Diffeological Spaces and Denotational Semantics for Differential Programming

Ohad Kammar, Sam Staton, and Matthijs Vákár

Domains 2018
Oxford
8 July 2018

What is differential programming?

∂

- PL in which all (some) constructs are differentiable...
- ...and derivatives can be computed mechanically.
- Compute derivatives **compositionally, using chain rule**, rather than ~~symbolically or using finite differences~~.
- AKA Automatic Differentiation.

Why study semantics of differential programming?

Important:

- Efficient optimisation (in high dim);
- Efficient integration/sampling (in high dim);
- Often combined with concurrency and probability $\rightsquigarrow$ correctness non-trivial and good test-coverage hard to obtain;

Why study semantics of differential programming?

Non-Trivial:

- How to differentiate through traditional language constructs?
  - conditionals
  - iteration and term recursion
  - higher order functions
  - probability, state and other effects
- How to differentiate at structured types?
  - inductive/recursive types
  - refinement types
  - function types
  - quotient types
- So need to move beyond traditional calculus!

Concrete example of higher-order differential prog

Finding Shortest Arc Length

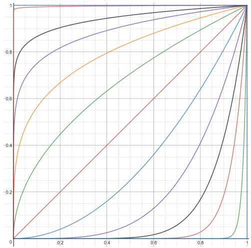
```
-- Integrate :: ([0,1] => real) => real
-- Differentiate :: ([0,1] => real) => [0,1] => real

ArcLength :: ([0,1] => real) => [0,infty)
ArcLength(f) = Integrate(Sqrt(1 + Differentiate(f)^2))

Power :: (0,infty) => [0,1] => real
Power(a)(t) = t^a

-- Minimise :: ((0,infty) => real) => (0,infty)
Minimal_a :: real
Minimal_a = Minimise(\a -> ArcLength(Power(a)))

-- Should be 1
```



Is this not an artificial problem?

Stan language:

For specifying a smooth objective function and sampling from it / optimising it:

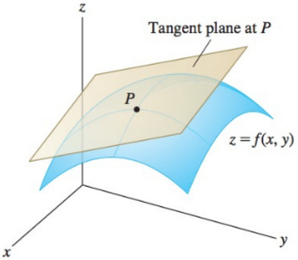
- conditionals (for custom combinations of objective functions), while loops, recursive functions (for iterative approximations);
- higher-order functions (ODE and algebraic solvers, used in pharmacokinetics);
- certain refinement types (constraints: intervals, unit vectors, simplices, symmetric positive definite matrices);
- fwd and rev mode AutoDiff through everything.

Calculus 101 Refresher

- Traditional derivative (Jacobian)

$$T_x f : \mathbb{R}^n \multimap \mathbb{R}^m$$

  of $f : U \to V$ at x, where
  U open in $\mathbb{R}^n$ and
  V open in $\mathbb{R}^m$



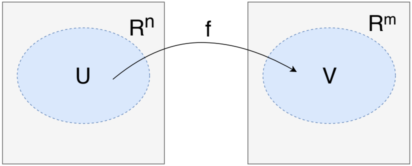
- Best linear approximation:

$$\lim_{v \to 0} \frac{||f(x+v) - f(x) - T_x f(v)||}{||v||} = 0$$

Categories for Smoothness?

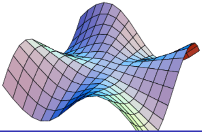
Category $\mathbf{Open}$:

- objects - opens U in some $\mathbb{R}^n$;
- morphisms $f : U \to V$ - smooth (C^∞) functions from U to V;



Category $\mathbf{Mfd}$ of fin. dim. manifolds and smooth maps:

- Adds surfaces (certain refinements);
- Whitney Embedding: idempotent completion of $\mathbf{Open}$;
- Great! If only there were more...



