

Physics-informed kernel learning

CRM Spring school, Montreal 2025

Work team

2 / 40



Nathan
Doumèche



Francis
Bach



Gérard
Biau



Claire
Boyer

1. Hybrid modeling
2. Survival kit on kernel learning
3. PIML as a kernel method
4. The PIKL algorithm

1. Hybrid modeling
2. Survival kit on kernel learning
3. PIML as a kernel method
4. The PIKL algorithm

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

Goal: estimate f^* using

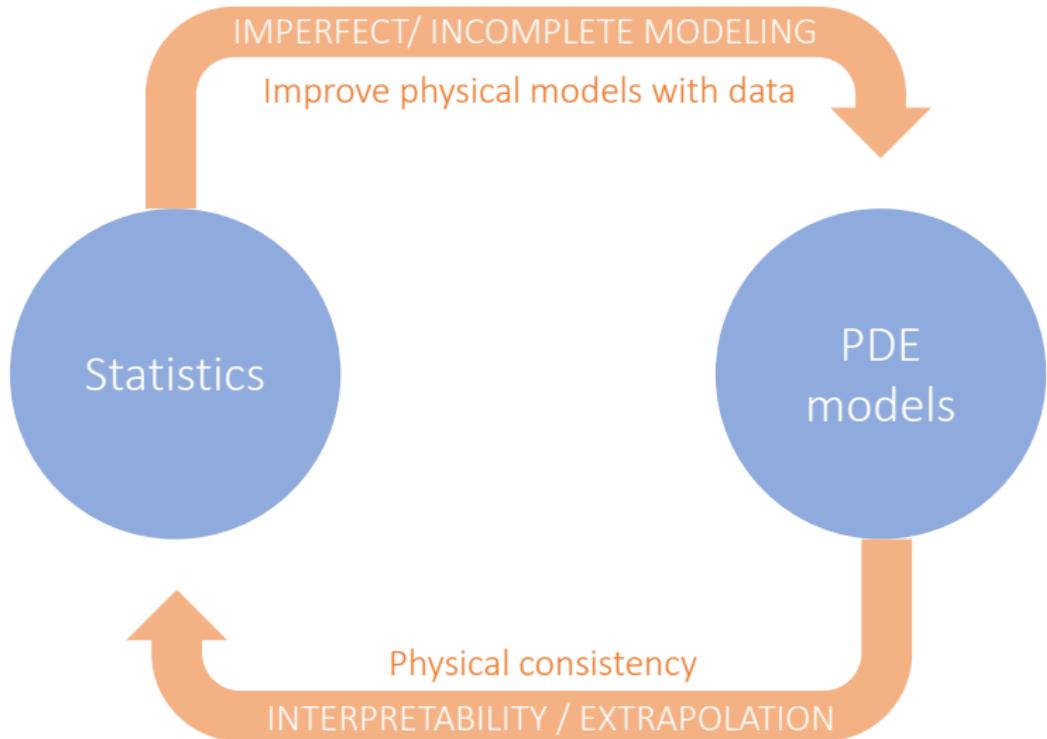
- ▶ Supervised learning: an i.i.d. training sample $(X_i, Y_i)_{1 \leq i \leq n}$
- ▶ Physical modeling: a prior knowledge

$$\mathcal{D}(f^*, \cdot) \simeq 0$$

with a known differential operator \mathcal{D}

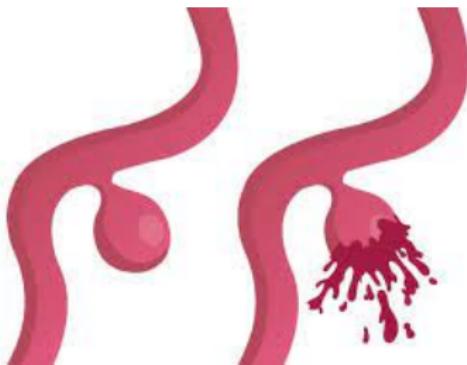
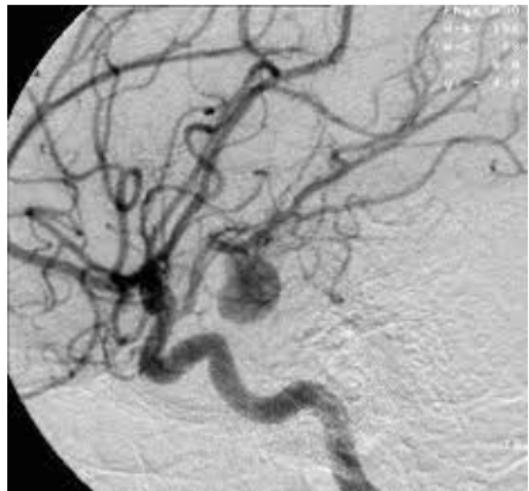
Why combining learning with physics?

