.1. Compute the differential $dd\,x^2 = d(dx^2)$, treating differential variables just like ordinary variables.

Solution. The differential $dx^2 = 2x\,dx$ involves two variables: x and dx. Thus, $d(dx^2)$ will have two terms, one where we differentiate with respect to x and multiply the result by dx, and one where we differentiate with respect to dx and multiply the result by ddx:

$$\begin{aligned} ddx^2 &= d(dx^2) \\ &= d(2x\,dx) \\ &= \frac{\partial(2x\,dx)}{\partial x}\,dx + \frac{\partial(2x\,dx)}{\partial dx}\,ddx \\ &= 2\,dx\,dx + 2x\,ddx. \end{aligned}$$

Note that when we compute $\frac{\partial(2x\,dx)}{\partial x}$, we treat dx as a variable independent from x, so dx can be treated as a constant. Likewise when we compute $\frac{\partial(2x\,dx)}{\partial dx}$, x is treated as a constant. □

Example 2.2. Compute the differential $dd\,e^x$, treating differential variables just like ordinary variables.

Solution. As in Example 2.1, since $d\,e^x = e^x\,dx$,

$$\begin{aligned} dd\,e^x &= d\,(e^x\,dx) \\ &= \frac{\partial(e^x\,dx)}{\partial x}\,dx + \frac{\partial(e^x\,dx)}{\partial dx}\,ddx \\ &= e^x\,dx\,dx + e^x\,ddx. \end{aligned}$$

□

Example 2.3. Compute $dd\,f(x)$, treating differential variables just like ordinary variables.

Solution. Just as above,

$$\begin{aligned} dd\,f(x) &= d(f'(x)\,dx) \\ &= \frac{\partial(f'(x)\,dx)}{\partial x}\,dx + \frac{\partial(f'(x)\,dx)}{\partial dx}\,ddx \\ &= f''(x)\,dx\,dx + f'(x)\,ddx. \end{aligned}$$

□

In a later section, we will formalize and prove a formal chain rule (Corollary 6.9). For now, we will state it informally:

Remark 2.4. (Abstract Chain Rule, stated informally) Let T and U be expressions and let x be a non-differential variable. Assume T, U, and all of their subexpressions are everywhere infinitely differentiable. Then

$$d(T[x|U]) = (dT)[x|U],$$

where the operator $[x|U]$ works by simultaneously replacing all occurrences of x by U, all occurrences of dx by dU, all occurrences of d^2x by d^2U, and so on.

The Abstract Chain Rule can be stated in English: "substituting first and then applying d gives the same result as applying d first and then substituting, provided that when one substitutes U for x, one also substitutes dU for dx and so on."

Example 2.5. Compute $(e^{x^2})''$.

Solution. By Example 2.3, $(e^{x^2})''$ is the $dx\,dx$-coefficient of $dd\,e^{x^2}$. We compute:

$$\begin{aligned}
dd\,e^{x^2} &= dd\,(e^x[x|x^2]) \\
&= (dd\,e^x)[x|x^2] && \text{(Abstract Chain Rule)} \\
&= (e^x\,dx\,dx + e^x\,ddx)[x|x^2] && \text{(Example 2.2)} \\
&= e^{x^2}\,d(x^2)\,d(x^2) + e^{x^2}\,dd(x^2) && \text{(Substituting)} \\
&= e^{x^2}\,(2x\,dx)^2 + e^{x^2}\,(2\,dx\,dx + 2x\,ddx) && \text{(Example 2.1)} \\
&= (4x^2+2)e^{x^2}\,dx\,dx + 2xe^{x^2}\,ddx.
\end{aligned}$$

The answer is the above $dx\,dx$-coefficient:

$$(e^{x^2})'' = (4x^2+2)e^{x^2}.$$

□

Our Abstract Chain Rule works for multivariable and higher-order derivatives, too.

Example 2.6. The iterated total derivative

$$d^3 \sin xy = d^3(\sin x\,[x|xy]) = (d^3 \sin x)[x|xy]$$

encodes:

- $\partial^3 \sin xy/\partial x^3$ as its $dx\,dx\,dx$-coefficient.
- $\partial^3 \sin xy/\partial y^3$ as its $dy\,dy\,dy$-coefficient.
- $\frac{\partial^3 \sin xy}{\partial x \partial y \partial y} = \frac{\partial^3 \sin xy}{\partial y \partial x \partial y} = \frac{\partial^3 \sin xy}{\partial y \partial y \partial x}$ times 3 as its $dx\,dy\,dy = dy\,dx\,dy = dy\,dy\,dx$-coefficient (the fact that there are three ways to write this coefficient is why we write "times 3").

In Sections 5–6 we will formalize and prove the Abstract Chain Rule. But first, we will connect these higher-order differentials to a more concrete higher-order chain rule known as Faà di Bruno's formula, and also show how the same ideas lead to a finite calculus chain rule.

3. Faà di Bruno's formula

Faà di Bruno's formula, named after the 19th century Italian priest Francesco Faà di Bruno, is a formula for the higher derivatives of $f(g(x))$. See [8] and [3] for the history of Faà di Bruno's formula (see also [10] for related work in category theory by another ACMS presenter). The formula can be stated combinatorially:

$$f(g(x))^{(n)} = \sum_{\pi \in \Pi_n} f^{(|\pi|)}(g(x)) \prod_{B \in \pi} g^{(|B|)}(x)$$

where π ranges over the set Π_n of all partitions of $\{1, \ldots, n\}$ (so for each such partition π, B ranges over the blocks in π).

