Some statistical insights into PINNs

Nathan Doumèche

SORBONNE UNIVERSITÉ

Joint work with

Gérard Biau

Claire Boyer

Hybrid modeling

Statistical model: $Y = u^\star(\boldsymbol{X}) + \varepsilon$

Goal: estimate $u^\star$ using

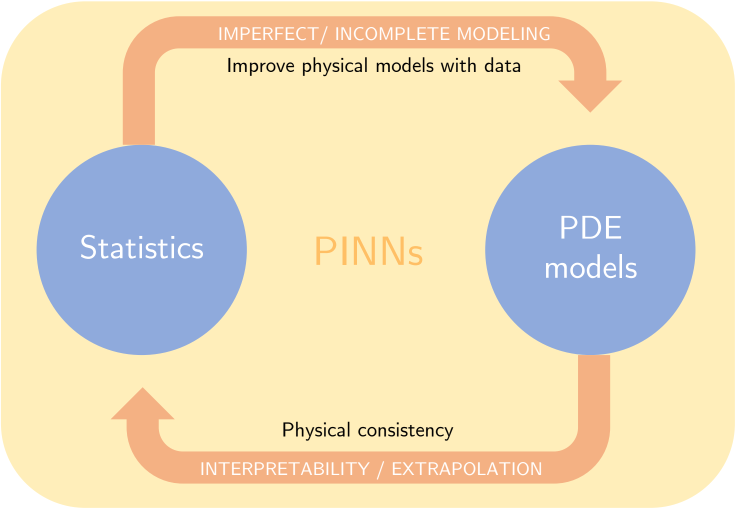
- Supervised learning: an i.i.d. training sample $(\boldsymbol{X}_i, Y_i)_{1 \leqslant i \leqslant n}$
- Physical modeling: a prior knowledge

$$\mathscr{F}_k(u^\star, \cdot) \simeq 0, \quad 1 \leqslant k \leqslant M,$$

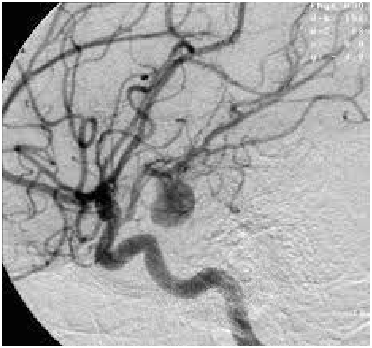
  + boundary/initial conditions for $u^\star$

- Neural networks

- Physics-Informed Neural Networks (PINNs)

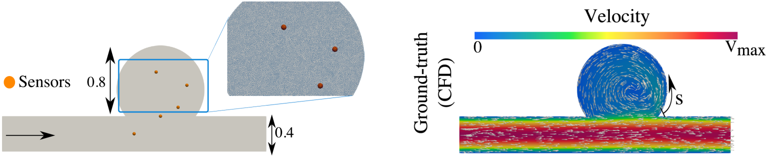

IMPERFECT/ INCOMPLETE MODELING
Improve physical models with data
Statistics
PINNs
PDE
models
Physical consistency
INTERPRETABILITY / EXTRAPOLATION

Example: Blood flow in an aneurysm



Modeling the blood flow

[Arzani et al., 2021]



Goal: estimate the blood flow $u = (u_x, u_y, P)$

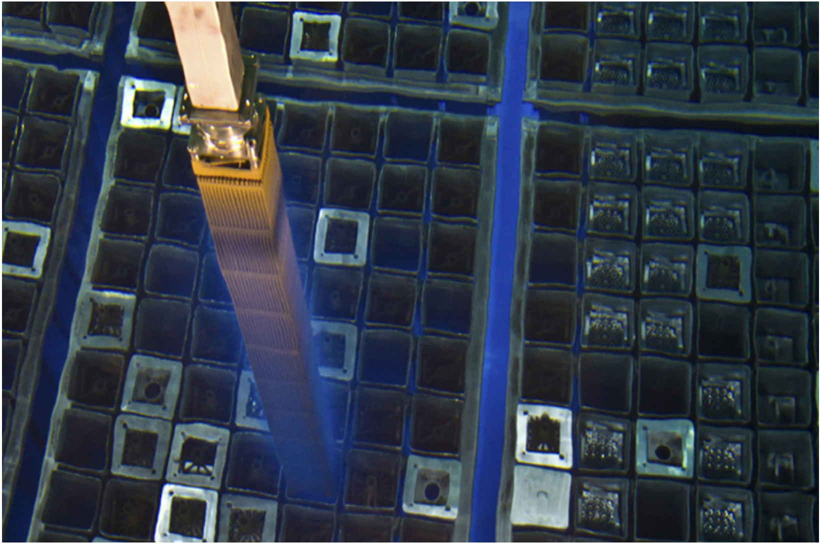
Navier-Stokes equations:

