Some statistical insights into PINNs

CRM Spring School, 2025

SCIENCES SORBONNE UNIVERSITÉ

université PARIS-SACLAY

Work team

Nathan Doumèche

Gérard Biau

Claire Boyer

Hybrid modeling

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

Goal: estimate $f^\star$ using

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

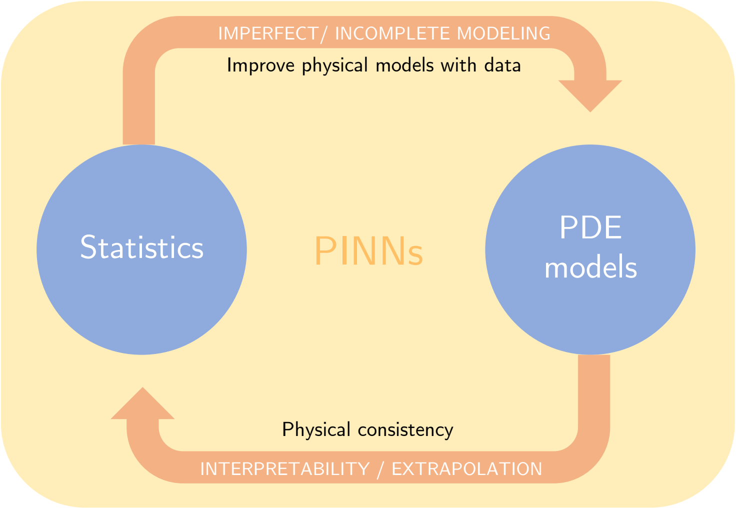
$$\mathscr{D}_k(f^\star, \cdot) \simeq 0, \quad 1 \leqslant k \leqslant M,$$

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

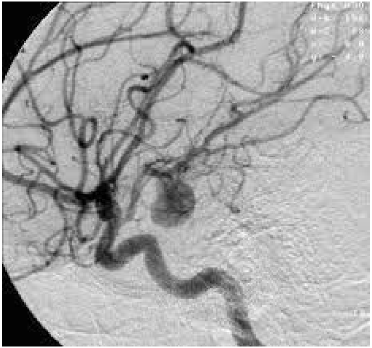
- Neural networks

- Physics-Informed Neural Networks (PINNs)

PINNs in a nutshell

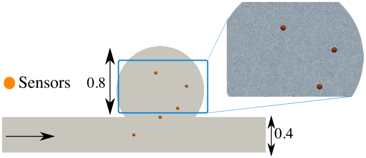


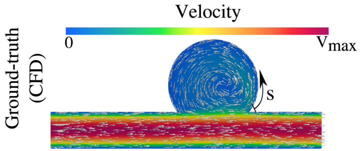
Example: Blood flow in an aneurysm



Modeling the blood flow

[Arzani et al., 2021]





Goal: estimate the blood flow $f = (f_x, f_y, P)$

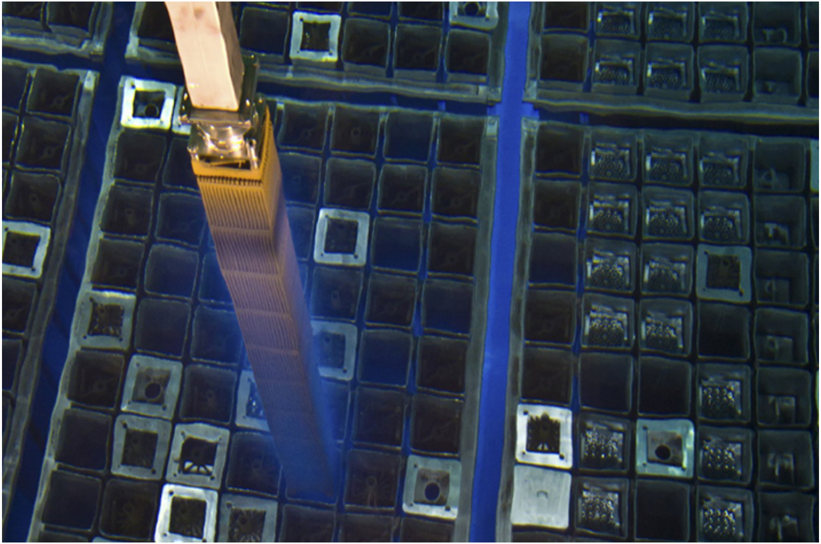
Navier-Stokes equations:

- $\mathscr{D}_1(f\cdot) = f_x \partial_x f_x + f_y \partial_y f_x - \partial^2_{x,x} f_x - \partial^2_{y,y} f_x + \partial_x P$
- $\mathscr{D}_2(f,\cdot) = f_x \partial_x f_y + f_y \partial_y f_y - \partial^2_{x,x} f_y - \partial^2_{y,y} f_y + \partial_y P$
- $\mathscr{D}_3(f,\cdot) = \partial_x f_x + \partial_y f_y$

(Incomplete) boundary conditions:

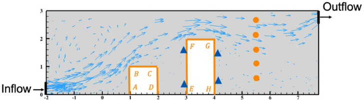
- $(f_x, f_y) = 0$ on the boundaries of the vessel
- Unknown inflow and outflow

Example: Heat transfer in uranium bundles

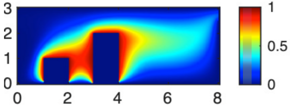


Modeling the heat transfer

[Cai et al., 2021]



▲ Temperature sensors
● Velocity sensors



Goal: estimate the temperature T on the bundles

Navier-Stokes and diffusion equations on $f = (f_x, f_y, P, T)$:

- $\mathscr{D}_1(f, \cdot) = f_x \partial_x f_x + f_y \partial_y f_x - \partial^2_{x,x} f_x - \partial^2_{y,y} f_x + \partial_x P$
- $\mathscr{D}_2(f, \cdot) = f_x \partial_x f_y + f_y \partial_y f_y - \partial^2_{x,x} f_y - \partial^2_{y,y} f_y + \partial_y P$
- $\mathscr{D}_3(f, \cdot) = \partial_x f_x + \partial_y f_y$
- $\mathscr{D}_4(f, \cdot) = f_x \partial_x T + f_y \partial_y T - \partial^2_{x,x} T - \partial^2_{y,y} T$

(Incomplete) boundary conditions:

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

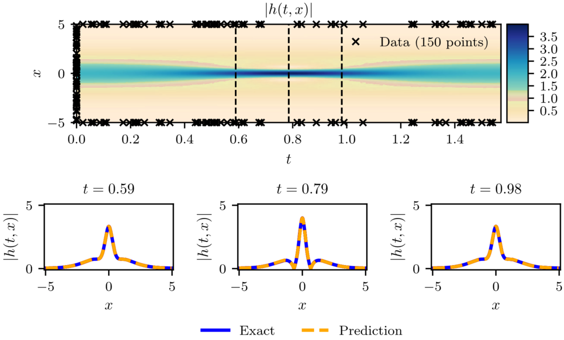
The PDE solver case

Specificity: no data Y_i and exact modeling

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

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

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



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

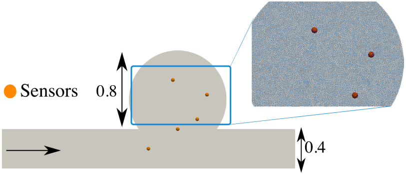
Our objective

To better understand the capabilities and limitations of PINNs

Summary

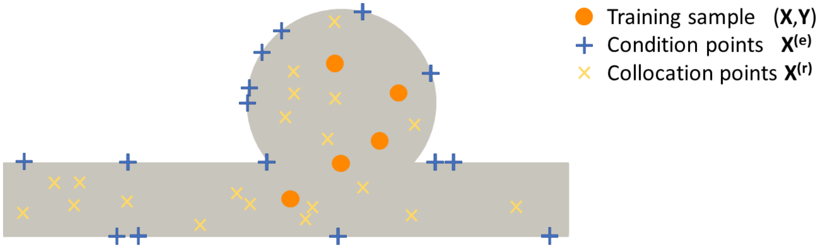
Summary

Geometry of the problem



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

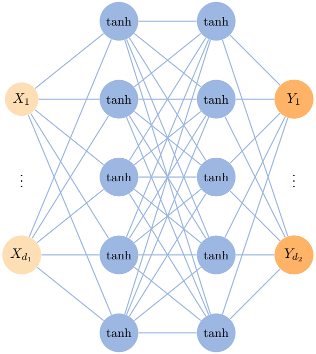
Empirical risk function

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

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

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
- $f_\theta \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$