The ideas of Section 2 offer an intuitive way to understand the above formula[1]. For any partition $\pi = \{B_1, \ldots, B_k\}$ of $\{1, \ldots, n\}$, let $I(\pi)$ be the expression

$$I(\pi) = f^{(k)}(x)\, d^{|B_1|}x\, d^{|B_2|}x\, \cdots\, d^{|B_k|}x$$

involving iterated differentials as in Section 2. By an inductive argument, one can check that

$$d^n f(x) = \sum_{\pi \in \Pi_n} I(\pi)$$

(for the inductive step, consider the different ways of obtaining a partition $\pi' \in \Pi_{n+1}$ from a partition $\pi \in \Pi_n$: one can either add $\{n+1\}$ as a new block, which corresponds to changing $f^{(k)}(x)$ to $f^{(k+1)}(x)dx$ when using the product rule to calculate $dI(\pi)$; or one can add $n+1$ to existing block B_i of π, which corresponds to changing $d^{|B_i|}x$ to $d^{|B_i|+1}x$ when using the product rule to calculate $dI(\pi)$).

By similar reasoning as in Examples 2.3 and 2.5, $f(g(x))^{(n)}$ is the $(dx)^n$-coefficient of $d^n f(g(x)) = d^n(f(x)[x|g(x)]) = (d^n f(x))[x|g(x)]$. Thus $f(g(x))^{(n)}$ is the $(dx)^n$-coefficient of

$$\sum_{\pi \in \Pi_n} I(\pi)[x|g(x)] = \sum_{\pi \in \Pi_n} f^{(|\pi|)}(g(x)) \prod_{B \in \pi} d^{|B|}g(x).$$

One can check that $d^{|B|}g(x) = g^{(|B|)}(x)d^{|B|}x + o$ where o is a sum of terms involving higher-order differentials (which can be ignored because they contribute nothing to the $(dx)^n$-coefficient we seek). Faà di Bruno's formula follows.

4. Application to finite calculus

The ideas in this paper also lead to a chain rule for the so-called finite calculus. The finite calculus is described in Section 2.6 of Graham, Knuth and Patashnik's *Concrete Mathematics* [4]. In finite calculus, one defines an operator Δ on functions by $\Delta f(x) = f(x+1) - f(x)$. This operator has many surprising analogies with differentiation, but Graham et al claim: "there's no corresponding chain rule of finite calculus, because there's no nice form for $\Delta f(g(x))$." To the contrary, since $\Delta x = (x+1) - x = 1$, an equivalent way to write $\Delta f(x)$ is

$$\Delta f(x) = f(x + \Delta x) - f(x).$$

One can then easily check that

$$\Delta f(g(x)) = f(g(x + \Delta x)) - f(g(x)) = f(g(x) + \Delta g(x)) - f(g(x)),$$

which can be expressed as a chain rule

$$\Delta(f(x)[\![x|g(x)]\!]) = (\Delta f(x))[\![x|g(x)]\!],$$

where $[\![x|g(x)]\!]$ operates by replacing x by $g(x)$ and Δx by $\Delta g(x)$.

---

[1]Shortly after presenting this argument at ACMS, we realized that the argument can actually be applied directly, without using iterated differentials at all, yielding a shockingly short elementary proof of Faà di Bruno's formula. Examining the literature, we found that the basic idea is already known [9] [6], but both published proofs which we found are actually proofs of more complicated multivariable generalizations of Faà di Bruno's formula. For the single-variable special case, the idea (essentially the same idea which we presented using iterated differentials at ACMS) is so simple that it can be written with a single sentence [2].

Of course, to make this rigorous, it would be necessary to work in a formal language so as to carefully track which "1"s are "Δx"s. For example, if $f(x) = 1/(1+x^2)$, we want $f(x)[\![x|g(x)]\!]$ to be $1/(1+g(x)^2)$, not $\Delta g(x)/(\Delta g(x)+g(x)^2)$, even though $\Delta x = 1$. We will not go through the necessary formalism in this paper, but it would be very similar to the formalism required for the d operator, which we devote the whole rest of the paper to.

5. Formalizing terms

In this section, we will formalize the terms (or expressions) of differential calculus. We attempt to make this formalization self-contained. The machinery we develop here is very similar to the machinery used to define terms in first-order logic, except that we assume more structure on the set of variables than is assumed in first-order logic.

Note that one could strongly argue that elementary calculus already implicitly operates on terms, abusing language to call terms "functions". For example, $x \mapsto x^2$ and $y \mapsto y^2$ are two names for the exact same function. Yet, nevertheless, in elementary calculus, the expressions x^2 and y^2 are *not* interchangeable [5]. Evidently, such discrepancies point to the fact that elementary calculus really is done using formal terms, implicitly. In the following, we make it explicit.

Definition 5.1. (Variables) We fix a set of *variables* defined inductively as follows.