- $\mathscr{F}_1(u, \cdot) = u_x \partial_x u_x + u_y \partial_y u_x - \partial^2_{x,x} u_x - \partial^2_{y,y} u_x + \partial_x P$
- $\mathscr{F}_2(u, \cdot) = u_x \partial_x u_y + u_y \partial_y u_y - \partial^2_{x,x} u_y - \partial^2_{y,y} u_y + \partial_y P$
- $\mathscr{F}_3(u, \cdot) = \partial_x u_x + \partial_y u_y$

(Incomplete) boundary conditions:

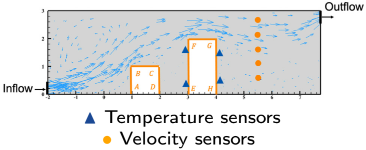
- $(u_x, u_y) = 0$ on the boundaries of the vessel
- Unknown inflow and outflow

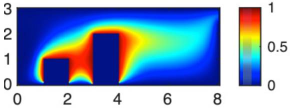
Example: Heat transfer in uranium bundles



Modeling the heat transfer

[Cai et al., 2021]





Goal: estimate the temperature T on the bundles

Navier-Stokes and diffusion equations on $u = (u_x, u_y, P, T)$:

- $\mathscr{F}_1(u, \cdot) = u_x \partial_x u_x + u_y \partial_y u_x - \partial^2_{x,x} u_x - \partial^2_{y,y} u_x + \partial_x P$
- $\mathscr{F}_2(u, \cdot) = u_x \partial_x u_y + u_y \partial_y u_y - \partial^2_{x,x} u_y - \partial^2_{y,y} u_y + \partial_y P$
- $\mathscr{F}_3(u, \cdot) = \partial_x u_x + \partial_y u_y$
- $\mathscr{F}_4(u, \cdot) = u_x \partial_x T + u_y \partial_y T - \partial^2_{x,x} T - \partial^2_{y,y} T$

(Incomplete) boundary conditions:

- $(u_x, u_y) = 0$ and $T = 0$ on the physical boundaries
- Inflow with $u_x = 1$, $u_y = 0$, and $T = 0$
- Outflow with $\partial_x T = 0$

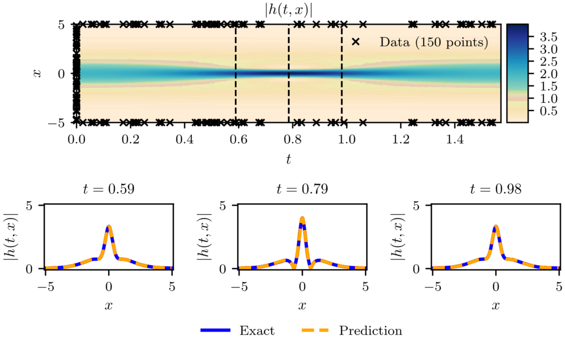
The PDE solver case

Specificity: no data Y_i and exact modeling

Example: the nonlinear Schrödinger PDE [Raissi et al., 2019]

$$i\partial_t u + 0.5\partial^2_{x,x} u + |u|^2 u = 0$$

Periodic boundary conditions and initial condition: $u(x, 0) = 2\,/\cosh(x)$



Challenges

Hybrid modeling problems:

- Improve imperfect/incomplete physical models with data
- Conversely, provide interpretability and extrapolation in ML

PDE solvers:

- Rely on complex triangulations of the domain
- Prone to the curse of the dimension

PINNs:

- A modern and efficient ML tool for both problems
- Natural implementation in the deep learning framework

