A statistical perspective into physics-informed learning:

from PINNs to kernels

CEA-EDF-Inria Spring school, June 2025

Work team

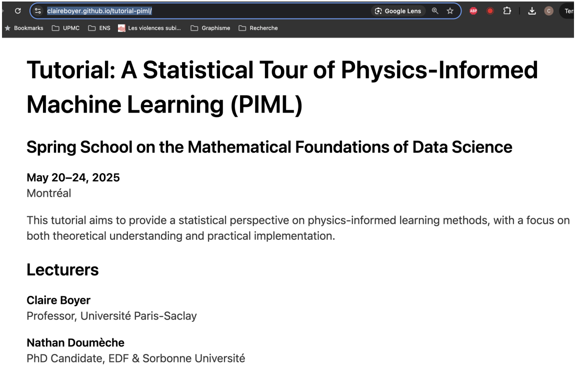
Nathan Doumèche

Francis Bach

Gérard Biau

Claire Boyer

Shameless Advertising



Tutorial: A Statistical Tour of Physics-Informed Machine Learning (PIML)

Spring School on the Mathematical Foundations of Data Science

May 20–24, 2025
Montréal

This tutorial aims to provide a statistical perspective on physics-informed learning methods, with a focus on both theoretical understanding and practical implementation.

Lecturers

Claire Boyer
Professor, Université Paris-Saclay

Nathan Doumèche
PhD Candidate, EDF & Sorbonne Université

https://claireboyer.github.io/tutorial-piml/

Summary

Summary

Hybrid modeling

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

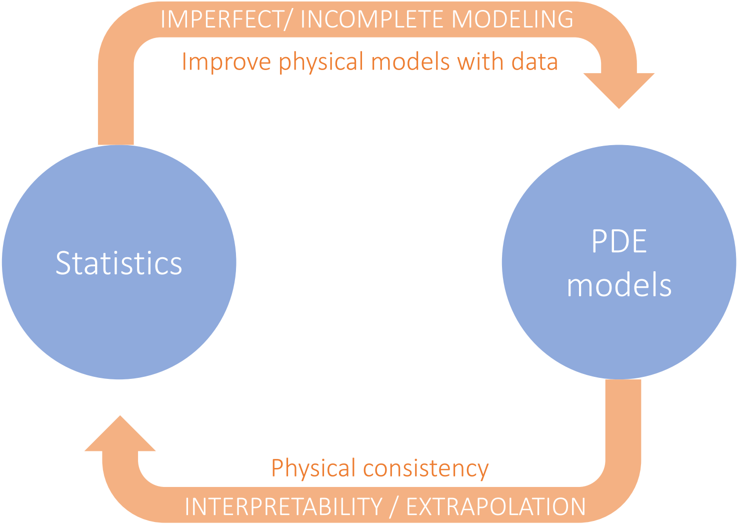
Goal: estimate $f^\star$ using

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

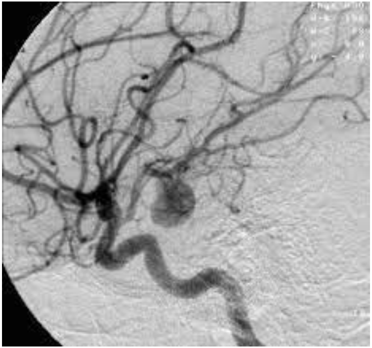
$$\mathscr{D}(f^\star, \cdot) \simeq 0$$

with a known differential operator $\mathscr{D}$

Why combining learning with physics?

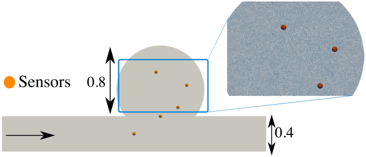


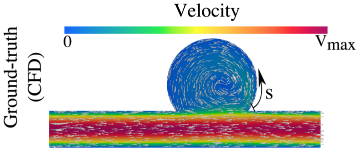
Example: Blood flow in an aneurysm



Modeling the blood flow

[Arzani et al., 2021]



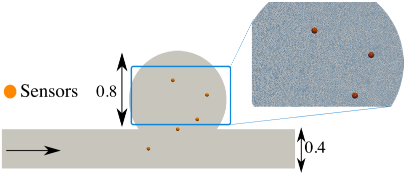


Goal: estimate the blood flow $f = (f_x, f_y, P)$

Navier-Stokes equations:

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

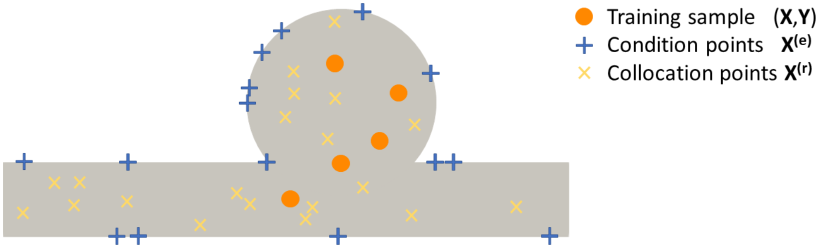
Geometry of the problem



- $\Omega \subseteq [-L, L]^d$: the bounded set on which the problem is posed
- $f^\star : \Omega \to \mathbb{R}$: the unknown target function
- Differential operator $\mathscr{D}(f^\star, \cdot) \simeq 0$ on Ω

