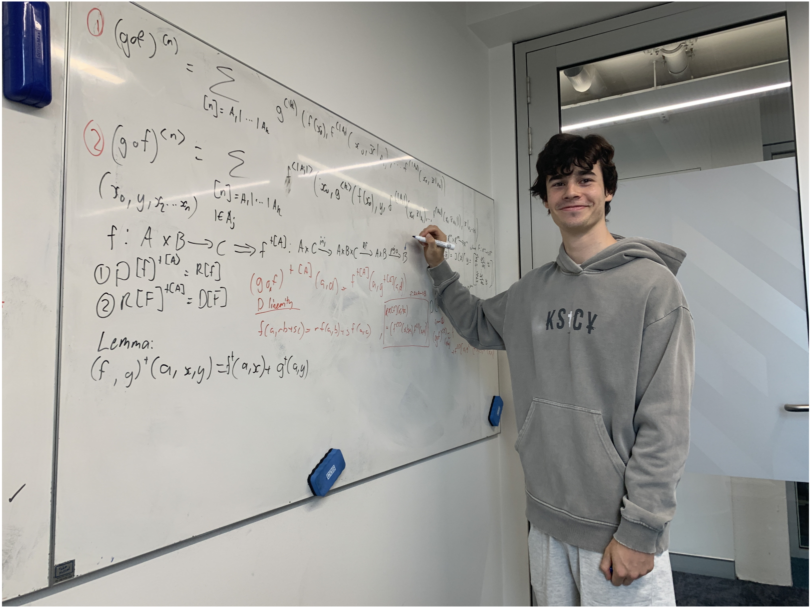
Reverse Faà di Bruno's Formula for Cartesian Reverse Differential Categories

JS PL (he/him), joint work with Aaron Biggin

Dedicated to Phil Scott.

Joint work with Aaron Biggin



Today's Story

- There are two types of derivative operations used in automatic differentiation:

forward differentiation **reverse differentiation**

Today's Story

- There are two types of derivative operations used in automatic differentiation:

  forward differentiation **reverse differentiation**

- Cartesian differential categories provide the categorial foundations of **forward differentiation**

  R. Blute, R. Cockett, R.A.G. Seely, **Cartesian Differential Categories**

  Cartesian differential categories have a rich literature and have been successful in formalizing important concepts from differential calculus.

Today's Story

- There are two types of derivative operations used in automatic differentiation:

  forward differentiation **reverse differentiation**

- Cartesian differential categories provide the categorial foundations of **forward differentiation**

  R. Blute, R. Cockett, R.A.G. Seely, **Cartesian Differential Categories**

  Cartesian differential categories have a rich literature and have been successful in formalizing important concepts from differential calculus.

- Cartesian reverse differential categories provide the categorial foundations of **reverse differentiation**

  Cockett, R., Cruttwell, G., Gallagher, J., Lemay, J. S. P., MacAdam, B., Plotkin, G., & Pronk, D. (2020). *Reverse derivative categories.* In the proceedings of CSL2020.

  and have become quite popular, but still in the early stages...

  G. Cruttwell,& J.-S. P. Lemay *Reverse Tangent Categories.* CSL2024 (2024)

  G. Cruttwell, J. Gallagher, J.-S. P. Lemay & D. Pronk *Monoidal reverse differential categories.* MSCS (2022)

  Wilson, P., & Zanasi, F. *Categories of Differentiable Polynomial Circuits for Machine Learning.* ACT2022 (2022)

  Cruttwell, G., Gavranović, B., Ghani, N., Wilson, P., & Zanasi, F.: *Categorical foundations of gradient-based learning.* ESOP2022 (2022)

  Wilson, P., & Zanasi, F. *Derivative Ascent: A Categorical Approach to Learning Boolean Circuits.* ACT2020

  Cruttwell, G., Gallagher, J., & Pronk, D. (2020) *Categorical semantics of a simple differential programming language.* ACT2020

Today's Story

- Cartesian differential categories come equipped a **forward differentiation**, whose axioms includes the **chain rule**, expressing the forward derivative of a composition.

- Cartesian reverse differential categories come equipped a **reverse differential combinator**, whose axioms includes the **reverse chain rule**, expressing the reverse derivative of a composition.

Today's Story

- Cartesian differential categories come equipped a **forward differentiation**, whose axioms includes the **chain rule**, expressing the forward derivative of a composition.

- Cartesian reverse differential categories come equipped a **reverse differential combinator**, whose axioms includes the **reverse chain rule**, expressing the reverse derivative of a composition.

- The main theorem about reverse differentiation is that:

  $$\textbf{reverse differentiation} = \textbf{forward differentiation} + \textbf{transpose operation}$$

  So every Cartesian reverse differential category is a Cartesian differential category with a linear transpose operator, and vice-versa. In particular, the reverse derivative is the transpose of the forward derivative, and vice-versa.

- So when developing and studying the theory of reverse differentiation, we can take the forward differentiation concepts and see what the transpose operation gives us.

Today's Story

- **Faà di Bruno's Formula** provides a higher-order chain rule for the formula of a higher-order forward derivative of a composition.

- Faà di Bruno's Formula also holds in a Cartesian differential category

Cockett, J.R.B. and Seely, R.A.G. **The Faa di bruno construction**. (2011)

Garner, R, and Lemay, J-S P. **Cartesian differential categories as skew enriched categories**.

Today's Story

- **Faà di Bruno's Formula** provides a higher-order chain rule for the formula of a higher-order forward derivative of a composition.

- Faà di Bruno's Formula also holds in a Cartesian differential category

Cockett, J.R.B. and Seely, R.A.G. **The Faa di bruno construction**. (2011)

Garner, R, and Lemay, J-S P. **Cartesian differential categories as skew enriched categories**.