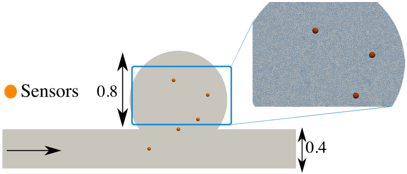
Our objective

To better understand the capabilities and limitations of PINNs

Summary

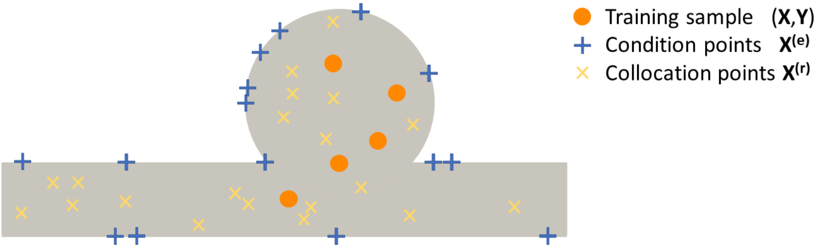
Summary

Geometry of the problem



- $\Omega \subseteq \mathbb{R}^{d_1}$: the bounded set on which the problem is posed
- $u^\star : \Omega \to \mathbb{R}^{d_2}$: the unknown target function
- Differential operators $\mathscr{F}_k(u^\star, \cdot) \simeq 0$ on Ω, $\quad 1 \leqslant k \leqslant M$
- $\partial\Omega$: the boundary of $\Omega \Rightarrow$ often not C^1 but Lipschitz
- Dirichlet conditions: $u^\star(\boldsymbol{x}) \simeq h(\boldsymbol{x})$ on $E \subseteq \partial\Omega$
- Possible extensions to other types of boundary/initial conditions

A general framework: 3 samplings



A general framework

- Training sample $(\boldsymbol{X}_1, Y_1), \dots, (\boldsymbol{X}_n, Y_n) \in \Omega \times \mathbb{R}^{d_2}$ (unknown distribution)
- Boundary/initial sample $\boldsymbol{X}_1^{(e)}, \dots, \boldsymbol{X}_{n_e}^{(e)} \in E \subseteq \partial\Omega$ (chosen distribution)
- Collocation points $\boldsymbol{X}_1^{(r)}, \dots, \boldsymbol{X}_{n_r}^{(r)} \in \Omega$ (uniform distribution)

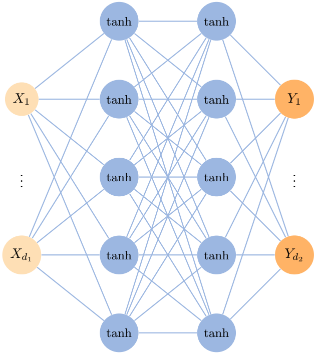
Empirical risk function

$$R_{n,n_e,n_r}(u_\theta) = \underbrace{\frac{\lambda_d}{n}\sum_{i=1}^{n} \|u_\theta(\boldsymbol{X}_i) - Y_i\|_2^2}_{\text{data-fidelity}} + \underbrace{\frac{\lambda_e}{n_e}\sum_{j=1}^{n_e} \|u_\theta(\boldsymbol{X}_j^{(e)}) - h(\boldsymbol{X}_j^{(e)})\|_2^2}_{\text{boundary conditions}}$$

$$+ \underbrace{\frac{1}{n_r}\sum_{k=1}^{M}\sum_{\ell=1}^{n_r} \mathscr{F}_k(u_\theta, \boldsymbol{X}_\ell^{(r)})^2}_{\text{PDEs}}$$

Neural architecture

- $\mathrm{NN}_H(D)$: the set of neural networks with H hidden layers of width D
- $\mathrm{NN}_H = \cup_D \mathrm{NN}_H(D)$
- θ: parameter of the neural network
- tanh: activation function
- $u_\theta \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$



Minimizing sequence

We denote by $(\hat{\theta}(p, n_e, n_r, D))_{p\in\mathbb{N}}$ any minimizing sequence, i.e.,

$$\lim_{p\to\infty} R_{n,n_e,n_r}(u_{\hat{\theta}(p,n_e,n_r,D)}) = \inf_{u_\theta \in \mathrm{NN}_H(D)} R_{n,n_e,n_r}(u_\theta).$$

- The training of PINNs relies on the backpropagation algorithm

Questions

Hybrid modeling

- Statistical properties of PINNs
- Impact of the physical model
- Tuning of the PINN hyperparameters

PDE solver

- Reconstruction of the solution $u^\star$ of a PDE system
- Curse of the dimension

Density of neural networks in Hölder spaces

Proposition

Let $\Omega \subseteq \mathbb{R}^{d_1}$ be a bounded Lipschitz domain and $K \in \mathbb{N}$. Then, for any function $u \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$, there exists a sequence $(u_p)_{p \in \mathbb{N}} \in \mathrm{NN}_H$ such that $\lim_{p \to \infty} \|u - u_p\|_{C^K(\Omega)} = 0$.