(1) For the base step, we fix a countably infinite set $\{x_0, x_1, \ldots\}$ of distinct elements called *precalculus variables*, and we declare them to be variables.
(2) Inductively, for every variable v, we fix a new variable dv, which we call a *differential variable*; we do this in such a way as to satisfy the following requirement (we write $d^n v$ for $ddd \cdots dv$ where d occurs n times):
   - (Unique Readability) For all $n, m \in \mathbb{N}$, for all variables v and w, if $d^n v$ is the same variable as $d^m w$, then $n = m$ and $v = w$.

We write $\mathscr{V}$ for the set of variables.

Examples of variables include x_1, x_{50}, dx_0, ddx_3, d^4x_{50} (shorthand for $ddddx_{50}$), and so on. The unique readability property guarantees that, for example, dx_1 is not the same variable as dx_2 or ddx_3 or $dddx_1$, etc. We allow n or m to be 0 in the unique readability requirement, so, for example, x_1 and dx_1 are not the same variable (since d^0x_1 denotes x_1). Every variable is either a precalculus variable (in which case it is x_n for some $n \in \mathbb{N}$) or a differential variable (in which case it is $d^m x_n$ for some $n, m \in \mathbb{N}$ with $m > 0$).

Definition 5.2. (Constant symbols and function symbols)

(1) We fix a distinct set $\{\overline{r}\}_{r\in\mathbb{R}}$ of *constant symbols* for the real numbers. For any $r \in \mathbb{R}$, $\overline{r}$ is the *constant symbol for* r.
(2) For every $n \in \mathbb{N}$ with $n > 0$, we fix a distinct set $\{\overline{f}\}_f$ of *n-ary function symbols*, where f ranges over the set of all functions from $\mathbb{R}^n$ to $\mathbb{R}$. For any such f, $\overline{f}$ is the *n-ary function symbol for* f.

We make these choices in such a way that no variable is a constant symbol, no variable is an n-ary function symbol (for any n), and no constant symbol is an n-ary function symbol (for any n).

For example, the exponential function exp gives rise to a 1-ary (or *unary*) function symbol $\overline{\exp}$. The addition function + gives rise to a 2-ary (or *binary*) function symbol $\overline{+}$.

Definition 5.3. (Terms) We define the *terms of differential calculus* (or simply *terms*) inductively as follows.

(1) Every variable v is a term.
(2) Every constant symbol is a term.
(3) For all $n \in \mathbb{N}$ $(n > 0)$, for every $f : \mathbb{R}^n \to \mathbb{R}$, for all terms $U_1, \ldots, U_n$, $\overline{f}(U_1, \ldots, U_n)$ is a term.

Examples of terms include $\overline{5}$, $\overline{\pi}$, x_1, dx_2, $\overline{\sin}(x_1)$, $\overline{+}(x_0, x_1)$, and so on. We often abuse notation and suppress the overlines and possibly parentheses when writing terms. For example, we might write $\sin x_0$ instead of $\overline{\sin}(x_0)$, $\cos \pi$ instead of $\overline{\cos}(\overline{\pi})$, and so on. For certain well-known functions, we sometimes abuse notation further, for example, writing:

- $x_0 + x_1$ instead of $\overline{+}(x_0, x_1)$;
- $2x_0$ instead of $\overline{\cdot}(\overline{2}, x_0)$;
- $x_0\, dx_1$ instead of $\overline{\cdot}(x_0, dx_1)$;
- x_0^2 instead of $\overline{x \mapsto x^2}(x_0)$;
- e^{x_1} instead of $\overline{exp}(x_1)$;
- $x_0\, dx_1 + x_1\, dx_0$ instead of $\overline{+}(\overline{\cdot}(x_0, dx_1), \overline{\cdot}(x_1, dx_0))$;
- and so on.

This should cause no confusion in practice.

Definition 5.4. (Term interpretation)

- By an *assignment*, we mean a function $s : \mathscr{V} \to \mathbb{R}$ (recall that $\mathscr{V}$ is the set of variables).
- Let s be an assignment. For every term T, we define the *interpretation* $T^s \in \mathbb{R}$ of T (according to s) by induction on term complexity as follows.
  (1) If T is a constant symbol $\overline{r}$, then $T^s = r$.
  (2) If T is a variable v, then $T^s = s(v)$.
  (3) If T is $\overline{f}(U_1, \ldots, U_n)$ for some $f : \mathbb{R}^n \to \mathbb{R}$ and terms $U_1, \ldots, U_n$, then $T^s = f(U_1^s, \ldots, U_n^s)$.

For example, if $s(x_0) = 5$, then $\overline{\exp}(x_0)^s = e^5$. If $s(x_0) = 9$ and $s(dx_0) = 0.1$, then $(x\, dx)^s = 9 \cdot 0.1 = 0.9$.

Definition 5.5. (Free variables) We define the *free variables* $\mathrm{FV}(T)$ of a term T as follows.

(1) If T is a constant symbol, then $\mathrm{FV}(T) = \emptyset$ (the empty set).
(2) If T is a variable v, then $\mathrm{FV}(T) = \{v\}$.
(3) If T is $\overline{f}(U_1, \ldots, U_n)$ for some $f : \mathbb{R}^n \to \mathbb{R}$ and terms $U_1, \ldots, U_n$, then

$$\mathrm{FV}(T) = \mathrm{FV}(U_1) \cup \cdots \cup \mathrm{FV}(U_n).$$

For example, $\mathrm{FV}(\overline{5}) = \emptyset$, $\mathrm{FV}(x_6) = \{x_6\}$, $\mathrm{FV}(dx_2) = \{dx_2\}$ (note that x_2 is *not* a free variable of dx_2), $\mathrm{FV}(e^{x_0+x_1}) = \{x_0, x_1\}$, $\mathrm{FV}(x_1\, dx_2) = \{x_1, dx_2\}$.

