A primer on physics-informed ML

Claire Boyer

université PARIS-SACLAY

institut universitaire de France

Joint work with

Nathan Doumèche

Francis Bach

Gérard Biau

Summary

Summary

Hybrid modeling

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

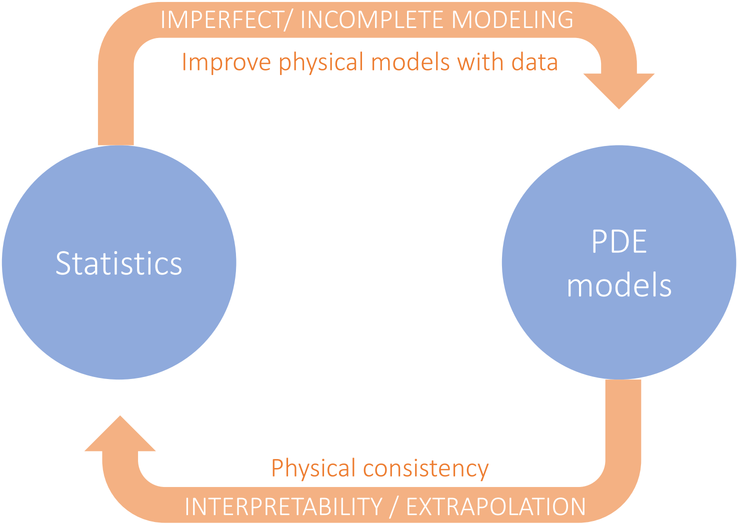
Goal: estimate $f^\star$ using

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

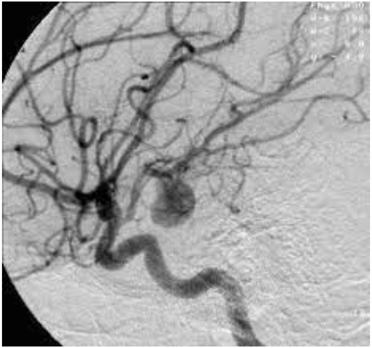
$$\mathscr{D}(f^\star, \cdot) \simeq 0 \quad \text{on} \quad \Omega$$

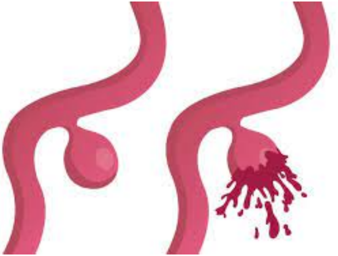
  with a known differential operator $\mathscr{D}$
  with possibly boundary/initial conditions

Why combining learning with physics?



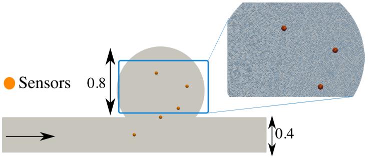
Example: Blood flow in an aneurysm

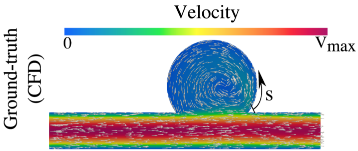




Example: Modeling the blood flow

[Arzani et al., 2021]





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

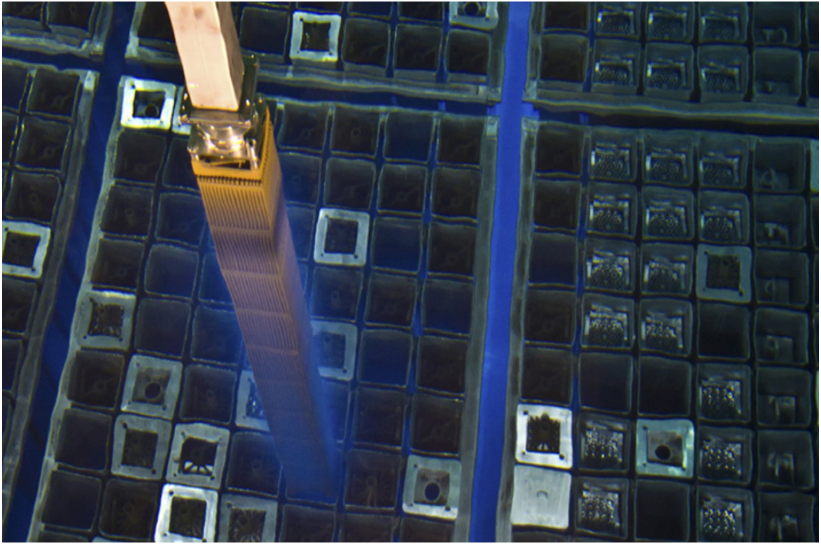
Navier-Stokes equations:

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

(Incomplete) boundary conditions:

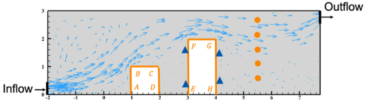
- $(f_x, f_y) = 0$ on the boundaries of the vessel
- Unknown inflow and outflow

Example: Heat transfer in uranium bundles

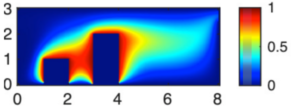


Example: Modeling the heat transfer

[Cai et al., 2021]



▲ Temperature sensors
● Velocity sensors



Goal: estimate the temperature T on the bundles

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

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

(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 $\partial_x T = 0$

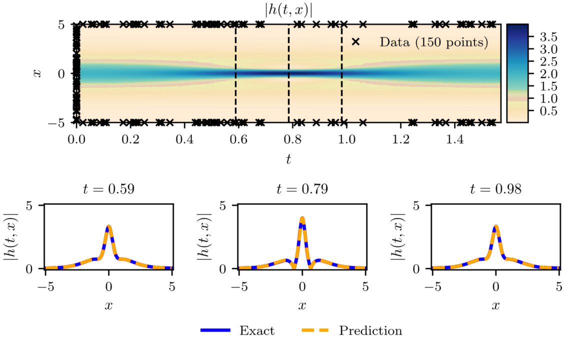
The PDE solver case

Specificity: no data Y_i and exact modeling

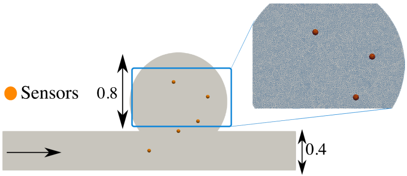
Example: the nonlinear Schrödinger PDE [Raissi et al., 2019]

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

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



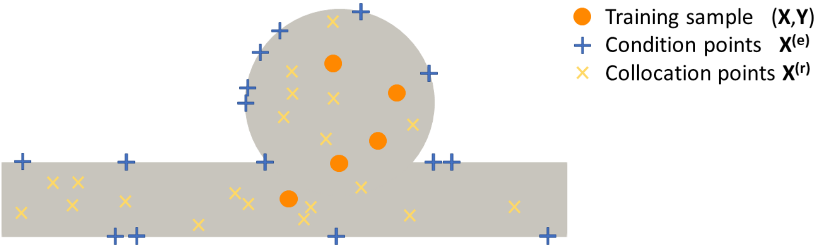
Geometry of the problem



- $\Omega \subseteq [-L, L]^d$: 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
- $\partial\Omega$: the boundary of $\Omega \Rightarrow$ often not C^1 but Lipschitz
- Dirichlet conditions: $f^\star(X) \simeq h(X)$ on $E \subseteq \partial\Omega$

Summary

A general framework: 3 samplings



A general framework

- Training sample $(X_1, Y_1), \dots, (X_n, Y_n) \in \Omega \times \mathbb{R}^{d_2}$ (unknown distribution)
- Boundary/initial sample $X_1^{(e)}, \dots, X_{n_e}^{(e)} \in E \subseteq \partial\Omega$ (chosen distribution)
- Collocation points $X_1^{(r)}, \dots, X_{n_r}^{(r)} \in \Omega$ (uniform distribution)

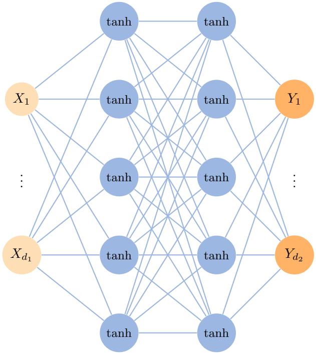
Empirical risk function

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

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

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

Theoretical risk and consistency

Theoretical risk

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

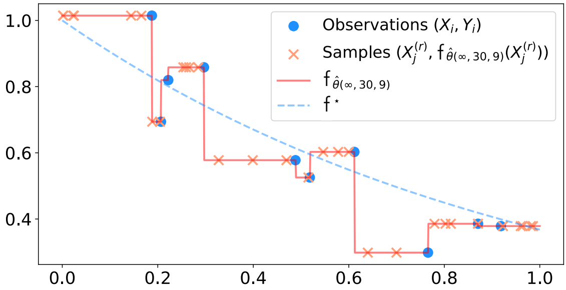
A natural requirement: Risk-consistency

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

- Warning: possible overfitting

Overfitting: hybrid modeling

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



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

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

Risk-consistency of ridge PINNs

Assumptions:

- The condition function h is Lipschitz
- $\mathscr{D}_1, \dots, \mathscr{D}_M$ 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,

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

and

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

Risk-consistency and strong convergence

✓ Ridge PINNs are risk-consistent

Question

Is this sufficient to have $\lim_{D, n_e, n_r, p \to \infty} f_{\hat{\theta}^{(\mathrm{ridge})}(p, n_e, n_r, D)} = f^\star$ 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 $f^\star$).

Risk-consistency and strong convergence

✓ Ridge PINNs are risk-consistent

Question

Is this sufficient to have $\lim_{D,n_e,n_r,p\to\infty} f_{\hat{\theta}^{(\mathrm{ridge})}(p,n_e,n_r,D)} = f^\star$ 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:

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

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

- Theoretical risk:

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

- 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 f is defined by

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

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

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

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* $\lambda_e = 1$, $\lambda_t = (\log n)^{-1}$, *and* $\lambda_d = n^{1/2}/(\log n)$. *Then*

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

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

- Conclusion: statistical accuracy + physical consistency

PINNs wrap-up

- 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
- Open question: Implicit regularizing effect of SGD to achieve risk-consistency?

References

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

☞ Code available here

Summary

Step back on physics-informed learning

- Naive discretization of the PDE can lead to PINN overfitting
- Regularity of the estimate in hybrid modeling is central
- Sobolev and ridge regularizations for PINNs enforce the regularity of the estimator

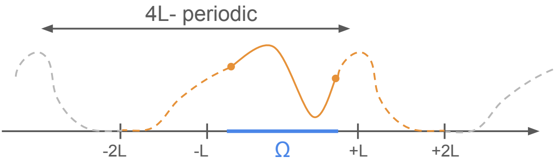
Step back on physics-informed learning

- Naive discretization of the PDE can lead to PINN overfitting
- Regularity of the estimate in hybrid modeling is central
- Sobolev and ridge regularizations for PINNs enforce the regularity of the estimator

- Why don't we use other types of estimator which are inherently regular?
- As for instance kernel methods?

Another empirical risk minimization

- 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 $H^s([-2L, 2L]^d)$ consisting 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

$$R_n(f) = \underbrace{\underbrace{\frac{1}{n}\sum_{i=1}^{n} |f(X_i) - Y_i|^2}_{\text{data-fidelity term}} + \underbrace{\lambda_n \|f\|^2_{H^s_{\mathrm{per}}([-2L,2L]^d)}}_{\text{target regularity}}}_{\text{classical loss for Sobolev kernel regression}} + \underbrace{\mu_n \|\mathscr{D}(f)\|^2_{L^2(\Omega)}}_{\text{PDE}}$$

Objective

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

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

Objective

Understand PIML as a kernel method

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

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

- How does the PDE penalty help for 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

Standard ridge regression

- Inputs/outputs: $X_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$
- Linear model with parameter $\beta^\star \in \mathbb{R}^d$: $Y = \langle X, \beta^\star \rangle + \text{noise}$
- Least-square estimator with a ridge regularization:

$$\hat{\beta} = \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle X_i, \beta \rangle)^2 + \lambda_n \|\beta\|_2^2$$
$$= \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \frac{1}{n} \|\mathbb{Y} - \mathbb{X}\beta\|^2 + \lambda_n \|\beta\|_2^2$$

with $\mathbb{Y} = (Y_1, \dots, Y_n)^\top \in \mathbb{R}^n$ and $\mathbb{X} = \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \in \mathbb{R}^{n \times d}$

- Explicit formula: $\hat{\beta} = (\underbrace{\mathbb{X}^\top \mathbb{X} + n\lambda_n I_d}_{\text{always invertible}})^{-1} \mathbb{X}^\top \mathbb{Y}$