- Valid for bounded Lipschitz domains + $C^K(\Omega)$ norm
- Generalization of De Ryck et al. (2021)
- In line with practical applications, where $D \gg H$
- Key property to solve PDE systems

Summary

Theoretical risk and consistency

Theoretical risk

$$\mathscr{R}_n(u) = \frac{\lambda_d}{n} \sum_{i=1}^{n} \|u(\boldsymbol{X}_i) - Y_i\|_2^2 + \lambda_e \mathbb{E}\|u(\boldsymbol{X}^{(e)}) - h(\boldsymbol{X}^{(e)})\|_2^2$$
$$+ \frac{1}{|\Omega|} \sum_{k=1}^{M} \int_{\Omega} \mathscr{F}_k(u, \boldsymbol{x})^2 dx$$

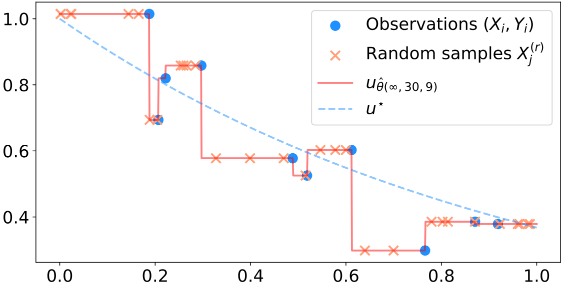
A natural requirement: Risk-consistency

$$\lim_{n_e, n_r \to \infty} \lim_{p \to \infty} \mathscr{R}_n(u_{\hat{\theta}(p, n_e, n_r, D)}) \stackrel{?}{=} \inf_{u \in \mathrm{NN}_H(D)} \mathscr{R}_n(u)$$

- Warning: possible overfitting

Overfitting: hybrid modeling

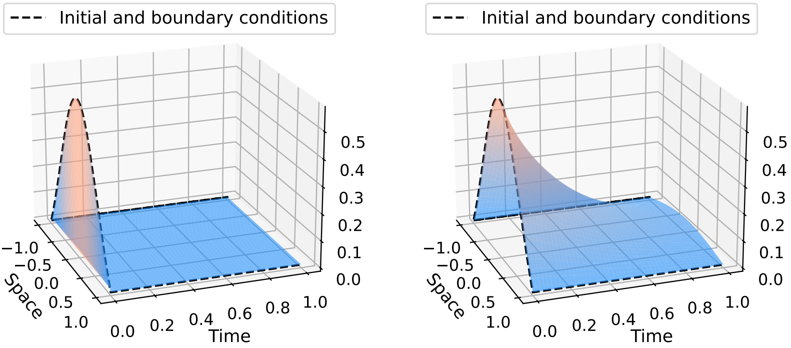
- Observations: $Y_i = u^\star(\boldsymbol{X}_i) + \varepsilon_i$
- Goal: estimate the trajectory $u^\star$ on $\Omega =]0, 1[$
- Model (dynamics with friction): $\mathscr{F}(u, \boldsymbol{x}) = u''(\boldsymbol{x}) + u'(\boldsymbol{x})$



- Overfitting: $R_{n,n_r} = 0$ but $\mathscr{R}_n = \infty$

Overfitting: PDE solver

- Heat equation: $\mathscr{F}(u, \boldsymbol{x}) = \partial_t u(\boldsymbol{x}) - \partial^2_{x,x} u(\boldsymbol{x})$ + boundary/initial conditions
- Goal: reconstruct the solution $u^\star$ on $\Omega =]-1, 1[\times]0, 1[$



- Overfitting: $R_{n_e,n_r} = 0$ but $\mathscr{R} = \infty$

Fighting overfitting: ridge regularization

Proposition

There exists a constant $C_{K,H} > 0$ such that

$$\|u_\theta\|_{C^K(\mathbb{R}^{d_1})} \leqslant C_{K,H}(D+1)^{HK+1}(1+\|\theta\|_2)^{HK}\|\theta\|_2.$$

Ridge PINNs

$$R^{(\mathrm{ridge})}_{n,n_e,n_r}(u_\theta) = R_{n,n_e,n_r}(u_\theta) + \lambda_{(\mathrm{ridge})}\|\theta\|_2^2$$

We denote by $(\hat{\theta}^{(\mathrm{ridge})}_{(p,n_e,n_r,D)})_{p\in\mathbb{N}}$ a minimizing sequence of this risk.

- Implemented in standard DL libraries via weight decay

Polynomial operators

Example: the Navier-Stokes equations on $u = (u_x, u_y, P)$:

- $\mathscr{F}_1(u, \cdot) = u_x \partial_x u_x + u_y \partial_y u_x - \partial^2_{x,x} u_x - \partial^2_{y,y} u_x + \partial_x P$
- $\mathscr{F}_1(u, \cdot) = \mathscr{P}(u_x, \partial_x u_x, \partial^2_{x,x} u_x, \partial_y u_x, \partial^2_{y,y} u_x, u_y, \partial_x P)$

- $\mathscr{P}(Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7) = Z_1 Z_2 + Z_6 Z_4 - Z_3 - Z_5 + Z_7$
- The coefficient in front of the monomial $Z_1 Z_2$ is $1 \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$
- Warning: $\deg \mathscr{F}_1 = 3$ but $\deg \mathscr{P} = 2$

Polynomial operator

An operator $\mathscr{F}(u, \cdot)$ is polynomial if it can be expressed as a polynomial in u and its derivatives, with smooth functions as coefficients.

