Higher-Order, Higher-Order Automatic Differentiation

(early, incomplete draft—comments invited[*]*)*

Conal Elliott

Draft of November 27, 2019 15:39

1 Introduction

This note picks up where Elliott (2018) left off, and I assume the reader to be familiar with that paper and have it close at hand. I am circulating this follow-on in fairly rough form for early feedback, to then evolve in to a full research paper. The main new contributions are two senses of "*higher-order* automatic differentiation":

- derivatives of higher-order functions, and
- higher-order derivatives of functions, and

The former has been addressed in a recent paper (Vytiniotis et al., 2019), but in a way I find dissatisfying for a variety of reasons described in Section 8 and discussed at length with the authors.

Begin with the category of computably differentiable functions from (Elliott, 2018, Section 4.1):

$$\textbf{newtype}\ D\ a\ b = D\ (a \to b \times (a \multimap b))$$

where $a \multimap b$ is the type of linear maps from a to b. The function around which the automatic differentiation (AD) algorithm is organized simply "zips" together a function $f : a \to b$ and its derivative $\mathcal{D}\ f : a \to a \multimap b$:[1,2]

$$\begin{aligned}
&\hat{\mathcal{D}} : (a \to b) \to D\ a\ b \\
&\hat{\mathcal{D}}\ f = D\ (\lambda\, a \to (f\ a, \mathcal{D}\ f\ a)) \\
&\phantom{\hat{\mathcal{D}}\ f} = D\ (f \vartriangle \mathcal{D}\ f)
\end{aligned}$$

[*]Latest version, questions, suggestions, bugs, etc: `https://github.com/conal/hoad`. Email: conal@conal.net.

[1]This paper generally uses Haskell notation but deviates slightly by using a single colon rather than double colon for type signatures. [Experimental.]

[2]The infix operators for function types ("$\to$") and linear maps ("$\multimap$") both associate to the right and have equal, very low precedence. For instance, "$a \to a \multimap b$" means $a \to (a \multimap b)$.

Note that this definition is not computable, since $\mathcal{D}$ is not (Pour-El and Richards, 1978, 1983). The whole specification of AD is then simply that $\hat{\mathcal{D}}$ is a homomorphism with respect to a standard compositional vocabulary of functions, namely that of cartesian categories, plus a collection of numeric primitives like (uncurried) addition and multiplication, *sin* and *cos*, etc. An example of such a homomorphism equation is $\hat{\mathcal{D}}\ g \circ \hat{\mathcal{D}}\ f = \hat{\mathcal{D}}\ (g \circ f)$, in which the only unknown is the meaning of the LHS ($\circ$), i.e., sequential composition in the category D. Solving the collection of such homomorphism equations yields correct-by-construction AD.

The function $\hat{\mathcal{D}}$ is invertible, i.e., $\hat{\mathcal{D}}^{-1} \circ \hat{\mathcal{D}} = id$, where $\hat{\mathcal{D}}^{-1}$ simply drops the derivative:[3]

$$\begin{aligned}
&\hat{\mathcal{D}}^{-1} : D\ a\ b \to (a \to b)\\
&\hat{\mathcal{D}}^{-1}\ (D\ h) = exl \circ h
\end{aligned}$$

Indeed, $\hat{\mathcal{D}}^{-1}$ is a left inverse of $\hat{\mathcal{D}}$:

$$\begin{aligned}
&\hat{\mathcal{D}}^{-1}\ (\hat{\mathcal{D}}\ f)\\
&= \hat{\mathcal{D}}^{-1}\ (D\ (f \vartriangle \mathcal{D}\ f)) && \text{-- } \hat{\mathcal{D}} \text{ definition}\\
&= exl \circ (f \vartriangle \mathcal{D}\ f) && \text{-- } \hat{\mathcal{D}}^{-1} \text{ definition}\\
&= f && \text{-- cartesian law: } exl \circ (g \vartriangle h) = g
\end{aligned}$$

As defined so far, $\hat{\mathcal{D}}^{-1}$ is *not* a right inverse to $\hat{\mathcal{D}}$, since the linear map portion might not be the true derivative. We will thus *restrict* the category D to be the image of $\hat{\mathcal{D}}$, which is to say that $\hat{\mathcal{D}}$ is surjective, i.e., the derivative is correct.[4] This restriction guarantees that $\hat{\mathcal{D}}^{-1}$ is indeed a right inverse of $\hat{\mathcal{D}}$. Given $\hat{f} : D\ a\ b$ (with the mentioned restriction), there is an $f : a \to b$ such that $\hat{f} = \hat{\mathcal{D}}\ f$, so[5]

$$\begin{aligned}
&\hat{\mathcal{D}}\ (\hat{\mathcal{D}}^{-1}\ \hat{f})\\
&= \hat{\mathcal{D}}\ (\hat{\mathcal{D}}^{-1}\ (\hat{\mathcal{D}}\ f)) && \text{-- } \hat{f} = \hat{\mathcal{D}}\ f\\
&= \hat{\mathcal{D}}\ f && \text{-- } \hat{\mathcal{D}}^{-1} \circ \hat{\mathcal{D}} = id\\
&= \hat{f} && \text{-- } \hat{\mathcal{D}}\ f = \hat{f}
\end{aligned}$$

Thus, $\hat{\mathcal{D}} \circ \hat{\mathcal{D}}^{-1} = id$ as well.

AD is often described as coming in forward and reverse "modes". For many practical applications (including deep learning and other high-dimensional optimization problems), reverse mode is much more efficient than forward mode. As typically presented, reverse mode is also much more complicated, but this difference appears to be due only to unfortunate choices in how to understand and implement AD. Instead, a single, simple algorithm works for forward, reverse, and other modes. Reverse mode is distinguished only by using a different linear map representation resulting from a simple classic trick (Elliott, 2018).

This general AD algorithm is justified by three main theorems about differentiation:

[3] This paper uses "*exl*" and "*exr*" to name left and right product projections (defined on cartesian categories), rather than Haskell's (function-only) "*fst*" and "*snd*".

[4] Haskell's type system is not expressive enough to capture this restriction by itself, so the restriction will be only implied in this draft. For more rigor, one could use a language (extension) with refinement types such as Liquid Haskell [ref] or a dependently-typed language such as Agda [ref] or Idris [ref].

[5] This reasoning hold for *any* surjective function with a left inverse.

Theorem 1 (compose/"chain" rule)**.**

$$\mathcal{D}\ (g \circ f)\ a = \mathcal{D}\ g\ (f\ \ a) \circ \mathcal{D}\ f\ \ a$$

Theorem 2 (cross rule)**.**

$$\mathcal{D}\ (f \times g)\ (a, b) = \mathcal{D}\ f\ \ a \times \mathcal{D}\ g\ b$$

Theorem 3 (linear rule)**.** *For all linear functions f,*

$$\mathcal{D}\ f\ \ a = f$$

In addition to these three theorems, we need a collection of facts about the derivatives of various mathematical operations, e.g., $\hat{\mathcal{D}}\ \mathit{sin}\ x = \mathit{scale}\ (\mathit{cos}\ x)$, where $\mathit{scale} : a \to a \multimap a$ is uncurried scalar multiplication (so $\mathit{scale}\ s$ is linear for all s).

2 Some Additional Properties of Differentiation

A few additional properties of differentiation will prove useful in extending Elliott (2018) to higher-order functions and higher-order derivatives.

2.1 Linearity and invertibility

2.2 Pair-Valued Domains

One half of the *curry*/*uncurry* isomorphism involves functions of pair-valued domains. The notion of partial derivatives is helpful for differentiating such functions.[6]

Theorem 4 (proved in Section A.1)**.** *Given a function $f : a \times b \to c$,*

$$\mathcal{D}\ f\ \ (a, b) = \mathcal{D}_l\ f\ \ (a, b) \triangledown \mathcal{D}_r\ f\ \ (a, b)$$

where $\mathcal{D}_l$ and $\mathcal{D}_r$ construct the ("first" and "second", or "left" and "right") "partial derivatives" of f at (a, b), defined as follows:

$$\begin{aligned} &\mathcal{D}_l : (a \times b \to c) \to a \times b \to (a \multimap c) \\ &\mathcal{D}_l\ f\ \ (a, b) = \mathcal{D}\ (f \circ (, b))\ a \end{aligned}$$

$$\begin{aligned} &\mathcal{D}_r : (a \times b \to c) \to a \times b \to (b \multimap c) \\ &\mathcal{D}_r\ f\ \ (a, b) = \mathcal{D}\ (f \circ (a,))\ b \end{aligned}$$

The notation "$(a,)$" and "$(b,)$" refers to right and left "sections" of pairing: $(, b)\ a = (a,)\ b = (a, b)$. Equivalently,

$$\begin{aligned} &\mathcal{D}_l\ f\ \ (a, b) = \mathcal{D}\ f\ \ (a, b) \circ \mathit{inl} \\ &\mathcal{D}_r\ f\ \ (a, b) = \mathcal{D}\ f\ \ (a, b) \circ \mathit{inr} \end{aligned}$$

[6]Recall that, on linear maps, $(f \triangledown g)\ (a, b) = f\ \ a + g\ \ b$, $\mathit{inl}\ a = (a, 0)$, and $\mathit{inr}\ b = (0, b)$

Note also that $f \circ (a,) = \mathit{curry}\ f\ a$ and $f \circ (,b) = \mathit{curry}'\ f\ b$, where

$$\begin{array}{l} \mathit{curry}\ \ f\ a\ b = f\ (a,b) \\ \mathit{curry}'\ f\ b\ a = f\ (a,b) \end{array}$$

As an example of how this decomposition of $\mathcal{D}\ f$ helps construct derivatives, suppose that f is *bilinear*, which is to say that f is linear in each argument, while holding the other constant. More formally *bilinearity* of f means that $f \circ (a,)$ and $f \circ (b,)$ (equivalently, $\mathit{curry}\ f\ a$ and $\mathit{curry}'\ f\ b$) are both linear for all a and b.

Corollary 4.1. *If $f : a \times b \to c$ is bilinear then*

$$\mathcal{D}\ f\ (a,b) = f \circ (,b) \triangledown f \circ (a,)$$

Proof.

$$\begin{array}{ll} \quad \mathcal{D}\ f\ (a,b) & \\ = \mathcal{D}_l\ f\ (a,b) \triangledown \mathcal{D}_r\ f\ (a,b) & \text{-- Theorem 4} \\ = \mathcal{D}\ (f \circ (,b))\ a \triangledown \mathcal{D}\ (f \circ (a,))\ b & \text{-- } \mathcal{D}_l \text{ and } \mathcal{D}_r \text{ definitions} \\ = f \circ (,b) \triangledown f \circ (a,) & \text{-- linearity} \end{array}$$

□

For instance, the derivative of uncurried multiplication is given by the Leibniz product rule:

$$\begin{array}{l} \quad \mathcal{D}\ (\mathit{uncurry}\ (*))\ (a,b) \\ = \mathit{uncurry}\ (*) \circ (,b) \triangledown \mathit{uncurry}\ (*) \circ (a,) \\ = (*\ b) \triangledown (a\ *) \\ = \lambda\,(da,db) \to da * b + a * db \end{array}$$

More generally, consider differentiating interacts with uncurrying:

Corollary 4.2 (proved in Section A.2)**.**

$$\mathcal{D}\ (\mathit{uncurry}\ g)\ (a,b) = \mathit{at}\ b \circ \mathcal{D}\ g\ a \triangledown \mathcal{D}\ (g\ a)\ b$$

As a special case, let g be curried multiplication:

$$\begin{array}{l} \quad \mathcal{D}\ (\mathit{uncurry}\ (*)) \\ = \mathit{at}\ b \circ \mathcal{D}\ (*)\ a \triangledown \mathcal{D}\ (a*)\ b \\ = \mathit{at}\ b \circ (*) \triangledown (a*) \\ = (*\ b) \triangledown (a\ *) \end{array}$$

which agrees with the calculation above.