Kernel ridge regression

- Inputs/outputs: $X_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$
- Feature map: $\varphi : \mathbb{R}^d \to \mathbb{R}^D$
  - Polynomial features: $d = 1$, $\varphi(x) = (1, x, x^2) \in \mathbb{R}^D$ with $D = 3$
  - RBF features: $d = 1$, $\varphi(x) = e^{-x^2/2}(1, x, \frac{1}{\sqrt{2!}}x^2, \ldots) \in \ell^2(\mathbb{R})$
- Model of parameter $\beta^\star$: $Y = \langle \varphi(X), \beta^\star \rangle + \text{noise}$
- Kernel ridge estimator:

$$\hat{\beta} = \operatorname*{argmin}_{\beta \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle \varphi(X), \beta \rangle)^2 + \lambda_n \|\beta\|^2$$

$$\text{“}=\text{''} \operatorname*{argmin}_{\beta \in \mathbb{R}^D} \frac{1}{n} \|\mathbb{Y} - \Phi\beta\|^2 + \lambda_n \|\beta\|^2 \qquad \text{(in finite dim)}$$

with $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$ and $\Phi = \begin{pmatrix} \varphi(X_1)^\top \\ \vdots \\ \varphi(X_n)^\top \end{pmatrix} \in \mathbb{R}^{n \times D}$

- Explicit formula: $\hat{\beta} = \underbrace{(\Phi^\top \Phi + n\lambda_n I_D)^{-1}}_{\text{always invertible of size } D \times D} \Phi^\top \mathbb{Y}$

Kernel ridge regression

- What if D is large/infinite?
- Trick $$\underbrace{(\Phi^\top \Phi + n\lambda_n I_D)^{-1}}_{\text{of size } D\times D} \Phi^\top = \Phi^\top \underbrace{(\Phi\Phi^\top + n\lambda_n I_n)^{-1}}_{\text{of size } n\times n}$$

  Proof of the trick:

  1. From $(\Phi^\top \Phi + n\lambda_n I_D)\Phi^\top = \Phi^\top(\Phi\Phi^\top + n\lambda_n I_n)$
  2. Multiply on the left by $(\Phi^\top \Phi + n\lambda_n I_D)^{-1}$
  3. Multiply on the right by $(\Phi\Phi^\top + n\lambda_n I_d)^{-1}$

- Tractable formula $\hat{\beta} = \Phi^\top(\underbrace{\Phi\Phi^\top}_{=:\mathbb{K}} + n\lambda_n I_n)^{-1}\mathbb{Y}$
- Kernel matrix $\mathbb{K} = \Phi\Phi^\top \in \mathbb{R}^{n\times n}$ with $\mathbb{K}_{ii'} = \langle \varphi(X_i), \varphi(X_{i'})\rangle_{\mathcal{H}}$

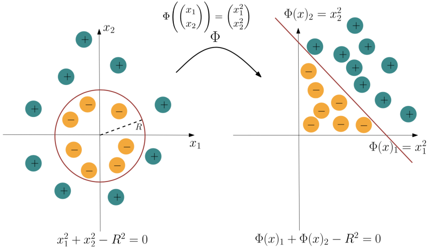
Kernel ridge regression in general

- Kernel: choose a 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 $h : \mathcal{X} \to \mathbb{R}$ such that

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

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

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



The kernel trick

- RKHS functions $h : \mathcal{X} \to \mathbb{R}$

$$\mathrm{RKHS} := \{h : x \mapsto \langle \eta, \varphi(x) \rangle_{\mathcal{T}},\ \eta \in \mathcal{T}\}$$

- Regularized empirical risk minimization

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

- Representer theorem

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

- The solution lives in a finite-dimensional subspace!

Kernel methods

- This amounts to solve a finite-dimensional problem

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

- Final predictor

$$
\hat{h}_n(x) = \langle \hat{h}_n, K(x, \cdot) \rangle_{\mathrm{RKHS}} = \sum_{i=1}^{n} \hat{\alpha}_i K(X_i, x)
$$

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

Effective dimension & convergence

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

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

- Limit theorem

$$\hat{h}_n = \sum_{i=1}^{n} \left[(\mathbb{K} + n\lambda_n I_n)^{-1} \mathbb{Y} \right]_i K(X_i, \cdot) \underset{n\to\infty}{\simeq} (L_K + \mathrm{id})^{-1} L_K h^\star$$

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

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

Underlying RKHS for linear PDEs

Lemma

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

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

Underlying RKHS for linear PDEs

Lemma

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

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

Morally,

$$f(x) \text{ “} = \text{''} \langle f, \delta_x \rangle = \langle \mathscr{O}_n^{-1/2}(f), \mathscr{O}_n^{1/2}(\delta_x)\rangle_{L^2([-2L,2L]^d)}$$