- ✓ Linear PDEs (e.g., advection, heat, and Maxwell)
- ✓ Some nonlinear PDEs (e.g., Blasius, Burger, and Navier-Stokes)

Risk-consistency of ridge PINNs

Assumptions:

- The condition function h is Lipschitz
- $\mathscr{F}_1, \dots, \mathscr{F}_M$ are polynomial operators

Theorem

With a ridge hyperparameter of the form

$$\lambda_{(\text{ridge})} = \min(n_e, n_r)^{-\kappa}, \qquad \kappa = \frac{1}{12 + 4H(1 + (2 + H)\max_k \deg(\mathscr{F}_k))},$$

one has, almost surely,

$$\lim_{n_e, n_r \to \infty} \lim_{p \to \infty} \mathscr{R}_n(u_{\hat{\theta}^{(\text{ridge})}(p, n_e, n_r, D)}) = \inf_{u \in \mathrm{NN}_H(D)} \mathscr{R}_n(u)$$

and

$$\lim_{D \to \infty} \lim_{n_e, n_r \to \infty} \lim_{p \to \infty} \mathscr{R}_n(u_{\hat{\theta}^{(\text{ridge})}(p, n_e, n_r, D)}) = \inf_{u \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})} \mathscr{R}_n(u).$$

Take-home message

- Ridge regularization prevents overfitting of PINNs
- The decay rate of $\lambda_{(\mathrm{ridge})} = \min(n_e, n_r)^{-\kappa}$ does not depend on the dimension d_1 of Ω
- $\lambda_{(\mathrm{ridge})}$ can be tuned by monitoring the overfitting gap

Summary

Risk-consistency and strong convergence

✓ Ridge PINNs are risk-consistent

Question

Is this sufficient to have $\lim_{D,n_e,n_r,p\to\infty} u_{\hat{\theta}^{(\mathrm{ridge})}(p,n_e,n_r,D)} = u^\star$ in $L^2(\Omega)$?

Answer: No

Let $\Omega =]0,1[^2$, $h(x,0) = 1$, $h(0,t) = 1$, and $\mathscr{F}(u,\cdot) = \partial_x u + \partial_t u$. Then, for any $(\boldsymbol{X}_i, Y_i)_{1\leqslant i\leqslant n}$, there exists $(u_p)_{p\in\mathbb{N}} \in \mathrm{NN}_H(2n)$ such that

$$\lim_{p\to\infty} \mathscr{R}_n(u_p) = 0,$$

but $\lim_{p\to\infty} u_p = 1$ in $L^2(\Omega)$ (independently of $u^\star$).

✗ KO if imperfect modeling

✓ Possible solution: Sobolev regularization

Sobolev spaces

Weak derivatives

A function $v \in L^2(\Omega, \mathbb{R}^{d_2})$ is the α-th weak derivative of $u \in L^2(\Omega, \mathbb{R}^{d_2})$ if, for any $\varphi \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$ with compact support in Ω, one has

$$\int_\Omega \langle v, \varphi \rangle = (-1)^{|\alpha|} \int_\Omega \langle u, \partial^\alpha \varphi \rangle.$$

Notation: $v = \partial^\alpha u$.

Sobolev spaces

$H^m(\Omega, \mathbb{R}^{d_2})$ is the space of all functions $u \in L^2(\Omega, \mathbb{R}^{d_2})$ such that $\partial^\alpha u$ exist for all $|\alpha| \leqslant m$. This space is naturally endowed with the norm

$$\|u\|^2_{H^m(\Omega)} = \frac{1}{|\Omega|} \sum_{|\alpha| \leqslant m} \int_\Omega \|\partial^\alpha u\|_2^2.$$

- $C^m(\bar{\Omega}, \mathbb{R}^{d_2}) \subseteq H^m(\Omega, \mathbb{R}^{d_2})$
- Standard derivatives $\leftrightarrow$ weak derivatives

Sobolev regularization

Sobolev-regularized risks

- Empirical risk:

$$R^{(\mathrm{reg})}_{n,n_e,n_r}(u_\theta) = R_{n,n_e,n_r}(u_\theta) + \lambda_{(\mathrm{ridge})}\|\theta\|_2^2 + \frac{\lambda_t}{n_r}\sum_{\ell=1}^{n_r}\sum_{|\alpha|\leqslant m+1}\|\partial^\alpha u_\theta(\mathbf{X}^{(r)}_\ell)\|_2^2$$

- Minimizing sequence: $(\hat{\theta}^{(\mathrm{reg})}(p, n_e, n_r, D))_{p\in\mathbb{N}}$