For cartesian closure, we'll need the derivative of another function with a pair-valued domain:

$$\begin{array}{l} \mathit{eval} : (a \to b) \times a \to b \\ \mathit{eval}\ (f,a) = f\ a \quad \text{-- on functions} \end{array}$$

(Note that *eval* is neither linear nor bilinear, so Theorem 3 and Corollary 4.1 are both inapplicable.) We'll need one more linear map operation, which is curried, reverse function application:[7]

$$\begin{aligned}&\mathit{at} : a \to (a \to b) \multimap b\\&\mathit{at}\ a\ \mathit{df} = \mathit{df}\ a\end{aligned}$$

Corollary 4.3 (proved in Section A.3)**.**

$$\mathcal{D}\ \mathit{eval}\ (f, a) = \mathit{at}\ a \triangledown \mathcal{D}\ f\ a$$

2.3 Function-Valued Codomains

It will also be useful to calculate derivatives of functions with higher-order codomains.[8] We'll need another linear map operation, which is the indexed variant of $(\vartriangle)$ (and a specialization of Haskell's *flip* function):

$$\begin{aligned}&\mathit{fork}_F : (b \to a \multimap c) \to (a \multimap b \to c)\\&\mathit{fork}_F\ h = \lambda\, \mathit{da}\ b \to h\ b\ \mathit{da}\end{aligned}$$

Theorem 5 (proved in Section A.4)**.** *Given a function* $g : a \to b \to c$*,*

$$\mathcal{D}\ g\ a = \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}\ (\mathit{at}\ b \circ g)\ a).$$

Corollary 5.1 (proved in Section A.5)**.**

$$\mathcal{D}\ (\mathit{curry}\ f)\ a = \mathit{fork}_F\ (\mathcal{D}_l\ f \circ (a,))$$

3 Cartesian Closure, first attempt

While D is a category and a *cartesian* category at that, as specified by $\hat{\mathcal{D}}$ being a cartesian functor, another question naturally arises. Can $\hat{\mathcal{D}}$ also be a cartesian *closed* functor? In other words, are there definitions of *eval*, *curry*, and *uncurry* on D such that

$$\begin{aligned}&\mathit{curry}\ (\hat{\mathcal{D}}\ f) = \hat{\mathcal{D}}\ (\mathit{curry}\ f)\\&\mathit{uncurry}\ (\hat{\mathcal{D}}\ g) = \hat{\mathcal{D}}\ (\mathit{uncurry}\ g)\\&\mathit{eval} = \hat{\mathcal{D}}\ \mathit{eval}\end{aligned}$$

As usual, we'll want to solve each homomorphism equation for its single unknown, which is a categorical operation on D (on the LHS).

[7]Linearity of *at a* follows from the usual definition of addition and scaling on functions.

[8][The previous section and this one provide "adjoint" techniques in a sense that currying is an adjunction from functions from products to functions to functions. Is there something else interesting to say here?]

3.1 Curry

Start with *curry*, simplifying the LHS:

$$\begin{aligned} & \mathit{curry}\ (\hat{\mathcal{D}}\ f) \\ = {} & \mathit{curry}\ (D\ (f \vartriangle \mathcal{D}\ f)) \quad \text{-- } \hat{\mathcal{D}} \text{ definition} \end{aligned}$$

Then the RHS:

$$\begin{aligned} & \hat{\mathcal{D}}\ (\mathit{curry}\ f) \\ = {} & D\ (\mathit{curry}\ f \vartriangle \mathcal{D}\ (\mathit{curry}\ f)) && \text{-- } \hat{\mathcal{D}} \text{ definition} \\ = {} & D\ (\lambda\, a \rightarrow (\mathit{curry}\ f\ \ a, \mathcal{D}\ (\mathit{curry}\ f)\ \ a)) && \text{-- } (\vartriangle) \text{ on functions} \\ = {} & D\ (\lambda\, a \rightarrow ((\lambda\, b \rightarrow f\ \ (a,b)), \mathit{fork}_F\ (\mathcal{D}_l\ f \circ (a,)))) && \text{-- } \mathit{curry} \text{ and } (a,)\text{; Corollary 5.1} \\ = {} & D\ (\lambda\, a \rightarrow ((\lambda\, b \rightarrow f\ \ (a,b)), \mathit{fork}_F\ (\lambda\, b \rightarrow \mathcal{D}_l\ f\ \ (a,b)))) && \text{-- } (\circ) \text{ on functions} \\ = {} & D\ (\lambda\, a \rightarrow ((\lambda\, b \rightarrow f\ \ (a,b)), \mathit{fork}_F\ (\lambda\, b \rightarrow \mathcal{D}\ f\ \ (a,b) \circ \mathit{inl}))) && \text{-- Section A.1} \end{aligned}$$

The last form uses f and $\mathcal{D}\ f$, which can be extracted from $\hat{\mathcal{D}}\ f = D\ (f \vartriangle \mathcal{D}\ f)$: Thus a sufficient condition for our homomorphic specification ($\mathit{curry}\ (\hat{\mathcal{D}}\ f) = \hat{\mathcal{D}}\ (\mathit{curry}\ f)$) is

$$\begin{aligned} & \mathit{curry}\ (D\ \mathit{ff}') = D\ (\lambda\, a \rightarrow ((\lambda\, b \rightarrow f\ \ (a,b)), \mathit{fork}_F\ (\lambda\, b \rightarrow f'\ \ (a,b) \circ \mathit{inl}))) \\ & \quad \textbf{where}\ (f, f') = \mathit{fork}^{-1}\ \mathit{ff}' \end{aligned}$$

The fork^{-1} function is half of an isomorphism that holds for all cartesian categories:

$$\begin{aligned} & \mathit{fork} : \mathit{Cartesian}\ k \Rightarrow (a \mathbin{`k`} c) \times (a \mathbin{`k`} d) \rightarrow (a \mathbin{`k`} (c \times d)) \\ & \mathit{fork} = \mathit{uncurry}\ (\vartriangle) \end{aligned}$$

$$\begin{aligned} & \mathit{fork}^{-1} : \mathit{Cartesian}\ k \Rightarrow (a \mathbin{`k`} (c \times d)) \rightarrow (a \mathbin{`k`} c) \times (a \mathbin{`k`} d) \\ & \mathit{fork}^{-1}\ h = (\mathit{exl} \circ h, \mathit{exr} \circ h) \end{aligned}$$

Lemma 6. *The pair of functions fork and fork^{-1} form a linear isomorphism.* (Proof: Exercise.)

Another such linear isomorphism can be found in cocartesian categories. The following types are specialized to biproduct categories (such as linear maps):

$$\begin{aligned} & \mathit{join} : \mathit{Cocartesian}\ k \Rightarrow (a \mathbin{`k`} c) \times (a \mathbin{`k`} d) \rightarrow (a \mathbin{`k`} (c \times d)) \\ & \mathit{join} = \mathit{uncurry}\ (\triangledown) \end{aligned}$$

$$\begin{aligned} & \mathit{join}^{-1} : \mathit{Cocartesian}\ k \Rightarrow (a \mathbin{`k`} (c \times d)) \rightarrow (a \mathbin{`k`} c) \times (a \mathbin{`k`} d) \\ & \mathit{join}^{-1}\ h = (h \circ \mathit{inl}, h \circ \mathit{inr}) \end{aligned}$$

Lemma 7. *The pair of functions join and join^{-1} form a linear isomorphism.* (Proof: Exercise.)

Another useful operation is the *uncurried* version of the monoidal $(\times)$:

$$\begin{aligned} & \mathit{cross} : \mathit{Monoidal}\ k \Rightarrow (a \mathbin{`k`} c) \times (b \mathbin{`k`} d) \rightarrow ((a \times b) \mathbin{`k`} (c \times d)) \\ & \mathit{cross} = \mathit{uncurry}\ (\times) \end{aligned}$$

Lemma 8. *The cross function is linear.* (Proof: Exercise.)

These two isomorphism pairs were used by Elliott (2018) to construct a correct-by-construction implementation of reverse-mode AD, by merely altering the representation of linear maps used in the simple, general AD algorithm.

Although $\mathit{fork}/\mathit{fork}^{-1}$ form an isomorphism and hence preserve information, fork^{-1} can result in a loss of efficiency, due to computation that can be (and often is) in common to a function f and its derivative $\mathcal{D}\ f$. Indeed, the definition of $\mathit{fork}^{-1}\ h$ above shows that h gets replicated. It's unclear how to avoid this redundancy problem in practice with currying when D is used to represent computably differentiable functions. Personal experience with compiling to categories Elliott (2017) suggests that most uses of *curry* generated during translation from the λ calculus (e.g., Haskell) are in fact transformed away at compile time using various equational CCC laws. Still, it does seem an important question to explore.

Intriguingly, curried functions can also help eliminate redundant computation suggested by uncurried counterparts functions. Given a function $g : a \to b \to c$, it is sometimes convenient to "partially apply" g to an argument u and then apply the resulting $g\ u : b \to c$ to many different $v : b$. In some cases, a considerable amount of work can be done based solely on u, saving residual work to be done for different b values. In such situations, *uncurry* g loses this performance advantage.

3.2 Uncurry

Next, let's tackle *uncurry*, whose defining homomorphism is

$$\mathit{uncurry}\ (\hat{\mathcal{D}}\ g) = \hat{\mathcal{D}}\ (\mathit{uncurry}\ g)$$

Simplify the LHS:

$$\begin{aligned}
&\mathit{uncurry}\ (\hat{\mathcal{D}}\ g) \\
=\ &\mathit{uncurry}\ (D\ (g \vartriangle \mathcal{D}\ g)) \quad \text{-- } \hat{\mathcal{D}} \text{ definition}
\end{aligned}$$

Then the RHS:

$$\begin{aligned}
&\hat{\mathcal{D}}\ (\mathit{uncurry}\ g) \\
=\ &D\ (\mathit{uncurry}\ g \vartriangle \mathcal{D}\ (\mathit{uncurry}\ g)) && \text{-- } \hat{\mathcal{D}} \text{ definition} \\
=\ &D\ (\lambda\,(a,b) \to (\mathit{uncurry}\ g\ (a,b), \mathcal{D}\ (\mathit{uncurry}\ g)\ (a,b)) && \text{-- } (\vartriangle) \text{ definition} \\
=\ &D\ (\lambda\,(a,b) \to (g\ a\ b, \mathcal{D}\ (\mathit{uncurry}\ g)\ (a,b)) && \text{-- } \mathit{uncurry} \text{ on functions} \\
=\ &D\ (\lambda\,(a,b) \to (g\ a\ b, \mathit{at}\ b \circ \mathcal{D}\ g\ a \triangledown \mathcal{D}\ (g\ a)\ b)) && \text{-- Corollary 4.2}
\end{aligned}$$

Now we have a problem with solving the defining homomorphism above. Although we can extract g and $\mathcal{D}\ g$ from $\hat{\mathcal{D}}\ g$, we cannot extract $\mathcal{D}\ (g\ a)$. Or rather we can, but not computably.

3.3 Eval

We don't need to work out both *uncurry* and *eval*, since each can be defined in terms of the other:

$$\begin{aligned} \mathit{eval} &= \mathit{uncurry}\ \mathit{id} \\ \mathit{uncurry}\ g &= \mathit{eval} \circ (g \times \mathit{id}) \\ &= \mathit{eval} \circ \mathit{first}\ g \end{aligned}$$

Since we got stuck on *uncurry*, let's try *eval* as well to see if we learn anything new.

The corresponding homomorphism equation has a particularly simple form:

$$\mathit{eval} = \hat{\mathcal{D}}\ \mathit{eval}$$

It might appear that we are done at the start, taking the equation to be a definition for *eval*. Recall, however, that $\hat{\mathcal{D}}$ is noncomputable, being defined via $\mathcal{D}$ (differentiation itself). Simplifying the RHS,

$$\begin{array}{ll} \quad \hat{\mathcal{D}}\ \mathit{eval} & \\ = D\ (\mathit{eval} \vartriangle \mathcal{D}\ \mathit{eval}) & \text{-- } \hat{\mathcal{D}} \text{ definition} \\ = D\ (\lambda\,(f,a) \rightarrow (\mathit{eval}\ (f,a), \mathcal{D}\ \mathit{eval}\ (f,a))) & \text{-- } (\vartriangle) \text{ on functions} \\ = D\ (\lambda\,(f,a) \rightarrow (f\ \ a, \mathcal{D}\ \mathit{eval}\ (f,a))) & \text{-- } \mathit{eval} \text{ on functions} \\ = D\ (\lambda\,(f,a) \rightarrow (f\ \ a, \mathit{at}\ a \triangledown \mathcal{D}\ f\ \ a)) & \text{-- Corollary 4.3} \end{array}$$

As with uncurrying (Section 3.2), the final form is well-defined but is not a computable recipe, leaving us in a pickle. Next, let's look for some wiggle room.

4 Object Mapping

The choice of category-associated products and exponentials is a degree of freedom not exercised in the development of AD in Elliott (2018) (or above) and is tied closely to another such choice available in the general notion of *cartesian closed functor* in category theory. In general, a functor has two aspects:

- a mapping from arrows to arrows, and
- a mapping from objects to objects.

The functor $\hat{\mathcal{D}}$ defined (noncomputably) above implicitly chooses an *identity object mapping*, as evident in its type signature $\hat{\mathcal{D}} : (a \rightarrow b) \rightarrow D\ a\ b$. The type of $\hat{\mathcal{D}}$ plus the requirement that it be a cartesian *closed* functor implies that the object mapping aspect of $\hat{\mathcal{D}}$ is the identity. More generally, we can define an object mapping $O : \mathit{Type} \rightarrow \mathit{Type}$ for a new functor $\mathcal{D}_o$:[9]

$$\mathcal{D}_o : (a \rightarrow b) \rightarrow D\ (O\ \ a)\ (O\ \ b)$$

Each cartesian category k has its own notion of categorical product $a \times_k b$ (satisfying a universality property), and similarly for cocartesian categories (with categorical products and coproducts coinciding for biproduct categories). Likewise, each cartesian *closed* category k has its own notion of *exponential* objects $a \Rightarrow_k b$.