- **TODAY'S STORY:** Provide a reverse differential version of Faà di Bruno's Formula for a higher-order reverse chain rule in a Cartesian reverse differential category.

How we will get there

- To express Faà di Bruno's Formula in a Cartesian differential category, we need:
  - Partial Forward Differentiation
  - Higher-Order Forward Derivatives

- So to express the reverse Faà di Bruno's Formula we will need to properly define:
  - Partial Reverse Differentiation
  - Higher-Order Reverse Derivatives

  in a Cartesian reverse differential category, such that they are indeed the transpose of their forward mode counterparts. So in our ACT paper, we developed partial reverse derivation which allowed us to revisit and give a better understanding of some of the results in:

Cockett, R., Cruttwell, G., Gallagher, J., Lemay, J. S. P., MacAdam, B., Plotkin, G., & Pronk, D. (2020). *Reverse derivative categories.* In the proceedings of CSL2020.

How we will get there

- To express Faà di Bruno's Formula in a Cartesian differential category, we need:
  - Partial Forward Differentiation
  - Higher-Order Forward Derivatives
- So to express the reverse Faà di Bruno's Formula we will need to properly define:
  - Partial Reverse Differentiation
  - Higher-Order Reverse Derivatives

  in a Cartesian reverse differential category, such that they are indeed the transpose of their forward mode counterparts. So in our ACT paper, we developed partial reverse derivation which allowed us to revisit and give a better understanding of some of the results in:

  Cockett, R., Cruttwell, G., Gallagher, J., Lemay, J. S. P., MacAdam, B., Plotkin, G., & Pronk, D. (2020). *Reverse derivative categories.* In the proceedings of CSL2020.

- The reverse Faà di Bruno's Formula is then computed by taking the transpose of the Faà di Bruno's Formula. Surprisingly the reverse Faà di Bruno's Formula involves both reverse derivatives and forward derivatives!

Cartesian Left k-Linear Categories – Definition

The underlying category of a Cartesian (reverse) differential is a Cartesian left k-linear category.

Definition

For a commutative semiring k, a **left k-linear category** is a category $\mathbb{X}$ with finite products such that each homset $\mathbb{X}(A, B)$ is a k-module such that pre-composition preserves the k-linear structure:

$$(s \cdot f + t \cdot g) \circ x = s \cdot (f \circ x) + t \cdot (g \circ x)$$

A map $f : A \to B$ is said to be k-**linear** if post-composition by f preserves the k-linear structure:

$$f \circ (s \cdot x + t \cdot y) = s \cdot (f \circ x) + t \cdot (f \circ y)$$

A **Cartesian left k-linear category** is a Cartesian left k-linear with finite products such that the projection maps $\pi_j : A_1 \times \ldots \times A_n \to A_j$ are k-linear.

Definition

A **Cartesian k-differential category** (CDC) is a Cartesian left k-linear category which comes equipped with a **forward differential combinator**:

$$\frac{f : A \to B}{\mathrm{D}[f] : A \times A \to B}$$

where $\mathrm{D}[f]$ is called the **forward derivative** of f, and satisfies seven axioms.

Cartesian Differential Categories – Definition

Definition

A **Cartesian k-differential category** (CDC) is a Cartesian left k-linear category which comes equipped with a **forward differential combinator**:

$$\frac{f : A \to B}{\mathrm{D}[f] : A \times A \to B}$$

where $\mathrm{D}[f]$ is called the **forward derivative** of f, and satisfies seven axioms.

There is a very practical term logic. So we write:

$$\mathrm{D}[f](a, b) := \frac{\mathrm{d}f(x)}{\mathrm{d}x}(a) \cdot b$$

In particular, the chain rule is:

$$\mathrm{D}[g \circ f](a, b) = \mathrm{D}[g]\left(f(a), \mathrm{D}[f](a, b)\right) \qquad \frac{\mathrm{d}g(f(x))}{\mathrm{d}x}(a) \cdot b = \frac{\mathrm{d}g(x)}{\mathrm{d}x}(f(a)) \cdot \left(\frac{\mathrm{d}f(x)}{\mathrm{d}x}(a) \cdot b\right)$$

Example

Let SMOOTH be the Lawvere theory of real smooth functions, that is, whose objects are Euclidean spaces $\mathbb{R}^n$ and whose maps are real smooth functions $F : \mathbb{R}^n \to \mathbb{R}^m$. Note we have that $F = \langle f_1, \dots, f_n \rangle$ for real smooth functions $f_i : \mathbb{R}^n \to \mathbb{R}$.

Then SMOOTH is a CDC, where for $F : \mathbb{R}^n \to \mathbb{R}^m$, its derivative $\mathsf{D}[F] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ is then defined as:

$$\mathsf{D}[F](\vec{x}, \vec{y}) = \left\langle \sum_{i=1}^{n} \frac{\partial f_1}{\partial x_i}(\vec{x}) y_i, \dots, \sum_{i=1}^{n} \frac{\partial f_m}{\partial x_i}(\vec{x}) y_i \right\rangle$$

which is the *total* derivative of F.

In particular, for a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, its derivative $\mathsf{D}[f] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is just the sum of its partial derivatives:

$$\mathsf{D}[f](\vec{x}, \vec{y}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{x}) y_i$$

Definition

A **Cartesian k-reverse differential category** (CRDC) is a Cartesian left k-linear category which comes equipped with a **reverse differential combinator**:

$$\frac{f : A \to B}{\mathrm{R}[f] : A \times B \to A}$$

where $\mathrm{R}[f]$ is called the **reverse derivative** of f, and satisfies seven axioms.

