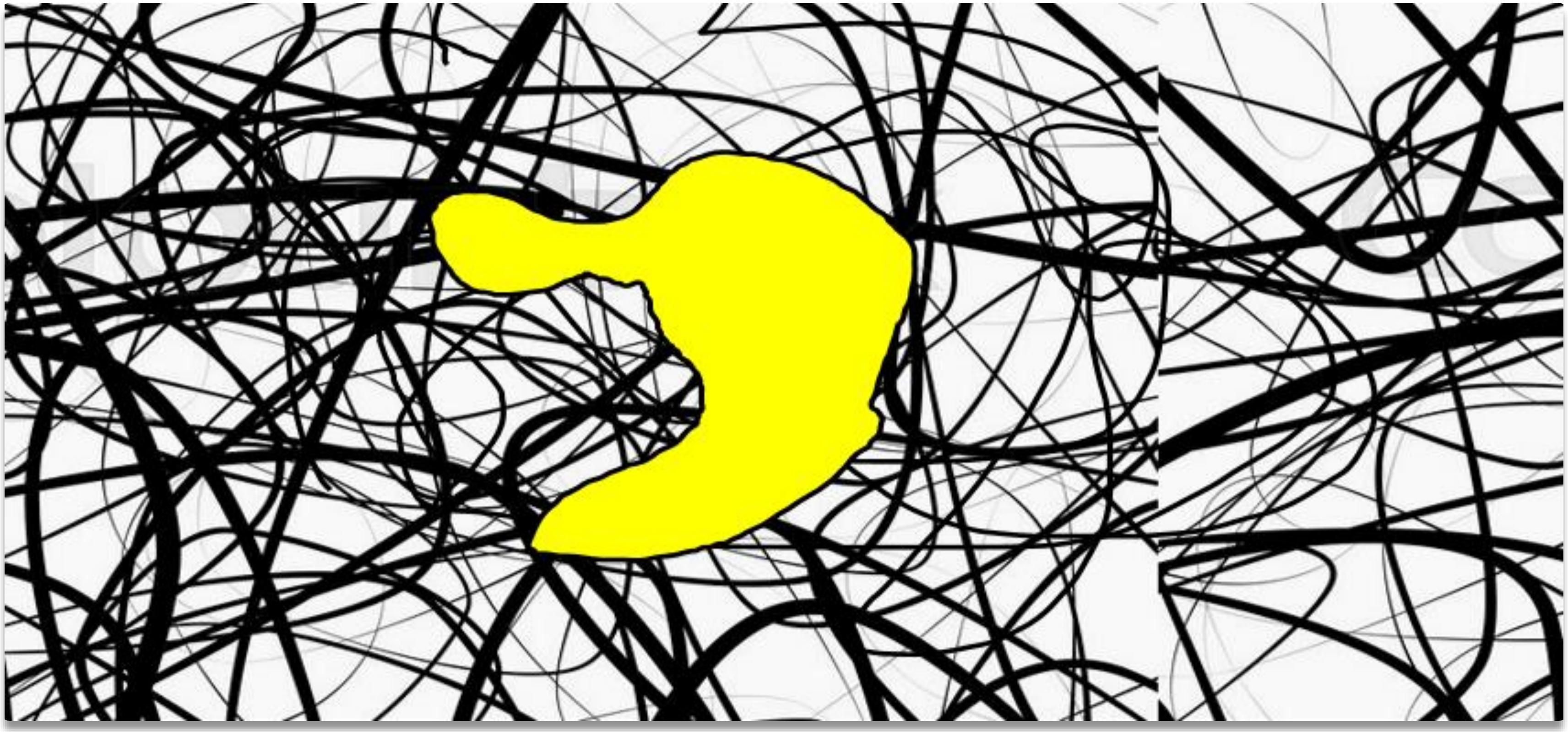
Differential Programming and Tangent Categories