Lemma 5.6. *Suppose T is a term, v is a variable, and s is an assignment. If $v \notin \mathrm{FV}(T)$, then T^s does not depend on $s(v)$.*

Proof. By induction. □

Definition 5.7. (Semantic equivalence) If T and U are terms, we declare $T \equiv U$ (and say that T and U are *semantically equivalent*) if for every assignment s, $T^s = U^s$.

For example, $\sin(x_0 + 2\pi) \equiv \sin x_0$, by which we mean $\overline{\sin}(\overline{+}(x_0, \overline{2\pi})) \equiv \overline{\sin}(x_0)$.

5.1. **Formal derivatives.**

Definition 5.8. (Ordered free variables) If T is a term, we define the *ordered free variables* $\mathrm{OFV}(T)$ to be the finite sequence whose elements are the free variables $\mathrm{FV}(T)$ of T (each appearing exactly one time in the sequence), ordered such that:

- Whenever $0 < n < m$ then $d^n x_i$ precedes $d^m x_j$.
- Whenever $0 < i < j$ then $d^n x_i$ precedes $d^n x_j$.

For example,

$$\mathrm{OFV}(e^{x_1+x_3+x_2+x_2+x_{99}}\, dx_1\, d^3x_1\, dx_2\, d^{50}x_0) = (x_1, x_2, x_3, x_{99}, dx_1, dx_2, d^3x_1, d^{50}x_0).$$

Definition 5.9. If s is an assignment, w is a variable, and $r \in \mathbb{R}$, we write $s(w|r)$ for the assignment defined by

$$s(w|r)(v) = \begin{cases} r & \text{if } v \text{ is } w \\ s(v) & \text{otherwise.} \end{cases}$$

In other words, $s(w|r)$ is the assignment which is identical to s except that it overrides s's output on w, mapping w to r instead.

Lemma 5.10. *For any assignment s and variable v, $s(v|s(v)) = s$.*

Proof. Trivial. □

Definition 5.11. (Everywhere-differentiability) Let T be a term, w a variable. We say that T *is everywhere-differentiable with respect to* w if for every assignment s, the limit

$$\lim_{h \to 0} \frac{T^{s(w|s(w)+h)} - T^s}{h}$$

converges to a finite real number.

Lemma 5.12. *Let T be a term with $\mathrm{OFV}(T) = (v_1, \ldots, v_n) \neq \emptyset$, and let w be a variable. Assume T is everywhere-differentiable with respect to w. For all $r_1, \ldots, r_n$, let*

$$f(r_1, \ldots, r_n) = \lim_{h \to 0} \frac{T^{s(w|s(w)+h)} - T^s}{h}$$

where s is some assignment such that each $s(v_i) = r_i$. Then $f : \mathbb{R}^n \to \mathbb{R}$ is well-defined.

Proof. In other words, for any $r_1, \ldots, r_n \in \mathbb{R}$, $f(r_1, \ldots, r_n)$ does not depend on the choice of s, as long as each $s(v_i) = r_i$. This follows from Lemma 5.6 since T has no free variables other than $v_1, \ldots, v_n$. □

Definition 5.13. If T is a term with $\mathrm{OFV}(T) = (v_1, \ldots, v_n)$, w is a variable, and T is everywhere-differentiable with respect to w, then we define the *derivative of* T *with respect to* w, a term, written $\frac{\partial T}{\partial w}$, as

$$\frac{\partial T}{\partial w} = \overline{f}(v_1, \ldots, v_n)$$

where f is as in Lemma 5.12. We define $\frac{\partial T}{\partial w}$ to be the term $\overline{0}$ if $\mathrm{FV}(T) = \emptyset$.

Example 5.14. (Some example term derivatives)

(1) $\partial x_0 / \partial x_0 \equiv 1$.
(2) $\partial x_0 / \partial x_1 \equiv 0$.
(3) $\partial x_0 / \partial dx_0 \equiv 0$.
(4) $\partial(e^{x_1 x_2}\, dx_1)/\partial x_1 \equiv x_2 e^{x_1 x_2}\, dx_1$.

Proof. (1) The function f of Lemma 5.12 is

$$f(r) = \lim_{h \to 0} \frac{x_0^{s(x_0|s(x_0)+h)} - x_0^s}{h}$$

(for any assignment s with $s(x_0) = r$). By Definitions 5.4 and 5.9 this simplifies to $f(r) = \lim_{h\to 0} \frac{s(x_0)+h-s(x_0)}{h} = 1$. The claim follows.

(2) The function f of Lemma 5.12 is

$$f(r) = \lim_{h \to 0} \frac{x_0^{s(x_1|s(x_1)+h)} - x_0^s}{h}$$

(where $s(x_0) = r$). This simplifies to $f(r) = \lim_{h\to 0} \frac{s(x_0)-s(x_0)}{h} = 0$. The claim follows.