Cartesian Reverse Differential Categories – Definition

Definition

A **Cartesian** k**-reverse differential category** (CRDC) is a Cartesian left k-linear category which comes equipped with a **reverse differential combinator**:

$$\frac{f : A \to B}{\mathrm{R}[f] : A \times B \to A}$$

where $\mathrm{R}[f]$ is called the **reverse derivative** of f, and satisfies seven axioms.

Inspired by the term logic for CDC, we use this notation[a]:

$$\mathrm{R}[f](a, b) := \frac{\mathsf{r} f(x)}{\mathsf{r} x}(a) \cdot b$$

In particular, the reverse chain rule is:

$$\mathrm{R}[g \circ f](a, b) = \mathrm{R}[f]\left(a, \mathrm{R}[g](f(a), b)\right) \qquad \frac{\mathsf{r} g\left(f(x)\right)}{\mathsf{r} x}(a) \cdot b = \frac{\mathsf{r} f(x)}{\mathsf{r} x}(a) \cdot \left(\frac{\mathsf{r} g(y)}{\mathsf{r} y}(f(a)) \cdot b\right)$$

[a] We leave it to future work to give a sound and complete term logic for CRDC

Cartesian Reverse Differential Categories – Smooth Functions

Example

SMOOTH is a CRDC, where for $F : \mathbb{R}^n \to \mathbb{R}^m$, its reverse derivative $\mathsf{R}[F] : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is defined as:

$$\mathsf{R}[F](\vec{x}, \vec{y}) = \left\langle \sum_{i=1}^{m} \frac{\partial f_i}{\partial x_1}(\vec{x}) y_i, \ldots, \sum_{i=1}^{m} \frac{\partial f_i}{\partial x_n}(\vec{x}) y_i \right\rangle$$

In particular, for a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, its reverse derivative $\mathsf{R}[f] : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is just the tuple of its partial derivatives:

$$\mathsf{R}[f](\vec{x}, y) = \left\langle \frac{\partial f}{\partial x_1}(\vec{x}) y, \ldots, \frac{\partial f}{\partial x_n}(\vec{x}) y \right\rangle$$

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger***.*

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation = **forward differentiation** + **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger***.*

To help express this, it will useful to talk about **partial reverse derivatives**!

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger***.*

To help express this, it will useful to talk about **partial reverse derivatives**!

But first the forward version.

Partial Forward Derivatives

In a CDC, for a map $f : A_1 \times \ldots \times A_n \to B$, we can take its partial forward derivative with respect to A_j while keeping the rest constant.

Partial Forward Derivatives

In a CDC, for a map $f : A_1 \times \ldots \times A_n \to B$, we can take its partial forward derivative with respect to A_j while keeping the rest constant. We do this by inserting zeroes into the total derivative.

Definition

In a CDC, for a map $f : A_1 \times \ldots \times A_n \to B$, its **$j$-th partial forward derivative** is the map $\mathsf{D}_j[f] : A_1 \times \ldots \times A_n \times A_j \to B$ defined as:

$$\frac{\mathsf{d}f(a_1, \ldots, a_{j-1}, x_j, a_{j+1}, \ldots, a_n)}{\mathsf{d}x_j}(a_j) \cdot b := \frac{\mathsf{d}f(x_1, \ldots, x_n)}{\mathsf{d}(x_1, \ldots, x_n)}(a_1, \ldots, a_n) \cdot (0, \ldots, 0, b, 0, \ldots, 0)$$

Partial Forward Derivatives

In a CDC, for a map $f : A_1 \times \ldots \times A_n \to B$, we can take its partial forward derivative with respect to A_j while keeping the rest constant. We do this by inserting zeroes into the total derivative.

Definition

In a CDC, for a map $f : A_1 \times \ldots \times A_n \to B$, its **$j$-th partial forward derivative** is the map $\mathsf{D}_j[f] : A_1 \times \ldots \times A_n \times A_j \to B$ defined as:

$$\frac{\mathsf{d}f(a_1, \ldots, a_{j-1}, x_j, a_{j+1}, \ldots, a_n)}{\mathsf{d}x_j}(a_j) \cdot b := \frac{\mathsf{d}f(x_1, \ldots, x_n)}{\mathsf{d}(x_1, \ldots, x_n)}(a_1, \ldots, a_n) \cdot (0, \ldots, 0, b, 0, \ldots, 0)$$

Lemma

In a CDC, for a map $f : A_1 \times \ldots \times A_n \to B$:

$$\frac{\mathsf{d}f(x_1, \ldots, x_n)}{\mathsf{d}(x_1, \ldots, x_n)}(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = \sum_{j=1}^{n} \frac{\mathsf{d}f(a_1, \ldots, a_{j-1}, x_j, a_{j+1}, \ldots, a_n)}{\mathsf{d}x_j}(a_j) \cdot b_j$$

Partial Reverse Derivatives **New**

In a CRDC, for a map $f : A_1 \times \ldots \times A_n \to B$, we wish to take its partial reverse derivative with respect to A_j while keeping the rest constant.

So we want a map of type $\mathsf{R}_j[f] : A_1 \times \ldots \times A_n \times B \to A_j$.

Partial Reverse Derivatives **New**

In a CRDC, for a map $f : A_1 \times \ldots \times A_n \to B$, we wish to take its partial reverse derivative with respect to A_j while keeping the rest constant.

So we want a map of type $\mathsf{R}_j[f] : A_1 \times \ldots \times A_n \times B \to A_j$.

Well consider the total reverse derivative $\mathsf{R}[f] : A_1 \times \ldots \times A_n \times B \to A_1 \times \ldots \times A_n$.

Partial Reverse Derivatives **New**

In a CRDC, for a map $f : A_1 \times \ldots \times A_n \to B$, we wish to take its partial reverse derivative with respect to A_j while keeping the rest constant.

So we want a map of type $\mathsf{R}_j[f] : A_1 \times \ldots \times A_n \times B \to A_j$.