Work with Geoff Cruttwell and Ben MacAdam

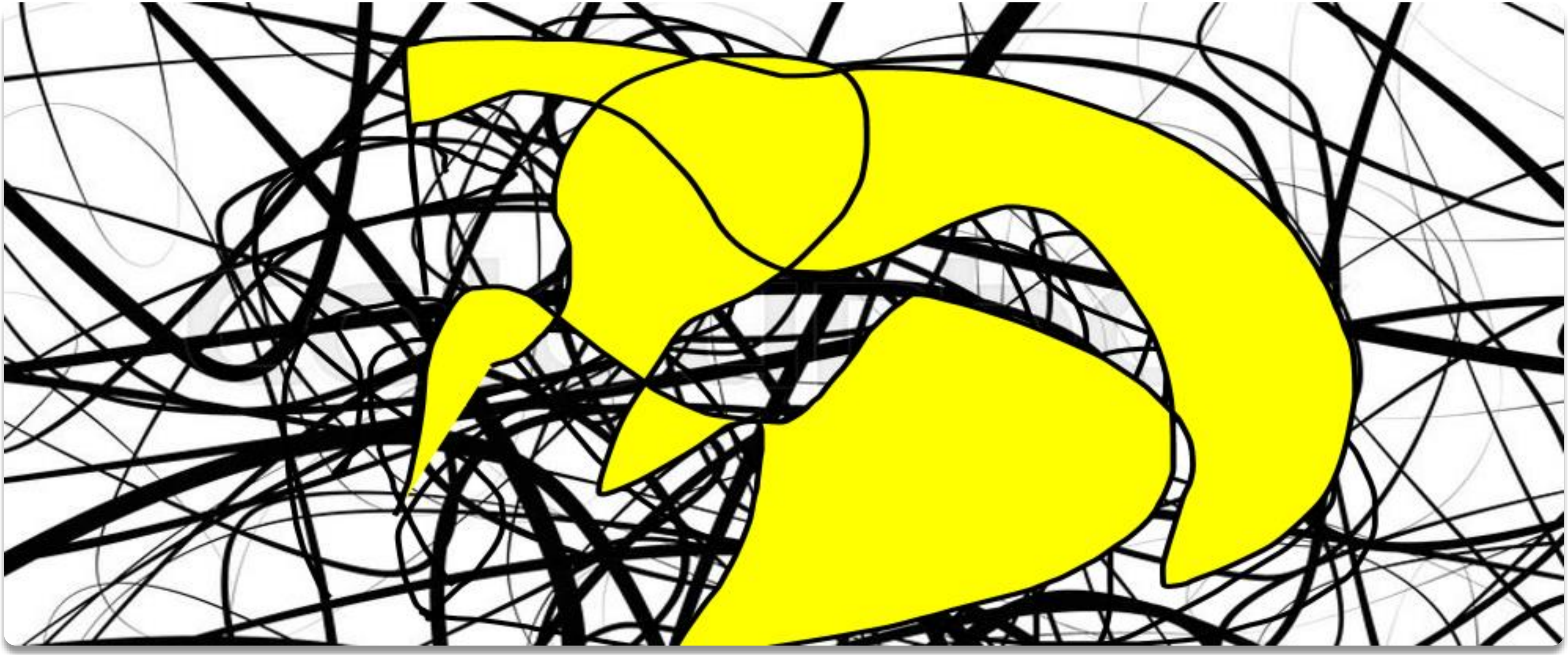
What is this talk about?

Apply techniques from synthetic differential geometry and tangent categories to understand internal homs in differential programming

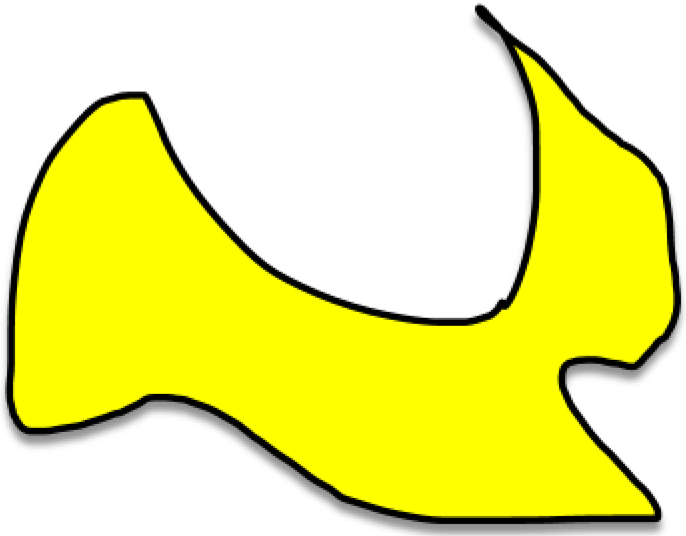


This is a wiggalump

Right, we’re gonna learn wiggalumps real quick like.

