Why I (fluid-PDE) started learning machine learning?

Closure problem

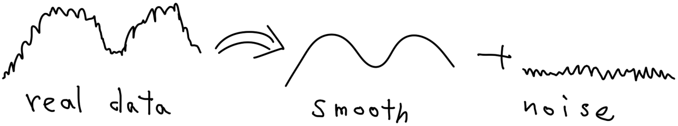
The Navier-Stokes equations

$$\partial_t u + (u \cdot \nabla) u = -\nabla p + \nu \Delta u + f, \quad \nabla \cdot u = 0, \quad u|_{t=0} = u_0$$

Let $\hat{\cdot}$ be the Fourier transoform and $\mathcal{F}^{-1}$ be its inverse.

- Mean flow (smooth): $\langle u(t,x) \rangle = \mathcal{F}_\xi^{-1}[\hat{u}(t,\xi)\chi_{\{|\xi| \leq 1\}}](t,x)$
- Turbulent fluctuation (noise): $u'(t,x) = 1 - \langle u(t,x) \rangle$

Note that the noise may NOT be white noise.



Apply $\langle \cdot \rangle$ to both sides of the Navier-Stokes equations:

$$\partial_t \langle u \rangle + (\langle u \rangle \cdot \nabla)\langle u \rangle + \nabla \cdot \tau = -\nabla \langle p \rangle + \nu \Delta \langle u \rangle + \langle f \rangle, \quad \nabla \cdot \langle u \rangle = 0.$$

$\tau = \langle u \otimes u \rangle - \langle u \rangle \otimes \langle u \rangle$ is called the Reynolds stress, depends on the turbulent fluctuation u'. The classical closure problem is as follows:

Can τ be approximated using only the mean flow $\langle u \rangle$?

Breakthrough in this study (at least I think)

The Google Science Team in 2023:
GraphCast: Learning skillful medium-range global weather forecasting

$\Rightarrow$ ML model for weather forecasting (NOT physics-informed anymore, i.e. without any use of fluid eqs.) that surpasses conventional physical models.

Implementation-driven mathematics

(my standpoint)

Libraries such as TensorFlow and PyTorch fundamentally rely on backpropagation (gradient descent), and modifying that core algorithm is VERY DIFFICULT at the user level!!

$\Rightarrow$ I take a math-first approach to model design.

The first thing for pure mathematicians to learn about ML

- input space $\mathcal{U}$, output space $\mathcal{D}$ and feature space $\mathcal{X}$
- $F : \mathcal{U} \to \mathcal{D}$: underlying system, might be disconti, really want to know
- $\dim(\mathcal{U}), \dim(\mathcal{D}) \ll \dim(\mathcal{X})$

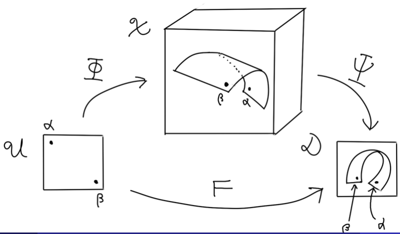
Example:

- u: picture ($\mathcal{U}$ has $1000px \times 1000px \times 3colors$ dimension)
- $\underline{d = 1\text{: pic is cat}}$, $\underline{d = 0\text{: pic is NOT cat}}$ ($\mathcal{D}$ has only one dimension)

Purpose of ML in pure math description

Find continuous maps $\Phi : \mathcal{U} \to \mathcal{X}$ and $\Psi : \mathcal{X} \to \mathcal{D}$ such that

$$|\Psi \circ \Phi(u) - d| \quad \text{is small enough for } \forall (u, d) \in \mathcal{U} \times F(\mathcal{U}).$$



The crucial structure of ML (black box is not black box)

There may exist $(u_1, d_1), (u_2, d_2) \in \mathcal{U} \times F(\mathcal{U})$ satisfying $|u_1 - u_2| \gtrsim 1$ such that, the corresponding feature vectors are similar: $|\Phi(u_1) - \Phi(u_2)| \ll 1$. How can we structure this into Φ?

