Correctness of Automatic Differentiation via Diffeologies and Categorical Gluing

Mathieu Huot[1] Sam Staton[1] Matthijs Vákár[2]

(1) University of Oxford, UK

(2) Utrecht University, The Netherlands

February 27, 2020, Pihoc, IRIF, Paris

Mathieu Huot, Sam Staton, and Matthijs Vákár. "Correctness of Automatic Differentiation via Diffeologies and Categorical Gluing". In: *arXiv preprint arXiv:2001.02209* (2020), to appear in FOSSACS 2020

Higher-order language, inductive types

↓

Denotational semantics

↓

Correctness of Automatic Differentiation via Diffeologies and Categorical Gluing

↑

Logical relations argument

⇒ Aim: Bridge the gap between syntactic AD and mathematics

1. Language for higher-order AD

2. Denotational semantics for a higher-order language

3. A gluing construction

4. Another example: Jet-based AD

Simple higher-order language

τ, σ, ρ	$::=$		types
	$\mid$	**real**	real numbers
	$\mid$	$(\tau_1 * \ldots * \tau_n)$	finite product
	$\mid$	$\{\ell_1\, \tau_1 \mid \ldots \mid \ell_n\, \tau_n\}$	variant
	$\mid$	$\tau \to \sigma$	function
	$\mid$	**list**(τ)	list

$\Rightarrow$ Higher-order

$\Rightarrow$ Inductive types

Simple higher-order language

t, s, r	$::=$	terms
$\mid$	x	variable
$\mid$	$\underline{c} \mid t + s \mid t * s \mid \varsigma(t)$	operations/constants
$\mid$	$\langle t_1, \ldots, t_n \rangle \mid \textbf{case}\ t\ \textbf{of}\ \langle x_1, \ldots, x_n \rangle \to s$	tuples/pattern matching
$\mid$	$\lambda x.t \mid t\, s$	function abstraction/app.
$\mid$	$\tau.\ell\ t$	variant constructor
$\mid$	$\textbf{case}\ t\ \textbf{of}\ \{\ell_1\, x_1 \to s_1 \mid \cdots \mid \ell_{\mathrm{n}}\, x_n \to s_n\}$	pattern matching: variants
$\mid$	$[\,] \mid t :: s$	empty list and cons
$\mid$	$\textbf{fold}\ (x_1, x_2).t\ \textbf{over}\ s\ \textbf{from}\ r$	list fold

Goal of traditional AD

Given $f \stackrel{\mathrm{def}}{=} f_1; f_2; \ldots; f_n : \mathbb{R}^k \to \mathbb{R}$,

compute $\nabla f \stackrel{\mathrm{def}}{=} \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_k}\right)$

using the chain rule,

assuming $\frac{\partial f_i}{\partial x_j}$ known.

AD as a source-code transformation

$$\overrightarrow{\mathcal{D}}(\textbf{real}) \stackrel{\text{def}}{=} (\textbf{real}*\textbf{real})$$

$$\overrightarrow{\mathcal{D}}(\tau \to \sigma) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\tau) \to \overrightarrow{\mathcal{D}}(\sigma)$$
$$\overrightarrow{\mathcal{D}}((\tau_1 * \cdots * \tau_n)) \stackrel{\text{def}}{=} (\overrightarrow{\mathcal{D}}(\tau_1) * \cdots * \overrightarrow{\mathcal{D}}(\tau_n))$$
$$\overrightarrow{\mathcal{D}}(\{\ell_1\,\tau_1 \mid \ldots \mid \ell_\mathsf{n}\,\tau_n\}) \stackrel{\text{def}}{=} \{\ell_1\,\overrightarrow{\mathcal{D}}(\tau_1) \mid \ldots \mid \ell_\mathsf{n}\,\overrightarrow{\mathcal{D}}(\tau_n)\}$$
$$\overrightarrow{\mathcal{D}}(\textbf{list}(\tau)) \stackrel{\text{def}}{=} \textbf{list}(\overrightarrow{\mathcal{D}}(\tau))$$

AD as a source-code transformation

$$\overrightarrow{\mathcal{D}}(\mathbf{real}) \stackrel{\text{def}}{=} (\mathbf{real} * \mathbf{real})$$

$$\overrightarrow{\mathcal{D}}(\tau \to \sigma) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\tau) \to \overrightarrow{\mathcal{D}}(\sigma)$$
$$\overrightarrow{\mathcal{D}}((\tau_1 * \cdots * \tau_n)) \stackrel{\text{def}}{=} (\overrightarrow{\mathcal{D}}(\tau_1) * \cdots * \overrightarrow{\mathcal{D}}(\tau_n))$$
$$\overrightarrow{\mathcal{D}}(\{\ell_1\, \tau_1 \mid \ldots \mid \ell_n\, \tau_n\}) \stackrel{\text{def}}{=} \{\ell_1\, \overrightarrow{\mathcal{D}}(\tau_1) \mid \ldots \mid \ell_n\, \overrightarrow{\mathcal{D}}(\tau_n)\}$$
$$\overrightarrow{\mathcal{D}}(\mathbf{list}(\tau)) \stackrel{\text{def}}{=} \mathbf{list}(\overrightarrow{\mathcal{D}}(\tau))$$

Non-trivial definition of $\overrightarrow{\mathcal{D}}$:

$$\overrightarrow{\mathcal{D}}(\underline{c}) \stackrel{\text{def}}{=} \langle \underline{c}, 0 \rangle$$
$$\overrightarrow{\mathcal{D}}(t + s) \stackrel{\text{def}}{=} \mathbf{case}\, \overrightarrow{\mathcal{D}}(t)\, \mathbf{of}\, \langle x, x' \rangle \to \mathbf{case}\, \overrightarrow{\mathcal{D}}(s)\, \mathbf{of}\, \langle y, y' \rangle \to \langle x + y, x' + y' \rangle$$
$$\overrightarrow{\mathcal{D}}(t * s) \stackrel{\text{def}}{=} \mathbf{case}\, \overrightarrow{\mathcal{D}}(t)\, \mathbf{of}\, \langle x, x' \rangle \to \mathbf{case}\, \overrightarrow{\mathcal{D}}(s)\, \mathbf{of}\, \langle y, y' \rangle \to \langle x * y, x * y' + x' * y \rangle$$
$$\overrightarrow{\mathcal{D}}(\varsigma(t)) \stackrel{\text{def}}{=} \mathbf{case}\, \overrightarrow{\mathcal{D}}(t)\, \mathbf{of}\, \langle x, x' \rangle \to \mathbf{let}\, y = \varsigma(x)\, \mathbf{in}\, \langle y, x' * y * (1 - y) \rangle$$