Well consider the total reverse derivative $\mathsf{R}[f] : A_1 \times \ldots \times A_n \times B \to A_1 \times \ldots \times A_n$.

Then we get $\mathsf{R}_j[f]$ simply by projection!

Definition

In a CRDC, for a map $f : A_1 \times \ldots \times A_n \to B$, its **$j$-th partial reverse derivative** is the map $\mathsf{R}_j[f] : A_1 \times \ldots \times A_n \times B \to A_j$ defined as:

$$\frac{\mathsf{r}f(a_1, \ldots, a_{j-1}, x_j, a_{j+1}, \ldots, a_n)}{\mathsf{r}x_j}(a_j) \cdot b := \pi_j \left(\frac{\mathsf{r}f(x_1, \ldots, x_n)}{\mathsf{r}(x_1, \ldots, x_n)}(a_1, \ldots, a_n) \cdot b \right)$$

Partial Reverse Derivatives **New**

In a CRDC, for a map $f : A_1 \times \ldots \times A_n \to B$, we wish to take its partial reverse derivative with respect to A_j while keeping the rest constant.

So we want a map of type $\mathsf{R}_j[f] : A_1 \times \ldots \times A_n \times B \to A_j$.

Well consider the total reverse derivative $\mathsf{R}[f] : A_1 \times \ldots \times A_n \times B \to A_1 \times \ldots \times A_n$.

Then we get $\mathsf{R}_j[f]$ simply by projection!

Definition

In a CRDC, for a map $f : A_1 \times \ldots \times A_n \to B$, its **$j$-th partial reverse derivative** is the map $\mathsf{R}_j[f] : A_1 \times \ldots \times A_n \times B \to A_j$ defined as:

$$\frac{\mathsf{r}f(a_1, \ldots, a_{j-1}, x_j, a_{j+1}, \ldots, a_n)}{\mathsf{r}x_j}(a_j) \cdot b := \pi_j \left(\frac{\mathsf{r}f(x_1, \ldots, x_n)}{\mathsf{r}(x_1, \ldots, x_n)}(a_1, \ldots, a_n) \cdot b \right)$$

Lemma

In a CRDC, for a map $f : A_1 \times \ldots \times A_n \to B$:

$$\frac{\mathsf{r}f(x_1, \ldots, x_n)}{\mathsf{r}(x_1, \ldots, x_n)}(a_1, \ldots, a_n) \cdot b = \left\langle \frac{\mathsf{r}f(x_1, a_2, \ldots, a_n)}{\mathsf{r}x_1}(a_1) \cdot b, \ldots, \frac{\mathsf{r}f(a_1, \ldots, x_n)}{\mathsf{r}x_n}(a_n) \cdot b \right\rangle$$

Partial reverse derivatives have been quite useful, and they have given us a useful new perspective on some of the identities in:

Cockett, R., Cruttwell, G., Gallagher, J., Lemay, J. S. P., MacAdam, B., Plotkin, G., & Pronk, D. (2020). *Reverse derivative categories.* In the proceedings of CSL2020.

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger***.*

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger**.

In a CRDC, the **forward differential combinator** is defined as follows:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x}(a)\cdot b := \frac{\mathrm{r}\frac{\mathrm{r}f(x)}{\mathrm{r}x}(a)\cdot y}{\mathrm{r}y}(0)\cdot b$$

So essentially the partial reverse derivative in the second argument of the total reverse derivative.

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger**.

In a CRDC, the **forward differential combinator** is defined as follows:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x}(a) \cdot b := \frac{\mathrm{r}\frac{\mathrm{r}f(x)}{\mathrm{r}x}(a) \cdot y}{\mathrm{r}y}(0) \cdot b$$

So essentially the partial reverse derivative in the second argument of the total reverse derivative.

To talk about the transpose, we first need to talk about **linearity**!

D-Linearity

Now a map is D-**linear** in an argument if when forward differentiating with respect to that argument, one gets back the starting map.

D-Linearity

Now a map is D-**linear** in an argument if when forward differentiating with respect to that argument, one gets back the starting map.

Definition

In a CDC, a map $f : A_1 \times \ldots \times A_n \to B$ is said to be **differential linear (**D-**linear)** in A_j if when taking its j-th partial forward derivative, the following equality holds:

$$\frac{\mathsf{d}f(a_1, \ldots, a_{j-1}, x_j, a_{j+1}, \ldots, a_n)}{\mathsf{d}x_j}(a_j) \cdot b = f(a_1, \ldots, a_{j-1}, b, a_{j+1}, \ldots, a_n)$$

Example

In SMOOTH, being D-linear in argument is the same as being $\mathbb{R}$-linear. (However, in an arbitrary CDC, while D-linearity implies k-linearity, the converse is not necessarily true.)

Contextual Linear Dagger

A contextual linear dagger is an involutive and contravariant operation on maps with a D-linear argument, which swaps the codomain with said D-linear argument.

Definition

A CDC is said to have a **contextual linear dagger** if it comes equipped with an operator † which for every map:

$$f : C_1 \times A \times C_2 \to B$$

which is D-linear in A, gets associated to a map of type:

$$f^{\dagger[C_1 \times \_ \times C_2]} : C_1 \times B \times C_2 \to A$$

which is D-linear in B, and is called the D-**linear transpose in** A of f. Moreover, † is required to be involutative, contravariant, and behaves well with the product structure.

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger**.