6 / 40



Example: Blood flow in an aneurysm

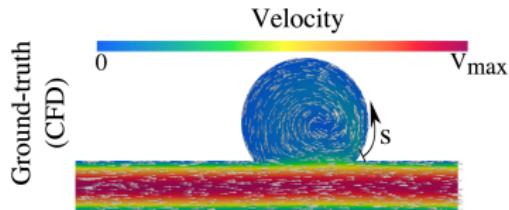
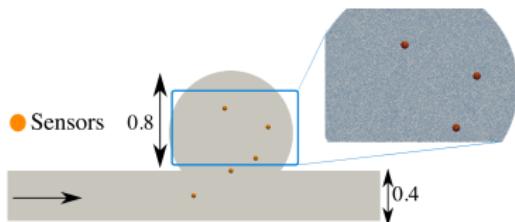
7 / 40



Modeling the blood flow

8 / 40

[Arzani et al., 2021]



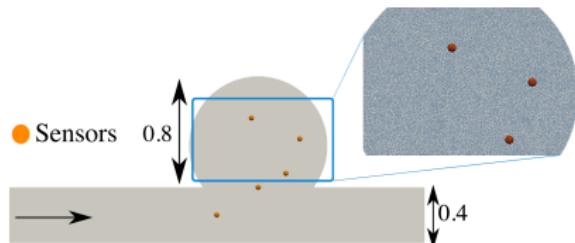
Goal: estimate the blood flow $f = (f_x, f_y, P)$

Navier-Stokes equations:

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

Geometry of the problem

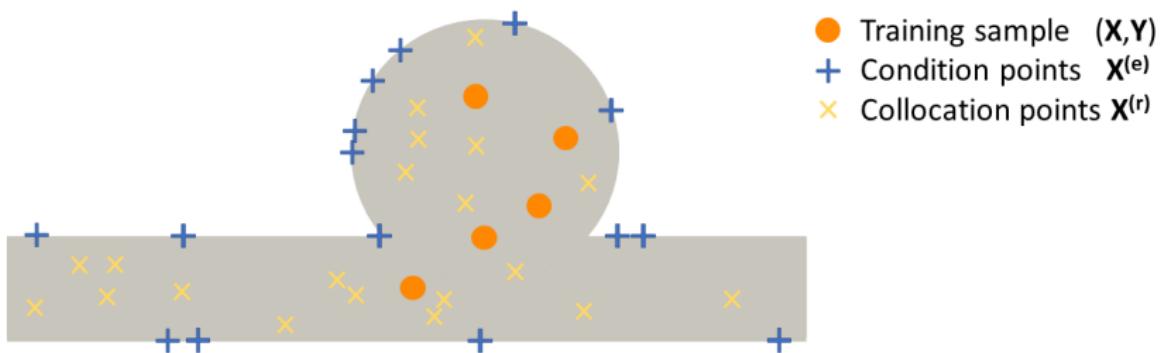
9 / 40



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

Three samplings

10 / 40



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 $\mathbf{X}_1^{(e)}, \dots, \mathbf{X}_{n_e}^{(e)} \in E \subseteq \partial\Omega$ (**chosen** distribution)
- ▶ Collocation points $\mathbf{X}_1^{(r)}, \dots, \mathbf{X}_{n_r}^{(r)} \in \Omega$ (**uniform** distribution)

Empirical risk function

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

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

- ▶ Physics-Informed Neural Networks: NN obtained after training

[Doumèche, Biau, Boyer, 2024]

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

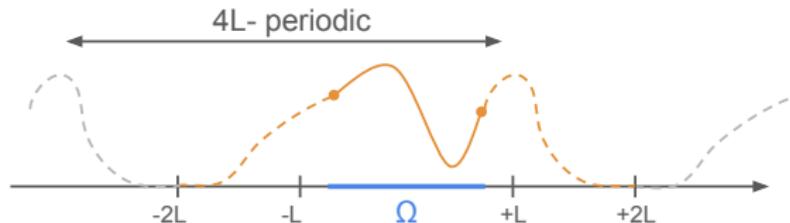
$$\widehat{\mathcal{NN}}_{\theta} \in \operatorname{argmin}_{\mathcal{NN}} \frac{1}{n} \sum_{i=1}^n \|f_{\theta}(X_i) - Y_i\|_2^2 + \frac{\lambda_{(\text{pde})}}{n_r} \sum_{\ell=1}^{n_r} \|\mathcal{D}(f_{\theta}, \mathbf{X}_{\ell}^{(r)})\|_2^2 \\ + \frac{\lambda_{(\text{sob})}}{n_r} \sum_{\ell=1}^{n_r} \left(\sum_{|\alpha| \leq s} \|\partial^{\alpha} f_{\theta}(X_i^{(r)})\|_2^2 \right) + \lambda_{(\text{ridge})} \|\theta\|_2^2$$