(3) Similar to (2).

(4) By Definition 5.8, $\mathrm{OFV}(e^{x_1 x_2}\, dx_1) = (x_1, x_2, dx_1)$. So, letting $v_1 = x_1$, $v_2 = x_2$, $v_3 = dx_1$, the function f of Definition 5.12 is

$$f(r_1, r_2, r_3) = \lim_{h \to 0} \frac{(e^{x_1 x_2}\, dx_1)^{s(v_1|s(v_1)+h)} - (e^{x_1 x_2}\, dx_1)^s}{h}$$

(where each $s(v_i) = r_i$). By Definitions 5.4 and 5.9 this simplifies to

$$f(r_1, r_2, r_3) = \lim_{h \to 0} \frac{e^{(r_1+h)r_2} r_3 - e^{r_1 r_2} r_3}{h},$$

which is $r_2 e^{r_1 r_2} r_3$ by calculus. The claim follows. □

Another way to prove Example 5.14 would be to use the following lemma.

Lemma 5.15. *For each term* T*, variable* w*, and assignment* t*, if* T *is everywhere-differentiable with respect to* w*, then*

$$\left(\frac{\partial T}{\partial w}\right)^t = \lim_{h \to 0} \frac{T^{t(w|t(w)+h)} - T^t}{h}.$$

Proof. If $\mathrm{FV}(T) = \emptyset$, the lemma is trivial. Assume not. Let $(v_1, \ldots, v_n) = \mathrm{OFV}(T)$. By definition, $\frac{\partial T}{\partial w} = \overline{f}(v_1, \ldots, v_n)$, where $f : \mathbb{R}^n \to \mathbb{R}$ is such that for all $r_1, \ldots, r_n \in \mathbb{R}$, for any assignment s with each $s(v_i) = r_i$,

$$f(r_1, \ldots, r_n) = \lim_{h \to 0} \frac{T^{s(w|s(w)+h)} - T^s}{h}.$$

In particular, let each $r_i = t(v_i)$. Then:

$$\begin{aligned}
\left(\tfrac{\partial T}{\partial w}\right)^t &= \overline{f}(v_1,\ldots,v_n)^t && \text{(Definition 5.13)}\\
&= f(t(v_1),\ldots,t(v_n)) && \text{(Definition 5.4)}\\
&= f(r_1,\ldots,r_n) && \text{(Choice of } r_1,\ldots,r_n)\\
&= \lim_{h\to 0}\frac{T^{t(w|t(w)+h)} - T^t}{h}, && \text{(Since each } t(v_i)=r_i)
\end{aligned}$$

as desired. □

Definition 5.16. (Term total differentials) Suppose T is a term. We say T is *everywhere totally differentiable* if T is everywhere-differentiable with respect to every variable. If so, we define the *total differential* $\mathbf{d}T$, a term, as follows. If $\mathrm{FV}(T) = \emptyset$ then we define $\mathbf{d}T = \overline{0}$. Otherwise, let $\mathrm{OFV}(T) = (v_1,\ldots,v_n)$ and define

$$\mathbf{d}T = \frac{\partial T}{\partial v_1}dv_1 + \cdots + \frac{\partial T}{\partial v_n}dv_n.$$

Furthermore, we inductively define $\mathbf{d}^1T$ to be $\mathbf{d}T$ and, whenever $\mathbf{d}^nT$ is defined and is everywhere totally differentiable, we define $\mathbf{d}^{n+1}T = \mathbf{d}\mathbf{d}^nT$.

For example,

$$\begin{aligned}
\mathbf{d}(x_1\,dx_2) &= \frac{\partial(x_1\,dx_2)}{\partial x_1}dx_1 + \frac{\partial(x_1\,dx_2)}{\partial dx_2}ddx_2\\
&\equiv dx_1\,dx_2 + x_1\,ddx_2.
\end{aligned}$$

Lemma 5.17. *If term T is everywhere totally differentiable and if $v_1,\ldots,v_n$ are distinct variables such that $\mathrm{FV}(T)\subseteq\{v_1,\ldots,v_n\}$, then*

$$\mathbf{d}T \equiv \frac{\partial T}{\partial v_1}dv_1 + \cdots + \frac{\partial T}{\partial v_n}dv_n.$$

Proof. Follows from the commutativity of addition and the fact that clearly $\frac{\partial T}{\partial v_i} \equiv \overline{0}$ if $v_i \notin \mathrm{FV}(T)$. □

In order to prove an abstract chain rule in Section 6, we will need a form of the classical multivariable chain rule, expressed for formal terms. For this purpose, we first introduce shorthand for finite summation notation[2].

Definition 5.18. If $m > 0$ is an integer and $T_1,\ldots,T_m$ are terms, we write $\sum_{i=1}^m T_i$ (or just $\sum_i T_i$ if no confusion results) as shorthand for $T_1 + \cdots + T_m$.