- Theoretical risk:

$$\mathscr{R}^{(\mathrm{reg})}_n(u) = \mathscr{R}_n(u) + \lambda_t\|u\|^2_{H^{m+1}(\Omega)}$$

- The Sobolev regularization is straightforward to implement in the PINN framework with $\mathscr{F}_\alpha(u,\cdot) = \partial^\alpha u$
- Computational scalability via the backpropagation algorithm
- Coercivity of the risk

Strong convergence: PDE solver

Theorem (Linear PDE systems)

Assume that there exists a unique solution $u^\star \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ *to the PDE system, where* $m \geqslant \max_k \deg(\mathscr{F}_k)$*. Thus, taking*

$$\lambda_{(\mathrm{ridge})} = \min(n_e, n_r)^{-\kappa}, \ \kappa = \frac{1}{12 + 4H(1 + (2 + H)(m + 2))},$$

one has, almost surely,

$$\lim_{\lambda_t \to 0} \lim_{D \to \infty} \lim_{n_e, n_r \to \infty} \lim_{p \to \infty} \| u_{\hat{\theta}^{(\mathrm{reg})}(p, n_e, n_r, D, \lambda_t)} - u^\star \|_{H^m(\Omega)} = 0.$$

- The parameters m and $\lambda_{(\mathrm{ridge})}$ do not depend on d_1
- The convergence is in $H^m(\Omega)$ for the penalty $\|u\|^2_{H^{m+1}(\Omega)}$
- Tools: Lax Milgram + functional analysis (weak topology)

Strong convergence: hybrid modeling

Physics inconsistency

For any $u \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, the physics inconsistency of u is defined by

$$\mathrm{PI}(u) = \lambda_e \mathbb{E}\|u(\boldsymbol{X}^{(e)}) - h(\boldsymbol{X}^{(e)})\|_2^2 + \frac{1}{|\Omega|}\sum_{k=1}^{M}\int_\Omega \mathscr{F}_k(u, \boldsymbol{x})^2 d\boldsymbol{x}.$$

Theorem (Linear PDE systems)

Assume that $u^\star \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ *for some* $m \geqslant \max(\lfloor d_1/2 \rfloor, K)$. *Let* $\lambda_e = 1$, $\lambda_t = (\log n)^{-1}$, *and* $\lambda_d = n^{1/2}/(\log n)$. *Then*

$$\lim_{D\to\infty} \lim_{n_e, n_r \to\infty} \lim_{p\to\infty} \mathbb{E}\int_\Omega \|u^{(n)}_{\hat{\theta}^{(\mathrm{reg})}(p, n_e, n_r, D)} - u^\star\|_2^2 d\mu_{\boldsymbol{X}} \lesssim \frac{\log^2(n)}{n^{1/2}}$$

and $$\lim_{D\to\infty} \lim_{n_e, n_r \to\infty} \lim_{p\to\infty} \mathbb{E}(\mathrm{PI}(u^{(n)}_{\hat{\theta}^{(\mathrm{reg})}(p, n_e, n_r, D)})) \leqslant \mathrm{PI}(u^\star) + \underset{n\to\infty}{\mathrm{o}}(1).$$

- Conclusion: statistical accuracy + physical consistency

Summary

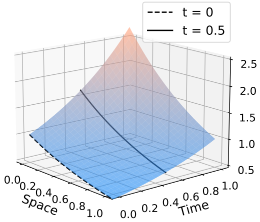
Setting

Regression model:

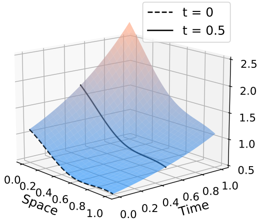
- $Y = u^\star(\boldsymbol{X}) + \mathcal{N}(0, 10^{-2})$
- $u^\star(x,t) = \exp(t-x) + 0.1\cos(2\pi x)$ on $\Omega =]0,1[^2$
- $((x_i, t_i), Y_i)_{1 \leqslant i \leqslant n}$ for $0 < t_i < 0.5$

Advection model:

- $\mathscr{F}(u, \cdot) = \partial_x u + \partial_t u$
- $h(x,0) = \exp(-x)$ and $h(0,t) = \exp(t)$
- $u_{\text{model}}(x,t) = \exp(t-x)$