- ☞ Ridge regularization to prevent overfitting
- ☞ Sobolev and ridge regularizations for strong convergence
 - + statistical accuracy on the support of the training data
 - + physical consistency anywhere else

Towards a kernel approach

13 / 40

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



Empirical risk

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

Framing PIML as a kernel method

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

with $\|f\|_{\text{RKHS}}^2 = \lambda_{(\text{sob})} \|f\|_{H_{\text{per}}^s([-2L, 2L]^d)}^2 + \lambda_{(\text{pde})} \|\mathcal{D}(f)\|_{L^2(\Omega)}^2$

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

Assumption

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

1. Hybrid modeling
2. Survival kit on kernel learning
3. PIML as a kernel method
4. The PIKL algorithm

Kernel ridge regression

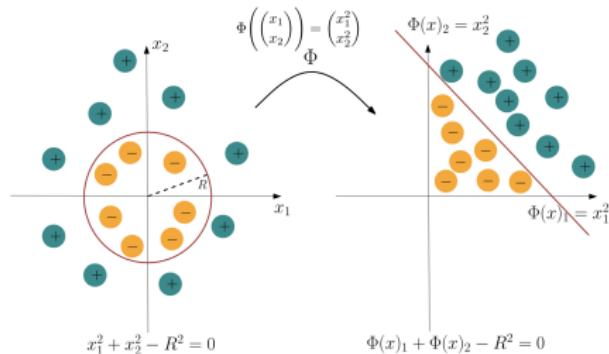
- ▶ Kernel: a symmetric positive definite function $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, i.e., $\sum_{i,i'=1}^n \alpha_i \alpha_{i'} K(x_i, x_{i'}) \geq 0$

- ▶ There exists a Hilbert space RKHS of functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that

$$\begin{aligned} (i) \quad & \forall x \in \mathcal{X}, K(\cdot, x) \in \text{RKHS} \\ (ii) \quad & \forall f \in \mathcal{H}, \langle f, K(\cdot, x) \rangle_{\text{RKHS}} = f(x) \end{aligned}$$

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

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



- ▶ Regularized empirical risk minimization

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

- ▶ 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!

- We solve a finite-dimensional problem

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

- 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!

- ▶ Integral/covariance operator $L_K : L^2(\mathcal{X}, \mathbb{P}_X) \rightarrow L^2(\mathcal{X}, \mathbb{P}_X)$, 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: $\text{tr}(L_K(\lambda \text{Id} + L_K)^{-1})$
- ▶ Convergence rate of the kernel method

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

1. Hybrid modeling
2. Survival kit on kernel learning
3. PIML as a kernel method
4. The PIKL algorithm

Lemma

There exists a positive operator \mathcal{O}_n on $L^2([-2L, 2L]^d)$ such that, for any $f \in H_{\text{per}}^s([-2L, 2L]^d)$,

$$\|\mathcal{O}_n^{-1/2}(f)\|_{L^2([-2L, 2L]^d)}^2 = \lambda_{(\text{sob})} \|f\|_{H_{\text{per}}^s([-2L, 2L]^d)}^2 + \lambda_{(\text{pde})} \|\mathcal{D}(f)\|_{L^2(\Omega)}^2.$$

☞ This suggests the inner product

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

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

$$\mathcal{O}_n(f)(x) = \sum_{m \in \mathbb{N}} \textcolor{brown}{a}_m \langle f, \textcolor{blue}{v}_m \rangle_{L^2([-2L, 2L]^d)} \textcolor{blue}{v}_m(x)$$

- ▶ Orthonormal basis of eigenfunctions $\textcolor{blue}{v}_m \in H_{\text{per}}^s([-2L, 2L]^d)$
- ▶ Eigenvalues $\textcolor{brown}{a}_m > 0$

Theorem

The space $H_{\text{per}}^s([-2L, 2L]^d)$, equipped with the inner product

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

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

$$\|f\|_{\text{RKHS}}^2 = \lambda_{(\text{sob})} \|f\|_{H_{\text{per}}^s([-2L, 2L]^d)}^2 + \lambda_{(\text{pde})} \|\mathcal{D}(f)\|_{L^2(\Omega)}^2,$$

and the associated kernel is $K(x, y) = \sum_{m \in \mathbb{N}} \textcolor{brown}{a}_m \textcolor{blue}{v}_m(x) \textcolor{blue}{v}_m(y)$.