In a CRDC, the **forward differential combinator** is defined as follows:

$$\frac{\mathsf{d}f(x)}{\mathsf{d}x}(a)\cdot b := \frac{\mathsf{r}\frac{\mathsf{r}f(x)}{\mathsf{r}x}(a)\cdot y}{\mathsf{r}y}(0)\cdot b$$

Cartesian Reverse Differential Categories – Fundamental Theorem

reverse differentiation $=$ **forward differentiation** $+$ **transpose operation**

Theorem

A Cartesian reverse differential category is the same as a Cartesian differential category equipped with a **contextual linear dagger***.*

In a CRDC, the **forward differential combinator** is defined as follows:

$$\frac{\mathsf{d}f(x)}{\mathsf{d}x}(a)\cdot b := \frac{\mathsf{r}\frac{\mathsf{r}f(x)}{\mathsf{r}x}(a)\cdot y}{\mathsf{r}y}(0)\cdot b$$

while the contextual linear dagger is defined as follows:

$$f^{\dagger[C_1\times\_\times C_2]}(c_1,b,c_2) := \frac{\mathsf{r}f(c_1,x,c_2)}{\mathsf{r}x}(0)\cdot b$$

So the partial reverse derivative in the D-linear variable.

Forward Differentiation and Reverse Differentiation are Transposes!

In a CRDC, we have that:

$$\mathsf{R}[f]^{\dagger[A \times \_]} = \mathsf{D}[f] \qquad\qquad \mathsf{D}[f]^{\dagger[A \times \_]} = \mathsf{R}[f]$$

So the total **forward derivative** is D-linear transpose of the **reverse derivative** (and vice-versa).

Forward Differentiation and Reverse Differentiation are Transposes!

In a CRDC, we have that:

$$\mathsf{R}[f]^{\dagger[A \times \_]} = \mathsf{D}[f] \qquad \mathsf{D}[f]^{\dagger[A \times \_]} = \mathsf{R}[f]$$

So the total **forward derivative** is D-linear transpose of the **reverse derivative** (and vice-versa). The same is true for the partial (reverse) differentiation!

Lemma

In a CRDC

$$\mathsf{R}_j[f]^{\dagger[A_1 \times \ldots \times A_n \times \_]} = \mathsf{D}_j[f] \qquad \mathsf{D}_j[f]^{\dagger[A_1 \times \ldots \times A_n \times \_]} = \mathsf{R}_j[f]$$

How we will get there

- To express Faà di Bruno's Formula in a Cartesian differential category, we need:
  - Partial Forward Differentiation
  - Higher-Order Forward Derivatives
- So to express the reverse Faà di Bruno's Formula we will need to properly define:
  - Partial Reverse Differentiation
  - Higher-Order Reverse Derivatives

  in a Cartesian reverse differential category, such that they are indeed the transpose of their forward mode counterparts.
- The reverse Faà di Bruno's Formula is then computed by taking the transpose of the Faà di Bruno's Formula. Surprisingly the reverse Faà di Bruno's Formula involves both reverse derivatives and forward derivatives!

Higher Order Forward Derivatives

In a CDC, we can apply the differential order n-times to get:

$$\mathsf{D}^n[f] : A^{\times^{2^n}} \to B$$

called the ***n*-th total forward derivative** of f.

Higher Order Forward Derivatives

In a CDC, we can apply the differential order n-times to get:

$$\mathrm{D}^n[f] : A^{\times^{2^n}} \to B$$

called the ***n*-th total forward derivative** of f.

However, by the axioms of a CDC there is a lot of redundant information in $\mathrm{D}^n[f]$.

For example, the second total forward derivative can be worked out to be:

$$\frac{\mathrm{d}\frac{\mathrm{d}f(x)}{\mathrm{d}x}(y) \cdot z}{\mathrm{d}(y,z)}(a,b) \cdot (c,d) = \frac{\mathrm{d}\frac{\mathrm{d}f(x)}{\mathrm{d}x}(y) \cdot b}{\mathrm{d}y} a \cdot c + \frac{\mathrm{d}f(x)}{\mathrm{d}x}(a) \cdot d$$

So we see that $\mathrm{D}^2[f]$ has a $\mathrm{D}[f]$ summand – which does not tell us any new information about f. Instead, all the new information comes from differentiating the first argument repeatedly.

Higher Order Forward Derivatives

For a map $f : A \to B$, its ***n*-th forward derivative** is defined by continuously deriving the first argument of the derivative:

$$A \to B$$

$$A \times A \to B$$

$$A \times A \times A \to B$$

$$\vdots$$

$$A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Higher Order Forward Derivatives

For a map $f : A \to B$, its n**-th forward derivative** is defined by continuously deriving the first argument of the derivative:

$$A \to B$$
$$A \times A \to B$$
$$A \times A \times A \to B$$
$$\vdots$$
$$A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Definition

In a CDC, for a map $f : A \to B$, its **n-th forward derivative** is the map $\partial^{(n)}[f] : A \times A^{\times^n} \to B$, which we write as:

$$\partial^{(n)}[f](a_0, a_1, \ldots, a_n) := \frac{\mathrm{d}^{(n)} f(x)}{\mathrm{d}x}(a_0) \cdot a_1 \cdot \ldots \cdot a_n$$

and is defined inductively as:

$$\frac{\mathrm{d}^{(0)} f(x)}{\mathrm{d}x}(a_0) = f(a_0) \qquad \frac{\mathrm{d}^{(n+1)} f(x)}{\mathrm{d}x}(a_0) \cdot a_1 \cdot \ldots \cdot a_n \cdot a_{n+1} = \frac{\mathrm{d}\frac{\mathrm{d}^{(n)} f(x)}{\mathrm{d}x}(y) \cdot a_1 \cdot \ldots \cdot a_n}{\mathrm{d}y}(a_0) \cdot a_{n+1}$$

The n-th derivative is D-linear in its last n-arguments and also symmetric in its last n-arguments.