[9] [Experiment with different notation for $O\ a$, e.g., "$\bar{a}$".]

The generalized interface for cartesian closed categories with per-category exponentials is as follows:[10]

$$
\begin{array}{l}
\textbf{class } \mathit{Cartesian}\ k \Rightarrow \mathit{CartesianClosed}\ k \textbf{ where} \\
\quad \textbf{type } (\Rightarrow_k) : \mathit{Type} \to \mathit{Type} \to \mathit{Type} \\
\quad \mathit{curry} \quad : ((a \times_k b)\ \text{`}k\text{`}\ c) \to (a\ \text{`}k\text{`}\ (b \Rightarrow_k c)) \\
\quad \mathit{uncurry} : (a\ \text{`}k\text{`}\ (b \Rightarrow_k c)) \to ((a \times_k b)\ \text{`}k\text{`}\ c) \\
\quad \mathit{eval} \quad\ : ((a \Rightarrow_k b) \times_k a)\ \text{`}k\text{`}\ b
\end{array}
$$

where $a \Rightarrow_k b$ is a type of "exponential objects" (first class functions/arrows) from a to b for the category k.

The property of being a closed cartesian functor requires O to preserve categorical products and exponentials, i.e.,

$$
\begin{array}{l}
O\ (a\ \times\ b) = O\ a \times_D\ O\ b \\
O\ (a \to b) = O\ a \Rightarrow_D O\ b
\end{array}
$$

The usual notion of cartesian products are working fine, so we'll continue to choose $a \times_D b = a \times b$. While $\hat{\mathcal{D}}$ being a closed cartesian functor (CCF) from $(\to)$ to D implies an noncomputable *eval* and *uncurry* (Section 3.2 and Section 3.3), our goal is to define $\Rightarrow_D$ and $\mathcal{D}_o$ such that $\mathcal{D}_o$ is a CCF with computable operations.

Consider again the homomorphic specification for *curry* (part of the CCF definition): $\mathit{eval} = \mathcal{D}_o\ \mathit{eval}$. The RHS *eval* (on functions) has type $(a \to b) \times a \to b$, while the RHS *eval* (on D) has type

$$
\begin{array}{l}
\quad D\ (O\ ((a \to b) \times a))\ (O\ b) \\
= D\ (O\ (a \to b) \times O\ a)\ (O\ b) \\
= D\ ((O\ a \Rightarrow_D O\ b) \times O\ a)\ (O\ b)
\end{array}
$$

The difficulty with our attempt at *eval* in Section 3.3 was that we were given a function f, but we also needed its derivative $\mathcal{D}\ f$. Similarly, with *uncurry* in Section 3.2, we were given $g : a \to b \to c$, and we needed not only $g\ a$ but also $\mathcal{D}\ (g\ a)$. In both cases the exponential object was a function, but we also needed its (computable) derivative.

This analysis suggests that we include a derivative in the exponential object, simply by choosing $\Rightarrow_D$ to be D itself. Additionally, map scalars to themselves and cartesian products to cartesian products:

$$
\begin{array}{l}
O\ \mathbb{R} = \mathbb{R} \\
O\ (a\ \times\ b) = O\ a \times_D\ O\ b = O\ a \times O\ b \\
O\ (a \to b) = O\ a \Rightarrow_D O\ b = D\ (O\ a)\ (O\ b)
\end{array}
$$

We will need to convert between a and $O\ a$, which we can do with a family of *linear isomorphisms*[11] indexed by a:[12]

[10]These operations support higher-order programming and arise during translation from a typed lambda calculus (e.g., Haskell) to categorical vocabulary (Elliott, 2017).

[11]The implicit requirements for all *HasO* instances are thus that $o \circ o^{-1} = id$, $o^{-1} \circ o = id$, and *to* and o^{-1} are linear.

[12][It may be more elegant to combine the functions o and o^{-1} into a single *isomorphism*.]

$$\begin{array}{l}
\textbf{class}\ \mathit{HasO}\ t\ \textbf{where} \\
\quad \textbf{type}\ O\ t \\
\quad o \quad : t \to O\ t \\
\quad o^{-1} : O\ t \to t
\end{array}$$

For scalar types a and the unit type, $O\ a = a$, the isomorphism is trivial:

$$\begin{array}{l}
\textbf{instance}\ \mathit{HasO}\ \mathbb{R}\ \textbf{where} \\
\quad \textbf{type}\ O\ \mathbb{R} = \mathbb{R} \\
\quad o \quad = \mathit{id} \\
\quad o^{-1} = \mathit{id}
\end{array}$$

$$\begin{array}{l}
\textbf{instance}\ \mathit{HasO}\ ()\ \textbf{where} \\
\quad \textbf{type}\ O\ () = () \\
\quad o \quad = \mathit{id} \\
\quad o^{-1} = \mathit{id}
\end{array}$$

For products, convert components independently:[13]

$$\begin{array}{l}
\textbf{instance}\ (\mathit{HasO}\ a, \mathit{HasO}\ b) \Rightarrow \mathit{HasO}\ (a \times b)\ \textbf{where} \\
\quad \textbf{type}\ O\ (a \times b) = O\ a \times O\ b \\
\quad o \quad = o \quad \times o \\
\quad o^{-1} = o^{-1} \times o^{-1}
\end{array}$$

The new functor $\mathcal{D}_o$ converts its given $a \to b$ to $O\ a \to O\ b$ and then applies the $\hat{\mathcal{D}}$ functor:[14]

$$\begin{array}{l}
(\Rightarrow) : (p' \to p) \to (q \to q') \to ((p \to q) \to (p' \to q')) \\
f \Rightarrow h = \lambda\, g \to h \circ g \circ f
\end{array}$$

$$\begin{array}{l}
\mathit{wrap}_o : (a \to b) \to (O\ a \to O\ b) \\
\mathit{wrap}_o = o^{-1} \Rightarrow o
\end{array}$$

$$\begin{array}{l}
\mathit{wrap}_o^{-1} : (O\ a \to O\ b) \to (a \to b) \\
\mathit{wrap}_o^{-1} = o \Rightarrow o^{-1}
\end{array}$$

$$\begin{array}{l}
\mathcal{D}_o : (a \to b) \to D\ (O\ a)\ (O\ b) \\
\mathcal{D}_o = \hat{\mathcal{D}} \circ \mathit{wrap}_o
\end{array}$$

$$\begin{array}{l}
\mathcal{D}_o^{-1} : D\ (O\ a)\ (O\ b) \to (a \to b) \\
\mathcal{D}_o^{-1} = \mathit{wrap}_o^{-1} \circ \hat{\mathcal{D}}^{-1}
\end{array}$$

Theorem 9 (proved in Section A.6)**.** *wrap_o and wrap_o^{-1} form a linear isomorphism.*

Theorem 10 (proved in Section A.7)**.** *$\mathcal{D}_o$ and $\mathcal{D}_o^{-1}$ form a linear isomorphism.*

[13] Recall that $(f \times g)\ (a, b) = (f\ a, g\ b)$, so $o\ (a, b) = (o \times o)\ (a, b) = (o\ a, o\ b)$, and similarly for o^{-1}.

[14] [Consider dropping the $(\Rightarrow)$ definition and uses here.]

Theorem 11 (proved in Section A.8)**.** *$wrap_o$ is a cartesian functor.*

[To do: reconsider theorems vs lemmas vs corollaries. I think more lemmas.]

The cartesian category operations already defined on D (Elliott, 2018) are solutions to homomorphism equations saying that $\hat{\mathcal{D}}$ is a cartesian functor. Thanks to the simple, regular structure of o and o^{-1},

Theorem 12. *$\mathcal{D}_o$ is a cartesian functor.*

Proof: $\hat{\mathcal{D}}$ is a cartesian functor (Elliott, 2018), as is $wrap_o$ (Theorem 11), so $\mathcal{D}_o = \hat{\mathcal{D}} \circ wrap_o$ is also.

What about exponentials and cartesian *closure*? As mentioned above, $O\ (a \to b) = O\ a \Rightarrow_D O\ b = D\ (O\ a)\ (O\ b)$, which suggests using $\mathcal{D}_o$ and $\mathcal{D}_o^{-1}$ for o and o^{-1}:

$$\begin{array}{l} \textbf{instance}\ (HasO\ a, HasO\ b) \Rightarrow HasO\ (a \to b)\ \textbf{where} \\ \quad \textbf{type}\ O\ (a \to b) = D\ (O\ a)\ (O\ b) \\ \quad o \quad\ = \mathcal{D}_o \\ \quad o^{-1} = \mathcal{D}_o^{-1} \end{array}$$

A useful consequence:

Theorem 13 (proved in Section A.9)**.**

$$wrap_o\ (curry\ f) = \hat{\mathcal{D}} \circ curry\ (wrap_o\ f)$$

Corollary 13.1.

$$curry\ (wrap_o\ f) = \hat{\mathcal{D}}^{-1} \circ wrap_o\ (curry\ f)$$

Proof. Left-compose $\hat{\mathcal{D}}^{-1}$ with both sides of Theorem 13; then simplify and reverse the resulting equation. □

Let's now try to solve the CCF equations for $\mathcal{D}_o$. This time begin with *eval*:

Theorem 14 (proved in Section A.10)**.** *With the following (effective) definition of eval on D, $eval = \mathcal{D}_o\ eval$:*

$$eval = D\ (\lambda\,(D\ h, a) \to \textbf{let}\ (b, f') = h\ a\ \textbf{in}\ (b, at\ a \circ \hat{\mathcal{D}}^{-1} \triangledown f'))$$

For *uncurry*, use the standard definition $uncurry\ g = eval \circ first\ g$.

The definition of *curry* in Section 3.1 worked fine, but we'll need to check again, as we did with the cartesian category operations (Theorem 12). The homomorphism equation is $curry\ (\mathcal{D}_o\ f) = \mathcal{D}_o\ (curry\ f)$, to be solved for the unknown LHS *curry* (on D), with $f : a \times b \to c$. First let $f_o = wrap_o\ f$. Simplify the LHS:

$$\begin{array}{lll} & curry\ (\mathcal{D}_o\ f) & \\ = & curry\ (\hat{\mathcal{D}}\ (wrap_o\ f)) & \text{-- } \mathcal{D}_o \text{ definition} \\ = & curry\ (\hat{\mathcal{D}}\ f_o) & \text{-- } f_o \text{ definition} \\ = & curry\ (D\ (f_o \vartriangle \mathcal{D}\ f_o)) & \text{-- } \hat{\mathcal{D}} \text{ definition} \end{array}$$

Then the RHS:[15]

[15][State, prove, and use a lemma about $\hat{\mathcal{D}}\ (g \circ f)\ a$ for linear g and another for linear f. Maybe also $\mathcal{D}_o\ (g \circ f)\ a$ for linear g or f.]

Theorem 15 (proved in Section A.11)**.**

$$\begin{aligned}\mathcal{D}_o\ (\mathit{curry}\ f) = D\ (\lambda\, a \rightarrow (D\ (\lambda\, b \rightarrow (f_o\ (a,b), \mathcal{D}_r\ f_o\ (a,b)))\\ ,\lambda\, da \rightarrow D\ (\lambda\, b \rightarrow (\mathcal{D}_l\ f_o\ (a,b)\ da, \mathit{at}\ da \circ \mathcal{D}_r\ (\mathcal{D}_l\ f_o)\ (a,b)))))\end{aligned}$$

where $f_o = \mathit{wrap}_o\, f$.

The RHS uses $f_o\ (a,b)$ and $\mathcal{D}\ f_o\ (a,b)$ (via its components $\mathcal{D}_l\ f_o\ (a,b)$ and $\mathcal{D}_r\ f_o\ (a,b)$), but it also uses a *second* partial derivative $\mathcal{D}_r\ (\mathcal{D}_l\ f_o)\ (a,b)$, which is not available from the *curry* argument $D\ (f_o \vartriangle \mathcal{D}\ f_o)$.

5 Where Are We?

Let's now reflect on what we've learned so far:

- The cartesian functor (CF) $\hat{\mathcal{D}} : (a \rightarrow b) \rightarrow D\ a\ b$ also forms a cartesian *closed* functor (CCF) with suitable definitions of *curry*, *uncurry*, and *eval*, but not computably (Section 3). More specifically, *curry* is computable, but *uncurry* and *eval* are not, since they need to synthesize derivatives of regular computable functions.