Summary

Three samplings



Physics-informed empirical risk

- Training sample $(X_1, Y_1), \dots, (X_n, Y_n) \in \Omega \times \mathbb{R}^{d_2}$ (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

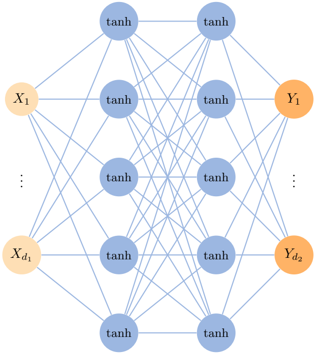
$$R_{n,n_e,n_r}(f_\theta) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \|f_\theta(X_i) - Y_i\|_2^2}_{\text{data-fidelity}} + \underbrace{\frac{\lambda_{(\mathrm{pde})}}{n_r}\sum_{\ell=1}^{n_r} \|\mathscr{D}(f_\theta, \boldsymbol{X}_\ell^{(r)})\|_2^2}_{\text{PDEs}}$$

$$+ \underbrace{\frac{\lambda_e}{n_e}\sum_{j=1}^{n_e} \|f_\theta(\boldsymbol{X}_j^{(e)}) - h(\boldsymbol{X}_j^{(e)})\|_2^2}_{\text{boundary conditions}}$$

- Physics-Informed Neural Networks: NN obtained after training

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

Theoretical risk and consistency

Theoretical risk

$$\mathscr{R}_n(f) = \frac{1}{n}\sum_{i=1}^{n} \|f(X_i) - Y_i\|_2^2 + \lambda_e \mathbb{E}\|f(\boldsymbol{X}^{(e)}) - h(\boldsymbol{X}^{(e)})\|_2^2 + \frac{\lambda_{(\mathrm{pde})}}{|\Omega|}\int_\Omega \|\mathscr{D}(f, \boldsymbol{x}^{(r)})\|_2^2 d\boldsymbol{x}^{(r)}$$

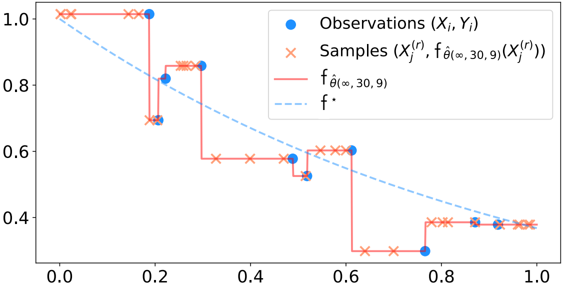
A natural requirement: Risk-consistency

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

- Warning: possible overfitting

Overfitting

- Observations: $Y_i = f^\star(X_i) + \varepsilon_i$
- 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 $\mathscr{R}_n = \infty$
- Can happen also in the PDE solver case

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

Risk-consistency of ridge PINNs

Assumptions:

- The condition function h is Lipschitz
- $\mathscr{D}$ polynomial operator

Theorem

With a ridge hyperparameter of the form

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

one has, almost surely,

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

and

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

Beyond risk-consistency: strong convergence?

✓ Ridge PINNs are risk-consistent

Question

Is this sufficient to have $\lim_{D, n_e, n_r, p \to \infty} f_{\hat{\theta}^{(\mathrm{ridge})}(p, n_e, n_r, D)} = f^\star$ 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 $(X_i, Y_i)_{1 \leqslant i \leqslant n}$, 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 regularization

Sobolev-regularized risks

- Empirical risk:

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

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

- Theoretical risk:

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

- 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: hybrid modeling

Physics inconsistency

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

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

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* $\lambda_{e/(\mathrm{pde})} = \log(n)/\sqrt{n}$ *and* $\lambda_{(\mathrm{sob})} = 1/\sqrt{n}$. *Then*

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

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

- Conclusion: statistical accuracy + physical consistency

Overall: statistical convergence of PINNs

[Doumèche, Biau, Boyer, 2024]

- Training of a NN with large width and a few hidden layers

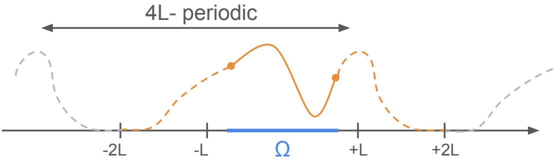
$$\widehat{NN}_\theta \in \mathrm{argmin}_{NN} \frac{1}{n}\sum_{i=1}^{n} \|f_\theta(X_i) - Y_i\|_2^2 + \frac{\lambda_e}{n_r}\sum_{\ell=1}^{n_r} \|\mathscr{D}(f_\theta, \boldsymbol{X}_\ell^{(r)})\|_2^2$$
$$+ \frac{\lambda_{(\mathrm{pde})}}{n_r}\sum_{\ell=1}^{n_r}\Big(\sum_{|\alpha|\leqslant s} \|\partial^\alpha f_\theta(X_i^{(r)})\|_2^2\Big) + \lambda^{\mathrm{ridge}}\|\theta\|_2^2$$

- ☞ Ridge regularization to prevent overfitting
- ☞ Sobolev and ridge regularizations for strong convergence
- ☞ Could we avoid the discretization issue and the delicate training of PINNs?