Higher Order Reverse Derivatives

In a CRDC, we can apply the reverse differential combinator:

$$A \to B$$
$$A \times B \to A$$
$$A \times B \times A \to A \times B$$
$$A \times B \times A \times A \times B \to A \times B \times A$$
$$\vdots$$

So the type of $\mathsf{R}^n[f]$, the ***n*-th total reverse derivative** of f, is quite complex with a large number of inputs and a large number of outputs.

Higher Order Reverse Derivatives

In a CRDC, we can apply the reverse differential combinator:

$$A \to B$$
$$A \times B \to A$$
$$A \times B \times A \to A \times B$$
$$A \times B \times A \times A \times B \to A \times B \times A$$
$$\vdots$$

So the type of $\mathsf{R}^n[f]$, the ***n*-th total reverse derivative** of f, is quite complex with a large number of inputs and a large number of outputs.

However, by the axioms of a CRDC there is a lot of redundant information in $\mathsf{R}^n[f]$ again.

For example, the second total reverse derivative can be worked out to be:

$$\frac{\mathsf{r}\frac{\mathsf{r}f(x)}{\mathsf{r}x}(y)\cdot z}{\mathsf{r}(y,z)}(a_1,b)\cdot a_2 = \left\langle \frac{\mathsf{r}\frac{\mathsf{r}f(x)}{\mathsf{r}x}(a_1)\cdot u}{\mathsf{r}u}(b)\cdot a_2, \frac{\mathsf{d}f(x)}{\mathsf{d}x}(a_1)\cdot a_2 \right\rangle$$

So we see that $\mathsf{R}^2[f]$ has a $\mathsf{R}[f]$ part...

Higher Order Reverse Derivatives **NEW**

For a map $f : A \to B$, its n-**th reverse derivative** is defined by continuously reverse deriving the first argument of the reverse derivative:

$$A \to B$$
$$A \times B \to A$$
$$A \times B \times A \to A$$
$$\vdots$$
$$A \times \times B \times \underbrace{A \times \ldots \times A}_{(n-1)\text{-times}} \to A$$

Higher Order Reverse Derivatives **NEW**

For a map $f : A \to B$, its n**-th reverse derivative** is defined by continuously reverse deriving the first argument of the reverse derivative:

$$A \to B$$
$$A \times B \to A$$
$$A \times B \times A \to A$$
$$\vdots$$
$$A \times \times B \times \underbrace{A \times \ldots \times A}_{(n-1)\text{-times}} \to A$$

Definition

In a CDC, for a map $f : A \to B$, its n**-th reverse derivative** is the map $\rho^{(n)}[f] : A \times B \times A^{\times^{(n-1)}} \to B$, which we write as:

$$\rho^{(n)}[f](a_0, b, a_2, \ldots, a_n) = \frac{\mathrm{r}^{(n)} f(x)}{\mathrm{r}x}(a_0) \cdot b \cdot a_2 \cdot \ldots \cdot a_n$$

and is defined inductively as:

$$\frac{\mathrm{r}^{(0)} f(x)}{\mathrm{r}x}(a_0) = f(a_0) \qquad \frac{\mathrm{r}^{(n+1)} f(x)}{\mathrm{r}x}(a_0) \cdot b \cdot a_1 \cdot \ldots \cdot a_{n+1} = \frac{\mathrm{r}\,\frac{\mathrm{r}^{(n)} f(x)}{\mathrm{d}x}(y) \cdot b \cdot \ldots \cdot a_n}{\mathrm{r}y}(a_0) \cdot a_{n+1}$$

Transpose?

We would like that the nth higher order forward derivative is the D-linear transpose of the nth higher order reverse derivative...

Transpose?

We would like that the *n*th higher order forward derivative is the D-linear transpose of the *n*th higher order reverse derivative...

The first thing to address is which D-linear argument should we transpose:

$$\partial^{(n)}[f] : A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Transpose?

We would like that the nth higher order forward derivative is the D-linear transpose of the nth higher order reverse derivative...

The first thing to address is which D-linear argument should we transpose:

$$\partial^{(n)}[f] : A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Well to fit with the type of:

$$\rho^{(n)}[f] : A \times \times B \times \underbrace{A \times \ldots \times A}_{(n-1)\text{-times}} \to A$$

We transpose the first A...

Transpose?

We would like that the *n*th higher order forward derivative is the D-linear transpose of the *n*th higher order reverse derivative...

The first thing to address is which D-linear argument should we transpose:

$$\partial^{(n)}[f] : A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Well to fit with the type of:

$$\rho^{(n)}[f] : A \times \times B \times \underbrace{A \times \ldots \times A}_{(n-1)\text{-times}} \to A$$

We transpose the first A... $\partial^{(n)}[f]$ but since is symmetric in its D-linear arguments, it does not matter which one we take really.

Transpose?

We would like that the *n*th higher order forward derivative is the D-linear transpose of the *n*th higher order reverse derivative...

The first thing to address is which D-linear argument should we transpose:

$$\partial^{(n)}[f] : A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Well to fit with the type of:

$$\rho^{(n)}[f] : A \times \times B \times \underbrace{A \times \ldots \times A}_{(n-1)\text{-times}} \to A$$

We transpose the first A... $\partial^{(n)}[f]$ but since is symmetric in its D-linear arguments, it does not matter which one we take really.

We want this:

$$\partial^{(n)}[f]^{\dagger[A \times \_ \times A^n]} = \rho^{(n)}[f]$$

Transpose?

We would like that the nth higher order forward derivative is the D-linear transpose of the nth higher order reverse derivative...

The first thing to address is which D-linear argument should we transpose:

$$\partial^{(n)}[f] : A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Well to fit with the type of:

$$\rho^{(n)}[f] : A \times \times B \times \underbrace{A \times \ldots \times A}_{(n-1)\text{-times}} \to A$$

We transpose the first A... $\partial^{(n)}[f]$ but since is symmetric in its D-linear arguments, it does not matter which one we take really.