- General categorical functors can remap objects (here, types) as well as morphisms (here, functions). Exploiting this degree of freedom, define $\mathcal{D}_o : (a \rightarrow b) \rightarrow D\ (O\ a)\ (O\ b)$, where $O : \mathit{Type} \rightarrow \mathit{Type}$ replaces regular functions with computably differentiable functions, i.e., $O\ (u \rightarrow v) = D\ (O\ u)\ (O\ v)$. This new function is defined in terms of the old one, $\mathcal{D}_o = \hat{\mathcal{D}} \circ \mathit{wrap}_o$, and indeed $\mathcal{D}_o$ is a CF as well. In the absence of higher-order functions, O is the identity mapping, and $\mathcal{D}_o$ coincides with $\hat{\mathcal{D}}$.

- Computably satisfying the required homomorphism properties of $\mathcal{D}_o$ for *uncurry* and *eval* becomes easy, since the operations are *given* the required derivatives rather than having to synthesize them. Unfortunately, now *curry* becomes noncomputable because it has to synthesize partial *second* derivatives.

6 Higher-Order Derivatives

Where can we go from here? An obvious next step is to add second order derivatives to the representation of computably differentiable functions. It seem likely, however, that the CCF specification would reveal that *curry* needs at least third order derivatives, and so on. In other words, differentiation of higher-order functions requires all higher-order derivatives of functions.

In order to construct higher-order derivatives, it will help to examine the linearity properties of our familiar categorical vocabulary, which turns out to be mostly linear with just a bit of bilinearity. As noted in Elliott (2018), the categorical operation *id*; the cartesian operations *exl*, *exr*, *dup*; and the cocartesian operations *inl*, *inr*, and *jam* are all linear. Theorem 6 and Theorem 7 have already noted that the functions *fork* and *join* (uncurried versions of $(\vartriangle)$ and $(\triangledown)$ (Section 3.1) are linear (as well as isomorphisms). Next, let *comp* be uncurried composition:[16]

$$\begin{array}{l} comp : Category\ k \Rightarrow (b\ `k`\ c) \times (a\ `k`\ b) \to (a\ `k`\ c) \\ comp = uncurry\ (\circ) \end{array}$$

Lemma 16 (proved in Section A.12)**.** *comp is bilinear.*

Lemma 17 (proved in Section A.13)**.** *Given any bilinear function h:*

a. $curry\ h\ a$ is linear for all a.

b. $curry'\ h\ b$ is linear for all b.

c. $curry\ h$ and $curry'\ h$ are linear.

d. $\mathcal{D}\ h$ is linear.

Corollary 17.1. *On linear maps,*

a. $(g\ \circ)$ is linear for all g.

b. $(\circ\ f)$ is linear for all f.

c. $(\circ)$ and $flip\ (\circ)$ are linear.

d. $\mathcal{D}\ comp$ is linear.

These properties will help re-express Theorem 1 and Theorem 2 and related facts in a form more amenable to constructing higher derivatives.

Lemma 18 (proved in Section A.14)**.**

a. $\mathcal{D}\ (g \circ f) = comp \circ (\mathcal{D}\ g \circ f \vartriangle \mathcal{D}\ f)$.

b. $\mathcal{D}\ (f \times g) = cross \circ (\mathcal{D}\ f \times \mathcal{D}\ g)$.

c. $\mathcal{D}\ (f \vartriangle g) = fork \circ (\mathcal{D}\ f \vartriangle \mathcal{D}\ g)$.

d. For a linear *function f, $\mathcal{D}\ f = const\ f$.*

e. For any function $f : a \times b \to c$, $\mathcal{D}\ f = join \circ (\mathcal{D}_l\ f \times \mathcal{D}_r\ f)$.

f. For a bilinear *function $f : a \times b \to c$, $\mathcal{D}\ f = join \circ (curry'\ f \times curry\ f) \circ swap$.*

g. On linear maps, $\mathcal{D}\ comp = join \circ (flip\ (\circ) \times (\circ)) \circ swap$.

[16] [Maybe define *comp* only for linear maps.]

Let us now consider the task of constructing *all* orders of derivatives. The D category encapsulates a function f and its first derivative, i.e., the zeroth and first derivatives of f, which we might write as "$\hat{\mathcal{D}}\ f = \mathcal{D}^0\ f \vartriangle \mathcal{D}^1\ f$". Our new category will encapsulate *all* derivatives of f, i.e.,[17]

$$\mathcal{D}^*\ f = \mathcal{D}^0\ f \vartriangle \mathcal{D}^1\ f \vartriangle \mathcal{D}^2\ f \vartriangle \cdots$$

where

$$\begin{aligned} &\mathcal{D}^0 \quad f = f \\ &\mathcal{D}^{n+1}\ f = \mathcal{D}^n\ (\mathcal{D}\ f) \end{aligned}$$

Then

$$\begin{aligned} &\ \ \mathcal{D}^*\ f \\ &= \mathcal{D}^0\ f \vartriangle \mathcal{D}^1\ f \vartriangle \mathcal{D}^2\ f \vartriangle \mathcal{D}^3\ f \vartriangle \cdots \\ &= f \vartriangle \mathcal{D}^1\ f \vartriangle \mathcal{D}^2\ f \vartriangle \mathcal{D}^3\ f \vartriangle \cdots \\ &= f \vartriangle \mathcal{D}^0\ (\mathcal{D}\ f) \vartriangle \mathcal{D}^1\ (\mathcal{D}\ f) \vartriangle \mathcal{D}^2\ (\mathcal{D}\ f) \vartriangle \cdots \\ &= f \vartriangle \mathcal{D}^*\ (\mathcal{D}\ f) \end{aligned}$$

which we can take as a recursive definition of $\mathcal{D}^*$. Define a corresponding type of infinitely differentiable functions:[18]

$$\textbf{type}\ D^*\ a\ b = a \to T\ a\ b$$

$$\textbf{type}\ T\ a\ b = b \times T\ a\ (a \multimap b)$$

$$\begin{aligned} &\mathcal{D}^* : (a \to b) \to D^*\ a\ b \\ &\mathcal{D}^*\ f = f \vartriangle \mathcal{D}^*\ (\mathcal{D}\ f) \end{aligned}$$

We will want to find cartesian category operations for D^* such that $\mathcal{D}^*$ is a cartesian functor (CF), which will be coinductively assumed at several points below.

Start with the constant-zero function[19]: $\mathit{zero} : a \to b$:

$$\begin{aligned} &\ \ \mathcal{D}^*\ \mathit{zero} \\ &= \mathit{zero} \vartriangle \mathcal{D}^*\ (\mathcal{D}\ \mathit{zero}) && \text{-- } \mathcal{D}^* \text{ definition} \\ &= \mathit{zero} \vartriangle \mathcal{D}^*\ \mathit{zero} && \text{-- } \mathcal{D}\ \mathit{zero} = \mathit{const}\ \mathit{zero} = \mathit{zero} \\ &= \mathit{zero} \vartriangle \mathit{zero} && \text{-- coinduction} \\ &= \mathit{zero} && \text{-- Zero on pairs} \end{aligned}$$

Then constant functions more generally:

$$\begin{aligned} &\ \ \mathcal{D}^*\ (\mathit{const}\ b) \\ &= \mathit{const}\ b \vartriangle \mathcal{D}^*\ (\mathcal{D}\ (\mathit{const}\ b)) && \text{-- } \mathcal{D}^* \text{ definition} \end{aligned}$$

[17] Take $\vartriangle$ to be *right*-associative.
[18] For notational simplicity, we'll drop the **newtype** isomorphisms.
[19] As usual, types are restricted to vector spaces over a common field, which we can take to be $\mathbb{R}$

$$\begin{aligned}
&= \mathit{const}\ b \vartriangle \mathcal{D}^*\ \mathit{zero} && \text{-- } \mathcal{D}\ (\mathit{const}\ b) = \mathit{zero}\\
&= \mathit{const}\ b \vartriangle \mathit{zero} && \text{-- above}
\end{aligned}$$

Next, linear functions f:

$$\begin{aligned}
&\mathcal{D}^*\ f\\
&= f \vartriangle \mathcal{D}^*\ (\mathcal{D}\ f) && \text{-- } \mathcal{D}^* \text{ definition}\\
&= f \vartriangle \mathcal{D}^*\ (\mathit{const}\ f) && \text{-- } f \text{ linearity}\\
&= f \vartriangle \mathit{const}\ f \vartriangle \mathit{zero} && \text{-- above}
\end{aligned}$$

We will have several uses of this formula, so name it:

$$\begin{aligned}
&\mathit{linear} : (a \multimap b) \to D^*\ a\ b\\
&\mathit{linear}\ f = f \vartriangle \mathit{const}\ f \vartriangle \mathit{zero}
\end{aligned}$$

For instance, the following definitions of *id*, *exl* and *exr* satisfy the associated homomorphism (cartesian functor) properties:

$$\begin{aligned}
\mathit{id} &= \mathit{linear}\ \mathit{id}\\
\mathit{exl} &= \mathit{linear}\ \mathit{exl}\\
\mathit{exr} &= \mathit{linear}\ \mathit{exr}
\end{aligned}$$

Next, uncurried linear map composition:

$$\begin{aligned}
&\mathcal{D}^*\ \mathit{comp}\\
&= \mathit{comp} \vartriangle \mathcal{D}^*\ (\mathcal{D}\ \mathit{comp}) && \text{-- } \mathcal{D}^* \text{ definition}\\
&= \mathit{comp} \vartriangle \mathit{linear}\ (\mathcal{D}\ \mathit{comp}) && \text{-- derivative of bilinear is linear}\\
&= \mathit{comp} \vartriangle \mathit{linear}\ (\mathit{join} \circ (\mathit{flip}\ (\circ) \times (\circ)) \circ \mathit{swap}) && \text{-- Theorem 18-g}
\end{aligned}$$

Name $\mathcal{D}^*$ *comp* for future use:

$$\mathit{comp}' = \mathit{comp} \vartriangle \mathit{linear}\ (\mathit{join} \circ (\mathit{flip}\ (\circ) \times (\circ)) \circ \mathit{swap})$$

Then function compositions:

$$\begin{aligned}
&\mathcal{D}^*\ (g \circ f)\\
&= g \circ f \vartriangle \mathcal{D}^*\ (\mathcal{D}\ (g \circ f)) && \text{-- } \mathcal{D}^* \text{ definition}\\
&= g \circ f \vartriangle \mathcal{D}^*\ (\mathit{comp} \circ (\mathcal{D}\ g \circ f \vartriangle \mathcal{D}\ f)) && \text{-- Theorem 18-a}\\
&= g \circ f \vartriangle \mathcal{D}^*\ \mathit{comp} \circ \mathcal{D}^*\ (\mathcal{D}\ g) \circ \mathcal{D}^*\ f \vartriangle \mathcal{D}^*\ (\mathcal{D}\ f)) && \text{-- coinduction}\\
&= g \circ f \vartriangle \mathit{comp}' \circ \mathcal{D}^*\ (\mathcal{D}\ g) \circ \mathcal{D}^*\ f \vartriangle \mathcal{D}^*\ (\mathcal{D}\ f)) && \text{-- above}
\end{aligned}$$

Note that all of the components here (g, f $\mathcal{D}^*\ (\mathcal{D}\ g)$, $\mathcal{D}^*\ f$, and $\mathcal{D}^*\ (\mathcal{D}\ f)$) are available in $\mathcal{D}^*\ g$ and $\mathcal{D}^*\ f$, so we have a computable recipe for $(\circ)$ on D^*. [Fill in the details.]

Finally, $f \vartriangle g$:

$$\begin{aligned}
&\mathcal{D}^*\ (f \vartriangle g)\\
&= (f \vartriangle g) \vartriangle \mathcal{D}^*\ (\mathcal{D}\ (f \vartriangle g)) && \text{-- } \mathcal{D}^* \text{ definition}\\
&= (f \vartriangle g) \vartriangle \mathcal{D}^*\ (\mathit{fork} \circ (\mathcal{D}\ f \vartriangle \mathcal{D}\ g)) && \text{-- Theorem 18-c}
\end{aligned}$$

$$= (f \vartriangle g) \vartriangle \mathcal{D}^* \, \mathit{fork} \circ (\mathcal{D}^* \, (\mathcal{D} \, f) \vartriangle \mathcal{D}^* \, (\mathcal{D} \, g)) \quad \text{-- coinduction}$$
$$= (f \vartriangle g) \vartriangle \mathit{linear} \; \mathit{fork} \circ (\mathcal{D}^* \, (\mathcal{D} \, f) \vartriangle \mathcal{D}^* \, (\mathcal{D} \, g)) \quad \text{-- } \mathit{fork} \text{ linearity (Theorem 6)}$$

Again, the components here (f, g, $\mathcal{D}^* \, (\mathcal{D} \, f)$, and $\mathcal{D}^* \, (\mathcal{D} \, g)$) are all available from $\mathcal{D}^* \, f$ and $\mathcal{D}^* \, g$, so we have a computable recipe for $(\vartriangle)$ on D^*.