Summary

Towards a kernel approach

- Assumption: $f^\star \in H^s(\Omega)$
- Extension: $H^s(\Omega) \hookrightarrow H^s_{\mathrm{per}}([-2L, 2L]^d)$
- $H^s_{\mathrm{per}}([-2L, 2L]^d)$ = subspace of $H^s([-2L, 2L]^d)$ of functions whose $4L$-periodic extension is still s-times weakly differentiable
- Important: $f^\star \in H^s(\Omega) \iff f^\star \in H^s([-2L, 2L]^d)$



Empirical risk

$$R_n(f) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} |f(X_i) - Y_i|^2}_{\text{data-fidelity term}} + \underbrace{\lambda_{(\mathrm{sob})} \|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)}}_{\text{target regularity}} + \underbrace{\lambda_{(\mathrm{pde})} \|\mathscr{D}(f)\|^2_{L^2(\Omega)}}_{\text{PDE}}$$

Objective

Framing PIML as a kernel method

$$\hat{f}_n = \underset{f \in H^s_{\mathrm{per}}([-2L,2L]^d)}{\operatorname{argmin}} \ \frac{1}{n}\sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \|f\|^2_{\mathrm{RKHS}},$$

with $\|f\|^2_{\mathrm{RKHS}} = \lambda_{(\mathrm{sob})}\|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)} + \lambda_{(\mathrm{pde})}\|\mathscr{D}(f)\|^2_{L^2(\Omega)}$

- How does the PDE penalty impact learning?
- How to leverage the kernel toolbox?
- How to define a tractable estimator?

Assumption

$\mathscr{D}$ is a linear operator of the derivatives of f.

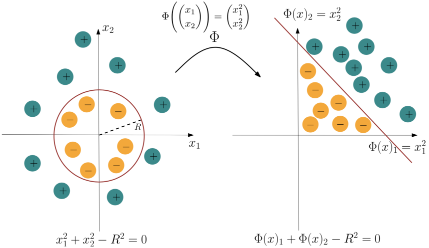
Survival kit: Kernel ridge regression

- Kernel: a symmetric positive definite function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, i.e., $\sum_{i,i'=1}^{n} \alpha_i \alpha_{i'} K(x_i, x_{i'}) \geqslant 0$
- There exists a Hilbert space RKHS of functions $f : \mathcal{X} \to \mathbb{R}$ such that

$$\begin{array}{ll} (i) & \forall x \in \mathcal{X},\ K(\cdot, x) \in \mathrm{RKHS} \\ (ii) & \forall f \in \mathrm{RKHS},\ \langle f, K(\cdot, x) \rangle_{\mathrm{RKHS}} = f(x) \end{array}$$

- Reproducing kernel Hilbert space with reproducing kernel K
- Example: polynomial kernel

$$K(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{T}} = \left\langle \begin{pmatrix} x \\ x^2 \end{pmatrix}, \begin{pmatrix} x' \\ (x')^2 \end{pmatrix} \right\rangle_{\mathcal{T}}$$



Survival kit: The kernel trick

- Regularized empirical risk minimization

$$\hat{f}_n = \operatorname*{argmin}_{f \in \mathrm{RKHS}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 + \lambda \|f\|^2_{\mathrm{RKHS}}$$

- Representer theorem

$$\hat{f}_n(x) = \sum_{i=1}^{n} \hat{\alpha}_i K(x, X_i)$$

- The solution lives in a finite-dimensional subspace!

Survival kit: Kernel methods

- We solve a finite-dimensional problem

$$
\begin{aligned}
\hat{\alpha} &= \operatorname*{argmin}_{\alpha \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^{n} \Big(Y_i - \sum_{j=1}^{n} \alpha_j K(X_i, X_j) \Big)^2 + \lambda \Big\| \sum_{j=1}^{n} \alpha_j K(\cdot, X_j) \Big\|_{\mathrm{RKHS}}^2 \\
&= \operatorname*{argmin}_{\alpha \in \mathbb{R}^n} \frac{1}{n} \|\mathbb{Y} - \mathbb{K}\alpha\|_2^2 + \lambda \alpha^\top \mathbb{K} \alpha \\
&= (\mathbb{K} + n\lambda I_n)^{-1} \mathbb{Y}
\end{aligned}
$$

- Final predictor

$$
\hat{f}_n(x) = \sum_{i=1}^{n} \hat{\alpha}_i K(x, X_i)
$$

☺ No need to explicitly use/know φ to train a kernel ridge regressor!