Structure preserving definition of $\overrightarrow{\mathcal{D}}$:

$$\overrightarrow{\mathcal{D}}(x) \stackrel{\text{def}}{=} x$$

$$\overrightarrow{\mathcal{D}}(\lambda x.t) \stackrel{\text{def}}{=} \lambda x.\overrightarrow{\mathcal{D}}(t) \qquad \overrightarrow{\mathcal{D}}(t\,s) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(t)\,\overrightarrow{\mathcal{D}}(s)$$

$$\overrightarrow{\mathcal{D}}(\langle t_1, \ldots, t_n\rangle) \stackrel{\text{def}}{=} \langle \overrightarrow{\mathcal{D}}(t_1), \ldots, \overrightarrow{\mathcal{D}}(t_n)\rangle$$

$$\overrightarrow{\mathcal{D}}(\textbf{case}\; t\; \textbf{of}\; \langle x_1, \ldots, x_n\rangle \to s) \stackrel{\text{def}}{=} \textbf{case}\; \overrightarrow{\mathcal{D}}(t)\; \textbf{of}\; \langle x_1, \ldots, x_n\rangle \to \overrightarrow{\mathcal{D}}(s)$$

$$\overrightarrow{\mathcal{D}}(\tau.\ell\; t) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\tau).\ell\, \overrightarrow{\mathcal{D}}(t) \qquad \overrightarrow{\mathcal{D}}([\,]) \stackrel{\text{def}}{=} [\,] \qquad \overrightarrow{\mathcal{D}}(t :: s) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(t) :: \overrightarrow{\mathcal{D}}(s)$$

$$\overrightarrow{\mathcal{D}}(\textbf{case}\; t\; \textbf{of}\; \{\ell_1\, x_1 \to s_1 \mid \cdots \mid \ell_n\, x_n \to s_n\}) \stackrel{\text{def}}{=}$$
$$\qquad \textbf{case}\; \overrightarrow{\mathcal{D}}(t)\; \textbf{of}\; \{\ell_1\, x_1 \to \overrightarrow{\mathcal{D}}(s_1) \mid \cdots \mid \ell_n\, x_n \to \overrightarrow{\mathcal{D}}(s_n)\}$$

$$\overrightarrow{\mathcal{D}}(\textbf{fold}\;(x_1, x_2).t\; \textbf{over}\; s\; \textbf{from}\; r) \stackrel{\text{def}}{=} \textbf{fold}\;(x_1, x_2).\overrightarrow{\mathcal{D}}(t)\; \textbf{over}\; \overrightarrow{\mathcal{D}}(s)\; \textbf{from}\; \overrightarrow{\mathcal{D}}(r)$$

Structure preserving definition of $\overrightarrow{\mathcal{D}}$:

$$\overrightarrow{\mathcal{D}}(x) \stackrel{\text{def}}{=} x$$

$$\overrightarrow{\mathcal{D}}(\lambda x.t) \stackrel{\text{def}}{=} \lambda x.\overrightarrow{\mathcal{D}}(t) \qquad \overrightarrow{\mathcal{D}}(t\,s) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(t)\,\overrightarrow{\mathcal{D}}(s)$$

$$\overrightarrow{\mathcal{D}}(\langle t_1, \ldots, t_n \rangle) \stackrel{\text{def}}{=} \langle \overrightarrow{\mathcal{D}}(t_1), \ldots, \overrightarrow{\mathcal{D}}(t_n) \rangle$$

$$\overrightarrow{\mathcal{D}}(\textbf{case}\; t\; \textbf{of}\; \langle x_1, \ldots, x_n \rangle \to s) \stackrel{\text{def}}{=} \textbf{case}\; \overrightarrow{\mathcal{D}}(t)\; \textbf{of}\; \langle x_1, \ldots, x_n \rangle \to \overrightarrow{\mathcal{D}}(s)$$

$$\overrightarrow{\mathcal{D}}(\tau.\ell\; t) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\tau).\ell\; \overrightarrow{\mathcal{D}}(t) \qquad \overrightarrow{\mathcal{D}}([\,]) \stackrel{\text{def}}{=} [\,] \qquad \overrightarrow{\mathcal{D}}(t :: s) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(t) :: \overrightarrow{\mathcal{D}}(s)$$

$$\overrightarrow{\mathcal{D}}(\textbf{case}\; t\; \textbf{of}\; \{\ell_1\, x_1 \to s_1 \mid \cdots \mid \ell_n\, x_n \to s_n\}) \stackrel{\text{def}}{=}$$
$$\qquad \textbf{case}\; \overrightarrow{\mathcal{D}}(t)\; \textbf{of}\; \{\ell_1\, x_1 \to \overrightarrow{\mathcal{D}}(s_1) \mid \cdots \mid \ell_n\, x_n \to \overrightarrow{\mathcal{D}}(s_n)\}$$

$$\overrightarrow{\mathcal{D}}(\textbf{fold}\; (x_1, x_2).t\; \textbf{over}\; s\; \textbf{from}\; r) \stackrel{\text{def}}{=} \textbf{fold}\; (x_1, x_2).\overrightarrow{\mathcal{D}}(t)\; \textbf{over}\; \overrightarrow{\mathcal{D}}(s)\; \textbf{from}\; \overrightarrow{\mathcal{D}}(r)$$

Lemma (Well-typedness and functoriality of $\overrightarrow{\mathcal{D}}$)

- *If* $\Gamma \vdash t : \tau$ *then* $\overrightarrow{\mathcal{D}}(\Gamma) \vdash \overrightarrow{\mathcal{D}}(t) : \overrightarrow{\mathcal{D}}(\tau)$
- *If* $\overrightarrow{\mathcal{D}}(t)[u/x] = \overrightarrow{\mathcal{D}}(t)[\overrightarrow{\mathcal{D}}(u)/x]$

AD as a source-code transformation (cont.)

Example (Inner product)

$$\Gamma \vdash t \cdot_n s \stackrel{\text{def}}{=} \quad \begin{array}{l} \textbf{case}\, t\, \textbf{of}\, \langle z_1, \ldots, z_n \rangle \to \\ \textbf{case}\, s\, \textbf{of}\, \langle y_1, \ldots, y_n \rangle \to z_1 * y_1 + \cdots + z_n * y_n : \textbf{real} \end{array}$$

AD as a source-code transformation (cont.)

Example (Inner product)

$$\Gamma \vdash t \cdot_n s \stackrel{\mathrm{def}}{=} \quad \mathbf{case}\, t\, \mathbf{of}\, \langle z_1, \ldots, z_n \rangle \to \\ \mathbf{case}\, s\, \mathbf{of}\, \langle y_1, \ldots, y_n \rangle \to z_1 * y_1 + \cdots + z_n * y_n : \mathbf{real}$$

Example ($x_1 : \mathbb{R}, x_2 : \mathbb{R} \vdash g : \mathbb{R}$)

- $\overrightarrow{\mathcal{D}}(g)((x_1, 1), (x_2, 0)) \stackrel{?}{=} (g(x_1, x_2), \frac{\partial g(x, x_2)}{\partial x}(x_1))$
- $\overrightarrow{\mathcal{D}}(g)((x_1, 0), (x_2, 1)) \stackrel{?}{=} (g(x_1, x_2), \frac{\partial g(x_1, x)}{\partial x}(x_2))$.

AD as a source-code transformation (cont.)