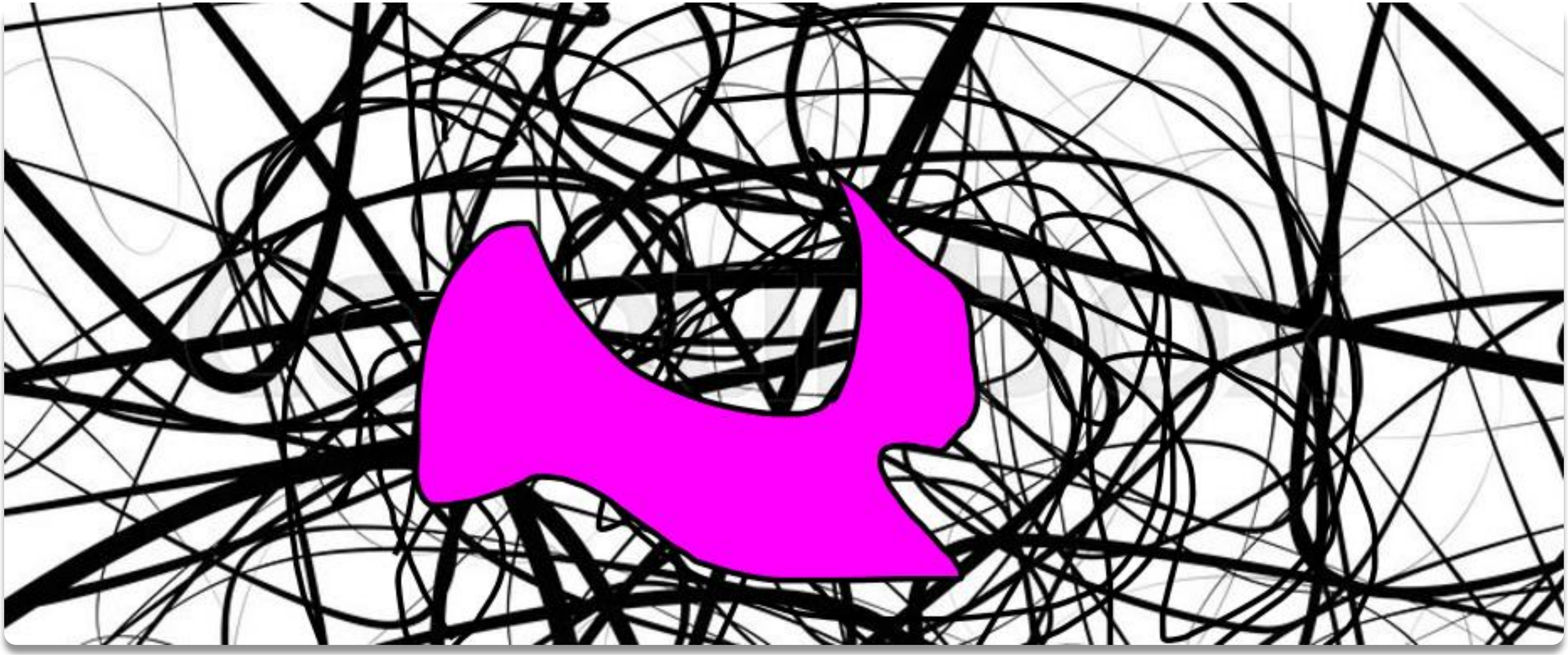


Is this a wiggalump?