Minimizing sequence

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

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

- 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 $f^\star$ 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 $f \in C^\infty(\bar{\Omega}, \mathbb{R}^{d_2})$, there exists a sequence $(f_p)_{p \in \mathbb{N}} \in \mathrm{NN}_H$ such that $\lim_{p\to\infty} \|f - f_p\|_{C^K(\Omega)} = 0$.

- 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

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

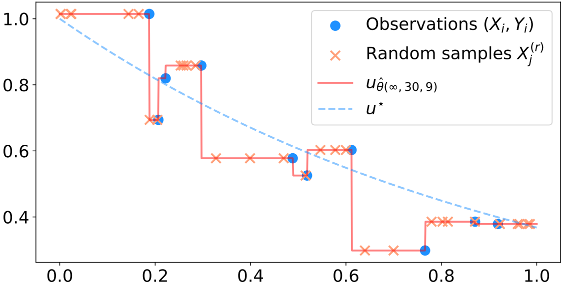
A natural requirement: Risk-consistency

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

- Warning: possible overfitting

Overfitting: hybrid modeling

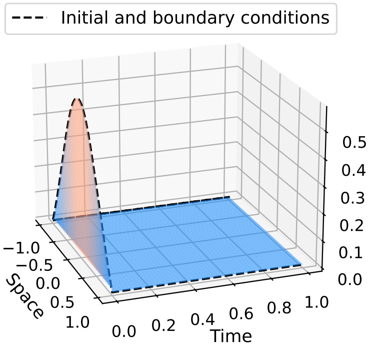
- Observations: $Y_i = f^\star(\boldsymbol{X}_i) + \varepsilon_i$
- Goal: estimate the trajectory $f^\star$ on $\Omega =]0, 1[$
- Model (dynamics with friction): $\mathscr{D}(f, \boldsymbol{x}) = f''(\boldsymbol{x}) + f'(\boldsymbol{x})$



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

Overfitting: PDE solver

- Heat equation: $\mathscr{D}(f, \boldsymbol{x}) = \partial_t f(\boldsymbol{x}) - \partial^2_{x,x} f(\boldsymbol{x})$ + boundary/initial conditions
- Goal: reconstruct the solution $f^\star$ 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

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

Ridge PINNs

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

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 $f = (f_x, f_y, P)$:

- $\mathscr{D}_1(f,\cdot) = f_x \partial_x f_x + f_y \partial_y f_x - \partial^2_{x,x} f_x - \partial^2_{y,y} f_x + \partial_x P$
- $\mathscr{D}_1(f,\cdot) = \mathscr{P}(f_x, \partial_x f_x, \partial^2_{x,x} f_x, \partial_y f_x, \partial^2_{y,y} f_x, f_y, \partial_x P)$

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

Polynomial operator

An operator $\mathscr{D}(u,\cdot)$ 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
- $\mathscr{D}_1, \ldots, \mathscr{D}_M$ are polynomial operators

Theorem

With a ridge hyperparameter of the form

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

one has, almost surely,

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

and

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

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 $\lim_{D, n_e, n_r, p \to \infty} f_{\hat{\theta}^{(\text{ridge})}(p, n_e, n_r, D)} = f^\star$ in $L^2(\Omega)$?