Example (Inner product)

$$\Gamma \vdash t \cdot_n s \stackrel{\text{def}}{=} \quad \textbf{case}\, t\, \textbf{of}\, \langle z_1, \ldots, z_n \rangle \to$$
$$\textbf{case}\, s\, \textbf{of}\, \langle y_1, \ldots, y_n \rangle \to z_1 * y_1 + \cdots + z_n * y_n : \textbf{real}$$

Example ($x_1 : \mathbb{R}, x_2 : \mathbb{R} \vdash g : \mathbb{R}$)

- $\overrightarrow{\mathcal{D}}(g)((x_1, 1), (x_2, 0)) \stackrel{?}{=} (g(x_1, x_2), \frac{\partial g(x, x_2)}{\partial x}(x_1))$
- $\overrightarrow{\mathcal{D}}(g)((x_1, 0), (x_2, 1)) \stackrel{?}{=} (g(x_1, x_2), \frac{\partial g(x_1, x)}{\partial x}(x_2))$.
- $\overrightarrow{\mathcal{D}}(g)((x_1, x_1'), (x_2, x_2')) \stackrel{?}{=} (g(x_1, x_2), x_1' \cdot \frac{\partial g(x, x_2)}{\partial x}(x_1) + x_2' \cdot \frac{\partial g(x_1, x)}{\partial x}(x_2))$

Main theorem: Correctness

Semantic correctness of $\overrightarrow{\mathcal{D}}$

τ : ground type $\qquad$ Γ : context with ground types only

$\Gamma \vdash t : \tau$ possibly with higher-order subterms

The syntactic translation $\overrightarrow{\mathcal{D}}$ coincides with the tangent bundle functor (modulo canonical isomorphisms):

$$
\begin{tikzcd}
[\![\overrightarrow{\mathcal{D}}(\Gamma)]\!] \arrow[r, "{[\![\overrightarrow{\mathcal{D}}(t)]\!]}"] \arrow[d, "\cong"'] & [\![\overrightarrow{\mathcal{D}}(\tau)]\!] \arrow[d, "\cong"] \\
\mathcal{T}([\![\Gamma]\!]) \arrow[r, "\mathcal{T}([\![t]\!])"'] & \mathcal{T}([\![\tau]\!])
\end{tikzcd}
$$

Tangent bundle $\qquad$ Denotational semantics

Denotational semantics for a first-order language

Definition (*CartSp*)

- *objects*(**CartSp**) : Euclidean spaces $\mathbb{R}^n$
- *morphisms*(**CartSp**) : smooth maps $\mathbb{R}^n \to \mathbb{R}^m$

$[\![\mathbf{real}]\!] = \mathbb{R}$
$[\![\tau_1 * \ldots * \tau_n]\!] = [\![\tau_1]\!] \times \ldots \times [\![\tau_n]\!]$
$[\![t + s]\!] = \langle [\![t]\!], [\![s]\!] \rangle ; +$
Similarly for the other operators $*, \varsigma, \ldots$

$\Gamma = (x_1 \colon \tau_1 \ldots x_n \colon \tau_n)$ is interpreted as $[\![\Gamma]\!] \stackrel{\text{def}}{=} \prod_{i=1}^{n} [\![\tau_i]\!]$.
Well typed terms $\Gamma \vdash t : \tau$ are interpreted as smooth functions $[\![t]\!] : [\![\Gamma]\!] \to [\![\tau]\!]$.

Denotational semantics for a first-order language

Definition (*CartSp*)

- *objects*(**CartSp**) : Euclidean spaces $\mathbb{R}^n$
- *morphisms*(**CartSp**) : smooth maps $\mathbb{R}^n \to \mathbb{R}^m$

$[\![\mathbf{real}]\!] = \mathbb{R}$
$[\![\tau_1 * \ldots * \tau_n]\!] = [\![\tau_1]\!] \times \ldots \times [\![\tau_n]\!]$
$[\![t + s]\!] = \langle [\![t]\!], [\![s]\!] \rangle; +$
Similarly for the other operators $*, \varsigma, \ldots$

$\Gamma = (x_1 : \tau_1 \ldots x_n : \tau_n)$ is interpreted as $[\![\Gamma]\!] \stackrel{\text{def}}{=} \prod_{i=1}^{n} [\![\tau_i]\!]$.
Well typed terms $\Gamma \vdash t : \tau$ are interpreted as smooth functions $[\![t]\!] : [\![\Gamma]\!] \to [\![\tau]\!]$.
What about variant and inductive types?

Denotational semantics for a first-order language

Definition

A smooth manifold is a space locally homeomorphic to $\mathbb{R}^n$ for some n.

Example

- A sphere $\mathbb{S}^2$ is locally homeomorphic to $\mathbb{R}^2$
- $\mathbb{R}^2 + \mathbb{R}^3$
- $\sum_{i \in \mathbb{N}} \mathbb{R}^i$

Denotational semantics for a first-order language

Definition (Man)

- ▻ *objects*(**Man**) : smooth (Hausdorff second-countable) manifolds
- ▻ *morphisms*(**Man**) : smooth maps

Denotational semantics for a first-order language

Definition (Man)

- *objects*(**Man**) : smooth (Hausdorff second-countable) manifolds
- *morphisms*(**Man**) : smooth maps

Example

$$[\![\tau + \sigma]\!] = [\![\tau]\!] + [\![\sigma]\!]$$
$$[\![\mathbf{list}(\mathbb{R})]\!] \cong \sum_{i \in \mathbb{N}} \mathbb{R}^i$$

Denotational semantics for a first-order language

Definition (Man)

- *objects*(**Man**) : smooth (Hausdorff second-countable) manifolds
- *morphisms*(**Man**) : smooth maps

Example

$$[\![\tau + \sigma]\!] = [\![\tau]\!] + [\![\sigma]\!]$$
$$[\![\textbf{list}(\mathbb{R})]\!] \cong \sum_{i \in \mathbb{N}} \mathbb{R}^i$$

Theorem

CartSp, **Man** *are not Cartesian-closed categories.*

Diffeological spaces

Definition

A *diffeological space* $(X, \mathcal{P}_X)$ consists of:

- a set X
- for each n and each open subset U of $\mathbb{R}^n$, a set $\mathcal{P}_X^U \subseteq [U \to X]$ of functions, called *plots*

Diffeological spaces

Definition

A *diffeological space* $(X, \mathcal{P}_X)$ consists of:

- a set X
- for each n and each open subset U of $\mathbb{R}^n$, a set $\mathcal{P}_X^U \subseteq [U \to X]$ of functions, called *plots*

such that

All constant functions are plots;

If $f \colon V \to U$ is a smooth function and $p \in \mathcal{P}_X^U$, then $f; p \in \mathcal{P}_X^V$;

If $\left(p_i \in \mathcal{P}_X^{U_i}\right)_{i \in I}$ is a compatible family of plots $(x \in U_i \cap U_j \Rightarrow p_i(x) = p_j(x))$ and $(U_i)_{i \in I}$ covers U, then the gluing $p : U \to X : x \in U_i \mapsto p_i(x)$ is a plot.

Definition

$f \colon X \to Y$ is *smooth* if, $\forall p \in \mathcal{P}_X^U$, we have that $p; f \in \mathcal{P}_Y^U$.

Diffeological spaces (cont.)

Example (Product diffeologies)

$(X_i)_{i \in I}$ diffeological spaces.
$\prod_{i \in I} X_i$ can be given the *product diffeology* in which U-plots are $(p_i)_{i \in I}$ for $p_i \in \mathcal{P}^U_{X_i}$.

Example (Functional diffeology)

$\mathbf{Diff}(X, Y)$ can be given the *functional diffeology* in which U-plots are $f \colon U \to \mathbf{Diff}(X, Y)$ such that $(u, x) \mapsto f(u)(x)$ is an element of $\mathbf{Diff}(U \times X, Y)$.

Diffeological spaces (cont.)

Theorem

The category **Diff** *of diffeological spaces is Cartesian-closed, cocomplete.*

Diffeological spaces (cont.)

Theorem

The category **Diff** *of diffeological spaces is Cartesian-closed, cocomplete.*

Proposition (Full embedding of Man into Diff)

A manifold $(X, \{\phi_U\}_U)$ is a diffeological space with:

the same carrier set X

plots $\mathcal{P}_X^U$ are the smooth functions in $\mathbf{Man}(U, X)$.

$X \to Y$ is **Man***-smooth iff it is* **Diff***-smooth.*
The embedding preserves finite products and countable coproducts.