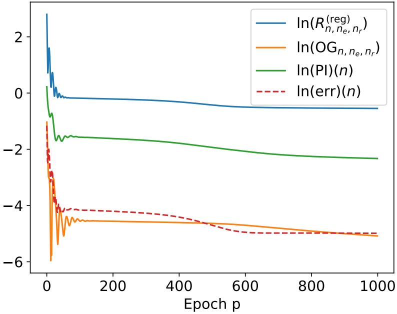
u_{model}



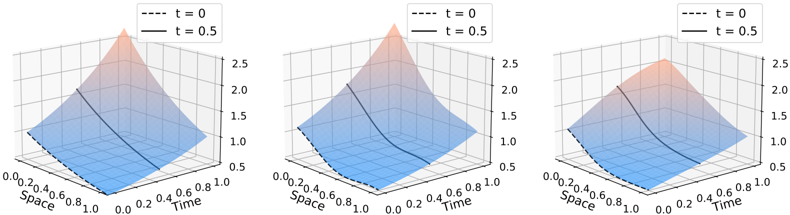
$u^\star$

Monitoring the risks

- Stability of the empirical risk $R^{(\mathrm{reg})}_{n,n_e,n_r} \Rightarrow p \simeq \infty$
- Overfitting gap $\mathrm{OG}_{n,n_e,n_r} = |R^{(\mathrm{ridge})}_{n,n_e,n_r} - \mathscr{R}_n|$
  $\Rightarrow$ choose the lowest possible $\lambda_{(\mathrm{ridge})}$
- Illustration with $n = 10$



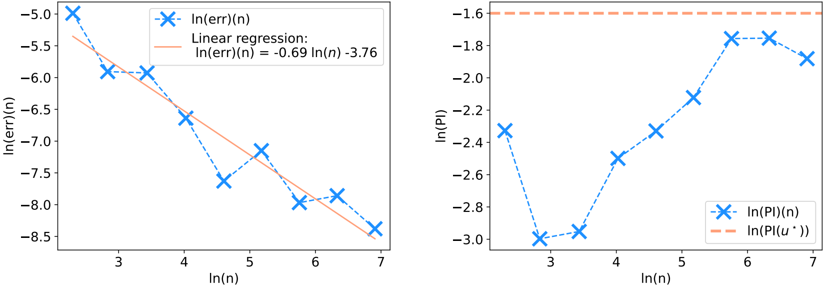
Result for $n = 10^3$



u_{model}, $u^\star$, and regularized PINN estimator

- Convergence on $\mathrm{supp}(\mu_{\boldsymbol{X}}) =]0,1[\times]0,0.5[$
- The regularized PINN follows the advection model (constant on the characteristics $x = t + \mathrm{cst}$)
- Flattening effect of the Sobolev regularization on $\Omega \backslash \mathrm{supp}(\mu_{\boldsymbol{X}})$

Asymptotic in n



As predicted by the theory:

- The convergence rate is less than -0.5
- The regularized PINN is more accurate than u_{model} for $n > 10$
- The physics inconsistency is bounded by the modeling error $\mathrm{PI}(u^\star)$

Thank you for your attention!

The slides and the corresponding paper are available

- on my website https://nathandoumeche.com
- on arXiv 2305.01240
  Convergence and error analysis of PINNs (Doumèche, Biau, Boyer)

The implementation of

- the numerical illustrations
- in particular the Sobolev regularization
- is available on my Github.

Summary

Degree of a polynomial operator

Degree of a monomial operator

The degree of a monomial operator $\mathscr{F}(u, \boldsymbol{x}) = \varphi(\boldsymbol{x}) \times \prod_{i=1}^{N_1} \partial^{\alpha_i} u(\boldsymbol{x})$, where $\varphi \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$, is $\deg \mathscr{F} = \sum_{i=1}^{N_1}(1 + |\alpha_i|)$.

Degree of a polynomial operator

The degree of a polynomial operator $\mathscr{F} = \sum_{i=1}^{N_2} \mathscr{F}_i$, where $\mathscr{F}_i$ is a monomial operator, is $\deg \mathscr{F} = \max_i \deg(\mathscr{F}_i)$.

- $\deg(\partial_x u) = 2$
- $\deg(u_y \partial_y u_x) = 3$
- $\deg(\sin(\boldsymbol{x}) u_x + \exp(\boldsymbol{x}) \partial^2_{x,y} u_y) = 3$

Lax-Milgram for regularized PINNs

Prop (Characterization of the unique minimizer of $\mathscr{R}_n^{(\mathrm{reg})}$)