Answer: No

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

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

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

✗ 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:

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

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

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

- The Sobolev regularization is straightforward to implement in the PINN framework with $\mathscr{D}_\alpha(f,\cdot) = \partial^\alpha f$
- 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 $f^\star \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ to the PDE system, where $m \geqslant \max_k \deg(\mathscr{D}_k)$. Thus, taking

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

one has, almost surely,

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

- The parameters m and $\lambda_{(\text{ridge})}$ 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 $f \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$, the physics inconsistency of f is defined by

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

Theorem (Linear PDE systems)

Assume that $f^\star \in H^{m+1}(\Omega, \mathbb{R}^{d_2})$ *for some* $m \geqslant \max(\lfloor d_1/2 \rfloor, K)$. *Let* $\lambda_{e/r} = \log(n)/\sqrt{n}$ *and* $\lambda_{(\mathrm{sob})} = 1/\sqrt{n}$. *Then*

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

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

- Conclusion: statistical accuracy + physical consistency

Summary

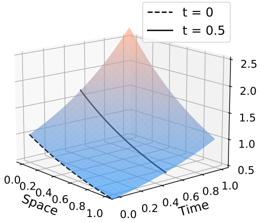
Setting

Regression model:

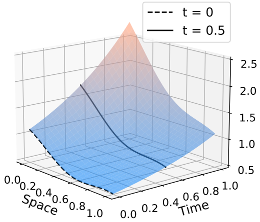
- $Y = f^\star(\boldsymbol{X}) + \mathcal{N}(0, 10^{-2})$
- $f^\star(x,t) = \exp(t-x) + 0.1\cos(2\pi x)$ on $\Omega =]0,1[^2$
- $((x_i, t_i), Y_i)_{1\leqslant i\leqslant n}$ for $0 < t_i < 0.5$

Advection model:

- $\mathscr{D}(f, \cdot) = \partial_x f + \partial_t f$
- $h(x,0) = \exp(-x)$ and $h(0,t) = \exp(t)$
- $f_{\mathrm{model}}(x,t) = \exp(t-x)$



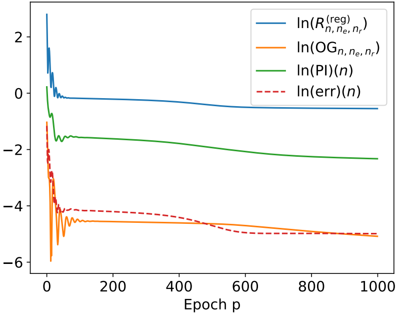
f_{model}



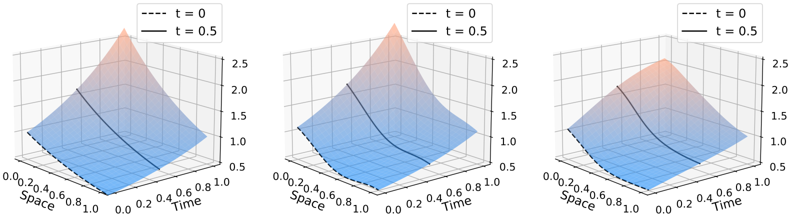
$f^\star$

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 $n = 10$



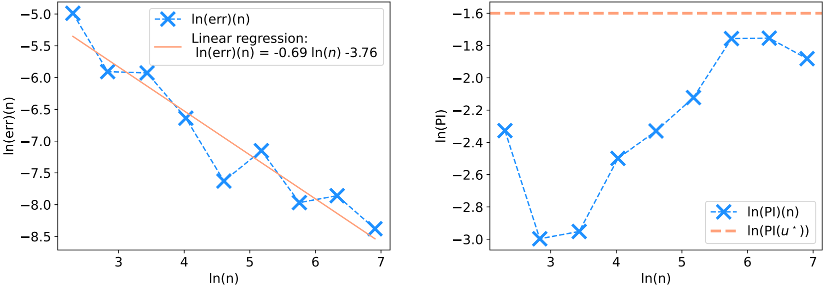
Result for $n = 10^3$



f_{model}, $f^\star$, 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 f_{model} for $n > 10$
- The physics inconsistency is bounded by the modeling error $\mathrm{PI}(f^\star)$

Conclusion

- Statistical and PDE models can successfully complement each other
- Risk-consistency and strong convergence are guaranteed, with an implementable regularization
- PINNs are an interesting approach to scientific computing

Thank you!

☞ Doumèche, Biau, Boyer - Bernoulli 2025 [arXiv] [pdf] [supp]

☞ Code available here