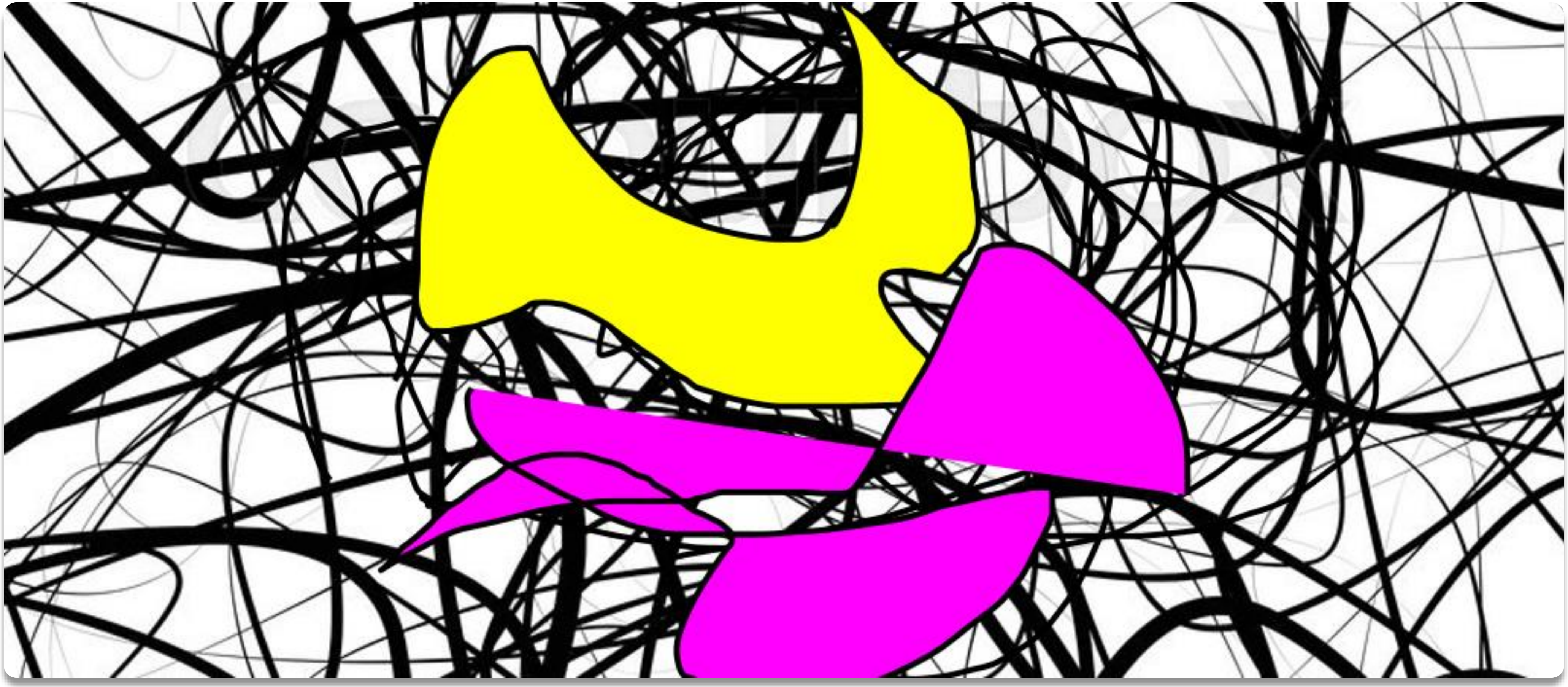


How about this?

Is *this* a wiggalump?



Is this a wiggalump?



Is *this a wiggalump?*

Okay, what's a wiggalump?