Working here

7 What's Next?

[Yet to come:

- Spell out the *Category* and *Cartesian* instances that result from solving the cartesian functor equations as in Section 6.
- Cartesian *closure* (*curry* and *eval*) for D^*, exploiting higher-order derivatives.
- Variation D_* of D^*, using $\mathcal{D}_* \, f = f \vartriangle \mathcal{D} \, (\mathcal{D}_* \, f)$.

]

8 Related Work

The most closely related work I'm aware of is by Vytiniotis et al. (2019), who also define an algorithm around the language of cartesian closed categories. There appear to be some significant shortcomings, however:

- Although the work is referred to as "differentiable programming", it appears to lack a specification and proof that match this claim, i.e., one defined by the mathematical operation of differentiation. As such, it's unclear to me whether the algorithm is about differentiation or something else. In contrast, the specification at the center of Elliott (2018) (and the extensions described above) is just (Fréchet) differentiation itself, combined with the original function as needed by the chain rule, or rather the requirement that the function-with-derivative satisfies a standard collection of homomorphism properties. Correctness of the algorithm was defined as faithfulness to this simple specification, and the algorithm is systematically derived from this specification and hence is correct by construction.
- Functions are already well-defined as a vector space, and thus linear maps (including derivatives) are as well, but the authors choose a different notion. The authors write

> [...] what should be the tangent space of a function type? Perhaps surprisingly, a function type itself is not the right answer. We provide two possible implementations for function tangents and differentiable currying, and explain the tradeoffs.

They do not, however, say what makes their answers "right" and the unsurprising answer wrong.

- The algorithm presented is limited to reverse mode rather than a general AD algorithm as in Elliott (2018) the work described here.

A second related paper is Brunel et al. (2019). Referring to Elliott (2018), the authors write

> However, Elliot's approach is still restricted to first-order programs (i.e., computational graphs): as far as we understand, the functor D is cartesian but not cartesian closed, so the higher-order primitives (λ-abstraction and application) lack a satisfactory treatment. This is implicit in Sect. 4.4 of Elliott (2018), where the author states that he only uses biproduct categories: it is well-known that non-trivial cartesian closed biproduct categories do not exist.

The confusion here—which was mistakenly encouraged by Elliott (2018)—is the idea that the category of differentiable functions itself is (or need be) a biproduct category. Rather, all that was needed is that the various representations *linear maps* (derivatives) are biproduct categories. This requirement is satisfied by construction, since the various representations are all calculated from their denotation (linear functions, itself a biproduct category) via simple cocartesian functors.

References

Aloïs Brunel, Damiano Mazza, and Michele Pagani. Backpropagation in the simply typed lambda-calculus with linear negation. *CoRR*, abs/1909.13768, 2019.

Conal Elliott. Compiling to categories. In *Proceedings of the ACM on Programming Languages (ICFP)*, 2017.

Conal Elliott. The essence of automatic differentiation. In *Proceedings of the ACM on Programming Languages (ICFP)*, 2018.

Jeremy Gibbons. Calculating functional programs. In *Algebraic and Coalgebraic Methods in the Mathematics of Program Construction*, volume 2297 of *Lecture Notes in Computer Science*. Springer-Verlag, 2002.

Marian Boykan Pour-El and Ian Richards. Differentiability properties of computable functions—A summary. *Acta Cybernetica*, 4(1):123–125, 1978.

Marian Boykan Pour-El and Ian Richards. Computability and noncomputability in classical analysis. *Transactions of the American Mathematical Society*, 275(2):539–560, 1983.

Dimitrios Vytiniotis, Dan Belov, Richard Wei, Gordon Plotkin, and Martin Abadi. The differentiable curry. October 2019. To appear in the Program Transformations workshop at NeurIPS 2019.

A Proofs

A.1 Theorem 4

Suppose we have a function $f : a \times b \to c$, and we want to compute its derivative at a point in its (pair-valued) domain. Because linear maps (derivatives) form a cocartesian category,[20]

$$\mathcal{D}\ f\ (a,b) = \mathcal{D}\ f\ (a,b) \circ \mathit{inl} \mathbin{\triangledown} \mathcal{D}\ f\ (a,b) \circ \mathit{inr}$$

Noting that $\mathit{inl}\ da = (da, 0)$ and $\mathit{inr}\ db = (0, db)$, we can see that the partial derivatives allow only one half of a pair to change.

Next, note that $\mathcal{D}\ f\ (a,b) \circ \mathit{inl} = \mathcal{D}\ (f \circ (,b))\ a$, by the following equational reasoning:

$$\begin{array}{ll}
\quad \mathcal{D}\ (f \circ (,b))\ a & \\
= \mathcal{D}\ f\ ((,b)\ a) \circ \mathcal{D}\ (,b)\ a & \text{-- chain rule} \\
= \mathcal{D}\ f\ (a,b) \circ \mathcal{D}\ (,b)\ a & \text{-- } (,b) \text{ definition} \\
= \mathcal{D}\ f\ (a,b) \circ \mathcal{D}\ (\mathit{inl} + \mathit{const}\ (0,b))\ a & \text{-- } \mathit{inl} \text{ on functions, and meaning of } (,b) \\
= \mathcal{D}\ f\ (a,b) \circ (\mathcal{D}\ \mathit{inl}\ a + \mathcal{D}\ (\mathit{const}\ (0,b))\ a) & \text{-- linearity of } (+) \\
= \mathcal{D}\ f\ (a,b) \circ \mathcal{D}\ \mathit{inl}\ a & \text{-- } \mathcal{D}\ (\mathit{const}\ z)\ a = 0 \\
= \mathcal{D}\ f\ (a,b) \circ \mathit{inl} & \text{-- linearity of } \mathit{inl}\text{; Theorem 3}
\end{array}$$

Likewise, $\mathcal{D}\ f\ (a,b) \circ \mathit{inr} = \mathcal{D}\ (f \circ (a,))\ b$.

A.2 Corollary 4.2

$$\begin{array}{ll}
\quad \mathcal{D}\ (\mathit{uncurry}\ g)\ (a,b) & \\
= \mathcal{D}_l\ (\mathit{uncurry}\ g)\ (a,b) \mathbin{\triangledown} \mathcal{D}_r\ (\mathit{uncurry}\ g)\ (a,b) & \text{-- Theorem 4} \\
= \mathcal{D}\ (\mathit{uncurry}\ g \circ (,b))\ a \mathbin{\triangledown} \mathcal{D}\ (\mathit{uncurry}\ g \circ (a,))\ b & \text{-- } \mathcal{D}_l \text{ and } \mathcal{D}_r \text{ definitions} \\
= \mathcal{D}\ (\lambda\, a' \to \mathit{uncurry}\ g\ (a',b))\ a \mathbin{\triangledown} & \text{-- } \eta \text{ expansion and simplification} \\
\quad \mathcal{D}\ (\lambda\, b' \to \mathit{uncurry}\ g\ (a,b'))\ b & \\
= \mathcal{D}\ (\lambda\, a' \to g\ a'\ b)\ a \mathbin{\triangledown} \mathcal{D}\ (\lambda\, b' \to g\ a\ b')\ b & \text{-- } \mathit{uncurry} \text{ on functions} \\
= \mathcal{D}\ (\mathit{at}\ b \circ g)\ a \mathbin{\triangledown} \mathcal{D}\ (g\ a)\ b & \text{-- } \mathit{at} \text{ definition and } \eta \text{ reduction} \\
= \mathcal{D}\ (\mathit{at}\ b)\ (g\ a) \circ \mathcal{D}\ g\ a \mathbin{\triangledown} \mathcal{D}\ (g\ a)\ b & \text{-- chain rule} \\
= \mathit{at}\ b \circ \mathcal{D}\ g\ a \mathbin{\triangledown} \mathcal{D}\ (g\ a)\ b & \text{-- linearity of } \mathit{at}
\end{array}$$

[20]The cocartesian law $h = h \circ \mathit{inl} \mathbin{\triangledown} h \circ \mathit{inr}$ is dual to the cartesian law $h = \mathit{exl} \circ h \mathbin{\triangle} \mathit{exr} \circ h$ (Gibbons, 2002).

A.3 Corollary 4.3

$$\begin{array}{ll}
\quad \mathcal{D}\ \mathit{eval}\ (f, a) & \\
= \mathcal{D}_l\ \mathit{eval}\ (f, a) \mathbin{\triangledown} \mathcal{D}_r\ \mathit{eval}\ (f, a) & \text{-- method of partial derivatives} \\
= \mathcal{D}\ (\mathit{eval} \circ (, a))\ f \mathbin{\triangledown} \mathcal{D}\ (\mathit{eval} \circ (f,))\ a & \text{-- } \mathcal{D}_l \text{ and } \mathcal{D}_r \text{ alternative definitions} \\
= \mathcal{D}\ (\mathit{at}\ a)\ f \mathbin{\triangledown} \mathcal{D}\ f\ a & \text{-- } \mathit{eval} \text{ on functions} \\
= \mathit{at}\ a \mathbin{\triangledown} \mathcal{D}\ f\ a & \text{-- linearity of } \mathit{at}\ a \\
= \lambda\,(\mathit{df}, \mathit{dx}) \to \mathit{df}\ a + \mathcal{D}\ f\ a\ \mathit{dx} & \text{-- } (\triangledown)\ \textit{on linear maps};\ \mathit{at} \text{ definition}
\end{array}$$

Alternatively, calculate $\mathcal{D}\ \mathit{eval}$ via *uncurry*:

$$\begin{array}{ll}
\quad \mathcal{D}\ \mathit{eval}\ (f, a) & \\
= \mathcal{D}\ (\mathit{uncurry}\ \mathit{id})\ (f, a) & \text{-- } \mathit{eval} = \mathit{uncurry}\ \mathit{id} \\
= \mathit{at}\ a \circ \mathcal{D}\ \mathit{id}\ a \mathbin{\triangledown} \mathcal{D}\ (\mathit{id}\ f)\ a & \text{-- Corollary 4.2} \\
= \mathit{at}\ a \circ \mathit{id} \mathbin{\triangledown} \mathcal{D}\ f\ a & \text{-- } \mathit{id} \text{ is linear} \\
= \mathit{at}\ a \mathbin{\triangledown} \mathcal{D}\ f\ a & \text{-- } \mathit{id} \text{ as identity}
\end{array}$$

A.4 Theorem 5

$$\begin{array}{ll}
\quad \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}\ (\mathit{at}\ b \circ g)\ a) & \\
= \lambda\, \mathit{da}\ b \to \mathcal{D}\ (\mathit{at}\ b \circ g)\ a\ \mathit{da} & \text{-- } \mathit{fork}_F \text{ definition} \\
= \lambda\, \mathit{da}\ b \to (\mathcal{D}\ (\mathit{at}\ b)\ (g\ a) \circ \mathcal{D}\ g\ a)\ \mathit{da} & \text{-- chain rule} \\
= \lambda\, \mathit{da}\ b \to (\mathit{at}\ b \circ \mathcal{D}\ g\ a)\ \mathit{da} & \text{-- } \mathit{at}\ b \text{ is linear} \\
= \lambda\, \mathit{da}\ b \to \mathit{at}\ b\ (\mathcal{D}\ g\ a\ \mathit{da}) & \text{-- } (\circ) \text{ on functions} \\
= \lambda\, \mathit{da}\ b \to \mathcal{D}\ g\ a\ \mathit{da}\ b & \text{-- } \mathit{at} \text{ definition} \\
= \mathcal{D}\ g\ a & \text{-- } \eta \text{ reduction (twice)}
\end{array}$$

A.5 Corollary 5.1

$$\begin{array}{ll}
\quad \mathcal{D}\ (\mathit{curry}\ f)\ a & \\
= \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}\ (\mathit{at}\ b \circ \mathit{curry}\ f))\ a & \text{-- Theorem 5} \\
= \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}\ (\lambda\, a \to \mathit{at}\ b\ (\mathit{curry}\ f\ a)))\ a & \text{-- } (\circ) \text{ on functions} \\
= \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}\ (\lambda\, a \to \mathit{curry}\ f\ a\ b))\ a & \text{-- } \mathit{at} \text{ definition} \\
= \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}\ (\lambda\, a \to f\ (a, b)))\ a & \text{-- } \mathit{curry} \text{ on functions} \\
= \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}\ (f \circ (, b)))\ a & \text{-- } (, b) \text{ definition} \\
= \mathit{fork}_F\ (\lambda\, b \to \mathcal{D}_l\ f\ (a, b)) & \text{-- } \mathcal{D}_l \text{ definition} \\
= \mathit{fork}_F\ (\mathcal{D}_l\ f \circ (a,)) & \text{-- } (a,) \text{ definition}
\end{array}$$

A.6 Theorem 9

The functions wrap_o and wrap_o^{-1} form an isomorphism:

$$\begin{array}{ll}
\quad \mathit{wrap}_o^{-1}\ (\mathit{wrap}_o\ f) & \\
= \mathit{wrap}_o^{-1}\ (o \circ f \circ o^{-1}) & \text{-- } \mathit{wrap}_o \text{ definition} \\
= o^{-1} \circ (o \circ f \circ o^{-1}) \circ o & \text{-- } \mathit{wrap}_o^{-1} \text{ definition} \\
= (o^{-1} \circ o) \circ f \circ (o^{-1} \circ o) & \text{-- } (\circ) \text{ associativity} \\
= \mathit{id} \circ f \circ \mathit{id} & \text{-- } o^{-1} \circ o = \mathit{id} \\
= f & \text{-- } \mathit{id} \text{ is identity for } (\circ)
\end{array}$$

$$\begin{array}{ll}
\quad \mathit{wrap}_o\ (\mathit{wrap}_o^{-1}\ h) & \\
= \mathit{wrap}_o\ (o^{-1} \circ h \circ o) & \text{-- } \mathit{wrap}_o^{-1} \text{ definition} \\
= o \circ (o^{-1} \circ h \circ o) \circ o^{-1} & \text{-- } \mathit{wrap}_o \text{ definition} \\
= (o \circ o^{-1}) \circ h \circ (o \circ o^{-1}) & \text{-- } (\circ) \text{ associativity} \\
= \mathit{id} \circ h \circ \mathit{id} & \text{-- } o \circ o^{-1} = \mathit{id} \\
= h & \text{-- } \mathit{id} \text{ is identity for } (\circ)
\end{array}$$

Linearity of wrap_o and wrap_o^{-1} follows from two facts:

- $(\circ f)$ is linear for all f.
- $(g\ \circ)$ is linear for all *linear* g.