Denotational semantics in **Diff**

$$[\![\textbf{real}]\!] \stackrel{\text{def}}{=} \mathbb{R}$$

$$[\![(\tau_1 * \ldots * \tau_n)]\!] \stackrel{\text{def}}{=} \prod_{i=1}^{n} [\![\tau_i]\!]$$

$$[\![\{\ell_1\, \tau_1 \mid \ldots \mid \ell_\mathsf{n}\, \tau_n\}]\!] \stackrel{\text{def}}{=} [\![\tau_1]\!] + \ldots + [\![\tau_n]\!]$$

$$[\![\tau \to \sigma]\!] \stackrel{\text{def}}{=} \textbf{Diff}([\![\tau]\!], [\![\sigma]\!])$$

Denotational semantics in **Diff**

$$[\![\textbf{real}]\!] \stackrel{\mathrm{def}}{=} \mathbb{R}$$

$$[\![(\tau_1 * \ldots * \tau_n)]\!] \stackrel{\mathrm{def}}{=} \prod_{i=1}^{n} [\![\tau_i]\!]$$

$$[\![\{\ell_1\, \tau_1 \mid \ldots \mid \ell_n\, \tau_n\}]\!] \stackrel{\mathrm{def}}{=} [\![\tau_1]\!] + \ldots + [\![\tau_n]\!]$$

$$[\![\tau \to \sigma]\!] \stackrel{\mathrm{def}}{=} \mathbf{Diff}([\![\tau]\!], [\![\sigma]\!])$$

$$[\![\varsigma(t)]\!](\rho) \stackrel{\mathrm{def}}{=} \varsigma([\![t]\!](\rho)), \text{ where } \rho \in [\![\Gamma]\!]$$

$$[\![\langle t_1, \ldots, t_n \rangle]\!](\rho) \stackrel{\mathrm{def}}{=} ([\![t_1]\!](\rho), \ldots, [\![t_n]\!](\rho))$$

$$[\![\lambda x{:}\tau.t]\!](\rho)(a) \stackrel{\mathrm{def}}{=} [\![t]\!](\rho, a) \; (a \in [\![\tau]\!])$$

$$[\![\textbf{case}\; t\; \textbf{of}\, \langle \ldots \rangle \to s]\!](\rho) \stackrel{\mathrm{def}}{=} [\![s]\!](\rho, [\![t]\!](\rho))$$

$$[\![t\, s]\!](\rho) \stackrel{\mathrm{def}}{=} [\![t]\!](\rho)([\![s]\!](\rho))$$

Initial syntactic category

Definition

- *Objects*(**Syn**) : types from our language
- Morphism $\tau \longrightarrow \sigma$: a term in context $x : \tau \vdash t : \sigma$ modulo $\beta\eta$-laws. Composition is by substitution.

Initial syntactic category

Definition

- *Objects*(**Syn**) : types from our language
- Morphism $\tau \to \sigma$: a term in context $x : \tau \vdash t : \sigma$ modulo $\beta\eta$-laws. Composition is by substitution.

Lemma (Universal Property of Syn)

$$\mathbf{Syn} \xrightarrow{\exists ! F \textit{ pres. } CC, +, Lists} \forall \mathcal{C}, \quad \textit{CC with } +, \textit{Lists}$$

given a choice of $F(\mathbf{real}) \in \mathcal{C}$, $F(\underline{c})$, $F(+)$, $F(*)$, $F(\varsigma)$.

Initial syntactic category

Example (Canonical definition of forward AD)

$\overrightarrow{\mathcal{D}}$ arises as a canonical CC functor on **Syn**.
Consider the unique CC functor $F : \mathbf{Syn} \to \mathbf{Syn}$ such that

$$F(\mathbf{real}) \stackrel{\text{def}}{=} \mathbf{real} * \mathbf{real}$$

$$F(\underline{c}) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\underline{c}) = (\underline{c}, \underline{0})$$

$$F(\varsigma) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\varsigma(x))$$

$$F(+) \stackrel{\text{def}}{=} z : F(\mathbf{real}) * F(\mathbf{real}) \vdash \mathbf{case}\, z\, \mathbf{of}\, \langle x, y \rangle \to \overrightarrow{\mathcal{D}}(x + y) : F(\mathbf{real})$$

Then $\forall \tau, x : \tau \vdash t : \sigma$, $F(\tau) = \overrightarrow{\mathcal{D}}(\tau)$ and $F(t) = \overrightarrow{\mathcal{D}}(t)$ as morphisms $F(\tau) \to F(\sigma)$ in **Syn**.

Initial syntactic category

Example (Canonical definition of forward AD)

$\overrightarrow{\mathcal{D}}$ arises as a canonical CC functor on **Syn**.
Consider the unique CC functor $F : \mathbf{Syn} \to \mathbf{Syn}$ such that

$$F(\mathbf{real}) \stackrel{\text{def}}{=} \mathbf{real} * \mathbf{real}$$

$$F(\underline{c}) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\underline{c}) = (\underline{c}, \underline{0})$$

$$F(\varsigma) \stackrel{\text{def}}{=} \overrightarrow{\mathcal{D}}(\varsigma(x))$$

$$F(+) \stackrel{\text{def}}{=} z : F(\mathbf{real}) * F(\mathbf{real}) \vdash \mathbf{case}\, z\, \mathbf{of}\, \langle x, y \rangle \to \overrightarrow{\mathcal{D}}(x + y) : F(\mathbf{real})$$

Then $\forall \tau, x : \tau \vdash t : \sigma$, $F(\tau) = \overrightarrow{\mathcal{D}}(\tau)$ and $F(t) = \overrightarrow{\mathcal{D}}(t)$ as morphisms $F(\tau) \to F(\sigma)$ in **Syn**.

$\Rightarrow$ Well-typedness, functoriality, definition on all type constructors of $\overrightarrow{\mathcal{D}}$ for free

No canonical tangent space on function spaces

Example

- $\overrightarrow{\mathcal{D}}((\mathbb{R} \to \mathbb{R}) \to \mathbb{R}) = (\mathbb{R}^2 \to \mathbb{R}^2) \to \mathbb{R}^2$
- $\overrightarrow{\mathcal{D}}(g)(f, \nabla f) = (f; g, \nabla(f; g))$

No canonical tangent space on function spaces

Example

- $\overrightarrow{\mathcal{D}}((\mathbb{R} \to \mathbb{R}) \to \mathbb{R}) = (\mathbb{R}^2 \to \mathbb{R}^2) \to \mathbb{R}^2$
- $\overrightarrow{\mathcal{D}}(g)(f, \nabla f) = (f; g, \nabla(f; g))$

No canonical tangent space on function spaces

No canonical tangent space on function spaces

Example

- $\overrightarrow{\mathcal{D}}((\mathbb{R} \to \mathbb{R}) \to \mathbb{R}) = (\mathbb{R}^2 \to \mathbb{R}^2) \to \mathbb{R}^2$
- $\overrightarrow{\mathcal{D}}(g)(f, \nabla f) = (f; g, \nabla(f; g))$

No canonical tangent space on function spaces

Idea

Pair a space X with a chosen tangent bundle X'.

No canonical tangent space on function spaces

Example

- $\overrightarrow{\mathcal{D}}((\mathbb{R} \to \mathbb{R}) \to \mathbb{R}) = (\mathbb{R}^2 \to \mathbb{R}^2) \to \mathbb{R}^2$
- $\overrightarrow{\mathcal{D}}(g)(f, \nabla f) = (f; g, \nabla(f; g))$

No canonical tangent space on function spaces

Idea

Pair a space X with a chosen tangent bundle X'.