Several Categories of Smooth Maps

- Many full subcategories:

$$
\begin{tikzcd}
\mathbf{EuclSp} \arrow[r, hook, "\subseteq"] \arrow[d, hook, "\subseteq"'] & \mathbf{Open} \arrow[r, hook, "\subseteq"] & \mathbf{Mfd} \arrow[d, hook, "\subseteq"'] & & & \\
\mathbf{FreSp} \arrow[rr, hook, "\subseteq"] \arrow[d, hook, "\subseteq"'] & & \mathbf{FreMfd} \arrow[d, hook, "\subseteq"'] & & & \\
\mathbf{ConvSp} \arrow[rr, hook, "\subseteq"] & & \mathbf{FroSp} \arrow[r, hook, "\subseteq"] & \mathbf{Diffeo} \arrow[r, hook, "\subseteq"] & \mathbf{Sh(Open)} \arrow[r, hook, "\subseteq"] & {[\mathbf{Open}^{op}, \mathbf{Set}]}
\end{tikzcd}
$$

- **Diffeo** Grothendieck quasi-topos and well-pointed;
- Interprets tuples, function types, variant types, (co)inductive types, dependent types;
- Conservative extension: embeddings into **Diffeo** preserve all limits and coproducts, colimits of open covers;

Diffeological spaces [Souriau]

Objects

A **diffeological space** $X = \big(|X|, S_X\big)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

Diffeological spaces [Souriau]

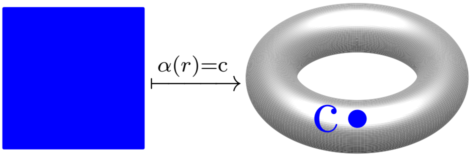
Objects

A **diffeological space** $X = (|X|, S_X)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

- constant functions $\underline{c}$;



Diffeological spaces [Souriau]

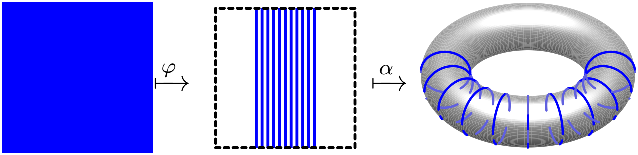
Objects

A **diffeological space** $X = (|X|, S_X)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

- constant functions $\underline{c}$;
- precomposition with a smooth $\varphi : U \to V$



Diffeological spaces [Souriau]

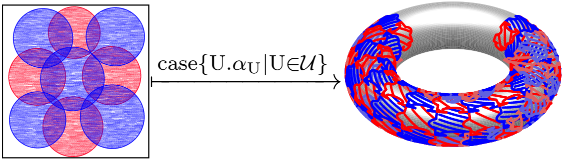
Objects

A **diffeological space** $X = (|X|, S_X)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

- constant functions $\underline{c}$;
- precomposition with a smooth $\varphi : U \to V$
- gluing of compatible families along open covers $\mathcal{U}$.



Diffeological spaces [Souriau]

Objects

A **diffeological space** $X = \big(|X|, S_X\big)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that...3 axioms...

Morphisms $f : X \to Y$

Functions $f : |X| \to |Y|$ such that:

$$\alpha \in S_X \qquad \Longrightarrow \qquad f \circ \alpha \in S_Y$$

Example: Traditional Spaces

- For manifold M, define (Yoneda) plots

$$S_M^U := \mathbf{Mfd}(U, M).$$

Categorical structure

Objects

A **diffeological space** $X = \big(|X|, S_X\big)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that...3 axioms...

Products

$$S_{X \times Y}^U := \left\{ r \mapsto \big(\alpha(r), \beta(r)\big) \middle| \alpha \in S_X^U, \beta \in S_Y^U \right\}$$

Categorical structure

Morphisms $f : X \to Y$

Functions $f : |X| \to |Y|$ such that:

$$\alpha \in S_X \qquad \Longrightarrow \qquad f \circ \alpha \in S_Y$$

Function spaces

$$\left|Y^X\right| := \mathbf{Diffeo}(X, Y)$$
$$S^U_{Y^X} := \left\{ f : U \to \left|Y^X\right| \middle| \text{uncurry } f \in \mathbf{Diffeo}(U \times X, Y) \right\}$$

NB: The exponential X^U is the space of U-plots.

Categorical structure

Morphisms $f : X \to Y$

Functions $f : |X| \to |Y|$ such that:

$$\alpha \in S_X \qquad \Longrightarrow \qquad f \circ \alpha \in S_Y$$

Function spaces

$$\left|Y^X\right| := \mathbf{Diffeo}(X, Y)$$
$$S^U_{Y^X} := \left\{ f : U \to \left|Y^X\right| \,\middle|\, \text{uncurry}\, f \in \mathbf{Diffeo}(U \times X, Y) \right\}$$

NB: The exponential X^U is the space of U-plots.

More structure

Coproducts, limits, colimits, ... (co)inductive types!