Lemma 5.19. *(Classic Multivariable Chain Rule for Terms) Suppose $f:\mathbb{R}^n\to\mathbb{R}$. Suppose $\vec{T} = (T_1,\ldots,T_n)$ are terms with each $\mathrm{FV}(T_i)\subseteq\{v_1,\ldots,v_m\}$ (where $v_1,\ldots,v_m$ are distinct). Assume that $\overline{f}(\vec{T})$ and $T_1,\ldots,T_n$ are everywhere totally differentiable. Then for all $1\le i\le m$,*

$$\frac{\partial(\overline{f}(\vec{T}))}{\partial v_i} \equiv \sum_{j=1}^n \overline{f_j}(\vec{T})\frac{\partial T_j}{\partial v_i},$$

[2] It is also possible to incorporate summation notation formally into Definition 5.3, but the details are complicated. See [1].

where $f_j = D_j f$ (the partial derivative of f (in the usual sense) with respect to its jth argument).

Proof. Let s be an assignment and fix $1 \leq i \leq m$. We must show (Definition 5.7) that

$$\left(\frac{\partial(\overline{f}(\vec{T}))}{\partial v_i}\right)^s = \left(\sum_{j=1}^n \overline{f_j}(\vec{T})\frac{\partial T_j}{\partial v_i}\right)^s.$$

Define functions $F, G_j : \mathbb{R} \to \mathbb{R}$ $(1 \leq j \leq n)$ by

$$F(z) = \overline{f}(\vec{T})^{s(v_i|z)},$$
$$G_j(z) = T_j^{s(v_i|z)}.$$

For all $1 \leq j \leq n$ and $z \in \mathbb{R}$,

$$\begin{aligned}
F(z) &= \overline{f}(\vec{T})^{s(v_i|z)} && \text{(Definition of } F_i\text{)}\\
&= f(T_1^{s(v_i|z)}, \ldots, T_n^{s(v_i|z)}) && \text{(Definition 5.4)}\\
&= f(G_1(z), \ldots, G_n(z)), && \text{(Definition of } G_j\text{)}\\
\text{so } (*)\ F'(z) &= \textstyle\sum_j f_j(G_1(z), \ldots, G_n(z))G_j'(z) && \text{(Classic multivar. chain rule)}
\end{aligned}$$

(the hypotheses of the classic multivariable chain rule are implied by the everywhere-total-differentiability of $\overline{f}(\vec{T})$ and each T_i, by Lemma 5.15). So armed, we compute:

$$\begin{aligned}
\left(\tfrac{\partial(\overline{f}(\vec{T}))}{\partial v_i}\right)^s &= \lim_{h\to 0}\frac{\overline{f}(\vec{T})^{s(v_i|s(v_i)+h)} - \overline{f}(\vec{T})^s}{h} && \text{(Lemma 5.15)}\\
&= \lim_{h\to 0}\frac{F(s(v_i)+h) - F(s(v_i))}{h} && \text{(Def. of } F\text{)}\\
&= F'(s(v_i)) && \text{(Def. of } F'\text{)}\\
&= \textstyle\sum_j f_j(G_1(s(v_i)), \ldots, G_n(s(v_i)))G_j'(s(v_i)) && \text{(By } (*)\text{)}\\
&= \textstyle\sum_j f_j(T_1^{s(v_i|s(v_i))}, \ldots, T_n^{s(v_i|s(v_i))})G_j'(s(v_i)) && \text{(Def. of } G_j\text{)}\\
&= \textstyle\sum_j f_j(T_1^s, \ldots, T_n^s)G_j'(s(v_i)) && \text{(Lemma 5.10)}\\
&= \textstyle\sum_j f_j(T_1^s, \ldots, T_n^s)\lim_{h\to 0}\frac{G_j(s(v_i)+h) - G_j(s(v_i))}{h} && \text{(Def. of } G_j'\text{)}\\
&= \textstyle\sum_j f_j(T_1^s, \ldots, T_n^s)\lim_{h\to 0}\frac{T^{s(v_i|s(v_i)+h)} - T^{s(v_i|s(v_i))}}{h} && \text{(Def. of } G_j\text{)}\\
&= \textstyle\sum_j f_j(T_1^s, \ldots, T_n^s)\lim_{h\to 0}\frac{T^{s(v_i|s(v_i)+h)} - T^s}{h} && \text{(Lemma 5.10)}\\
&= \textstyle\sum_j f_j(T_1^s, \ldots, T_n^s)\left(\frac{\partial T_j}{\partial v_i}\right)^s && \text{(Lemma 5.15)}\\
&= \left(\textstyle\sum_{j=1}^n \overline{f_j}(\vec{T})\frac{\partial T_j}{\partial v_i}\right)^s, && \text{(Def. 5.4)}
\end{aligned}$$

as desired. □

Note that in Lemma 5.19 the assumption that $\overline{f}(\vec{T})$ is everywhere totally differentiable does not automatically imply that $T_1, \ldots, T_n$ are everywhere totally differentiable. For example, f could be the function $f(x, y) = x$ in which case $f(T_1, T_2)$ would be everywhere totally differentiable iff T_1 is everywhere totally differentiable, regardless of the behavior of T_2.

6. An Abstract Chain Rule

Recall that $\mathscr{V}$ denotes the set of all variables. Let $\mathscr{T}$ denote the set of all terms.