Survival kit: Effective dimension & convergence

- Integral/covariance operator $L_K : L^2(\mathcal{X}, \mathbb{P}_X) \to L^2(\mathcal{X}, \mathbb{P}_X)$, defined by

$$\forall f \in L^2(\mathcal{X}, \mathbb{P}_X), \forall x \in \mathcal{X}, \quad L_K f(x) = \int_{\mathcal{X}} K(x, y) f(y) d\mathbb{P}_X(y)$$

- Effective dimension: $\mathrm{tr}(L_K(\lambda \mathrm{Id} + L_K)^{-1})$
- Convergence rate of the kernel method

$$\mathbb{E} \int_{\mathcal{X}} |\hat{f}_n - f^\star|^2 d\mathbb{P}_X = \mathcal{O}\Big(\frac{\text{Effective dimension}}{n}\Big)$$

Underlying RKHS for linear PDEs

Lemma

There exists a positive operator $\mathscr{O}_n$ on $L^2([-2L, 2L]^d)$ such that, for any $f \in H^s_{\mathrm{per}}([-2L, 2L]^d)$,

$$\|\mathscr{O}_n^{-1/2}(f)\|^2_{L^2([-2L,2L]^d)} = \lambda_{(\mathrm{sob})}\|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)} + \lambda_{(\mathrm{pde})}\|\mathscr{D}(f)\|^2_{L^2(\Omega)}.$$

☞ This suggests the inner product

$$\langle f, g\rangle_{\mathrm{RKHS}} = \langle \mathscr{O}_n^{-1/2}(f), \mathscr{O}_n^{-1/2}(g)\rangle_{L^2([-2L,2L]^d)}$$

Underlying RKHS for linear PDEs II

- For any $f \in L^2([-2L, 2L]^d)$ and $x \in [-2L, 2L]^d$,

$$\mathscr{O}_n(f)(x) = \sum_{m \in \mathbb{N}} a_m \langle f, v_m \rangle_{L^2([-2L,2L]^d)} v_m(x)$$

- Orthonormal basis of eigenfunctions $v_m \in H^s_{\mathrm{per}}([-2L, 2L]^d)$
- Eigenvalues $a_m > 0$

Theorem

The space $H^s_{\mathrm{per}}([-2L, 2L]^d)$, equipped with the inner product

$$\langle f, g \rangle_{\mathrm{RKHS}} = \langle \mathscr{O}_n^{-1/2} f, \mathscr{O}_n^{-1/2} g \rangle_{L^2([-2L,2L]^d)},$$

is a reproducing kernel Hilbert space. For $f \in H^s_{\mathrm{per}}([-2L, 2L]^d)$,

$$\|f\|^2_{\mathrm{RKHS}} = \lambda_{(\mathrm{sob})} \|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)} + \lambda_{(\mathrm{pde})} \|\mathscr{D}(f)\|^2_{L^2(\Omega)},$$

and the associated kernel is $K(x, y) = \sum_{m \in \mathbb{N}} a_m v_m(x) v_m(y)$.

PIML as a kernel method

$$\begin{aligned}\hat{f}_n &= \operatorname*{argmin}_{f \in H^s_{\mathrm{per}}([-2L,2L]^d)} \frac{1}{n}\sum_{i=1}^n |f(X_i) - Y_i|^2 + \lambda_{(\mathrm{sob})}\|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)} + \lambda_{(\mathrm{pde})}\|\mathscr{D}(f)\|^2_{L^2(\Omega)} \\ &= \operatorname*{argmin}_{f \in H^s_{\mathrm{per}}([-2L,2L]^d)} \frac{1}{n}\sum_{i=1}^n |f(X_i) - Y_i|^2 + \|f\|^2_{\mathrm{RKHS}}\end{aligned}$$

- ✓ $\hat{f}_n$ is a kernel method
- ✗ Computing the kernel is not straightforward

Proposition (Characterization of the kernel)

The kernel K is the unique solution to the following weak formulation, valid for all test functions $\varphi \in H^s_{\mathrm{per}}([-2L,2L]^d)$: for all $x \in \Omega$,

$$\lambda_{(\mathrm{sob})} \sum_{|\alpha| \leqslant s} \int_{[-2L,2L]^d} \partial^\alpha K(x,\cdot)\, \partial^\alpha \varphi + \lambda_{(\mathrm{pde})} \int_\Omega \mathscr{D}(K(x,\cdot))\, \mathscr{D}(\varphi) = \varphi(x).$$

Convergence rate of the PIML kernel method

- Integral operator $L_K : L^2(\Omega, \mathbb{P}_X) \to L^2(\Omega, \mathbb{P}_X)$, defined by

$$\forall f \in L^2(\Omega, \mathbb{P}_X), \forall x \in \Omega, \quad L_K f(x) = \int_\Omega K(x,y) f(y) d\mathbb{P}_X(y)$$