Proof: exercise.

A.7 Theorem 10

The functions $\mathcal{D}_o$ and $\mathcal{D}_o^{-1}$ form an isomorphism:

$$\begin{array}{ll}
\quad \mathcal{D}_o^{-1} \circ \mathcal{D}_o & \\
= \mathit{wrap}_o^{-1} \circ \hat{\mathcal{D}}^{-1} \circ \hat{\mathcal{D}} \circ \mathit{wrap}_o & \text{-- } \mathcal{D}_o^{-1} \text{ and } \mathcal{D}_o \text{ definitions} \\
= \mathit{wrap}_o^{-1} \circ \mathit{wrap}_o & \text{-- } \hat{\mathcal{D}}^{-1} \circ \hat{\mathcal{D}} = \mathit{id} \\
= \mathit{id} & \text{-- } \mathit{wrap}_o^{-1} \circ \mathit{wrap}_o = \mathit{id}
\end{array}$$

$$\begin{array}{ll}
\quad \mathcal{D}_o \circ \mathcal{D}_o^{-1} & \\
= \hat{\mathcal{D}} \circ \mathit{wrap}_o \circ \mathit{wrap}_o^{-1} \circ \mathcal{D}_o^{-1} & \text{-- } \mathcal{D}_o \text{ and } \mathcal{D}_o^{-1} \text{ definitions} \\
= \hat{\mathcal{D}} \circ \mathcal{D}_o^{-1} & \text{-- } \mathit{wrap}_o^{-1} \circ \mathit{wrap}_o = \mathit{id} \\
= \mathit{id} & \text{-- } \hat{\mathcal{D}}^{-1} \circ \hat{\mathcal{D}} = \mathit{id}
\end{array}$$

Linearity of $\mathcal{D}_o$ and $\mathcal{D}_o^{-1}$ follows from linearity of $\hat{\mathcal{D}}$ and $\hat{\mathcal{D}}^{-1}$ and Theorem 9.

A.8 Theorem 11

The proof that wrap_o is a cartesian functor mainly exploit the regular structure of o and o^{-1}:

$$\begin{array}{ll}
\quad \mathit{wrap}_o\ \mathit{id} & \\
= o \circ \mathit{id} \circ o^{-1} & \text{-- } \mathit{wrap}_o \text{ definition}
\end{array}$$

$$\begin{aligned}
&= o \circ o^{-1} && \text{-- } \mathit{id} \text{ is identity for } (\circ) \\
&= \mathit{id} && \text{-- } o \circ o^{-1} = \mathit{id}
\end{aligned}$$

$$\begin{aligned}
&\ \mathit{wrap}_o\ (g \circ f) && \\
&= o \circ g \circ f \circ o^{-1} && \text{-- } \mathit{wrap}_o \text{ definition} \\
&= o \circ g \circ o^{-1} \circ o \circ f \circ o^{-1} && \text{-- } o^{-1} \circ o = \mathit{id} \\
&= (o \circ g \circ o^{-1}) \circ (o \circ f \circ o^{-1}) && \text{-- } \hat{\mathcal{D}} \text{ is a functor} \\
&= \mathit{wrap}_o\ g \circ \mathit{wrap}_o\ f && \text{-- } \mathit{wrap}_o \text{ definition}
\end{aligned}$$

$$\begin{aligned}
&\ \mathit{wrap}_o\ (f \times g) && \text{-- } \mathcal{D}_o \text{ definition} \\
&= o \circ (f \times g) \circ o^{-1} && \text{-- } \mathit{wrap}_o \text{ definition} \\
&= (o \times o) \circ (f \times g) \circ (o^{-1} \times o^{-1}) && \text{-- } o \text{ on products} \\
&= o \circ f \circ o^{-1} \times o \circ g \circ o^{-1} && \text{-- monoidal category law} \\
&= o \circ f \circ o^{-1} \times o \circ g \circ o^{-1} && \text{-- } \hat{\mathcal{D}} \text{ is a monoidal functor} \\
&= \mathit{wrap}_o\ f \times \mathit{wrap}_o\ g && \text{-- } \mathit{wrap}_o \text{ definition}
\end{aligned}$$

$$\begin{aligned}
&\ \mathit{wrap}_o\ \mathit{exl} && \text{-- } \mathcal{D}_o \text{ definition} \\
&= o \circ \mathit{exl} \circ o^{-1} && \text{-- } \mathit{wrap}_o \text{ definition} \\
&= o \circ \mathit{exl} \circ (o^{-1} \times o^{-1}) && \text{-- } o^{-1} \text{ on products} \\
&= o \circ o^{-1} \circ \mathit{exl} && \text{-- } \mathit{exl} \circ (f \times g) = f \circ \mathit{exl} \text{ for cartesian categories} \\
&= \mathit{exl} && \text{-- } o \circ o^{-1} = \mathit{id}
\end{aligned}$$

$$\begin{aligned}
&\ \mathit{wrap}_o\ \mathit{exr} && \text{-- } \mathcal{D}_o \text{ definition} \\
&= o \circ \mathit{exr} \circ o^{-1} && \text{-- } \mathit{wrap}_o \text{ definition} \\
&= o \circ \mathit{exr} \circ (o^{-1} \times o^{-1}) && \text{-- } o^{-1} \text{ on products} \\
&= o \circ o^{-1} \circ \mathit{exr} && \text{-- } \mathit{exr} \circ (f \times g) = g \circ \mathit{exr} \text{ for cartesian categories} \\
&= \mathit{exr} && \text{-- } o \circ o^{-1} = \mathit{id}
\end{aligned}$$

$$\begin{aligned}
&\ \mathit{wrap}_o\ \mathit{dup} && \text{-- } \mathcal{D}_o \text{ definition} \\
&= o \circ \mathit{dup} \circ o^{-1} && \text{-- } \mathit{wrap}_o \text{ definition} \\
&= o \circ (o^{-1} \times o^{-1}) \circ \mathit{dup} && \text{-- } \mathit{dup} \circ f = (f \times f) \circ \mathit{dup} \text{ for cartesian categories} \\
&= o \circ o^{-1} \circ \mathit{dup} && \text{-- } o^{-1} \text{ on products} \\
&= \mathit{dup} && \text{-- } o \circ o^{-1} = \mathit{id}
\end{aligned}$$

A.9 Theorem 13

$$\begin{aligned}
&\ \mathit{wrap}_o\ (\mathit{curry}\ f) && \\
&= o \circ \mathit{curry}\ f \circ o^{-1} && \text{-- } \mathit{wrap}_o \text{ definition} \\
&= \mathcal{D}_o \circ \mathit{curry}\ f \circ o^{-1} && \text{-- } o \text{ on functions} \\
&= \hat{\mathcal{D}} \circ \mathit{wrap}_o \circ \mathit{curry}\ f \circ o^{-1} && \text{-- } \mathcal{D}_o \text{ definition} \\
&= \hat{\mathcal{D}} \circ \mathit{curry}\ (\mathit{wrap}_o\ f) && \text{-- below}
\end{aligned}$$

For this last step,

$$\begin{aligned}
&\ \mathit{wrap}_o \circ \mathit{curry}\ f \circ o^{-1} && \\
&= \lambda\, a \rightarrow \mathit{wrap}_o\ (\mathit{curry}\ f\ (o^{-1}\ a)) && \text{-- } \eta \text{ expansion}
\end{aligned}$$

$$\begin{array}{ll}
= \lambda\, a \to o \circ \mathit{curry}\ f\ (o^{-1}\ a) \circ o^{-1} & \text{-- } \mathit{wrap}_o \text{ definition} \\
= \lambda\, a\ b \to o\ (\mathit{curry}\ f\ (o^{-1}\ a)\ (o^{-1}\ b)) & \text{-- } \eta \text{ expansion} \\
= \lambda\, a\ b \to o\ (f\ (o^{-1}\ a, o^{-1}\ b)) & \text{-- } \mathit{curry} \text{ on functions} \\
= \lambda\, a\ b \to o\ (f\ (o^{-1}\ (a, b))) & \text{-- } o^{-1} \text{ on pairs} \\
= \lambda\, a\ b \to \mathit{wrap}_o\ f\ (a, b) & \text{-- } \mathit{wrap}_o \text{ definition} \\
= \mathit{curry}\ (\mathit{wrap}_o\ f) & \text{-- } \mathit{curry} \text{ on functions}
\end{array}$$

Equivalently, $\mathit{curry}\ (\mathit{wrap}_o\ f) = \hat{\mathcal{D}}^{-1} \circ \mathit{wrap}_o\ (\mathit{curry}\ f)$.[21]

A.10 $\mathcal{D}_o$ and *eval*

The homomorphism equation is $\mathit{eval} = \mathcal{D}_o\ \mathit{eval}$. Simplifying the RHS,

$$\begin{array}{ll}
\ \mathcal{D}_o\ \mathit{eval} & \\
= \hat{\mathcal{D}}\ (\mathit{wrap}_o\ \mathit{eval}) & \text{-- } \mathcal{D}_o \text{ definition} \\
= \hat{\mathcal{D}}\ (o \circ \mathit{eval} \circ o^{-1}) & \text{-- } \mathit{wrap}_o \text{ definition} \\
= \hat{\mathcal{D}}\ (o \circ \mathit{eval} \circ (\mathcal{D}_o^{-1} \times o^{-1})) & \text{-- } o^{-1} \text{ on } (a \to b) \times a \\
= \hat{\mathcal{D}}\ (\lambda\, (\hat{f}, a) \to (o \circ \mathit{eval} \circ (\mathcal{D}_o^{-1} \times o^{-1}))\ (\hat{f}, a)) & \text{-- } \eta \text{ expansion} \\
= \hat{\mathcal{D}}\ (\lambda\, (\hat{f}, a) \to o\ (\mathit{eval}\ (\mathcal{D}_o^{-1}\ \hat{f}, o^{-1}\ a))) & \text{-- } (\circ) \text{ and } (\times) \text{ on functions} \\
= \hat{\mathcal{D}}\ (\lambda\, (\hat{f}, a) \to o\ (\mathcal{D}_o^{-1}\ \hat{f}\ (o^{-1}\ a))) & \text{-- } \mathit{eval} \text{ on functions} \\
= \hat{\mathcal{D}}\ (\lambda\, (\hat{f}, a) \to o\ (\mathit{wrap}_o^{-1}\ (\hat{\mathcal{D}}^{-1}\ \hat{f})\ (o^{-1}\ a))) & \text{-- } \mathcal{D}_o^{-1} \text{ definition} \\
= \hat{\mathcal{D}}\ (\lambda\, (\hat{f}, a) \to o\ ((o^{-1} \circ \hat{\mathcal{D}}^{-1}\ \hat{f} \circ o)\ (o^{-1}\ a))) & \text{-- } \mathit{wrap}_o^{-1} \text{ definition} \\
= \hat{\mathcal{D}}\ (\lambda\, (\hat{f}, a) \to o\ (o^{-1}\ (\hat{\mathcal{D}}^{-1}\ \hat{f}\ (o\ (o^{-1}\ a))))) & \text{-- } (\circ) \text{ on functions} \\
= \hat{\mathcal{D}}\ (\lambda\, (\hat{f}, a) \to \hat{\mathcal{D}}^{-1}\ \hat{f}\ a) & \text{-- } o \circ o^{-1} = \mathit{id} \\
= \hat{\mathcal{D}}\ (\mathit{uncurry}\ \hat{\mathcal{D}}^{-1}) & \text{-- } \mathit{uncurry} \text{ on functions} \\
= D\ (\lambda\, (\hat{f}, a) \to (\mathit{uncurry}\ \hat{\mathcal{D}}^{-1}\ (\hat{f}, a), \mathcal{D}\ (\mathit{uncurry}\ \hat{\mathcal{D}}^{-1})\ (\hat{f}, a))) & \text{-- } \hat{\mathcal{D}} \text{ definition} \\
= D\ (\lambda\, (\hat{f}, a) \to (\hat{\mathcal{D}}^{-1}\ \hat{f}\ a, \mathcal{D}\ (\mathit{uncurry}\ \hat{\mathcal{D}}^{-1})\ (\hat{f}, a))) & \text{-- } \mathit{uncurry} \text{ on functions} \\
= D\ (\lambda\, (\hat{f}, a) \to (\hat{\mathcal{D}}^{-1}\ \hat{f}\ a, \mathit{at}\ a \circ \mathcal{D}\ \hat{\mathcal{D}}^{-1}\ \hat{f} \triangledown \mathcal{D}\ (\hat{\mathcal{D}}^{-1}\ \hat{f})\ a)) & \text{-- Section A.2} \\
= D\ (\lambda\, (\hat{f}, a) \to (\hat{\mathcal{D}}^{-1}\ \hat{f}\ a, \mathit{at}\ a \circ \hat{\mathcal{D}}^{-1} \triangledown \mathcal{D}\ (\hat{\mathcal{D}}^{-1}\ \hat{f})\ a)) & \text{-- } \hat{\mathcal{D}}^{-1} \text{ is linear}
\end{array}$$