We have three examples, convolutional NN, ESP, attention mechanism.

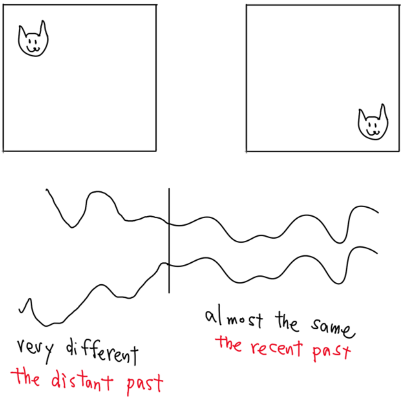


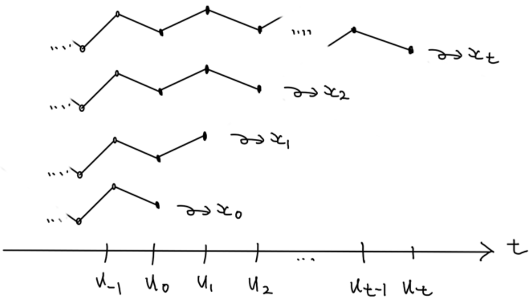
Figure: Left cat is α, Right cat is β.

ESP: Echo state property (for time series data)

The tricky point is the following definition: For $\Phi : \mathcal{U} \times \mathcal{X} \to \mathcal{X}$, let

$$x_t^1 = \Phi(u_t, \cdot) \circ \cdots \circ \Phi(u_2, \cdot) \circ \Phi(u_1, x_0^1),$$
$$x_t^2 = \Phi(u_t, \cdot) \circ \cdots \circ \Phi(u_2, \cdot) \circ \Phi(u_1, x_0^2).$$

Feature vectors x_0^1 and x_0^2 are expressing time series before $t = 0$.



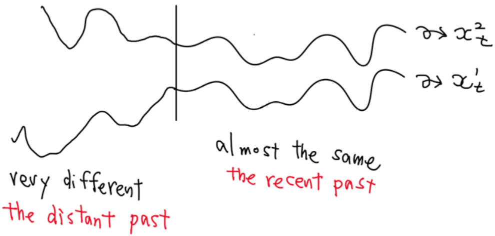
Theorem

Assume there is $\lambda \in (0,1)$ such that

$$|\Phi(u, x^1) - \Phi(u, x^2)| \leq \lambda |x^1 - x^2|$$

for $(u, x^1), (u, x^2) \in \mathcal{U} \times \mathcal{X}$ with $x_1 \neq x_2$. Then for any $x_0^1, x_0^2 \in \mathcal{X}$, we have

$$|x_t^1 - x_t^2| \to 0 \quad (t \to \infty).$$



The proof is obvious. We have the following for $t \to \infty$:

$$|x_{t+1}^1 - x_{t+1}^2| \leq |\Phi(u_t, x_t^1) - \Phi(u_t, x_t^2)| \leq |x_t^1 - x_t^2| \leq \cdots \leq \lambda^t |x_0^1 - x_0^2| \to 0.$$

Remark

ML framework

Recall that, the underlying system F could be discontinuous, on the other hand, Φ and Ψ are continuous maps.

From this ML framework, We see that Ψ and Φ may not be appropriate to be expressed as Fourier sums (or, more generally, any kind of sums). The evidence is the following: For the Fourier series of the indicator functions of several dimensional balls, we observe

- Gibbs phenomena,
- Pinsky phenomena 1993 (diverging at the origin),
- Kuratsubo phenomena 2010 (preventing pointwise convergence).

$$\Rightarrow |\Psi \circ \Phi(u) - d| \quad \text{cannot be small for} \quad \forall (u, d) \in \mathbb{Q} \cap \mathcal{U} \times F(\mathbb{Q} \cap \mathcal{U})$$

To understand Kuratsubo phenomena intuitively, see numerical computations in Section 7 in Kuratsubo-Nakai-Ootsubo (2010).