We want this:

$$\partial^{(n)}[f]^{\dagger[A \times \_ \times A^n]} = \rho^{(n)}[f]$$

Unfortunately, this does not seem to follow from just the axioms of the reverse differential combinator...

Transpose?

We would like that the nth higher order forward derivative is the D-linear transpose of the nth higher order reverse derivative...

The first thing to address is which D-linear argument should we transpose:

$$\partial^{(n)}[f] : A \times \underbrace{A \times \ldots \times A}_{n\text{-times}} \to B$$

Well to fit with the type of:

$$\rho^{(n)}[f] : A \times \times B \times \underbrace{A \times \ldots \times A}_{(n-1)\text{-times}} \to A$$

We transpose the first A... $\partial^{(n)}[f]$ but since is symmetric in its D-linear arguments, it does not matter which one we take really.

We want this:

$$\partial^{(n)}[f]^{\dagger[A \times \_ \times A^n]} = \rho^{(n)}[f]$$

Unfortunately, this does not seem to follow from just the axioms of the reverse differential combinator...

So we introduce an extra compatibility between the forward differential combinator and the reverse differential combinator.

Stable Rule

Definition

A CRDC is said to satisfy the **stable rule** if:

$$\frac{\mathsf{r}\frac{\mathsf{d}f(x)}{x}(y)\cdot a_2}{\mathsf{r}y}(a_1)\cdot b = \frac{\mathsf{r}\frac{\mathsf{r}f(x)}{x}(y)\cdot b}{\mathsf{d}y}(a_1)\cdot a_2$$

All examples of CRDC that we have satisfy the stable rule, so in particular SMOOTH satisfies the stable rule. (There is a linear closed reason for this...)

And yet this rule does not seem to follow from the reverse differential combinator axioms! (I could be wrong... I hope I'm hope wrong!)

Stable Rule

Definition

A CRDC is said to satisfy the **stable rule** if:

$$\frac{\mathsf{r}\dfrac{\mathsf{d}f(x)}{x}(y)\cdot a_2}{\mathsf{r}y}(a_1)\cdot b = \frac{\mathsf{r}\dfrac{\mathsf{r}f(x)}{x}(y)\cdot b}{\mathsf{d}y}(a_1)\cdot a_2$$

All examples of CRDC that we have satisfy the stable rule, so in particular SMOOTH satisfies the stable rule. (There is a linear closed reason for this...)

And yet this rule does not seem to follow from the reverse differential combinator axioms! (I could be wrong... I hope I'm hope wrong!)

Lemma

In a CRDC which satisfies the stable rule:

$$\partial^{(n)}[f]^{\dagger[A\times\_\times A^n]} = \rho^{(n)}[f]$$

Faà di Bruno Formula

The Faà di Bruno Formula is given by a sum indexed by non-empty partitions of finite sets.

Faà di Bruno Formula

The Faà di Bruno Formula is given by a sum indexed by non-empty partitions of finite sets.

- For every $n \in \mathbb{N}$, define the totally ordered set $[n+1] = \{1 < \ldots < n+1\}$.
- For every subset $I = \{i_1 < \ldots < i_m\} \subseteq [n+1]$, for a vector $\vec{x} = (x_1, \ldots, x_{n+1})$, define $\vec{x}|_I = (x_{i_1}, \ldots, x_{i_m})$
- We denote a *non-empty* partition of $[n+1]$ as $[n+1] = A_1|\ldots|A_k$, and let $|A_j|$ be the cardinality of A_j.

Faà di Bruno Formula

The Faà di Bruno Formula is given by a sum indexed by non-empty partitions of finite sets.

- For every $n \in \mathbb{N}$, define the totally ordered set $[n+1] = \{1 < \ldots < n+1\}$.
- For every subset $I = \{i_1 < \ldots < i_m\} \subseteq [n+1]$, for a vector $\vec{x} = (x_1, \ldots, x_{n+1})$, define $\vec{x}|_I = (x_{i_1}, \ldots, x_{i_m})$
- We denote a *non-empty* partition of $[n+1]$ as $[n+1] = A_1|\ldots|A_k$, and let $|A_j|$ be the cardinality of A_j.

Theorem

In a CDC,

$$\frac{\mathrm{d}^{(n+1)}g(f(x))}{\mathrm{d}x}(a_0) \cdot a_1 \cdot a_2 \cdot \ldots \cdot a_{n+1}$$

$$= \sum_{[n+1]=A_1|\ldots|A_k} \frac{\mathrm{d}^{(k)}g(z)}{\mathrm{d}z}(f(a_0)) \cdot \left(\frac{\mathrm{d}^{(|A_1|)}f(x)}{\mathrm{d}x}(a_0) \cdot \vec{a}|_{A_1} \right) \cdot \ldots \cdot \left(\frac{\mathrm{d}^{(|A_k|)}f(x)}{\mathrm{d}x}(a_0) \cdot \vec{a}|_{A_k} \right)$$

Reverse (Mode?) Faà di Bruno Formula

We are finally in a position to give the reverse Faà di Bruno's Formula.

Reverse (Mode?) Faà di Bruno Formula

We are finally in a position to give the reverse Faà di Bruno's Formula.

- For convenience, we will assume that in a non-empty partition $[n+1] = A_1 | \ldots | A_k$, that $1 \in [n+1]$ is always in $1 \in A_1$.

Reverse (Mode?) Faà di Bruno Formula

We are finally in a position to give the reverse Faà di Bruno's Formula.

- For convenience, we will assume that in a non-empty partition $[n+1] = A_1|\ldots|A_k$, that $1 \in [n+1]$ is always in $1 \in A_1$.