One can recognize here a reproducing property

$$f(x) = \langle f, \mathscr{O}_n(\delta_x)\rangle,$$

where the inner product would be given by

$$\langle g, h \rangle = \langle \mathscr{O}_n^{-1/2}(g), \mathscr{O}_n^{-1/2}(h)\rangle_{L^2([-2L,2L]^d)}.$$

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, such that for $f \in H^s_{\mathrm{per}}([-2L, 2L]^d)$,

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

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

PIML as a kernel method

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

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

✓ $\hat{f}_n$ is a kernel method

✗ Evaluation of the kernel not straightforward

PIML as a kernel method

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

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

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

Proposition (Characterization of the kernel)

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

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

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 $\mathscr{N}(\lambda_n, \mu_n) = \mathrm{tr}(L_K(\mathrm{Id} + L_K)^{-1})$

Theorem (Convergence rate)

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

$$\mathbb{E} \int_\Omega |\hat{f}_n - f^\star|^2 d\mathbb{P}_X \lesssim \log^2(n) \Big(\lambda_n \|f^\star\|^2_{H^s(\Omega)} + \mu_n \|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)} + \frac{M^2}{n^2 \lambda_n} + \frac{\sigma^2 \mathscr{N}(\lambda_n, \mu_n)}{n} \Big).$$

Convergence rate of the PIML kernel method

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

- A simple bound on $\mathscr{N}(\lambda_n, \mu_n)$ show that the PIML estimator converges at least at the Sobolev minimax rate over the class $H^s(\Omega)$, with $\lambda_n = n^{-2s/(2s+d)}\sqrt{\log(n)}$ and $\mu_n = \lambda_n\sqrt{\log(n)}$,

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

- Can we do better?

A toy example

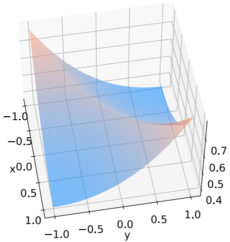
- $d = 1$, $\Omega = [-L, L]$, $\Omega^{\text{aug}} = [-2L, 2L]$
- $f^\star \in H^1(\Omega)$
- $\mathscr{D} = \frac{d}{dx}$ ($f^\star$ is approximately constant)
- $\hat{f}_n = \underset{f \in H^1_{\text{per}}(\Omega^{\text{aug}})}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_n \|f\|^2_{H^1_{\text{per}}(\Omega^{\text{aug}})} + \mu_n \|\mathscr{D}(f)\|^2_{L^2(\Omega)}$

A toy example

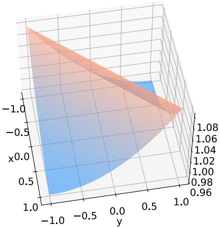
- $d = 1$, $\Omega = [-L, L]$, $\Omega^{\text{aug}} = [-2L, 2L]$
- $f^\star \in H^1(\Omega)$
- $\mathscr{D} = \frac{d}{dx}$ ($f^\star$ is approximately constant)
- $\hat{f}_n = \underset{f \in H^1_{\text{per}}(\Omega^{\text{aug}})}{\operatorname{argmin}} \; \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2 + \lambda_n \|f\|^2_{H^1_{\text{per}}(\Omega^{\text{aug}})} + \mu_n \|\mathscr{D}(f)\|^2_{L^2(\Omega)}$
- Explicit kernel

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

with $\gamma_n = \sqrt{\lambda_n/(\lambda_n + \mu_n)}$



(a) K with $\mu_n = \lambda_n = 1$



(b) Sobolev kernel of $H^1_{\text{per}}([-2, 2])$

Speed-up of the physical penalty

- We can show that for all $m \geqslant 3$,

$$\frac{4L^2}{(\lambda_n + \mu_n)(m+4)^2\pi^2} \leqslant \text{eigenvalue}_m(\mathrm{Proj}_\Omega\, \mathscr{O}_n\, \mathrm{Proj}_\Omega) \leqslant \frac{4L^2}{(\lambda_n + \mu_n)(m-2)^2\pi^2}$$