Learning via programs

Computing wins

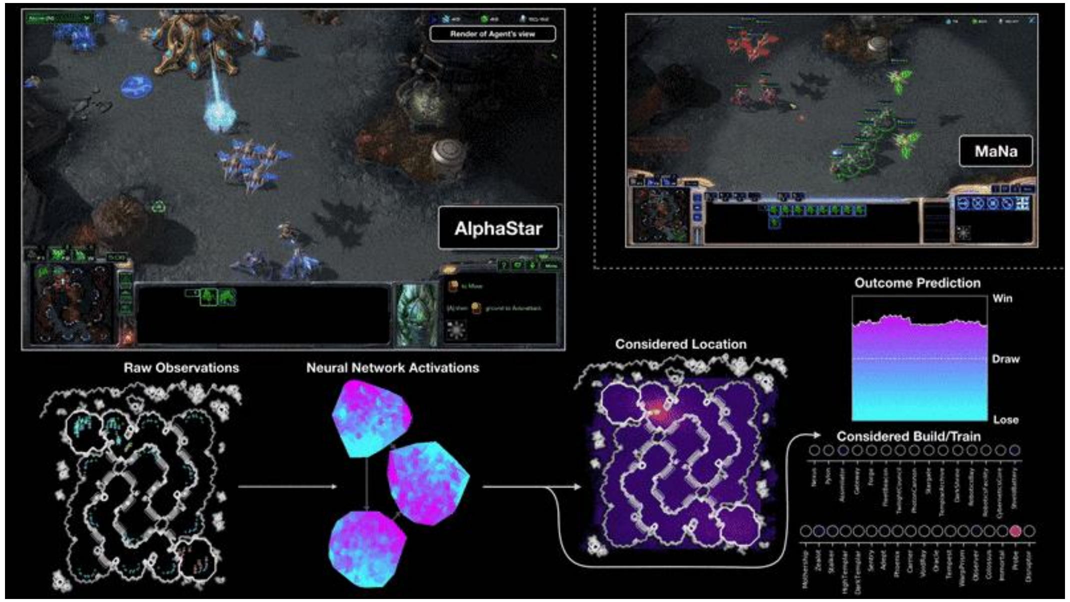
- Input and output
- Information processing and memory

Closely matched

- Complex movement
- Vision
- Language
- Structured problem solving

Brain still wins

- Creativity
- Emotion and Empathy
- Planning and Executive Function
- Consciousness



Deep mind beats grandmaster SCII player.

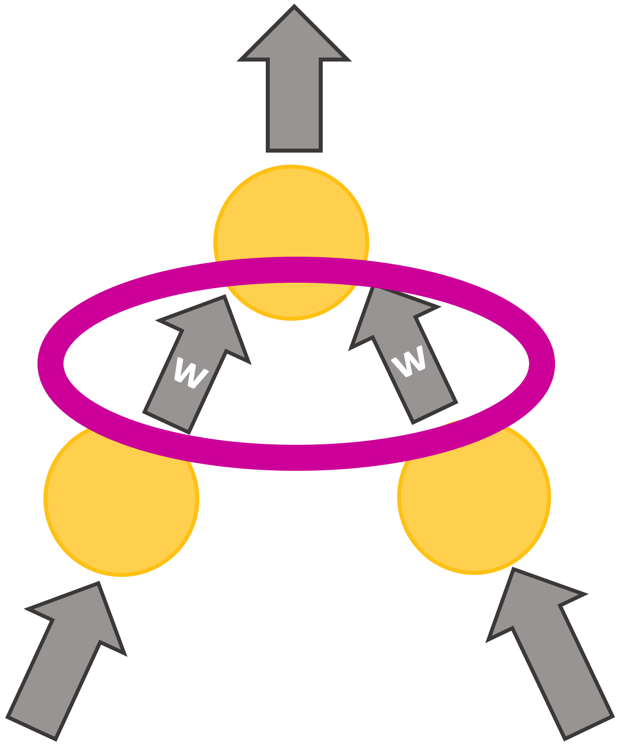
Learning via differentiable programs

- Differentiable neural computers:
  - Solve deep hard tasks
  - Solve tasks that were believed beyond “computers”
  - Make use of **recurrent** neural networks
  - Need a derivative, as in calculus?