Example

$(\mathbb{R}, \mathbb{R}^2, S)$ where $S \subseteq [\mathbb{R} \to \mathbb{R}] \times [\mathbb{R} \to \mathbb{R}^2]$ such that
$S = \{(f, (f, \nabla f)) \mid f : \mathbb{R} \to \mathbb{R} \text{ smooth}\}$.

Glueing category

Definition

- *Objects*($\mathbf{Gl}$) : triples (X, X', S) where
  X and X' are diffeological spaces
  $S \subseteq [\mathbb{R} \to X] \times [\mathbb{R} \to X']$ is a relation between curves.

Glueing category

Definition

- *Objects*($\mathbf{Gl}$) : triples (X, X', S) where

  X and X' are diffeological spaces
  $S \subseteq [\mathbb{R} \to X] \times [\mathbb{R} \to X']$ is a relation between curves.

- Morphism $(X, X', S) \to (Y, Y', T)$:

  smooth functions $f \colon X \to Y$, $f' : X' \to Y'$
  such that if $(g, g') \in S$ then $(g; f, g'; f') \in T$.

Glueing category

Definition

- $\triangleright$ *Objects*($\mathbf{Gl}$) : triples (X, X', S) where

  X and X' are diffeological spaces

  $S \subseteq [\mathbb{R} \to X] \times [\mathbb{R} \to X']$ is a relation between curves.
- $\triangleright$ Morphism $(X, X', S) \to (Y, Y', T)$:

  smooth functions $f \colon X \to Y$, $f' : X' \to Y'$

  such that if $(g, g') \in S$ then $(g; f, g'; f') \in T$.

Proposition

$\mathbf{Gl}$ *is cartesian closed with coproducts, has list objects, and the projection functor* $\mathrm{proj} : \mathbf{Gl} \to \mathbf{Diff} \times \mathbf{Diff}$ *preserves this structure.*

Proof of main theorem (sconing)

$$\begin{tikzcd} \mathbf{Syn} \arrow[r, "{(\mathrm{id}, \overrightarrow{\mathcal{D}}(-))}"] \arrow[d, "{\color{red}(\!|-|\!)}"'] & \mathbf{Syn} \times \mathbf{Syn} \arrow[d, "{[\![-]\!] \times [\![-]\!]}"] \\ \mathbf{Gl} \arrow[r, "\mathrm{proj}"'] & \mathbf{Diff} \times \mathbf{Diff} \end{tikzcd}$$

Proof of main theorem (sconing)

$$\begin{tikzcd}
\mathbf{Syn} \arrow[r, "(\mathrm{id}, \overrightarrow{\mathcal{D}}(-))"] \arrow[d, "(\!|-|\!)"'] & \mathbf{Syn} \times \mathbf{Syn} \arrow[d, "[\![-]\!] \times [\![-]\!]"] \\
\mathbf{Gl} \arrow[r, "\mathrm{proj}"'] & \mathbf{Diff} \times \mathbf{Diff}
\end{tikzcd}$$

where $(\!|-|\!) = ((\!|-|\!)_0, (\!|-|\!)_1, S_-)$, interpreting types τ as objects $((\!|\tau|\!)_0, (\!|\tau|\!)_1, S_\tau)$, and terms as morphisms:

- $(\!|\mathbf{real}|\!)_0 \stackrel{\mathrm{def}}{=} \mathbb{R}$, $(\!|\mathbf{real}|\!)_1 \stackrel{\mathrm{def}}{=} \mathbb{R}^2$ with the relation $S_{\mathbf{real}} \stackrel{\mathrm{def}}{=} \{(f, (f, \nabla f)) \mid f : \mathbb{R} \to \mathbb{R} \text{ smooth}\}$.

Proof of main theorem (sconing)

$$\begin{tikzcd} \mathbf{Syn} \arrow[r, "(\mathrm{id}, \overrightarrow{\mathcal{D}}(-))"] \arrow[d, "(\!|-|\!)"'] & \mathbf{Syn} \times \mathbf{Syn} \arrow[d, "[\![-]\!] \times [\![-]\!]"] \\ \mathbf{Gl} \arrow[r, "\mathrm{proj}"'] & \mathbf{Diff} \times \mathbf{Diff} \end{tikzcd}$$

where $(\!|-|\!) = ((\!|-|\!)_0, (\!|-|\!)_1, S_-)$, interpreting types τ as objects $((\!|\tau|\!)_0, (\!|\tau|\!)_1, S_\tau)$, and terms as morphisms:

- $(\!|\mathbf{real}|\!)_0 \stackrel{\text{def}}{=} \mathbb{R}$, $(\!|\mathbf{real}|\!)_1 \stackrel{\text{def}}{=} \mathbb{R}^2$ with the relation $S_{\mathbf{real}} \stackrel{\text{def}}{=} \{(f, (f, \nabla f)) \mid f : \mathbb{R} \to \mathbb{R} \text{ smooth}\}$.
- $(\!|\underline{c}|\!)_0 \stackrel{\text{def}}{=} \underline{c}$ and $(\!|\underline{c}|\!)_1 \stackrel{\text{def}}{=} (\underline{c}, 0)$
- Interpret $+, \times, \varsigma$ in the standard way (meaning, like $[\![-]\!]$) in $(\!|-|\!)_0$, but according to the derivatives in $(\!|-|\!)_1$, e.g.

$$(\!|*|\!)_1((x, x'), (y, y')) \stackrel{\text{def}}{=} (xy, xy' + x'y).$$

Elementary logical relations argument

Definition (Logical relation S_τ between curves in $[\![\tau]\!]$ and curves in $[\![\overrightarrow{\mathcal{D}}(\tau)]\!]$)

- $S_{\mathbf{real}} \stackrel{\text{def}}{=} \{(f, (f, \nabla f)) \mid f : \mathbb{R} \to \mathbb{R} \text{ smooth}\}$
- $S_{(\tau * \sigma)} \stackrel{\text{def}}{=} \{((f_1, g_1), (f_2, g_2)) \mid (f_1, f_2) \in S_\tau, (g_1, g_2) \in S_\sigma\}$
- $S_{\tau \to \sigma} \stackrel{\text{def}}{=} \{(f_1, f_2) \mid \forall (g_1, g_2) \in S_\tau . (x \mapsto f_1(x)(g_1(x)), x \mapsto f_2(x)(g_2(x))) \in S_\sigma\}$

Gilles Barthe et al. "On the Versatility of Open Logical Relations: Continuity, Automatic Differentiation, and a Containment Theorem". In: *arXiv preprint arXiv:2002.08489* (2020), to appear in ESOP 2020

Elementary logical relations argument

Definition (Logical relation S_τ between curves in $[\![\tau]\!]$ and curves in $[\![\overrightarrow{\mathcal{D}}(\tau)]\!]$)

- $S_{\mathbf{real}} \stackrel{\text{def}}{=} \{(f, (f, \nabla f)) \mid f : \mathbb{R} \to \mathbb{R} \text{ smooth}\}$
- $S_{(\tau * \sigma)} \stackrel{\text{def}}{=} \{((f_1, g_1), (f_2, g_2)) \mid (f_1, f_2) \in S_\tau, (g_1, g_2) \in S_\sigma\}$
- $S_{\tau \to \sigma} \stackrel{\text{def}}{=} \{(f_1, f_2) \mid \forall (g_1, g_2) \in S_\tau . (x \mapsto f_1(x)(g_1(x)), x \mapsto f_2(x)(g_2(x))) \in S_\sigma\}$