Definition 6.1. For any $\phi_0 : \mathscr{V} \to \mathscr{T}$, the *extension of ϕ_0 to all terms* is the function $\phi : \mathscr{T} \to \mathscr{T}$ defined by induction as follows:

(1) If T is a constant symbol then $\phi(T) = T$.
(2) If T is a variable then $\phi(T) = \phi_0(T)$.
(3) If T is $\overline{f}(S_1, \ldots, S_n)$ then $\phi(T) = \overline{f}(\phi(S_1), \ldots, \phi(S_n))$.

Lemma 6.2. *Let $\phi_0 : \mathscr{V} \to \mathscr{T}$ and let ϕ be the extension of ϕ_0 to all terms. Then:*

(1) *(The Substitution Lemma) For any assignment s, if $\phi(s)$ is the assignment defined by $\phi(s)(v) = \phi(v)^s$, then for every term T, $\phi(T)^s = T^{\phi(s)}$.*
(2) *For all terms T and U, if $T \equiv U$ then $\phi(T) \equiv \phi(U)$.*

Proof. (1) By induction on T. If T is a constant symbol or variable, the claim is trivial. Otherwise, T is $\overline{f}(U_1, \ldots, U_n)$. Then

$$\begin{aligned}
\phi(T)^s &= \overline{f}(\phi(U_1), \ldots, \phi(U_n))^s && \text{(Definition 6.1)}\\
&= f(\phi(U_1)^s, \ldots, \phi(U_n)^s) && \text{(Definition 5.4)}\\
&= f(U_1^{\phi(s)}, \ldots, U_n^{\phi(s)}) && \text{(Induction)}\\
&= T^{\phi(s)}. && \text{(Definition 5.4)}
\end{aligned}$$

(2) Assume $T \equiv U$. For any assignment s, if $\phi(s)$ is as in (1), then $T^{\phi(s)} = U^{\phi(s)}$ by Definition 5.7. Thus $\phi(T)^s = \phi(U)^s$ by (1). By arbitrariness of s, $\phi(T) \equiv \phi(U)$. □

Definition 6.3. Say $\phi_0 : \mathscr{V} \to \mathscr{T}$ *respects* d if for each variable v, $\phi_0(dv) \equiv \mathbf{d}\phi_0(v)$.

Definition 6.4. (Strong differentiability)

(1) We define the *subterms* of a term T by induction as follows. If T is a variable or constant symbol, then T is its own lone subterm. If T is $\overline{f}(U_1, \ldots, U_n)$, then the subterms of T are T itself along with the subterms of each U_i.
(2) A term T is *strongly differentiable* if every subterm of T is everywhere totally differentiable.

Thus, a term is strongly differentiable if it is built up from pieces which are everywhere totally differentiable. An example of a term which is everywhere totally differentiable but not strongly differentiable is $|x_0|^2$, which is everywhere totally differentiable despite having a subterm $|x_0|$ which is not. Note that the ordinary chain rule for $f(g(x))'$ fails when $f(x) = x^2$ and $g(x) = |x|$ (these functions fail the chain rule's hypotheses): $(|x|^2)' = 2x$, but $|x|'$ is undefined at $x = 0$. We avoid such traps in the following theorem by requiring strong differentiability.

Theorem 6.5. *(General Abstract Chain Rule) Let $\phi_0 : \mathscr{V} \to \mathscr{T}$ and assume that $\phi_0(v)$ is strongly differentiable for every variable v. Let ϕ be the extension of ϕ_0 to all terms. If T is strongly differentiable and ϕ_0 respects d, then $\mathbf{d}\phi(T) \equiv \phi(\mathbf{d}T)$.*

Proof. By induction on T. If T is a constant symbol, the theorem is trivial. If T is a variable, the theorem reduces to the statement that ϕ_0 respects d, which is one of the hypotheses. It remains to consider the case when T is $\overline{f}(\vec{T})$ where $f : \mathbb{R}^m \to \mathbb{R}$ and $\vec{T} = T_1, \ldots, T_m$ are simpler terms. Then $T_1, \ldots, T_m$ are subterms of T, so, since T is strongly differentiable, it follows that $T_1, \ldots, T_m$ are strongly differentiable. By induction, each $\mathbf{d}\phi(T_i) \equiv \phi(\mathbf{d}T_i)$. Let $\{v_1, \ldots, v_\ell\} = \mathrm{FV}(\phi(T_1)) \cup \cdots \cup \mathrm{FV}(\phi(T_m))$. For the rest of the proof, whenever S is a term and v is a variable, we will write S_v for $\frac{\partial S}{\partial v}$. Let $\overrightarrow{\phi(T)}$ denote $\phi(T_1), \ldots, \phi(T_m)$. We calculate:

$$\begin{aligned}
&\mathbf{d}\phi(\overline{f}(\vec{T})) \\
&\equiv \textstyle\sum_{i=1}^{\ell} \phi(\overline{f}(\vec{T}))_{v_i} dv_i && \text{(Lemma 5.17)} \\
&= \textstyle\sum_i \overline{f}(\overrightarrow{\phi(T)})_{v_i} dv_i && \text{(Definition 6.1)} \\
&\equiv \textstyle\sum_i \sum_{j=1}^{m} \overline{f_j}(\overrightarrow{\phi(T)}) \phi(T_j)_{v_i} dv_i && \text{(Lemma 5.19)} \\
&\equiv \textstyle\sum_j \overline{f_j}(\overrightarrow{\phi(T)}) \sum_i \phi(T_j)_{v_i} dv_i && \text{(Basic algebra)} \\
&\equiv \textstyle\sum_j \overline{f_j}(\overrightarrow{\phi(T)}) \mathbf{d}\phi(T_j) && \text{(Lemma 5.17)} \\
&\equiv \textstyle\sum_j \overline{f_j}(\overrightarrow{\phi(T)}) \phi(\mathbf{d}T_j) && \text{(Induction Hypothesis)} \\
&= \phi\left(\textstyle\sum_j \overline{f_j}(\vec{T}) \mathbf{d}T_j\right) && \text{(Definition 6.1)} \\
&\equiv \phi\left(\textstyle\sum_j \overline{f_j}(\vec{T}) \sum_{i=1}^{\ell} (T_j)_{v_i}\, dv_i\right) && \text{(Lemma 5.17)} \\
&\equiv \phi\left(\textstyle\sum_i \sum_j \overline{f_j}(\vec{T}) (T_j)_{v_i}\, dv_i\right) && \text{(Basic algebra)} \\
&\equiv \phi(\textstyle\sum_i \overline{f}(\vec{T})_{v_i} dv_i) && \text{(Lemma 5.19)} \\
&\equiv \phi(\mathbf{d}\overline{f}(\vec{T})) && \text{(Lemma 5.17)}
\end{aligned}$$

(in the last few lines, we use Lemma 6.2 part 2). □

A weakness of the familiar chain rule is that it does not iterate. The following corollary shows that the abstract chain rule does iterate.

Corollary 6.6. *For all ϕ_0, ϕ and T as in Theorem 6.5, for all $k \in \mathbb{N}$ ($k > 0$), if $\mathbf{d}^\ell T$ exists and is strongly differentiable for all $\ell < k$, then*

$$\mathbf{d}^k \phi(T) \equiv \phi(\mathbf{d}^k T).$$

Proof. By repeated applications of Theorem 6.5. □

In Sections 2 and 3 we used a special case of Theorem 6.5 which we will now formalize. Recall that a precalculus variable is one that is not of the form dv for any variable v.

Definition 6.7. (Variable substitution respecting differentials) Let v be a precalculus variable, U a term such that $\mathbf{d}^k U$ is strongly differentiable for all k. For every term T, we will define the *result of substituting U for v in T while respecting differentials*, written $T[v|U]$, as follows. First, we define $\phi_0 : \mathscr{V} \to \mathscr{T}$ so that:

1. $\phi_0(v) = U$.
2. For every $k > 0$, $\phi_0(d^k v) = \mathbf{d}^k U$.
3. For all variables w not of either of the above two forms, $\phi_0(w) = w$.

We define $T[v|U]$ to be $\phi(T)$ where ϕ is the extension of ϕ_0 to all terms (Definition 6.1).

Corollary 6.8. *(Abstract Chain Rule) Let U, v be as in Definition 6.7. If term T is strongly differentiable, then*

$$\mathbf{d}(T[v|U]) \equiv (\mathbf{d}T)[v|U].$$

Proof. If ϕ_0 is as in Definition 6.7 then evidently ϕ_0 satisfies the hypotheses of Theorem 6.5. The corollary then immediately follows from Theorem 6.5. □

Corollary 6.9. *(Iterated Abstract Chain Rule) Let v, T, U be as in Corollary 6.8. For all $k > 0$, if $\mathbf{d}^\ell T$ is strongly differentiable for all $\ell < k$, then*

$$\mathbf{d}^k(T[v|U]) \equiv (\mathbf{d}^k T)[v|U].$$

Proof. By repeated applications of Corollary 6.8. □

Acknowledgments

We gratefully acknowledge Bryan Dawson, Tevian Dray, and the reviewers for generous comments and feedback.

References

[1] Samuel Alexander. The first-order syntax of variadic functions. *Notre Dame Journal of Formal Logic*, 54(1):47–59, 2013.

[2] Samuel Alexander. A one-sentence elementary proof of the combinatorial Faà di Bruno's formula. *arXiv preprint 2206.02031*, 2022.

[3] Alex Craik. Prehistory of Faà di Bruno's formula. *The American Mathematical Monthly*, 112(2):119–130, 2005.

[4] Ronald L Graham, Donald E Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, 2nd edition, 1994.

[5] Joel David Hamkins. The differential operator $\frac{d}{dx}$ binds variables. In *Joel David Hamkins: mathematics and philosophy of the infinite (blog)*. 2012.

[6] Michael Hardy. Combinatorics of partial derivatives. *The Electronic Journal of Combinatorics*, 2006.

[7] HN Huang, SAM Marcantognini, and NJ Young. Chain rules for higher derivatives. *The Mathematical Intelligencer*, 28(2):61–69, 2006.

[8] Warren P Johnson. The curious history of Faà di Bruno's formula. *The American Mathematical Monthly*, 109(3):217–234, 2002.

[9] Tsoy-Wo Ma. Higher chain formula proved by combinatorics. *The Electronic Journal of Combinatorics*, 2009.

[10] Christina Osborne and Amelia Tebbe. A first step toward higher order chain rules in abelian functor calculus. In *Association for Women in Mathematics Research Symposium*, pages 97–119. Springer, 2017.

[11] James Stewart. *Calculus*. Brooks Cole, 8th edition, 2015.