- Effective dimension $\mathscr{N}(\lambda_{(\mathrm{sob})}, \lambda_{(\mathrm{pde})}) = \mathrm{tr}(L_K(\mathrm{Id} + L_K)^{-1})$

Theorem (Convergence rate)

Assume that $f^\star \in H^s(\Omega)$, $s > d/2$, $\frac{d\mathbb{P}_X}{dx} \leqslant \kappa$, and the noise ε is (M, σ)-sub-Gamma. Then, for all n large enough,

$$\begin{aligned}
&\mathbb{E} \int_\Omega |\hat{f}_n - f^\star|^2 d\mathbb{P}_X \\
&\quad \lesssim \log^2(n) \Big(\lambda_{(\mathrm{sob})} \|f^\star\|^2_{H^s(\Omega)} + \lambda_{(\mathrm{pde})} \|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)} + \\
&\qquad\qquad \frac{M^2}{n^2 \lambda_{(\mathrm{sob})}} + \frac{\sigma^2 \mathscr{N}(\lambda_{(\mathrm{sob})}, \lambda_{(\mathrm{pde})})}{n} \Big).
\end{aligned}$$

Convergence rate of the PIML kernel method

✗ Not easy to characterize the eigenvalues of L_K for the PIML estimator

- A simple bound on $\mathscr{N}(\lambda_{(\mathrm{sob})}, \lambda_{(\mathrm{pde})})$ shows that

$$\mathbb{E}\int_\Omega |\hat{f}_n - f^\star|^2 d\mathbb{P}_X = \mathcal{O}_n\big(n^{-2s/(2s+d)} \log^3(n)\big)$$

- Can we do better?

A toy example

- $d = 1$, $\Omega = [-L, L]$, $\Omega^{\text{aug}} = [-2L, 2L]$
- $f^\star \in H^1(\Omega)$
- $\mathscr{D} = \frac{d}{dx}$ ($f^\star$ is approximately constant)

$$\hat{f}_n = \underset{f \in H^1_{\text{per}}(\Omega^{\text{aug}})}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_{(\text{sob})}\|f\|^2_{H^1_{\text{per}}(\Omega^{\text{aug}})} + \lambda_{(\text{pde})}\|\mathscr{D}(f)\|^2_{L^2(\Omega)}$$

- Explicit kernel

$$K(x,y) = \frac{\gamma_n}{2\lambda_{(\text{sob})}\sinh(2\gamma_n L)}\Big((\cosh(2\gamma_n L) + \cosh(2\gamma_n x))\cosh(\gamma_n(x-y)) + ((1 - 2\times 1_{x>y})\sinh(2\gamma_n L) - \sinh(2\gamma_n x))\sinh(\gamma_n(x-y))\Big)$$

with $\gamma_n = \sqrt{\lambda_{(\text{sob})}/(\lambda_{(\text{sob})} + \lambda_{(\text{pde})})}$

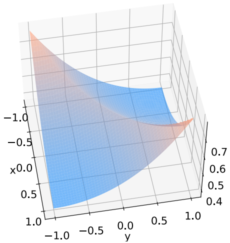


Fig.: K with $\lambda_{(\text{sob})} = \lambda_{(\text{pde})} = 1$

Speed-up of the physical penalty

Theorem (Kernel speed-up)

Let $\lambda_{(\mathrm{sob})} = n^{-1}\log(n)$ *and*

$$\lambda_{(\mathrm{pde})} = \begin{cases} n^{-2/3}/\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} & \text{if} \quad \|\mathscr{D}(f^\star)\|_{L^2(\Omega)} \neq 0 \\ 1/\log(n) & \text{if} \quad \|\mathscr{D}(f^\star)\|_{L^2(\Omega)} = 0. \end{cases}$$

Then

$$\mathbb{E}\int_{[-L,L]} |\hat{f}_n - f^\star|^2 d\mathbb{P}_X = \|\mathscr{D}(f^\star)\|_{L^2(\Omega)}\, \mathcal{O}_n\big(n^{-2/3}\log^3(n)\big) + \big(\|f^\star\|^2_{H^s(\Omega)} + \underbrace{\sigma^2 + M^2}_{\text{noise param}}\big)\, \mathcal{O}_n\big(n^{-1}\log^3(n)\big).$$

Speed-up of the physical penalty

Theorem (Kernel speed-up)

Let $\lambda_{(\mathrm{sob})} = n^{-1}\log(n)$ *and*

$$\lambda_{(\mathrm{pde})} = \begin{cases} n^{-2/3}/\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} & \text{if} \quad \|\mathscr{D}(f^\star)\|_{L^2(\Omega)} \neq 0 \\ 1/\log(n) & \text{if} \quad \|\mathscr{D}(f^\star)\|_{L^2(\Omega)} = 0. \end{cases}$$