Our example revisited

Finding Shortest Arc Length

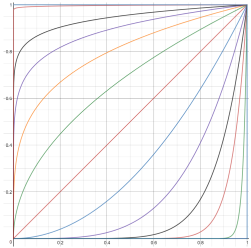
```
-- Integrate :: ([0,1] => real) => real
-- Differentiate :: ([0,1] => real) => [0,1] => real

ArcLength :: ([0,1] => real) => [0,infty)
ArcLength(f) = Integrate(Sqrt(1 + Differentiate(f)^2))

Power :: (0,infty) => [0,1] => real
Power(a)(t) = t^a

-- Minimise :: ((0,infty) => real) => (0,infty)
Minimal_a :: real
Minimal_a = Minimise(\a -> ArcLength(Power(a)))

-- Should be 1
```



- Uncurry to see `Power` is smooth;
- `Integrate` sends smooth functions of two arguments to smooth functions, so it is smooth.

Diffeo-morphisms are continuous!

D-Topology

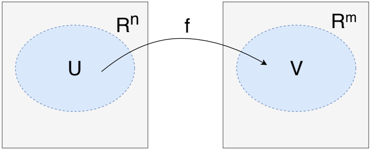
- Adjunction with topological spaces ($M \in \mathbf{Top}$, $X \in \mathbf{Diffeo}$):

$$\begin{aligned} S^U_{\mathbf{Diffeo}M} &:= \mathbf{Top}(U, M) \\ \mathcal{O}_{\mathbf{Top}X} &:= \left\{ B \subseteq X \middle| \forall U \in \mathbf{Open}.\alpha \in S^U_X, \alpha^{-1}[X] \in \mathcal{O}_U \right\} \end{aligned}$$

$$\begin{tikzcd} \mathbf{Diffeo} \arrow[rr, bend left, "\mathbf{Top}"] \arrow[rr, phantom, "\perp"] \arrow[dr] & & \mathbf{Top} \arrow[ll, bend left, "\mathbf{Diffeo}"] \arrow[dl] \\ & \mathbf{Set} \arrow[u, phantom, "="] & \end{tikzcd}$$

- Generalises Euclidean topology.

Covariant Derivative – Fwd Mode



- Write $TU := U \times \mathbb{R}^n \stackrel{\mathsf{fst}}{\to} U$
  $TV := V \times \mathbb{R}^m \stackrel{\mathsf{fst}}{\to} V$;
- And write
$$TU \xrightarrow{\quad Tf \quad} TV$$
$$\langle x, v\rangle \longmapsto \langle f(x), T_x f(v)\rangle;$$
- This defines a functor
$$\begin{tikzcd} \mathbf{Open} \arrow[r, hook, "T"] \arrow[dr, "\mathrm{id}"'] & \mathbf{Open}^{\to} \arrow[d, "\mathrm{cod}"] \\ & \mathbf{Open}. \end{tikzcd}$$
- Functoriality is chain rule!

Derivatives on $\mathbf{Diffeo}$

- Observe that $\mathbf{Diffeo}^{\rightarrow}$ and $\mathbf{Poly}(\mathbf{Diffeo})$ are cocomplete (and complete);
- Define through left Kan extension

$$\begin{tikzcd}
\mathbf{Open} \arrow[r, hook, "T"] \arrow[d, hook, "\mathbf{y}"'] \arrow[dr, phantom, "\Downarrow"] & \mathbf{Open}^{\rightarrow} \arrow[d, hook, "\mathbf{y}^{\rightarrow}"'] \\
\mathbf{Diffeo} \arrow[r, hook, dashed, "\mathsf{Lan}_{\mathbf{y}}(T^*;\mathbf{y}^{\rightarrow})"'] & \mathbf{Diffeo}^{\rightarrow}
\end{tikzcd}$$

- Functor by construction $\rightsquigarrow$ chain rule!
- $\mathbf{y}$ is full and faithful $\rightsquigarrow$ extends usual definition;
- But many other ways to define in literature! Do they coincide..?

Contravariant Derivative – Rev Mode

- Write $T^*_x f : \mathbb{R}^m \cong \mathbb{R}^m \multimap \mathbb{R} \overset{T_x f \multimap \mathbb{R}}{\to} \mathbb{R}^n \multimap \mathbb{R} \cong \mathbb{R}^n$;
- Define polynomials/containers $T^*U := U \times \mathbb{R}^n \overset{\mathsf{fst}}{\to} U$ $T^*V := V \times \mathbb{R}^m \overset{\mathsf{fst}}{\to} V$ and

$$f^*T^*V \xrightarrow{T^*f} T^*U$$

$$\langle x, \xi \rangle \longmapsto \langle x, T^*_x f(\xi) \rangle;$$

- This defines a functor

$$\begin{tikzcd} \mathbf{Open} \arrow[r, hook, "T^*"] \arrow[dr, "\mathrm{id}"'] & \mathbf{Poly}(\mathbf{Open}) \arrow[d, "\mathrm{cod}"] \\ & \mathbf{Open}. \end{tikzcd}$$

- Functoriality is chain rule!
- Left Kan extension: $\mathbf{Diffeo} \to \mathbf{Poly}(\mathbf{Diffeo})$

Derivatives are dependently typed!