Learning via differentiable programs

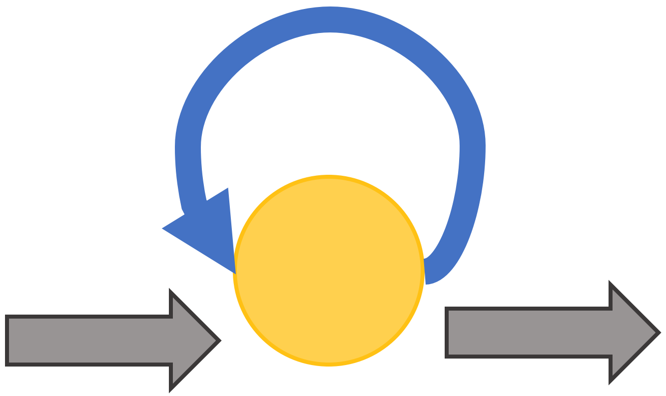
- Backpropagation uses the chain rule to push errors backwards in a **simple neural network**.
  - Treat a neural network
  as a function of the weights

  - Take the derivative
  of the error function
  and use the chain rule
  to pass error backwards



Learning via differentiable programs

- But for recurrent networks, it's a bit less straightforward.



- How do you take the derivative over a looping construct?

Differential Programming a la Plotkin

- A generalization of differential neural computers
    - Arbitrary programs with control structures etc
    - Encode smooth functions as programs
    - The derivative can be applied to programs

$$\mathcal{L} := \mathsf{FunDef}^*$$

$$\mathsf{FunDef} := \mathsf{Ident}\,(\,\mathsf{Ident}^*\,) \;:=\; M$$

$$M := x \in \mathsf{Var} \mid r \in \mathcal{R} \mid \sum_{i=1}^{n} M \mid f(M, \cdots, M)$$

$$\mid (M, \cdots, M) \mid \mathtt{let}(x_1, \ldots, x_n) = M\,\mathtt{in} M$$

$$\mid \mathtt{if}\, M\, \mathtt{then}\, M\, \mathrm{else}\, M \mid \mathtt{while}\, M \,.\, M$$

$$\mid \frac{\partial^R M}{\partial(x_1, \ldots, x_n)}(M) \cdot M$$

Today: Differential programming with tangent categories and SDG

Differential Programming → **Join restriction Tangent categories** → **SDG**

If-then-else

$$\frac{\Gamma \vdash b : \mathtt{Bool} \quad \Gamma \vdash f : A \quad \Gamma \vdash g : A}{\Gamma \vdash \mathtt{if}\, b\, \mathtt{then}\, f\, \mathtt{else}\, g : A}$$

Commandment: Guards should be continuous.

$b^{-1}(\mathtt{tt})$ and $b^{-1}(\mathtt{ff})$

Should be open and disjoint

Can be problematic:

$$g(x) = \begin{cases} -x & x < 0 \text{ or } x = 0 \\ x & x > 0 \end{cases}$$

$$f(x) = \begin{cases} x^2 & x \neq 0 \\ x & x = 0 \end{cases} \rightsquigarrow f(x) = x^2$$

Join Restriction Categories

- Join restriction categories allow packaging up partiality into a categorical framework.

  - Allows expressing domain of definition and detecting disjointness of domains

  - Both if-then-else and while require partiality

  - Allows expressing iteration using the join of disjoint domains
    - This is the trace of a coproduct in the idempotent splitting.

The category Smooth_P has

Objects: $\mathcal{R}^n$

Arrows: $(f, U) : \mathcal{R}^n \to \mathcal{R}^m$ is smooth when restricted to the open $U \subseteq \mathcal{R}^n$.

Domains/Restrictions:

$$\overline{f, U}(x) \begin{cases} x & x \in U \\ \uparrow & \text{else} \end{cases}$$

Joins: Suppose $f(x) = g(x)$ when $x \in U \cap V$;

$$[(f, U) \vee (g, V)](x) = \begin{cases} f(x) & x \in U \\ g(x) & \text{else} \end{cases}$$

The category Smooth_P also has:

Derivatives:

$$\frac{\mathcal{R}^n \xrightarrow{(f,U)} \mathcal{R}^m}{\mathcal{R}^n \times \mathcal{R}^n \xrightarrow{(Df,\mathcal{R}^n \times U)} \mathcal{R}^m}$$

In fact the above structure forms a differential join restriction category.

- All Faa di Bruno formulae for higher order chain rules hold.

In fact the above category is the category of differential objects of a Cartesian join restriction tangent category.

Join Restriction Tangent Categories

- Join restriction categories have all Weil_1 prolongations.