Now note that

$$\begin{array}{ll}
\ \hat{f} & \\
= \hat{\mathcal{D}}\ (\hat{\mathcal{D}}^{-1}\ \hat{f}) & \text{-- } \hat{\mathcal{D}} \circ \hat{\mathcal{D}}^{-1} = \mathit{id} \\
= D\ (\hat{\mathcal{D}}^{-1}\ \hat{f} \vartriangle \mathcal{D}\ (\hat{\mathcal{D}}^{-1}\ \hat{f})) & \text{-- } \hat{\mathcal{D}} \text{ definition}
\end{array}$$

Letting $D\ h = \hat{f}$, we have

$$\begin{aligned}
h\ a &= (\hat{\mathcal{D}}^{-1}\ \hat{f} \vartriangle \mathcal{D}\ (\hat{\mathcal{D}}^{-1}\ \hat{f}))\ a \\
&= (\hat{\mathcal{D}}^{-1}\ \hat{f}\ a, \mathcal{D}\ (\hat{\mathcal{D}}^{-1}\ \hat{f})\ a)
\end{aligned}$$

A bit of refactoring then replaces $\hat{\mathcal{D}}^{-1}\ \hat{f}\ a$ and (the noncomputable) $\mathcal{D}\ (\hat{\mathcal{D}}^{-1}\ \hat{f}\ a)$, yielding a *computable* form:

[21] [Maybe this form will help simplify another proof.]

$$\begin{aligned}
&\mathcal{D}_o\ eval\\
=\ &...\\
=\ &D\ (\lambda\,(D\ h, a) \rightarrow \mathbf{let}\ (b, f') = h\ a\ \mathbf{in}\ (b, at\ a \circ \hat{\mathcal{D}}^{-1} \triangledown f'))
\end{aligned}$$

Since this calculation was fairly involved, let's get a sanity check on the types in the final form:

$$\begin{aligned}
(D\ h, a) &: O\ ((a \rightarrow b) \times a)\\
&: D\ (O\ a)\ (O\ b) \times O\ a\\
D\ h &: D\ (O\ a)\ (O\ b)\\
a &: O\ a\\
h &: O\ a \rightarrow O\ b \times (O\ a \multimap O\ b)\\
h\ a &: O\ b \times (O\ a \multimap O\ b)
\end{aligned}$$

$$\begin{aligned}
(b, f') &: O\ b \times (O\ a \multimap O\ b)\\
b &: O\ b\\
f' &: O\ a \multimap O\ b
\end{aligned}$$

$$\begin{aligned}
\hat{\mathcal{D}}^{-1} &: D\ (O\ a)\ (O\ b) \multimap (O\ a \rightarrow O\ b)\\
at\ a &: (O\ a \rightarrow O\ b) \multimap O\ b\\
at\ a \circ \hat{\mathcal{D}}^{-1} &: D\ (O\ a)\ (O\ b) \multimap O\ b\\
at\ a \circ \hat{\mathcal{D}}^{-1} \triangledown f' &: D\ (O\ a)\ (O\ b) \times O\ a \multimap O\ b\\
(b, at\ a \circ \hat{\mathcal{D}}^{-1} \triangledown f') &: O\ b \times (D\ (O\ a)\ (O\ b) \times O\ a \multimap O\ b)\\
\lambda\,(D\ h, a) \rightarrow ...\mathbf{in}\ (b, at\ a \circ \hat{\mathcal{D}}^{-1} \triangledown f') &: O\ ((a \rightarrow b) \times a) \rightarrow O\ b \times (D\ (O\ a)\ (O\ b) \times O\ a \multimap O\ b\\
D\ (\lambda\,(D\ h, a) \rightarrow ...\mathbf{in}\ (b, at\ a \circ \hat{\mathcal{D}}^{-1} \triangledown f')) &: D\ (O\ ((a \rightarrow b) \times a))\ (O\ b)
\end{aligned}$$

$$\begin{aligned}
eval &: (a \rightarrow b) \times a \rightarrow b\\
\mathcal{D}_o\ \ eval &: D\ (O\ ((a \rightarrow b) \times a))\ (O\ b)
\end{aligned}$$

A.11 Theorem 15

Letting $f_o = wrap_o\, f$,

$$\begin{aligned}
&\mathcal{D}_o\ (curry\ f)\\
=\ &\hat{\mathcal{D}}\ (wrap_o\ (curry\ f)) && \text{-- } \mathcal{D}_o \text{ definition}\\
=\ &\hat{\mathcal{D}}\ (\hat{\mathcal{D}} \circ curry\ (wrap_o\, f)) && \text{-- Theorem 13}\\
=\ &\hat{\mathcal{D}}\ (\hat{\mathcal{D}} \circ curry\ f_o) && \text{-- } f_o \text{ definition}\\
=\ &D\ ((\hat{\mathcal{D}} \circ curry\ f_o) \vartriangle \mathcal{D}\ (\hat{\mathcal{D}} \circ curry\ f_o)) && \text{-- } \hat{\mathcal{D}} \text{ definition}\\
=\ &D\ (\lambda\, a \rightarrow \hat{\mathcal{D}}\ (curry\ f_o\ a), \mathcal{D}\ (\hat{\mathcal{D}} \circ curry\ f_o)\ a) && \text{-- } (\vartriangle) \text{ definition}\\
=\ &D\ (\lambda\, a \rightarrow \hat{\mathcal{D}}\ (curry\ f_o\ a), \hat{\mathcal{D}} \circ \mathcal{D}\ (curry\ f_o)\ a) && \text{-- chain rule; linearity of } \hat{\mathcal{D}}\\
=\ &D\ (\lambda\, a \rightarrow \hat{\mathcal{D}}\ (curry\ f_o\ a), \hat{\mathcal{D}} \circ fork_F\ (\mathcal{D}_l\ f_o \circ (a,))) && \text{-- Corollary 5.1}
\end{aligned}$$

Now, separately simplify the two main parts of this last form.

$$
\begin{array}{ll}
\quad \hat{\mathcal{D}}\ (\mathit{curry}\ f_o\ a) & \\
= D\ (\lambda\, b \to (f_o\ (a,b), \mathcal{D}_r\ f_o\ (a,b))) & \text{-- } \hat{\mathcal{D}} \text{ definition and Theorem 4}
\end{array}
$$

$$
\begin{array}{ll}
\quad \hat{\mathcal{D}} \circ \mathit{fork}_F\ (\mathcal{D}_l\ f_o \circ (a,)) & \\
= \hat{\mathcal{D}} \circ (\lambda\, da\ b \to (\mathcal{D}_l\ f_o \circ (a,))\ b\ da) & \text{-- } \mathit{fork}_F \text{ definition} \\
= \hat{\mathcal{D}} \circ (\lambda\, da\ b \to \mathcal{D}_l\ f_o\ (a,b)\ da) & \text{-- } (\circ) \text{ on functions} \\
= \lambda\, da \to \hat{\mathcal{D}}\ (\lambda\, b \to \mathcal{D}_l\ f_o\ (a,b)\ da) & \text{-- } (\circ) \text{ on functions} \\
= \lambda\, da \to \hat{\mathcal{D}}\ (\lambda\, b \to \mathit{at}\ da\ (\mathcal{D}_l\ f_o\ (a,b))) & \text{-- } \mathit{at} \text{ definition} \\
= \lambda\, da \to \hat{\mathcal{D}}\ (\mathit{at}\ da \circ \mathcal{D}_l\ f_o \circ (a,)) & \text{-- } (\circ) \text{ on functions} \\
= \lambda\, da \to D\ (\lambda\, b \to ((\mathit{at}\ da \circ \mathcal{D}_l\ f_o \circ (a,))\ b, \mathcal{D}\ (\mathit{at}\ da \circ \mathcal{D}_l\ f_o \circ (a,))\ b)) & \text{-- } \hat{\mathcal{D}} \text{ definition} \\
= \lambda\, da \to D\ (\lambda\, b \to (\mathcal{D}_l\ f_o\ (a,b)\ da, \mathcal{D}\ (\mathit{at}\ da \circ \mathcal{D}_l\ f_o \circ (a,))\ b)) & \text{-- } (\circ) \text{ on functions}
\end{array}
$$

Now simplify the remaining differentiated composition:

$$
\begin{array}{ll}
\quad \mathcal{D}\ (\mathit{at}\ da \circ \mathcal{D}_l\ f_o \circ (a,))\ b & \\
= \mathit{at}\ da \circ \mathcal{D}\ (\mathcal{D}_l\ f_o \circ (a,))\ b & \text{-- chain rule; linearity of } \mathit{at}\ da \\
= \mathit{at}\ da \circ \mathcal{D}_r\ (\mathcal{D}_l\ f_o)\ (a,b) & \text{-- Theorem 4}
\end{array}
$$

Putting the pieces back together,

$$
\begin{array}{l}
\mathcal{D}_o\ (\mathit{curry}\ f) = D\ (\lambda\, a \to (D\ (\lambda\, b \to (f_o\ (a,b), \mathcal{D}_r\ f_o\ (a,b))) \\
\qquad\qquad\qquad\qquad ,\lambda\, da \to D\ (\lambda\, b \to (\mathcal{D}_l\ f_o\ (a,b)\ da, \mathit{at}\ da \circ \mathcal{D}_r\ (\mathcal{D}_l\ f_o)\ (a,b)))))
\end{array}
$$

A.12 Theorem 16

To show that $\mathit{comp} = \mathit{uncurry}\ (\circ)$ is bilinear, we can show that it is linear in each argument, which is to say $\mathit{curry}\ \mathit{comp}\ g = (g\circ)$ and $\mathit{curry}'\ \mathit{comp}\ f = (\circ f)$ are linear for all g and f.