Theorem

In a CRDC which satisfies the stable rule:

$$\frac{\mathsf{r}^{(n+1)} g(f(x))}{\mathsf{r}x}(a_0) \cdot b \cdot a_2 \cdot \ldots \cdot a_{n+1} =$$

$$\sum_{\substack{[n+1]=A_1|\ldots|A_k \\ 1 \in A_1}} \frac{\mathsf{r}^{(|A_1|)} f(x)}{\mathsf{r}x}(a_0) \cdot \left(\frac{\mathsf{r}^{(k)} g(y)}{\mathsf{r}y}(a_0) \cdot b \cdot \left(\frac{\mathsf{d}^{(|A_2|)} f(x)}{\mathsf{d}x}(a_0) \cdot \vec{a}|_{A_2} \right) \cdot \ldots \cdot \left(\frac{\mathsf{d}^{(|A_k|)} f(x)}{\mathsf{d}x}(a_0) \cdot \vec{a}|_{A_k} \right) \right)$$

The proof is by applying the linear transpose to the forward mode Faà di Bruno Formula.

Reverse (Mode?) Faà di Bruno Formula

We are finally in a position to give the reverse Faà di Bruno's Formula.

- For convenience, we will assume that in a non-empty partition $[n+1] = A_1|\ldots|A_k$, that $1 \in [n+1]$ is always in $1 \in A_1$.

Theorem

In a CRDC which satisfies the stable rule:

$$\frac{\mathsf{r}^{(n+1)} g(f(x))}{\mathsf{r}x}(a_0) \cdot b \cdot a_2 \cdot \ldots \cdot a_{n+1} =$$

$$\sum_{\substack{[n+1]=A_1|\ldots|A_k \\ 1 \in A_1}} \frac{\mathsf{r}^{(|A_1|)} f(x)}{\mathsf{r}x}(a_0) \cdot \left(\frac{\mathsf{r}^{(k)} g(y)}{\mathsf{r}y}(a_0) \cdot b \cdot \left(\frac{\mathsf{d}^{(|A_2|)} f(x)}{\mathsf{d}x}(a_0) \cdot \vec{a}|_{A_2} \right) \cdot \ldots \cdot \left(\frac{\mathsf{d}^{(|A_k|)} f(x)}{\mathsf{d}x}(a_0) \cdot \vec{a}|_{A_k} \right) \right)$$

The proof is by applying the linear transpose to the forward mode Faà di Bruno Formula.

The formula can be given without assuming $1 \in A_1$, but it's much more convenient!

This can be full written using only reverse differential operators!

Reverse (Mode?) Faà di Bruno Formula

- When $n = 0$ in the reverse Faà di Bruno's Formula, we get back precisely the reverse chain rule! The only non-empty partition of $[1]$ is $[1] = A_1 = \{1\}$, which is why no forward derivatives appear in the reverse chain rule.

$$\frac{\mathrm{r}g\left(f(x)\right)}{\mathrm{r}x}(a) \cdot b = \frac{\mathrm{r}f(x)}{\mathrm{r}x}(a) \cdot \left(\frac{\mathrm{r}g(y)}{\mathrm{r}y}(f(a)) \cdot b\right)$$

Reverse (Mode?) Faà di Bruno Formula

- When $n = 0$ in the reverse Faà di Bruno's Formula, we get back precisely the reverse chain rule! The only non-empty partition of [1] is $[1] = A_1 = \{1\}$, which is why no forward derivatives appear in the reverse chain rule.

$$\frac{\mathrm{r}g\,(f(x))}{\mathrm{r}x}(a)\cdot b = \frac{\mathrm{r}f(x)}{\mathrm{r}x}(a)\cdot\left(\frac{\mathrm{r}g(y)}{\mathrm{r}y}(f(a))\cdot b\right)$$

- When $n = 1$, the non-empty partitions of [2] are $[1] = A_1 = \{1\}|A_2 = \{2\}$ and $[2] = A_1 = \{1, 2\}$. So the reverse Faà di Bruno's Formula for the second reverse derivative is:

$$\begin{aligned}
&\frac{\mathrm{r}^{(2)}g\,(f(x))}{\mathrm{r}x}(a_0)\cdot b\cdot a_2 \\
&= \frac{\mathrm{r}^{(1)}f(x)}{\mathrm{r}x}(a_0)\cdot\left(\frac{\mathrm{r}^{(2)}g(y)}{\mathrm{r}y}(f(a_0))\cdot b\cdot\left(\frac{\mathrm{d}f(x)}{\mathrm{d}x}(a_0)\cdot a_2\right)\right) \\
&\quad + \frac{\mathrm{r}^{(2)}f(x)}{\mathrm{r}x}(a_0)\cdot\left(\frac{\mathrm{r}^{(1)}g(y)}{\mathrm{r}y}(f(a_0))\cdot b\right)\cdot a_2
\end{aligned}$$

- Try working out the higher levels! (make sure you have lots of space!)

Future Work: Cofree Cartesian Reverse Differential Categories?

- For every Cartesian k-left linear category, there exists a cofree CDC over it via what's called the **Faà di Bruno construction**, where composition in this cofree CDC is given by Faà di Bruno's formula.

Cockett, J.R.B. and Seely, R.A.G. **The Faa di bruno construction**. (2011)

Garner, R, and Lemay, J-S P. **Cartesian differential categories as skew enriched categories**.

Lemay, J-S P. **A Tangent Category Alternative to the Faa di Bruno Construction**.

Lemay, J-S P. **Properties and Characterisations of Cofree Cartesian Differential Categories**.

- Can the reverse Faà di Bruno's formula be used to defined cofree CRDC?

Thanks for listening!

Oktober Fest, Kellogg College, University of Oxford – 2017

HOPE YOU ENJOYED MY TALK!

THANK YOU FOR LISTENING!

MERCI!

js.lemay@mq.edu.au

https://sites.google.com/view/jspl-personal-webpage/