Theorem 1. *If $\mathcal{E}$ is a join restriction tangent category, and*

$$\mathsf{Smooth}_P \to \mathcal{E}$$

is a join restriction tangent functor, then $\mathcal{E}$ admits an interpretation of differential programming.

Adding function spaces: SDG

- Idea: Use techniques from algebraic geometry
  - Augment Smooth manifolds to have function spaces
  - And formal tangent bundles of infinitesimal curves
- Well adapted:

$$\mathsf{SMan} \to \mathcal{E}$$

  - Full and faithful, and `commutes with construction of manifolds.’

Partial map category of a well adapted model

Maps are spans with the left leg monic:

$$\begin{tikzcd} & A' \arrow[dl, tail] \arrow[dr, "f"] & \\ A & & B \end{tikzcd}$$

- This gives a join restriction category.
- However, there is an issue with microlinearity or "good objects."
  - The good objects are not a join restriction tangent category.

Partial map category of a well adapted model

Maps are spans with the left leg etale monic:

$$\begin{tikzcd} & A' \arrow[dl, tail] \arrow[dr, "f"] & \\ A & & B \end{tikzcd}$$

- This gives a join restriction category.
- The good objects are ~~not~~ a join restriction tangent category.

$$\mathcal{R}^n \xrightarrow{f} \mathcal{R}^m \quad \longmapsto \quad \begin{tikzcd} & \iota(U) \arrow[dl, tail] \arrow[dr, "\iota(f)"] & \\ \iota(\mathcal{R})^n & & \iota(R)^m \end{tikzcd}$$

Etale partial maps a well adapted model, model differential programming

Etale Monics

Formal Étale maps have the right lifting property for all

$$\begin{tikzcd} A \times 1 \arrow[r, "f"] \arrow[d, "1\times 0"'] & B \arrow[d, "b"] \\ A \times D(n,k) \arrow[r, "g"'] \arrow[ur, dotted, "\exists!"'] & C \end{tikzcd}$$

Properties:

- Closed to composition;
- Stable under pullback
- Formal \'etale monics are
- Downclosed to microlinearity
- Closed to joins or pushouts of matching diagrams:

$$\begin{tikzcd} U_1 \cap U_2 \arrow[r] \arrow[d] & U_1 \arrow[d] \\ U_2 \arrow[r] & U_1 \cup U_2 \arrow[ul, phantom, "\ulcorner", very near start] \end{tikzcd}$$

Application: Sequence spaces and codata

- The Dubuc topos has a natural numbers object.
- Microlinear spaces are an exponential ideal.

  $V^{\mathbb{N}}$

  - is microlinear when V is.
  - inhabitants are defined by the universal property of NNOs.
  - behave as sequences (so we can do RNNs).
- Total smooth functions between manifolds are microlinear, so we have a functional language with `total currying'
- We now know how to do more general codata and comonads for tangent categories.
  - Data is still a bit hard.

Bibliography

Baez, J. C., & Hoffnung, A. E. (2008). Convenient Categories of Smooth Spaces. *ArXiv e-prints*.

Batubenge, A., Iglesias-Zemmour, P., Karshon, Y., & Watts, J. (2017, 12). Diffeological, Fr``{o}licher, and Differential Spaces. *ArXiv e-prints*.

Bauer, K., Johnson, B., Osborne, C., Riehl, E., & Tebbe, A. (2018). Directional derivatives and higher order chain rules for Abelianfunctor calculus. *Topology and its Applications, 235*, 375-427.

Beck, T., & Fischer, H. (1994). The if-problem in automatic differentiation. *Journal of Computational and Applied Mathematics, 50*, 119-131. doi:https://doi.org/10.1016/0377-0427(94)90294-1

Bertram, W. (2014, 2). Weil Spaces and Weil-Lie Groups. *ArXiv e-prints*.

Bertram, W., Glöckner, H., & Neeb, K. H. (2004). Differential calculus over general base fields and rings. *Expositiones Mathematicae, 22*, 213-282.

Blute, R., Cockett, R., & Seely, R. (2009). Cartesian Differential Categories. *Theory and Applications of Categories, 22*, 622-672.