Then

$$\mathbb{E}\int_{[-L,L]} |\hat{f}_n - f^\star|^2 d\mathbb{P}_X = \|\mathscr{D}(f^\star)\|_{L^2(\Omega)}\, \mathcal{O}_n\big(n^{-2/3}\log^3(n)\big) + \big(\|f^\star\|^2_{H^s(\Omega)} + \underbrace{\sigma^2 + M^2}_{\text{noise param}}\big)\, \mathcal{O}_n\big(n^{-1}\log^3(n)\big).$$

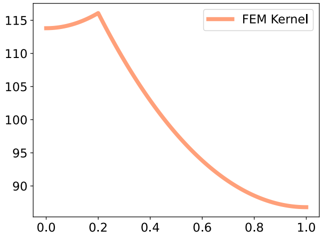
- ✓ When $\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} = 0 \to$ parametric rate of n^{-1}
- ✓ When $\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} > 0 \to$ Sobolev minimax rate in $H^1(\Omega)$ of $n^{-2/3}$

Summary

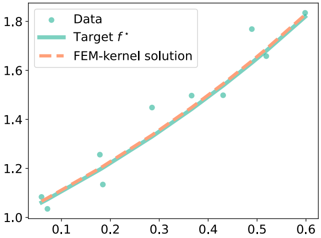
In practice

- Kernel estimator: $\hat{f}_n(x) = (K(x, X_1), \ldots, K(x, X_n))(\mathbb{K} + nI_n)^{-1}\mathbb{Y}$
- Problem: what to do when K is not explicit?
- Solution 1: finite element methods to solve a weak PDE

$$\lambda_{(\mathrm{sob})} \int_\Omega \Big[K(x,\cdot)\,\varphi + \sum_{|\alpha|=s} \partial^\alpha K(x,\cdot)\,\partial^\alpha\varphi\Big] + \lambda_{(\mathrm{pde})} \int_\Omega \mathscr{D}(K(x,\cdot))\,\mathscr{D}(\varphi) = \varphi(x)$$



Kernel function $K(0.4,\cdot)$ estimated by the FEM



Kernel method $\hat{f}_n$ combined with the FEM

Fig.: Example with $\mathscr{D}(f) = \frac{d}{dx}f - f$

In practice

- Kernel estimator $\hat{f}_n(x) = (K(x, X_1), \ldots, K(x, X_n))(\mathbb{K} + nI_n)^{-1}\mathbb{Y}$
- Problem: what to do when K is not explicit?
- Solution 2: exploit Fourier series

1. Periodize

$$\bar{R}_n(f) = \frac{1}{n}\sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_{(\mathrm{sob})}\|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)} + \lambda_{(\mathrm{pde})}\|\mathscr{D}(f)\|^2_{L^2(\Omega)}$$

2. Restrict the minimization to

$$H_m = \mathrm{Span}((\varphi_k)_{\|k\|_\infty \leqslant m}), \quad \text{with} \quad \varphi_k(x) = (4L)^{-d/2} e^{\frac{i\pi}{2L}\langle k, x\rangle}$$

3. PIKL estimator:

$$\hat{f}^{\mathrm{PIKL}} = \operatorname*{argmin}_{f \in H_m} \bar{R}_n(f)$$

Properties of the PIKL estimator

- Assumption: linear operator $\mathscr{D}(f) = \sum_{|\alpha| \leqslant s} a_\alpha \partial^\alpha f$

☞ Both the Sobolev norm $\|f\|_{H^s_{\mathrm{per}}([-2L,2L]^d)}$ and the PDE penalty $\|\mathscr{D}(f)\|_{L^2(\Omega)}$ are bilinear functions of the Fourier coefficients z of f

$$\|f\|^2_{\mathrm{RKHS}} = \langle z, M_m z\rangle_{\mathbb{C}^{(2m+1)^d}} \text{ on } H_m$$

- $M_m \in \mathbb{C}^{(2m+1)^d \times (2m+1)^d}$

$$(M_m)_{j,k} = \lambda_{(\mathrm{sob})} \underbrace{\left(1 + \left(\frac{\|k\|_2^2}{(2L)^d}\right)^s\right)\delta_{j,k}}_{\text{Sobolev norm}} + \lambda_{(\mathrm{pde})} \underbrace{\frac{P(j)\bar{P}(k)}{(4L)^d} \int_\Omega e^{\frac{i\pi}{2L}\langle k-j, x\rangle} dx}_{\text{PDE norm}},$$

where $P(k) = \sum_{|\alpha| \leqslant s} a_\alpha (\frac{-i\pi}{2L})^{|\alpha|} \prod_{\ell=1}^d (k_\ell)^{\alpha_\ell}$

- Computation of the integrals possible by numerical integration but also by closed-form formulas:
  - ✓ When $\Omega = [-L, L]^d$, integral $\propto \prod_{j=1}^d \frac{\sin(\pi k_j/2)}{\pi k_j}$
  - ✓ When $\Omega = B_2^2$, integral $\propto \frac{\text{Bessel function}_1(\pi\|k\|_2/2)}{4\|k\|_2}$