(Co)tangent bundles not all trivial (product projections): take S^2.
$\rightsquigarrow$ dependent types!

Type Formers:

- Given $\Gamma \vdash X$ type, we get

$$\Gamma, x : X \vdash T_x X \text{ type} \qquad \text{and} \qquad \Gamma, x : X \vdash T^*_x X \text{ type.}$$

- TX is syntactic sugar for $\Sigma_{x:X} T_x X$ and $T^* X$ for $\Sigma_{x:X} T^*_x X$.

Term Formers:

- Given $\Gamma, x : X \vdash f(x) : Y$, we get

$$\Gamma, x : X, v : T_x X \vdash T_x f(v) : T_{f(x)} Y \qquad \text{and}$$

$$\Gamma, x : X, \xi : T^*_{f(x)} Y \vdash T^*_x f(\xi) : T^*_x X.$$

Diffeological Domains [Kammar, Staton, Vákár]

Category $\omega\mathbf{Diffeo}$

- Objects: diffeological spaces X with an ω-cpo structure $\leq_X$, s.t. S_X is closed under lubs of ω-chains;
- Morphisms: Scott continuous $\mathbf{Diffeo}$-morphisms;
- Gives a (co)complete ccc $\omega\mathbf{Diffeo}$ (locally presentable);
- Recursive types through Fiore's axiomatic domain theory!
- Define T, T^* as enriched Kan extensions, but working with $\omega\mathbf{Diffeo}^{\rightarrow}_{fib}$ and its dual fibration?

3 More Equivalent Characterisations

- $\omega\mathbf{Cpo}$-Enriched separated sheaves on $\mathbf{Open}$;
- Models of an essentially algebraic theory;
- Internal language ω-cpos in $\mathbf{Diffeo}$.

Conclusion

Summary

- $\mathbf{Diffeo}$: very simple, well-behaved and rich setting for semantics of differential languages
- 11AM - Ohad Kammar - A Domain Theory for Probability

Some Questions

- Derivative of a recursive function and T, T^* on $\omega\mathbf{Diffeo}$?
- An adequate semantics for differential programming?
- Prove corrects NUTS?
- Analyse the (co)tangent bundle of function types?
- C.f (co)tangent bundles in literature?
- Smooth probability monad?
- Relationship cotangent bundle and CPS?
- Boman's theorem: why not restrict to $S_X^{\mathbb{R}}$?

Thanks!

Recall - definition

A **diffeological space** $X = \big(|X|, S_X\big)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

Thanks!

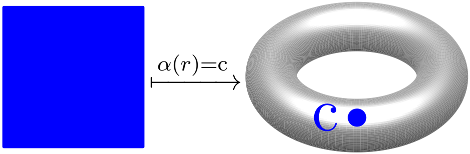
Recall - definition

A **diffeological space** $X = (|X|, S_X)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

- constant functions $\underline{c}$;



Thanks!

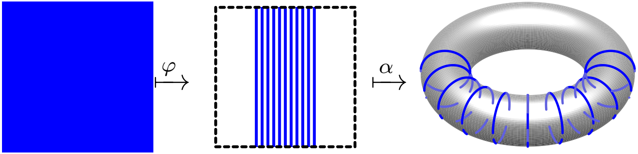
Recall - definition

A **diffeological space** $X = (|X|, S_X)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

- constant functions $\underline{c}$;
- precomposition with a smooth $\varphi : U \to V$



Thanks!

Recall - definition

A **diffeological space** $X = (|X|, S_X)$ consists of:

- a **carrier set** $|X|$;
- a set of **plots** $S_X^U \subseteq |X|^U$ for all $U \in \mathbf{Open}$

such that the plots are closed under:

- constant functions $\underline{c}$;
- precomposition with a smooth $\varphi : U \to V$
- gluing of compatible families along open covers $\mathcal{U}$.