Fundamental Lemma

If $x_1{:}\tau_1, \ldots, x_n{:}\tau_n \vdash t : \sigma$ and, for all $1 \leqslant i \leqslant n$, $y_1 \ldots y_m : \mathbf{real} \vdash s_i : \tau_i$ is such that $((f_1, \ldots, f_m); [\![s_i]\!], (f_1, \nabla f_1), \ldots, f_m, \nabla f_m); [\![\overrightarrow{\mathcal{D}}(s_i)]\!]) \in S_{\tau_i}$ for all smooth $f_i : \mathbb{R} \to \mathbb{R}$, then
$\big((f_1, \ldots, f_m); [\![t[{}^{s_1}/_{x_1}, \ldots, {}^{s_n}/_{x_n}]]\!], (f_1, \nabla f_1, \ldots, f_m, \nabla f_m); [\![\overrightarrow{\mathcal{D}}(t[{}^{s_1}/_{x_1}, \ldots, {}^{s_n}/_{x_n}])]\!]\big)$ is in S_σ for all smooth $f_i : \mathbb{R} \to \mathbb{R}$.

Gilles Barthe et al. "On the Versatility of Open Logical Relations: Continuity, Automatic Differentiation, and a Containment Theorem". In: *arXiv preprint arXiv:2002.08489* (2020), to appear in ESOP 2020

Main theorem: Correctness

Semantic correctness of $\overrightarrow{\mathcal{D}}$

τ : ground type $\quad$ Γ : context with ground types only

$\Gamma \vdash t : \tau$ possibly with higher-order subterms

The syntactic translation $\overrightarrow{\mathcal{D}}$ coincides with the tangent bundle functor (modulo canonical isomorphisms):

$$\begin{tikzcd}
[\![\overrightarrow{\mathcal{D}}(\Gamma)]\!] \arrow[r, "{[\![\overrightarrow{\mathcal{D}}(t)]\!]}"] \arrow[d, "\cong"'] & [\![\overrightarrow{\mathcal{D}}(\tau)]\!] \arrow[d, "\cong"] \\
\mathcal{T}([\![\Gamma]\!]) \arrow[r, "{\mathcal{T}([\![t]\!])}"'] & \mathcal{T}([\![\tau]\!])
\end{tikzcd}$$

Tangent bundle

Denotational semantics

1. Language for higher-order AD
2. Denotational semantics for a higher-order language
3. A gluing construction
4. Another example: Jet-based AD

Jet bundle functor

Definition (Tangent space)

$x \in M$ manifold.
$\mathcal{T}_x M \stackrel{\text{def}}{=} \{\gamma \in \mathbf{Man}(\mathbb{R}, M) \mid \gamma(0) = x\}/\sim$ where $\gamma_1 \sim \gamma_2$ iff $\nabla(\gamma_1; f)(0) = \nabla(\gamma_2; f)(0)$ for all smooth $f\colon M \to \mathbb{R}$.

Definition ((1,2)-Jet space)

$J_x^{(1,2)} M \stackrel{\text{def}}{=} \{\gamma \in \mathbf{Man}(\mathbb{R}, M) \mid \gamma(0) = x\}/\sim$
where $\gamma_1 \sim \gamma_2$ iff $\nabla(\gamma_1; f)(0) = \nabla(\gamma_2; f)(0)$ and $\nabla^2(\gamma_1; f)(0) = \nabla^2(\gamma_2; f)(0)$ for all smooth $f\colon M \to \mathbb{R}$.

Definition ((1,2)-Jet bundle of M)

$J^{(1,2)} M \stackrel{\text{def}}{=} \biguplus_{x \in M} J_x^{(1,2)}(M)$.
$J^{(1,2)} : \mathbf{Man} \to \mathbf{Man}$ is a functor.

Jet-based AD

Interested in transforming a function $g : \mathbb{R}^n \to \mathbb{R}$ into a function $h : (\mathbb{R} \times \mathbb{R} \times \mathbb{R})^n \to \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ in such a way that for any $f_1 \dots f_n : \mathbb{R} \to \mathbb{R}$,

$$
\begin{aligned}
&(f_1, \nabla f_1, \nabla^2 f_1, \dots, f_n, \nabla f_n, \nabla^2 f_n); h = \\
&((f_1 \dots f_n); g, \nabla((f_1, \dots, f_n); g), \nabla^2((f_1, \dots, f_n); g)).
\end{aligned}
$$

An intuition for h can be given in terms of triple numbers.

Jet-based AD

Example

$g\colon \mathbb{R}^2 \to \mathbb{R}$ is a smooth function, then
$h((x_1,1,0),(x_2,0,0)) = (g(x_1,x_2), \frac{\partial g(x,x_2)}{\partial x}(x_1), \frac{\partial^2 g(x,x_2)}{\partial x^2}(x_1))$,
$h((x_1,0,0),(x_2,1,0)) = (g(x_1,x_2), \frac{\partial g(x_1,x)}{\partial x}(x_2), \frac{\partial^2 g(x_1,x)}{\partial x^2}(x_2))$ and
$h((x_1,1,0),(x_2,1,0)) =$
$(g(x_1,x_2), \frac{\partial g(x_1,x)}{\partial x}(x_2), \frac{\partial^2 g(x,x_2)}{\partial x^2}(x_1) + \frac{\partial^2 g(x_1,x)}{\partial x^2}(x_2) + 2\frac{\partial^2 g(x,y)}{\partial x \partial y}(x_1,x_2))$.

Jet-based AD

$$\overrightarrow{\mathcal{D}}^2(\textbf{real}) \stackrel{\mathrm{def}}{=} (\textbf{real}*\textbf{real}*\textbf{real})$$

$$\overrightarrow{\mathcal{D}}^2(\tau \to \sigma) \stackrel{\mathrm{def}}{=} \overrightarrow{\mathcal{D}}^2(\tau) \to \overrightarrow{\mathcal{D}}^2(\sigma)$$
$$\overrightarrow{\mathcal{D}}^2((\tau_1 * \cdots * \tau_n)) \stackrel{\mathrm{def}}{=} (\overrightarrow{\mathcal{D}}^2(\tau_1) * \cdots * \overrightarrow{\mathcal{D}}^2(\tau_n))$$
$$\overrightarrow{\mathcal{D}}^2(\{\ell_1\ \tau_1 \mid \ldots \mid \ell_n\ \tau_n\}) \stackrel{\mathrm{def}}{=} \{\ell_1\ \overrightarrow{\mathcal{D}}^2(\tau_1) \mid \ldots \mid \ell_n\ \overrightarrow{\mathcal{D}}^2(\tau_n)\}$$
$$\overrightarrow{\mathcal{D}}^2(\textbf{list}(\tau)) \stackrel{\mathrm{def}}{=} \textbf{list}(\overrightarrow{\mathcal{D}}^2(\tau))$$