Computing the PIKL estimator

For $f \in H_m$:

- $\Phi_m(x) = (x \mapsto (4L)^{-d/2} e^{\frac{i\pi}{2L}\langle k,x\rangle})_{\|k\|_\infty \leqslant m}$
- $K_m(x,y) = \langle M_m^{-1/2}\Phi_m(x), M_m^{-1/2}\Phi_m(y)\rangle_{\mathbb{C}^{(2m+1)^d}}$ (kernel)
- $\mathbb{K}_m \in \mathbb{C}^{n\times n}$ with $(\mathbb{K}_m)_{i,j} = K_m(X_i, X_j)$ (kernel matrix)
- $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top$

PIKL estimator

$$\begin{aligned} \hat{f}^{\mathrm{PIKL}}(x) &= (K_m(x,X_1), \ldots, K_m(x,X_n))(\mathbb{K}_m + nI_n)^{-1}\mathbb{Y} \\ &= \Phi_m(x)^\star(\Phi\!\!\!\Phi^\star \Phi\!\!\!\Phi + nM_m)^{-1}\Phi\!\!\!\Phi^\star\mathbb{Y} \end{aligned}$$

with $\Phi\!\!\!\Phi = \begin{pmatrix} \Phi_m(X_1)^\star \\ \vdots \\ \Phi_m(X_n)^\star \end{pmatrix} \in \mathbb{C}^{n\times(2m+1)^d}$

Computing the PIKL estimator

PIKL estimator

$$\hat{f}^{\mathrm{PIKL}}(x) = (K_m(x, X_1), \dots, K_m(x, X_n)) \underbrace{(\mathbb{K}_m + nI_n)^{-1}}_{n \times n} \mathbb{Y}$$

$$= \Phi_m(x)^\star \underbrace{(\mathbb{\Phi}^\star \mathbb{\Phi} + nM_m)^{-1}}_{(2m+1)^d \times (2m+1)^d} \mathbb{\Phi}^\star \mathbb{Y}$$

- Complexity/storage: $n \times n$ vs. $(2m+1)^d \times (2m+1)^d$
- Possible computation of $\mathbb{\Phi}^\star \mathbb{\Phi}$ and $\mathbb{\Phi}^\star \mathbb{Y}$ online and in parallel
- Training longer than evaluation
- Interpretability of Fourier modes

XP: Perfect modeling with closed-form PDE solutions

- Harmonic oscillator differential prior
  - $d = 1$, $\Omega = [-\pi, \pi]$
  - $\mathscr{D}(f) = \frac{d^2 f}{dx^2} + \frac{df}{dx} + f$
- PDE solutions $f = a_1 f_1 + a_2 f_2$, where $(a_1, a_2) \in \mathbb{R}^2$, $f_1(x) = \exp(-x/2)\cos(\sqrt{3}x/2)$, and $f_2(x) = \exp(-x/2)\sin(\sqrt{3}x/2)$
- Comparison with OLS via $(\hat{a}_1, \hat{a}_2)$

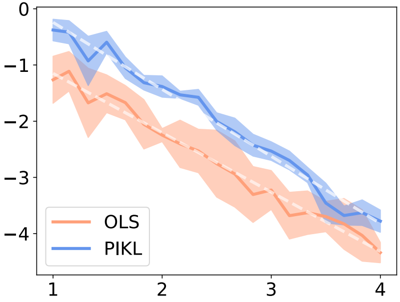


Fig.: L^2-error (mean $\pm$ std over 5 runs) w.r.t. n in $\log_{10} - \log_{10}$ scale.

- Expected parametric rate of n^{-1}

☞ The PIKL estimator ($m = 300$) performs as well as the OLS estimator specifically designed to explore the space of PDE solutions

XP: Imperfect modeling

- Heat equation
  - $d = 2$, $\Omega = [-\pi, \pi]^2$
  - $\mathscr{D}(f) = \partial_1 f - \partial^2_{2,2} f$
  - $f^\star(t, x) = \exp(-t)\cos(x) + 0.5\sin(2x)$
- Imperfect modeling: $\|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)} = \pi > 0$ and $\frac{\|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)}}{\|f^\star\|^2_{L^2(\Omega)}} \simeq 4 \times 10^{-3}$

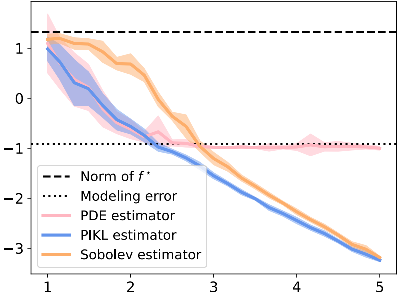


Fig.: L^2-error w.r.t. n in $\log_{10} - \log_{10}$ scale.