First, $(\circ\, f)$ is linear for all functions f (not just linear):

$$
\begin{array}{ll}
\quad (\circ\, f)\ (g + g') & \\
= (g + g') \circ f & \text{-- left section definition} \\
= \lambda\, a \to (g + g')\ (f\ a) & \text{-- } \eta \text{ expansion} \\
= \lambda\, a \to g\ (f\ a) + g'\ (f\ a) & \text{-- addition on functions} \\
= (\lambda\, a \to g\ (f\ a)) + (\lambda\, a \to g'\ (f\ a)) & \text{-- addition on functions} \\
= (g \circ f) + (g' \circ f) & \text{-- } (\circ) \text{ on functions} \\
= (\circ\, f)\ g + (\circ\, f)\ g' & \text{-- left section definition}
\end{array}
$$

$$
\begin{array}{ll}
\quad s \cdot (\circ\, f)\ g & \\
= s \cdot (g \circ f) & \text{-- left section definition} \\
= \lambda\, a \to s \cdot g\ (f\ a) & \text{-- scaling on functions} \\
= \lambda\, a \to (s \cdot g)\ (f\ a & \text{-- scaling on functions} \\
= (s \cdot g) \circ f & \text{-- } (\circ) \text{ on functions} \\
= (\circ\, f)\ (s \cdot g) & \text{-- left section definition}
\end{array}
$$

Second, $(g\circ)$ is linear for all linear functions g.

$$\begin{array}{ll}
\quad (g\circ)\ (f+f') & \\
= g\circ(f+f') & \text{-- right section definition}\\
= \lambda\, a \rightarrow g\ ((f+f')\ a) & \text{-- } \eta \text{ expansion}\\
= \lambda\, a \rightarrow g\ (f\ \ a + f'\ a) & \text{-- addition on functions}\\
= \lambda\, a \rightarrow g\ (f\ \ a) + g\ (f'\ a) & \text{-- linearity of } g\\
= (\lambda\, a \rightarrow g\ (f\ \ a)) + (\lambda\, a \rightarrow g\ (f'\ a)) & \text{-- addition on functions}\\
= (g\circ f) + (g\circ f') & \text{-- } (\circ) \text{ on functions}\\
= (g\circ)\ f + (g\circ)\ f' & \text{-- right section definition}
\end{array}$$

$$\begin{array}{ll}
\quad s\cdot(g\circ)\ f & \\
= s\cdot(g\circ f) & \text{-- right section definition}\\
= \lambda\, a \rightarrow s\cdot g\ (f\ \ a) & \text{-- scaling on functions}\\
= \lambda\, a \rightarrow g\ (s\cdot f\ \ a) & \text{-- linearity of } g\\
= \lambda\, a \rightarrow g\ ((s\cdot f)\ a) & \text{-- scaling on functions}\\
= g\circ(s\cdot f) & \text{-- } (\circ) \text{ on functions}\\
= (g\circ)\ (s\cdot f) & \text{-- right section definition}
\end{array}$$

comp (

A.13 Theorem 17

Given any bilinear function h,

a. *curry* h a is linear for all linear functions g.

$$\begin{array}{ll}
\quad \mathit{curry}\ h\ a\ (b+b') & \\
= h\ (a, b+b') & \text{-- } \mathit{curry} \text{ on functions}\\
= h\ (a,b) + h\ (a,b') & \text{-- bilinearity of } h\\
= \mathit{curry}\ h\ a\ b + \mathit{curry}\ h\ a\ b' & \text{-- } \mathit{curry} \text{ on functions}
\end{array}$$

$$\begin{array}{ll}
\quad \mathit{curry}\ h\ a\ (s\cdot b) & \\
= h\ (a, s\cdot b) & \text{-- } \mathit{curry} \text{ on functions}\\
= s\cdot h\ (a,b) & \text{-- bilinearity of } h
\end{array}$$

b. *curry'* h b is linear for all functions b. Proof similar to *curry* h a.

c. *curry* h and *curry'* h are linear.

$$\begin{array}{ll}
\quad \mathit{curry}\ h\ (a+a') & \\
= \lambda\, b \rightarrow \mathit{curry}\ h\ (a+a')\ b & \text{-- } \eta \text{ expansion}\\
= \lambda\, b \rightarrow h\ (a+a', b) & \text{-- } \mathit{curry} \text{ on functions}\\
= \lambda\, b \rightarrow h\ (a,b) + h\ (a',b) & \text{-- bilinearity of } h\\
= (\lambda\, b \rightarrow h\ (a,b)) + (\lambda\, b \rightarrow h\ (a',b)) & \text{-- addition on functions}
\end{array}$$

$$
\begin{array}{ll}
= \mathit{curry}\ h\ a + \mathit{curry}\ h\ a' & \text{-- } \mathit{curry} \text{ on functions}
\end{array}
$$

$$
\begin{array}{ll}
\quad \mathit{curry}\ h\ (s \cdot a) & \\
= \lambda\, b \rightarrow \mathit{curry}\ h\ (s \cdot a)\ b & \text{-- } \eta \text{ expansion} \\
= \lambda\, b \rightarrow h\ (s \cdot a', b) & \text{-- } \mathit{curry} \text{ on functions} \\
= \lambda\, b \rightarrow s \cdot h\ (a, b) & \text{-- bilinearity of } h \\
= s \cdot (\lambda\, b \rightarrow h\ (a, b)) & \text{-- scaling on functions} \\
= s \cdot \mathit{curry}\ h\ a & \text{-- } \mathit{curry} \text{ on functions}
\end{array}
$$

Similarly for $\mathit{curry}'\ h$.

d. $\mathcal{D}\ h$ is linear.

$$
\begin{array}{ll}
\quad \mathcal{D}\ h\ ((a, b) + (a', b')) & \\
= \mathcal{D}\ h\ (a + a', b + b') & \text{-- } (+) \text{ on functions} \\
= h \circ (, b + b') \triangledown h \circ (a + a') & \text{-- Corollary 4.1} \\
= \lambda\, (da, db) \rightarrow h\ (da, b + b') + h\ (a + a', db) & \text{-- } (\triangledown) \text{ on functions} \\
= \lambda\, (da, db) \rightarrow h\ (da, b) + h\ (da, b') + h\ (a, db) + h\ (a', db) & \text{-- bilinearity of } h \\
= \lambda\, (da, db) \rightarrow h\ (da, b) + h\ (a, db) + h\ (da, b') + h\ (a', db) & \text{-- commutativity of } (+) \\
= (\lambda\, (da, db) \rightarrow h\ (da, b) + h\ (a, db)) + & \\
\quad (\lambda\, (da, db) \rightarrow h\ (da, b') + h\ (a', db)) & \text{-- } (+) \text{ on functions} \\
= \mathcal{D}\ h\ (a, b) + \mathcal{D}\ h\ (a', b') & \text{-- Corollary 4.1}
\end{array}
$$

Similarly for scaling.

A.14 Theorem 18

$$
\begin{array}{ll}
\quad \mathcal{D}\ (g \circ f) & \\
= \lambda\, a \rightarrow \mathcal{D}\ (g \circ f)\ a & \text{-- } \eta \text{ expansion} \\
= \lambda\, a \rightarrow \mathcal{D}\ g\ (f\ a) \circ \mathcal{D}\ f\ a & \text{-- chain rule (Theorem 1)} \\
= \lambda\, a \rightarrow (\mathcal{D}\ g \circ f)\ a \circ \mathcal{D}\ f\ a & \text{-- } (\circ) \text{ on functions} \\
= \lambda\, a \rightarrow (\circ)\ ((\mathcal{D}\ g \circ f)\ a)\ (\mathcal{D}\ f\ a) & \text{-- alternative notation} \\
= \lambda\, a \rightarrow \mathit{uncurry}\ (\circ)\ ((\mathcal{D}\ g \circ f)\ a, \mathcal{D}\ f\ a) & \text{-- } \mathit{uncurry} \text{ on functions} \\
= \lambda\, a \rightarrow \mathit{comp}\ ((\mathcal{D}\ g \circ f)\ a, \mathcal{D}\ f\ a) & \text{-- } \mathit{comp} \text{ definition} \\
= \mathit{comp} \circ (\lambda\, a \rightarrow ((\mathcal{D}\ g \circ f)\ a, \mathcal{D}\ f\ a)) & \text{-- } (\circ) \text{ on functions} \\
= \mathit{comp} \circ (\mathcal{D}\ g \circ f \vartriangle \mathcal{D}\ f) & \text{-- } (\vartriangle) \text{ definition}
\end{array}
$$

$$
\begin{array}{ll}
\quad \mathcal{D}\ (f \times g) & \\
= \lambda\, (a, b) \rightarrow \mathcal{D}\ (f \times g)\ (a, b) & \text{-- } \eta \text{ expansion} \\
= \lambda\, (a, b) \rightarrow \mathcal{D}\ f\ a \times \mathcal{D}\ g\ b & \text{-- cross rule (Theorem 2)} \\
= \lambda\, (a, b) \rightarrow \mathit{uncurry}\ (\times)\ (\mathcal{D}\ f\ a, \mathcal{D}\ g\ b) & \text{-- } \mathit{uncurry} \text{ on functions} \\
= \lambda\, (a, b) \rightarrow \mathit{cross}\ (\mathcal{D}\ f\ a, \mathcal{D}\ g\ b) & \text{-- } \mathit{cross} \text{ definition} \\
= \lambda\, (a, b) \rightarrow \mathit{cross}\ ((\mathcal{D}\ f \times \mathcal{D}\ g)\ (a, b)) & \text{-- } (\times) \text{ on functions} \\
= \mathit{cross} \circ (\mathcal{D}\ f \times \mathcal{D}\ g) & \text{-- } (\circ) \text{ on functions}
\end{array}
$$

$$\begin{aligned}
&\mathcal{D}\ (f \mathbin{\triangle} g) && \\
&= \mathcal{D}\ ((f \times g) \circ \mathit{dup}) && \text{-- cartesian law} \\
&= \lambda\, a \to \mathcal{D}\ ((f \times g) \circ \mathit{dup})\ a && \text{-- } \eta \text{ expansion} \\
&= \lambda\, a \to \mathcal{D}\ (f \times g)\ (\mathit{dup}\ a) \circ \mathcal{D}\ \mathit{dup}\ a && \text{-- chain rule (Theorem 1)} \\
&= \lambda\, a \to \mathcal{D}\ (f \times g)\ (a, a) \circ \mathcal{D}\ \mathit{dup}\ a && \text{-- } \mathit{dup} \text{ for functions} \\
&= \lambda\, a \to \mathcal{D}\ f\ a \times \mathcal{D}\ g\ a \circ \mathcal{D}\ \mathit{dup}\ a && \text{-- cross rule (Theorem 2)} \\
&= \lambda\, a \to \mathcal{D}\ f\ a \times \mathcal{D}\ g\ a \circ \mathit{dup} && \text{-- } \mathit{dup} \text{ linearity} \\
&= \lambda\, a \to \mathcal{D}\ f\ a \mathbin{\triangle} \mathcal{D}\ g\ a && \text{-- cartesian law} \\
&= \mathit{fork} \circ (\mathcal{D}\ f \mathbin{\triangle} \mathcal{D}\ g) && \text{-- } \mathit{fork} \text{ definition}
\end{aligned}$$

For a linear function f,

$$\begin{aligned}
&\mathcal{D}\ f && \\
&= \lambda\, a \to \mathcal{D}\ f\ a && \text{-- } \eta \text{ expansion} \\
&= \lambda\, a \to f && \text{-- Theorem 3} \\
&= \mathit{const}\ f && \text{-- } \mathit{const} \text{ definition}
\end{aligned}$$

For any function $f : a \times b \to c$,

$$\begin{aligned}
&\mathcal{D}\ f && \\
&= \lambda\, (a, b) \to \mathcal{D}\ f\ (a, b) && \text{-- } \eta \text{ expansion} \\
&= \lambda\, (a, b) \to \mathcal{D}_l\ f\ (a, b) \mathbin{\triangledown} \mathcal{D}_r\ f\ (a, b) && \text{-- Theorem 4} \\
&= \mathit{join} \circ (\mathcal{D}_l\ f \mathbin{\triangle} \mathcal{D}_r\ f) && \text{-- } \mathit{join} \text{ definition}
\end{aligned}$$

For a *bilinear* function $f : a \times b \to c$,

$$\begin{aligned}
&\mathcal{D}\ f && \\
&= \lambda\, (a, b) \to \mathcal{D}\ f\ (a, b) && \text{-- } \eta \text{ expansion} \\
&= \lambda\, (a, b) \to f \circ (, b) \mathbin{\triangledown} f \circ (a,) && \text{-- Corollary 4.1} \\
&= \lambda\, (a, b) \to \mathit{curry}'\ f\ b \mathbin{\triangledown} \mathit{curry}\ f\ a && \text{-- section definitions} \\
&= \lambda\, (a, b) \to \mathit{join}\ (\mathit{curry}'\ f\ b, \mathit{curry}\ f\ a) && \text{-- } \mathit{join} = \mathit{uncurry}\ (\triangledown) \\
&= \lambda\, (a, b) \to \mathit{join}\ ((\mathit{curry}'\ f \times \mathit{curry}\ f)\ (b, a)) && \text{-- } (\times) \text{ on functions} \\
&= \lambda\, (a, b) \to \mathit{join}\ ((\mathit{curry}'\ f \times \mathit{curry}\ f)\ (\mathit{swap}\ (a, b))) && \text{-- } \mathit{swap} \text{ on functions} \\
&= \mathit{join} \circ (\mathit{curry}'\ f \times \mathit{curry}\ f) \circ \mathit{swap} && \text{-- } (\circ) \text{ on functions}
\end{aligned}$$

For uncurried composition on linear maps,

$$\begin{aligned}
&\mathcal{D}\ \mathit{comp} && \\
&= \mathit{join} \circ (\mathit{curry}'\ \mathit{comp} \times \mathit{curry}\ \mathit{comp}) \circ \mathit{swap} && \text{-- previous (}\mathit{comp} \text{ is bilinear)} \\
&= \mathit{join} \circ (\mathit{flip}\ (\circ) \times (\circ)) \circ \mathit{swap} && \text{-- } \mathit{comp} \text{ definition}
\end{aligned}$$