Assume that $\mathscr{F}_1, \dots, \mathscr{F}_M$ are affine operators of order K. i.e., $\mathscr{F}_k = \mathscr{F}_k^{(\mathrm{lin})} + B_k$, that $\lambda_t > 0$ and $m \geqslant \max(\lfloor d_1/2 \rfloor, K)$. Then the regularized theoretical risk $\mathscr{R}_n^{(\mathrm{reg})}$ has a unique minimizer $\hat{u}_n$ over $H^{m+1}(\Omega, \mathbb{R}^{d_2})$, satisfying

$$\forall v \in H^{m+1}(\Omega, \mathbb{R}^{d_2}), \quad \mathcal{A}_n(\hat{u}_n, v) = \mathcal{B}_n(v), \qquad \text{where}$$

$$\begin{aligned}
\mathcal{A}_n(\hat{u}_n, v) = \frac{\lambda_d}{n} \sum_{i=1}^{n} \langle \tilde{\Pi}(\hat{u}_n)(\mathbf{X}_i), \tilde{\Pi}(v)(\mathbf{X}_i) \rangle + \lambda_e \mathbb{E} \langle \tilde{\Pi}(\hat{u}_n)(\mathbf{X}^{(e)}), \tilde{\Pi}(v)(\mathbf{X}^{(e)}) \rangle \\
+ \frac{1}{|\Omega|} \sum_{k=1}^{M} \int_\Omega \mathscr{F}_k^{(\mathrm{lin})}(\hat{u}_n, \mathbf{x}) \mathscr{F}_k^{(\mathrm{lin})}(v, \mathbf{x}) d\mathbf{x} \\
+ \frac{\lambda_t}{|\Omega|} \sum_{|\alpha| \leqslant m+1} \int_\Omega \langle \partial^\alpha \hat{u}_n(\mathbf{x}), \partial^\alpha v(\mathbf{x}) \rangle d\mathbf{x},
\end{aligned}$$

$$\begin{aligned}
\mathcal{B}_n(v) = \frac{\lambda_d}{n} \sum_{i=1}^{n} \langle Y_i, \tilde{\Pi}(v)(\mathbf{X}_i) \rangle + \lambda_e \mathbb{E} \langle \tilde{\Pi}(v)(\mathbf{X}^{(e)}), h(\mathbf{X}^{(e)}) \rangle \\
- \frac{1}{|\Omega|} \sum_{k=1}^{M} \int_\Omega B_k(\mathbf{x}) \mathscr{F}_k^{(\mathrm{lin})}(v, \mathbf{x}) d\mathbf{x},
\end{aligned}$$

Lax-Milgram for regularized PINNs 2/2

- Sobolev embedding $\tilde{\Pi} : H^{m+1}(\Omega, \mathbb{R}^{d_2}) \to C^0(\Omega, \mathbb{R}^{d_2})$, i.e., $\tilde{\Pi}(u)$ is the unique continuous function that coincides with u almost everywhere.
- Minimizing $\mathscr{R}_n^{(\mathrm{reg})}$ amounts to minimizing $\mathcal{A}_n - 2\mathcal{B}_n$
- Weak formulation on $H^{m+1}(\Omega, \mathbb{R}^{d_2})$: if $\hat{u}_n \in H^{2(m+1)}(\Omega, \mathbb{R}^{d_2})$, then almost everywhere,

$$\sum_{k=1}^{M} (\mathscr{F}_k^{(\mathrm{lin})})^* \mathscr{F}_k(\hat{u}_n, \boldsymbol{x}) + \lambda_t \sum_{|\alpha| \leqslant m+1} (-1)^{|\alpha|} (\partial^\alpha)^2 \hat{u}_n(\boldsymbol{x}) = 0.$$

$(\mathscr{F}_k^{(\mathrm{lin})})^*$: adjoint operator of $\mathscr{F}_k^{(\mathrm{lin})}$, i.e., for all $u, v \in C^\infty(\Omega, \mathbb{R})$ with $v|_{\partial\Omega} = 0$,

$$\int_\Omega u \mathscr{F}^{(\mathrm{lin})}(v, \boldsymbol{x}) d\boldsymbol{x} = \int_\Omega (\mathscr{F}_k^{(\mathrm{lin})})^*(u, \boldsymbol{x}) v d\boldsymbol{x}.$$

- In the regime $\lambda_t \to 0$, the solution of the PINN problem does not satisfy the constraints $\mathscr{F}_k(u, \boldsymbol{x}) = 0$, but the slightly different ones $\sum_{k=1}^{M} (\mathscr{F}_k^{(\mathrm{lin})})^* \mathscr{F}_k(u, \boldsymbol{x}) = 0$.

Summary

PINNs v.s. data assimilation

PINNs

- Estimate the function $u^\star$ such that $Y = u^\star(\boldsymbol{X}) + \varepsilon$

Data assimilation (Cressman analysis, optimal interpolation, Kalman...)

- Propagate a forecast with a model, then apply corrections from new observations (innovation)

Similarities:

- Enhancing a statistical model with physics
- Noisy observations + imperfect model

Differences:

- Data assimilation has an inherent time-series structure
- Data assimilation are used for non-reproducible experiments
- Data assimilation needs to have a complete system of equations