Jet-based AD

$$\begin{aligned}
&\overrightarrow{\mathcal{D}}^2(\underline{c}) \stackrel{\text{def}}{=} \langle \underline{c}, 0, 0 \rangle \\
&\overrightarrow{\mathcal{D}}^2(t + s) \stackrel{\text{def}}{=} \textbf{case}\ \overrightarrow{\mathcal{D}}^2(t)\ \textbf{of}\ \langle x, x', x'' \rangle \to \textbf{case}\ \overrightarrow{\mathcal{D}}^2(s)\ \textbf{of}\ \langle y, y', y'' \rangle \to \\
&\qquad\qquad\qquad \langle x + y, x' + y', x'' + y'' \rangle \\
&\overrightarrow{\mathcal{D}}^2(t * s) \stackrel{\text{def}}{=} \textbf{case}\ \overrightarrow{\mathcal{D}}^2(t)\ \textbf{of}\ \langle x, x', x'' \rangle \to \textbf{case}\ \overrightarrow{\mathcal{D}}^2(s)\ \textbf{of}\ \langle y, y', y'' \rangle \to \\
&\qquad\qquad\qquad \langle x * y, x * y' + x' * y, x * y'' + 2 * x' * y' + x'' * y \rangle \\
&\overrightarrow{\mathcal{D}}^2(\varsigma(t)) \stackrel{\text{def}}{=} \textbf{case}\ \overrightarrow{\mathcal{D}}^2(t)\ \textbf{of}\ \langle x, x'x'' \rangle \to \textbf{let}\ y = \varsigma(x)\ \textbf{in}\ \textbf{let}\ z = y * (1 - y)\ \textbf{in} \\
&\qquad\qquad\qquad \langle y, x' * z, z * x'' + z * (1 - 2 * y) * x' * x' \rangle
\end{aligned}$$

Gluing category for Jet-based AD

Definition (Same as before)

Gl whose objects are triples (X, X', S) where X and X' are diffeological spaces and $S \subseteq [\mathbb{R} \to X] \times [\mathbb{R} \to X']$ is a relation between their curves. A morphism $(X, X', S) \to (Y, Y', T)$ is a pair of smooth functions $f \colon X \to Y$, $f' : X' \to Y'$, such that if $(g, g') \in S$ then $(g; f, g'; f') \in T$.

Proposition

The category **Gl** *is cartesian closed with coproducts, has list objects, and the projection functor* $\mathrm{proj} : \mathbf{Gl} \to \mathbf{Diff} \times \mathbf{Diff}$ *preserves this structure.*

Gluing category for Jet-based AD

Definition (Same as before)

Gl whose objects are triples (X, X', S) where X and X' are diffeological spaces and $S \subseteq [\mathbb{R} \to X] \times [\mathbb{R} \to X']$ is a relation between their curves. A morphism $(X, X', S) \to (Y, Y', T)$ is a pair of smooth functions $f \colon X \to Y$, $f' : X' \to Y'$, such that if $(g, g') \in S$ then $(g; f, g'; f') \in T$.

Proposition

The category **Gl** *is cartesian closed with coproducts, has list objects, and the projection functor* $\mathrm{proj} : \mathbf{Gl} \to \mathbf{Diff} \times \mathbf{Diff}$ *preserves this structure.*

$$
\begin{tikzcd}
\mathbf{Syn} \arrow[r, "(\mathrm{id}, \overrightarrow{\mathcal{D}}^2(-))"] \arrow[d, "(\!|-|\!)"'] & \mathbf{Syn} \times \mathbf{Syn} \arrow[d, "[\![-]\!] \times [\![-]\!]"] \\
\mathbf{Gl} \arrow[r, "\mathrm{proj}"'] & \mathbf{Diff} \times \mathbf{Diff}
\end{tikzcd}
$$

Gluing category for Jet-based AD

Definition (Same as before)

Gl whose objects are triples (X, X', S) where X and X' are diffeological spaces and $S \subseteq [\mathbb{R} \to X] \times [\mathbb{R} \to X']$ is a relation between their curves. A morphism $(X, X', S) \to (Y, Y', T)$ is a pair of smooth functions $f\colon X \to Y$, $f' : X' \to Y'$, such that if $(g, g') \in S$ then $(g; f, g'; f') \in T$.

Proposition

The category **Gl** *is cartesian closed with coproducts, has list objects, and the projection functor* $\mathrm{proj} : \mathbf{Gl} \to \mathbf{Diff} \times \mathbf{Diff}$ *preserves this structure.*

$$\begin{tikzcd}
\mathbf{Syn} \arrow[r, "{(\mathrm{id}, \overrightarrow{\mathcal{D}}^2(-))}"] \arrow[d, "(\!|-|\!)"'] & \mathbf{Syn} \times \mathbf{Syn} \arrow[d, "{[\![-]\!] \times [\![-]\!]}"] \\
\mathbf{Gl} \arrow[r, "\mathrm{proj}"'] & \mathbf{Diff} \times \mathbf{Diff}
\end{tikzcd}
\qquad
\begin{tikzcd}
{[\![\overrightarrow{\mathcal{D}}^2(\Gamma)]\!]} \arrow[r, "{[\![\overrightarrow{\mathcal{D}}^2(t)]\!]}"] \arrow[d, "\cong"'] & {[\![\overrightarrow{\mathcal{D}}^2(\tau)]\!]} \arrow[d, "\cong"] \\
J^{(1,2)}[\![\Gamma]\!] \arrow[r, "{J^{(1,2)}[\![t]\!]}"'] & J^{(1,2)}[\![\tau]\!]
\end{tikzcd}$$

Conclusion

Summary

- Higher-order language with inductive types with smooth primitives
- AD as a source-code transformation
- Denotational semantics in **Diff**
- Proof via categorical gluing
- Easily adapted, e.g. for jet-based AD

Mathieu Huot, Sam Staton, and Matthijs Vákár. "Correctness of Automatic Differentiation via Diffeologies and Categorical Gluing". In: *arXiv preprint arXiv:2001.02209* (2020)

Conclusion

Summary

- Higher-order language with inductive types with smooth primitives
- AD as a source-code transformation
- Denotational semantics in **Diff**
- Proof via categorical gluing
- Easily adapted, e.g. for jet-based AD

$\Rightarrow$ Bridge between AD and mathematics

Mathieu Huot, Sam Staton, and Matthijs Vákár. "Correctness of Automatic Differentiation via Diffeologies and Categorical Gluing". In: *arXiv preprint arXiv:2001.02209* (2020)

Conclusion

Summary

- Higher-order language with inductive types with smooth primitives
- AD as a source-code transformation
- Denotational semantics in **Diff**
- Proof via categorical gluing
- Easily adapted, e.g. for jet-based AD

$\Rightarrow$ Bridge between AD and mathematics

Future Work

- Partiality and recursion (e.g. for recursive NN)
- Probabilities (e.g. for HMC, VI)
- Efficient jet-based or reverse-mode implementations (e.g. for full or approximate Hessian computations)

Mathieu Huot, Sam Staton, and Matthijs Vákár. "Correctness of Automatic Differentiation via Diffeologies and Categorical Gluing". In: *arXiv preprint arXiv:2001.02209* (2020)

Reverse-mode AD

$$\overleftarrow{\mathcal{D}}_\rho(\textbf{real}) = (\textbf{real} * (\textbf{real} \to \rho)) \qquad \overleftarrow{\mathcal{D}}_k(\underline{c}) \stackrel{\text{def}}{=} \langle \underline{c}, \lambda z.\langle \underline{0}, \ldots, \underline{0} \rangle \rangle$$

$$\begin{aligned}
\overleftarrow{\mathcal{D}}_k(t + s) &\stackrel{\text{def}}{=} \textbf{case}\, \overleftarrow{\mathcal{D}}_k(t)\, \textbf{of}\, \langle x, x' \rangle \to \\
&\qquad \textbf{case}\, \overleftarrow{\mathcal{D}}_k(s)\, \textbf{of}\, \langle y, y' \rangle \to \langle x + y, \lambda z.x'\, z + y'\, z \rangle \\
\overleftarrow{\mathcal{D}}_k(t * s) &\stackrel{\text{def}}{=} \textbf{case}\, \overleftarrow{\mathcal{D}}_k(t)\, \textbf{of}\, \langle x, x' \rangle \to \\
&\qquad \textbf{case}\, \overleftarrow{\mathcal{D}}_k(s)\, \textbf{of}\, \langle y, y' \rangle \to \\
&\qquad\qquad \langle x * y, \lambda z.x'\, (y * z) + y'\, (x * z) \rangle \\
\overleftarrow{\mathcal{D}}_k(\varsigma(t)) &\stackrel{\text{def}}{=} \textbf{case}\, \overleftarrow{\mathcal{D}}_k(t)\, \textbf{of}\, \langle x, x' \rangle \to \textbf{let}\, y = \varsigma(x)\, \textbf{in}\, \langle y, \lambda z.x'\, (y * (1 - y) * z) \rangle.
\end{aligned}$$

Reverse-mode AD (cont.)