- Sobolev minimax rate of $n^{-2/3}$

☞ The PIKL estimator successfully combines the strengths of hybrid modeling

- using the PDE when data is scarce
- relying more on data when it becomes abundant

PDE solving: PINNs vs. PIKL

- Using PIKL as PDE solvers means
  - no noise (i.e., $\varepsilon = 0$)
  - no modeling error (i.e., $\mathscr{D}(f^\star) = 0$)
  - data = uniform samples of $\partial\Omega$ (boundary & init. conditions)
    $n = 100$
- Convection equation: on $\Omega = [0, 1] \times [0, 2\pi]$

$$\mathscr{D}(f) = \partial_t f + \beta \partial_x f \quad \text{with} \quad \begin{cases} \forall x \in [0, 2\pi], & f(0, x) = \sin(x), \\ \forall t \in [0, 1], & f(t, 0) = f(t, 2\pi) = 0 \end{cases}$$

- Solution: $f^\star(t, x) = \sin(x - \beta t) \notin H_m$

$^\diamond$ Krishnapriyan et al. (2021)

	Vanilla PINNs$^\diamond$	Curriculum-trained PINNs$^\diamond$	PIKL estimator
$\beta = 20$	7.50×10^{-1}	9.84×10^{-3}	$(1.56 \pm 3.46) \times 10^{-8}$
$\beta = 30$	8.97×10^{-1}	2.02×10^{-2}	$(0.91 \pm 2.20) \times 10^{-7}$
$\beta = 40$	9.61×10^{-1}	5.33×10^{-2}	$(7.31 \pm 6.44) \times 10^{-9}$

☞ PIKL ($m = 20$) improves the solution accuracy without being sensitive to β

PDE solving: PINNs vs. PIKL

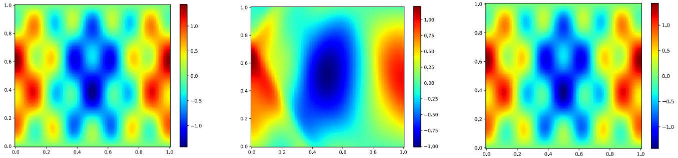
- 1d-wave equation: on $\Omega = [0,1]^2$

$$\mathscr{D}(f) = \partial^2_{t,t} f - 4\partial^2_{x,x} f \text{ with } \begin{cases} \forall x \in [0,1], & f(0,x) = \sin(\pi x) + \sin(4\pi x)/2, \\ \forall x \in [0,1], & \partial_t f(0,x) = 0, \\ \forall t \in [0,1], & f(t,0) = f(t,1) = 0. \end{cases}$$

- Solution: $f^\star(t,x) = \sin(\pi x)\cos(2\pi t) + \sin(4\pi x)\cos(8\pi t)/2$
- Significant variation $\|\partial_t f^\star\|_2^2 / \|f^\star\|_2^2 = 16\pi^2$

$^\diamond$ Wang et al. (2022)

	Vanilla PINNs$^\diamond$	NTK-optimized PINNs$^\diamond$	PIKL estimator
L^2 relative error	4.52×10^{-1}	1.73×10^{-3}	$(8.70{\pm}0.08) \times 10^{-4}$
Training data (n)	2.4×10^{6}	2.4×10^{6}	10^5
# parameters	5.03×10^{5}	5.03×10^{5}	1.68×10^{3}



Ground truth PINNs PIKL

☞ PIKL more accurate, requiring fewer data points and parameters

Opening: PIKL vs. PDE solvers

Performance of traditional PDE solvers for the wave equation on $\Omega = [0,1]^2$

	Euler explicit	RK4	CN	PIKL
L^2 relative error	3.8×10^{-6}	6.8×10^{-6}	5.6×10^{-3}	8.70×10^{-4}
Training data (n)	10^4	10^4	10^4	10^3

☞ Traditional PDE solvers outperform PIKL (even with fewer data)

Performance for the wave equation with noisy boundary conditions

	PINNs	Euler explicit	RK4	CN	PIKL estimator
L^2 relative error	4.61×10^{-1}	1.25×10^{-1}	6.05×10^{-2}	2.01×10^{-2}	1.87×10^{-2}
Training data (n)	2.4×10^6	4×10^4	4×10^4	4×10^4	4×10^4

☞ PIKL outperforms PDE solvers under noisy conditions

Conclusion

- Minimizing the empirical risk regularized by a PDE can be viewed as a kernel method
- Physical information can be beneficial to the statistical performance of the estimators
- PIKL: kernel toolbox for physics-informed learning
- Stay tuned: fast PIKL is coming!

```
In [ ]: pip install pikernel
```

Thank you!

☞ [Doumèche, Bach, Biau, Boyer] PIKL paper on arXiv 2409.13786

☞ [Doumèche, Bach, Biau, Boyer - COLT 2024] Kernel paper on arXiv 2402.07514

☞ [Doumèche, Biau, Boyer - Bernoulli 2024] PINN paper on arXiv 2305.01240