Speed-up of the physical penalty

- We can show that for all $m \geqslant 3$,

$$\frac{4L^2}{(\lambda_n+\mu_n)(m+4)^2\pi^2} \leqslant \text{eigenvalue}_m(\mathrm{Proj}_\Omega\, \mathscr{O}_n\, \mathrm{Proj}_\Omega) \leqslant \frac{4L^2}{(\lambda_n+\mu_n)(m-2)^2\pi^2}$$

Theorem (Kernel speed-up)

Let $\lambda_n = n^{-1}\log(n)$ *and*

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

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 $\lambda_n = n^{-1}\log(n)$ *and*

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

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

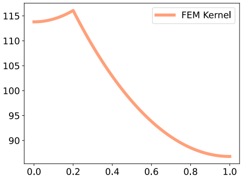
An estimator in practice

- Kernel estimator $\hat{f}_n = \Big(x \mapsto (K(x, X_1), \dots, K(x, X_n))(\mathbb{K} + nI_n)^{-1}\mathbb{Y}\Big)$
- Problem: What to do when K is not explicit?

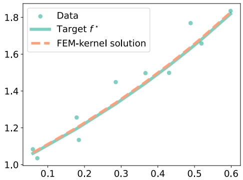
An estimator in practice

- Kernel estimator $\hat{f}_n = \Big(x \mapsto (K(x, X_1), \ldots, K(x, X_n))(\mathbb{K} + nI_n)^{-1}\mathbb{Y}\Big)$
- Problem: What to do when K is not explicit?
- Solution 1: Finite element methods to solve a weak PDE $\forall \varphi \in H^s(\Omega)$,

$$\lambda_n \int_\Omega \Big[K(x,\cdot)\,\varphi + \sum_{|\alpha|=d} \partial^\alpha K(x,\cdot)\,\partial^\alpha \varphi\Big] + \mu_n \int_\Omega \mathscr{D}(K(x,\cdot))\,\mathscr{D}(\varphi) = \varphi(x)$$



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



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

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

An estimator in practice

- Kernel estimator $\hat{f}_n = \left(x \mapsto (K(x, X_1), \ldots, K(x, X_n))(\mathbb{K} + nI_n)^{-1}\mathbb{Y}\right)$
- Problem: What to do when K is not explicit?
- Solution 2: Exploit Fourier tools
  1. Periodize

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

  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

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

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

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

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

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

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

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

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 (evaluation of $\Phi_m(x)$ straightforward)
- Interpretability of Fourier modes

XP: Perfect modeling with closed-form PDE solutions

- Harmonic oscillator differential prior
  - $d = 1$
  - $\mathscr{D}(f) = \frac{d^2 f}{dx^2} + \frac{df}{dx} + f$
- 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 $(\hat{a}_1, \hat{a}_2)$

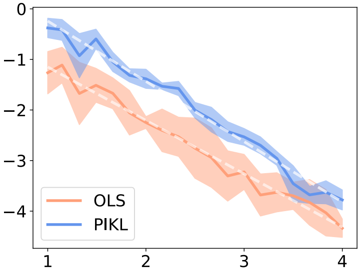


Fig.: L^2-error (mean $\pm$ std over 5 runs) w.r.t. n in $\log_{10} - \log_{10}$ scale.

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

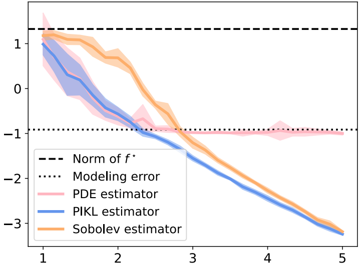


Fig.: L^2-error w.r.t. n in $\log_{10} - \log_{10}$ scale. The PDE error is the L^2-norm between $f^\star$ and the closest PDE solution to $f^\star$.

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

◇ Krishnapriyan et al. (2021)

	Vanilla PINNs◇	Curriculum-trained PINNs◇	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

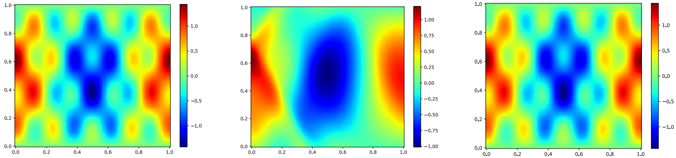
- 1d-wave equation: on $\Omega = [0,1]^2$

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

- 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

✗ Numerical strategies for PIML in general (non-linear PDEs)

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
- ✗ Numerical strategies for PIML in general (non-linear PDEs)

Thank you!

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

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

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