Blute, R., Ehrhard, T., & Tasson, C. (2010). A convenient differential category. *CoRR, abs/1006.3140*. Retrieved from http://arxiv.org/abs/1006.3140

Cockett, J. R., & Lack, S. (2002). Restriction categories I: categories of partial maps. *Theoretical Computer Science, 270*, 223-259. doi:https://doi.org/10.1016/S0304-3975(00)00382-0

Cockett, J. R., Cruttwell, G. S., & Gallagher, J. D. (2011). Differential restriction categories. *Theory and applications of categories, 25*, 537-613.

Cockett, R. & Lack, S.. (2007). Restriction categories III: colimits, partial limits and extensivity. *Mathematical Structures in Computer Science, 17*, 775–817. doi:10.1017/S0960129507006056

Cockett, R., & Cruttwell, G. (2014). Differential structure, tangent structure, and SDG. *Applied Categorical Structures, 22*, 331-417.

Cockett, R., & Seely, R. (2011). The Faá Di Bruno Construction. *Theory and Applications of Categories, 25*, 294-425.

Day, B. (1972). A reflection theorem for closed categories. *Journal of Pure and Applied Algebra, 2*, 1-11.

Dubuc, E., & Kock, A. (1984). On 1-Form Classifiers. *Communications in Algebra, 12*, 1471-1531.

Ehrhard, T., & Regnier, L. (2003). The differential lambda-calculus. *Theoretical Computer Science, 309*, 1-41.

Frölicher, A., & Kriegl, A. (1988).

Gallagher, J. (2009). *The differential lambda-calculus: syntax and semantics for differential geometry.* Ph.D. dissertation, University of Calgary.

Garner, R. (2018). An embedding theorem for tangent categories. *Advances in Mathematics, 323*, 668-687.

Janelidze, G., & Kelly, G. M. (2001). A note on actions of a monoidal category. *Theory and Applications of Categories, 4*, 61-91.

Kammar, O., Staton, S., & Vákár, M. (n.d.). Diffeological Spaces and Denotational Semantics for Differential Programming.

Kearfott, R. B. (1996). Chapter 1 Automatic Differentiation of Conditional Branches in an Operator Overloading Context.

Kock, A. (1978). Taylor series calculus for ring objects of line type. *Journal of Pure and Applied Algebra, 12*, 271-293.

Kock, A. (1981). *Synthetic differential geometry.* Cambridge University Press.

Kock, A., & Reyes, G. E. (1979). Manifolds in formal differential geometry. In M. Fourman, C. Mulvey, & D. Scott (Eds.), *Applications of Sheaves: Proceedings of the Research Symposium on Applications of Sheaf Theory to Logic, Algebra, and Analysis, Durham, July 9--21, 1977* (pp. 514-533). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/BFb0061832

Kriegl, A., & Michor, P. (1997). *The Convenient Setting of Global Analysis.* American Mathematical Society.

Laubinger, M. (2008). *Differential Geometry in Cartesian Closed Categories of Smooth Spaces.* Ph.D. dissertation, Louisiana State University and Agricultural and Mechanical College.

Lavendhomme, R. (1996). *Basic Concepts of Synthetic Differential Geometry.* Kluwer Academic Publishers.

Lawvere, B. (1979). Categorical dynamics. *Topos theoretic methods in geometry* (pp. 1-28). Aarhus University.

Lin, D. (2019, 6 01). Presheaves Over a Join Restriction Category. *Applied Categorical Structures, 27*, 289-310. doi:10.1007/s10485-018-09555-5

Mulry, P. S. (1994). Partial map classifiers and partial cartesian closed categories. *Theoretical Computer Science, 136*, 109-123. doi:https://doi.org/10.1016/0304-3975(94)00124-2

Plotkin, G. (2018, 6). *A Simple Differential Programming Language.* unpublished.

PyTorch. (n.d.).

Reyes, G. E., & Wraith, G. E. (1978). A note on the tangent bundle in a category with a ring object. *Math. Scand., 42*, 53-63.

Rosický, J. (1984). Abstract Tangent Functors. *Diagrammes, 12*, JR1--JR11.

TensorFlow. (n.d.).