Theorem (Semantic correctness of $\overleftarrow{\mathcal{D}}_k$)

For any ground τ, any first order context Γ and any term $\Gamma \vdash t : \tau$, syntactic translation $t \mapsto \mathrm{evR}^k_\tau(\overleftarrow{\mathcal{D}}_k(t)[{}^{\mathrm{lamR}^k_\Gamma(z)}/_{\ldots}])$ coincides with the tangent bundle functor, modulo these canonical isomorphisms:

$$\begin{tikzcd}
[\![\overrightarrow{\mathcal{D}}_k(\Gamma)]\!] \arrow[r, "{[\![\mathrm{lamR}^k_\Gamma;\overleftarrow{\mathcal{D}}_k(t);\mathrm{evR}^k_\tau]\!]}"] \arrow[d, "\phi^{\overrightarrow{\mathcal{D}}\,\mathcal{T}}_{\Gamma,k}"', "\cong"] & [\![\overrightarrow{\mathcal{D}}_k(\tau)]\!] \arrow[d, "\cong"', "\phi^{\overrightarrow{\mathcal{D}}\,\mathcal{T}}_{\tau,k}"] \\
\mathcal{T}^k([\![\Gamma]\!]) \arrow[r, "{\mathcal{T}^k([\![t]\!])}"'] & \mathcal{T}^k([\![\tau]\!])
\end{tikzcd}$$

Reverse-mode AD (cont.)

$\tau = \textbf{real}$, $\Gamma = x, y : \textbf{real}$
$x, y : \textbf{real} \vdash t : \textbf{real}$ at values $x = V, y = W$:

$$\mathrm{evR}^2_{\textbf{real}}\,(\overleftarrow{\mathcal{D}}_2(t)[{}^{(\mathrm{lamR}^2_{x:\textbf{real}}\ v)}/_{x}, {}^{(\mathrm{lamR}^2_{y:\textbf{real}}\ w)}/_{y}])[{}^{\langle V,\langle \underline{1},\underline{0}\rangle\rangle}/_{v}, {}^{\langle W,\langle \underline{0},\underline{1}\rangle\rangle}/_{w}].$$

Indeed,

$$[\![\mathrm{evR}^2_{\textbf{real}}\,(\overleftarrow{\mathcal{D}}_2(t)[{}^{(\mathrm{lamR}^2_{x:\textbf{real}}\ v)}/_{x}, {}^{(\mathrm{lamR}^2_{y:\textbf{real}}\ w)}/_{y}])[{}^{\langle V,\langle \underline{1},\underline{0}\rangle\rangle}/_{v}, {}^{\langle W,\langle \underline{0},\underline{1}\rangle\rangle}/_{w}]]\!] =$$
$$\big([\![t]\!]([\![V]\!],[\![W]\!]), \partial_1[\![t]\!]([\![V]\!],[\![W]\!]), \partial_2[\![t]\!]([\![V]\!],[\![W]\!])\big).$$

Diff as a category of concrete sheaves

A concrete site $\mathcal{C}$ is a subcanonical site with a terminal 1 such that

- $Hom(1, -) : \mathcal{C} \to \mathbf{Set}$ is faithful
- For each covering $(f_i : D_i \to D)_{i \in I}$, the family $(Hom(1, f_i) : Hom(1, D_i) \to Hom(1, D))_{i \in I}$ is jointly surjective

A sheaf X is concrete if for every $D \in \mathcal{C}$, the function $X(D) \ni \phi \mapsto \underline{\phi} : Hom(1, D) \to X(1)$ is one-to-one.

Diff as a category of concrete sheaves (cont.)

Let **Open** be the category of open subsets of $\mathbb{R}^n$ for all n, and smooth maps.

Theorem

Diff *arises as the category of concrete sheaves on* **Open** *whose Grothendieck topology is given by the subcanonical coverage given by open coverings.*

Theorem

Diff *is a Grothendieck quasi-topos. In particular it is Cartesian-closed, completete, cocomplete.*

CartSp, **Man** are not CCC

Lemma

There is no continuous injection $\mathbb{R}^{d+1} \to \mathbb{R}^d$.

Proof.

If there were, it would restrict to a continuous injection $S^d \to \mathbb{R}^d$. The Borsuk-Ulam theorem, however, tells us that every continuous $f \colon S^d \to \mathbb{R}^d$ has some $x \in S^d$ such that $f(x) = f(-x)$, which is a contradiction. $\square$

Let us define the terms:

$$x_0 : \mathbf{real}, \ldots, x_n : \mathbf{real} \vdash t_n = \lambda y. x_0 + x_1 * y + \cdots + x_n * y^n : \mathbf{real} \to \mathbf{real}$$

Assuming that **CartSp**/**Man** is cartesian closed, observe that these get interpreted as injective continuous (because smooth) functions $\mathbb{R}^n \to [\![\mathbf{real} \to \mathbf{real}]\!]$ in **CartSp** and **Man**.

CartSp, **Man** are not CCC (cont.)

Theorem

CartSp *is not cartesian closed.*

Proof.

In case **CartSp** were cartesian closed, we would have $[\![\mathbf{real} \to \mathbf{real}]\!] = \mathbf{real}^n$ for some n. Then, we would get, in particular a continuous injection $[\![t_{n+1}]\!] : \mathbb{R}^{n+1} \to \mathbb{R}^n$, which contradicts the lemma. $\square$

CartSp, **Man** are not CCC (cont.)

Theorem

Man *is not cartesian closed.*

Proof.

Observe that we have $\iota_n : \mathbb{R}^n \to \mathbb{R}^{n+1}; \langle a_0, \ldots, a_n \rangle \mapsto \langle a_0, \ldots, a_n, 0 \rangle$ and that $\iota_n ; [\![t_{n+1}]\!] = [\![t_n]\!]$. Let us write A_n for the image of $[\![t_n]\!]$ and $A = \cup_{n \in \mathbb{N}} A_n$. Then, A_n is connected because it is the continuous image of a connected set. Similarly, A is connected because it is the non-disjoint union of connected sets. This means that A lies in a single connected component of $[\![\textbf{real} \to \textbf{real}]\!]$, which is a manifold with some finite dimension, say d.

Take some $x \in \mathbb{R}^{d+1}$ (say, 0), take some open d-ball U around $[\![t_{d+1}]\!](x)$, and take some open $d+1$-ball V around x in $[\![t_{d+1}]\!]^{-1}(U)$. Then, $[\![t_{d+1}]\!]$ restricts to a continuous injection from V to U, or equivalently, $\mathbb{R}^{d+1}$ to $\mathbb{R}^d$, which contradicts the lemma. $\square$