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

- ✓ \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 $\varphi \in H_{\text{per}}^s([-2L, 2L]^d)$: for all $x \in \Omega$,

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

- Integral operator $L_K : L^2(\Omega, \mathbb{P}_X) \rightarrow L^2(\Omega, \mathbb{P}_X)$, 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 $\mathcal{N}(\lambda_{(\text{sob})}, \lambda_{(\text{pde})}) = \text{tr}(L_K(\text{Id} + L_K)^{-1})$

Theorem (Convergence rate)

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

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

- ✗ Not easy to characterize the eigenvalues of L_K for the PIML estimator
- ▶ A simple bound on $\mathcal{N}(\lambda_{(\text{sob})}, \lambda_{(\text{pde})})$ shows that

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

- ▶ Can we do better?

A toy example

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

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

- ▶ Explicit kernel

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

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

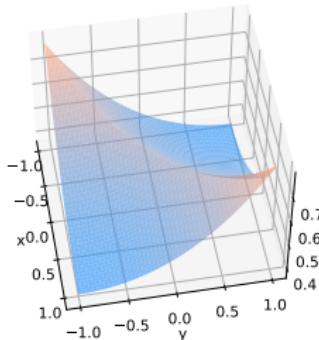


Fig.: K with $\lambda_{(\text{sob})} = \lambda_{(\text{pde})} = 1$

Theorem (Kernel speed-up)

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

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

Then

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

Theorem (Kernel speed-up)

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

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

Then

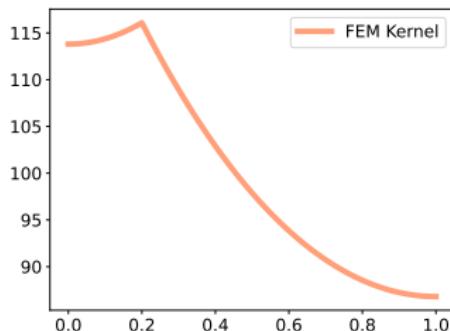
$$\begin{aligned} \mathbb{E} \int_{[-L, L]} |\hat{f}_n - f^*|^2 d\mathbb{P}_X &= \|\mathcal{D}(f^*)\|_{L^2(\Omega)} \mathcal{O}_n(n^{-2/3} \log^3(n)) \\ &\quad + (\|f^*\|_{H^s(\Omega)}^2 + \underbrace{\sigma^2 + M^2}_{\text{noise param}}) \mathcal{O}_n(n^{-1} \log^3(n)). \end{aligned}$$

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

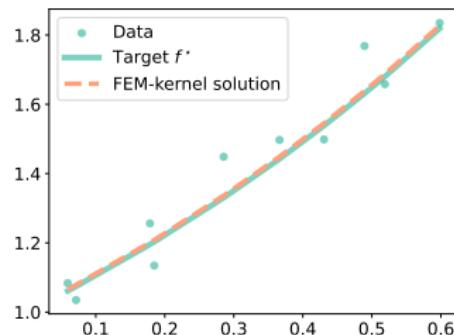
1. Hybrid modeling
2. Survival kit on kernel learning
3. PIML as a kernel method
4. The PIKL algorithm

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

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



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



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

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

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

1. Periodize

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

2. Restrict the minimization to

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

3. PIKL estimator:

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

Properties of the PIKL estimator

32 / 40

- ▶ Assumption: linear operator $\mathcal{D}(f) = \sum_{|\alpha| \leq s} a_\alpha \partial^\alpha f$
- ▶ Both the Sobolev norm $\|f\|_{H_{\text{per}}^s([-2L, 2L]^d)}$ and the PDE penalty $\|\mathcal{D}(f)\|_{L^2(\Omega)}$ are bilinear functions of the Fourier coefficients z of f

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

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

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

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

- ▶ 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 $f \in H_m$:

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

PIKL estimator

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

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

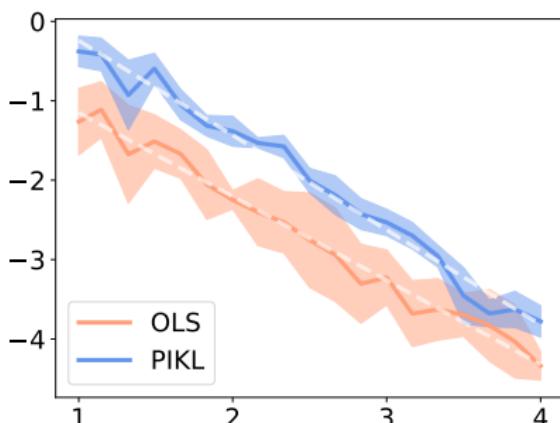
PIKL estimator

$$\begin{aligned}\hat{f}^{\text{PIKL}}(x) &= (K_m(x, X_1), \dots, K_m(x, X_n)) \underbrace{(\mathbb{K}_m + nI_n)^{-1}}_{n \times n} \mathbb{Y} \\ &= \Phi_m(x)^* \underbrace{(\Phi^* \Phi + nM_m)^{-1}}_{(2m+1)^d \times (2m+1)^d} \Phi^* \mathbb{Y}\end{aligned}$$

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

XP: Perfect modeling with closed-form PDE solutions^{35 / 40}

- ▶ Harmonic oscillator differential prior
 - ▶ $d = 1, \Omega = [-\pi, \pi]$
 - ▶ $\mathcal{D}(f) = \frac{d^2 f}{dx^2} + \frac{df}{dx} + f$
- ▶ 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)



- ▶ 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

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

XP: Imperfect modeling

- ▶ Heat equation

- ▶ $d = 2, \Omega = [-\pi, \pi]^2$
- ▶ $\mathcal{D}(f) = \partial_1 f - \partial_{2,2}^2 f$
- ▶ $f^*(t, x) = \exp(-t) \cos(x) + 0.5 \sin(2x)$

- ▶ Imperfect modeling: $\|\mathcal{D}(f^*)\|_{L^2(\Omega)}^2 = \pi > 0$ and $\frac{\|\mathcal{D}(f^*)\|_{L^2(\Omega)}^2}{\|f^*\|_{L^2(\Omega)}^2} \simeq 4 \times 10^{-3}$

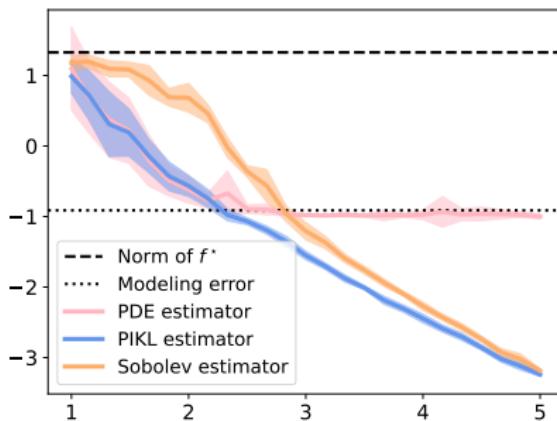


Fig.: L^2 -error w.r.t. n in $\log_{10} - \log_{10}$ scale.

- ▶ 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

- ▶ Using PIKL as PDE solvers means

- ▶ no noise (i.e., $\varepsilon = 0$)
- ▶ no modeling error (i.e., $\mathcal{D}(f^*) = 0$)
- ▶ data = uniform samples of $\partial\Omega$ (boundary & init. conditions)
 $n = 100$

- ▶ Convection equation: on $\Omega = [0, 1] \times [0, 2\pi]$

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

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

◊ 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

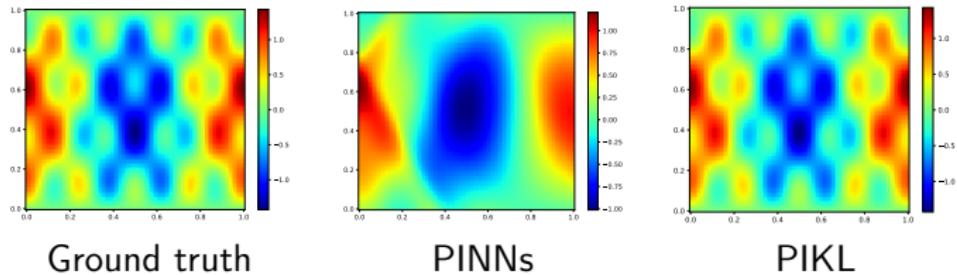
- 1d-wave equation: on $\Omega = [0, 1]^2$

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

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

◊ Wang et al. (2022)

	Vanilla PINNs [◊]	NTK-optimized PINNs [◊]	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



☞ PIKL more accurate, requiring fewer data points and parameters

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**

- ▶ 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

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