Physics-informed machine learning with kernels

Joachim Bona-Pellissier
MaLGa; Università degli studi di Genova

Joint work with Giacomo Meanti, Matteo Santacesaria, Lorenzo Rosasco

October 8, 2025

Outline

The physics-informed learning problem

Physics-Informed Neural Networks (PINNs)

Kernel methods for PIML

The "Physics-Informed" Learning Problem

We want to learn an unknown function $u^* : \Omega \to \mathbb{R}$, but we have two sources of information:

1. Observational Data

We have a (potentially small) set of measurements:

$$(x_i, y_i) \quad \text{where} \quad y_i \approx u^*(x_i)$$

2. Physical Laws

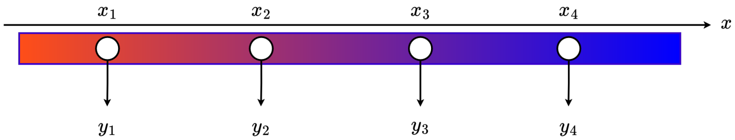
We know u^* must satisfy a governing partial differential equation (PDE):

$$\mathcal{D}u(x) = f(x) \quad \text{for } x \in \Omega$$

How can we leverage both the **data** and the **physics**?

Example: Heat Transfer in a 1D Rod

Temperature distribution $u(x)$ along a rod of length L.



The Data

Measurements y_i of the temperatures at the sensors x_i.

The Physics: steady-state 1D heat Equation

$$\underbrace{\frac{d^2u}{dx^2}}_{\mathcal{D}u}(x) = 0 \quad \text{for } x \in (0, L).$$

Outline

The physics-informed learning problem

Physics-Informed Neural Networks (PINNs)

Kernel methods for PIML

UniGe | MaLGa

Physics-Informed Neural Networks (PINNs)

[Raissi et al '17]

General principle

- Neural network $u_\theta(x)$, parameterized by weights θ.
- Composite loss function:

$$L(\theta) = L_{data}(\theta) + L_{physics}(\theta) + R(\theta).$$

- Train u_θ by minimizing L.

There are no physics-informed neural networks, only physics-informed losses!

Data-driven physics

Remember $\mathcal{D}u^*(x) = f(x)$.

1. Data Loss (L_{data})

$$L_{data}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(u_\theta(x_i) - y_i)^2$$

- Classical data:

$$y_i \approx u^*(x_i)$$

Data-driven physics

Remember $\mathcal{D}u^*(x) = f(x)$.

1. Data Loss (L_{data})

$$L_{data}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(u_\theta(x_i) - y_i)^2$$

- Classical data:

$$y_i \approx u^*(x_i)$$

2. Physics Loss ($L_{physics}$)

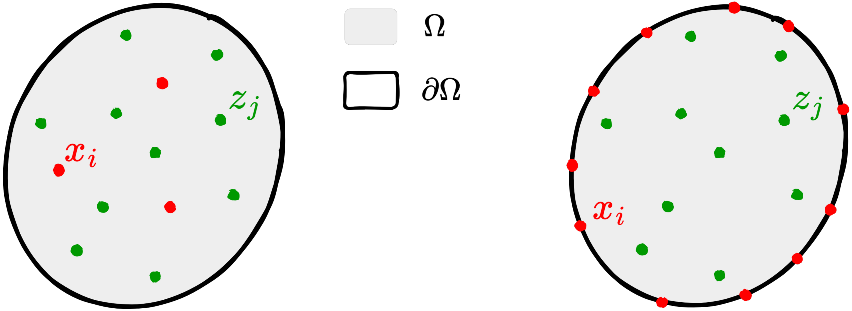
$$L_{physics}(\theta) = \frac{1}{m}\sum_{j=1}^{m}(\mathcal{D}u_\theta(z_j) - d_j)^2$$

- Physics-informed data:

$$d_j = f(z_j) = \mathcal{D}u^*(z_j)$$

- Derivatives in $\mathcal{D}$ are computed via **automatic differentiation**.

Different settings



- $y_i = u^*(x_i)$
- $d_j = f(z_j) = \mathcal{D}u^*(z_j)$

Università di Genova | MaLGa

PINNs: Summary

Strengths

- Approximation capabilities
- Flexible
- Fast inference
- Hard problems: nonlinearity, dimension

PINNs: Summary

Strengths

- Approximation capabilities
- Flexible
- Fast inference
- Hard problems: nonlinearity, dimension

Limitations

- Training difficulties
- Tend to underperform in forward problems (vs classical solvers)
- Spectral bias
- Theoretically hard to analyze

Outline

Kernel Ridge Regression

- $x_1, \dots, x_n \in \Omega$ sampled according to P_X;
- $y_i = u^*(x_i)$.

We want to minimize the true risk

$$L(u) = \mathbb{E}[(u(x) - u^*(x))^2] = \|u - u^*\|^2_{L^2(P_X)}.$$

Instead we consider the empirical risk

$$\widehat{L}(u) = \frac{1}{n}\sum_{i=1}^{n}(u(x_i) - y_i)^2.$$

Kernel Ridge Regression

Hypothesis space: space $\mathcal{H}$ of functions $\Omega \to \mathbb{R}$.
Assumption: $\mathcal{H}$ is a **Reproducing Kernel Hilbert Space** (RKHS).

Reproducing property:

For all $x \in \Omega$, there exists $K_x \in \mathcal{H}$ such that

$$u(x) = \langle u, K_x \rangle_{\mathcal{H}} \qquad \forall u \in \mathcal{H}.$$

Kernel Ridge Regression

Regularized Empirical Risk

$$\widehat{L}_\lambda(u) = \frac{1}{n}\sum_{i=1}^{n}(u(x_i) - y_i)^2 + \lambda\|u\|_{\mathcal{H}}^2.$$

The Representer Theorem

The minimizer $\widehat{u}$ has a simple finite-dimensional form:

$$\widehat{u}(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x)$$

We only need to find the coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

This reduces an infinite-dimensional problem to a finite one!

Kernel Ridge Regression

Substituting the form of $\widehat{u}$ into the objective leads to a simple matrix equation.

The Kernel Matrix

Let $\mathbf{K} \in \mathbb{R}^{n \times n}$ be the Gram matrix of the data points:

$$\mathbf{K}_{ij} = \langle K_{x_i}, K_{x_j} \rangle_{\mathcal{H}} = K(x_i, x_j)$$

Closed-Form Solution

The vector of coefficients $\boldsymbol{\alpha}$ is given by:

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda n \mathbf{I})^{-1} y.$$

Theoretical guarantees for KRR

Consider $u^* \in L^2(P_X)$.
If $K(x, y)$ is **universal**, i.e. if the hypothesis space $\mathcal{H}$ is rich enough
Then we have the asymptotic convergence

$$L(\widehat{u}) = \|\widehat{u} - u^*\|^2_{L^2(P_X)} \xrightarrow[n\to\infty]{} 0 \quad a.s.$$

If further assumptions on K, P_x and $u^* \implies$ rates
e.g. effective dimension, source condition...

Theoretical guarantees for KRR

Consider $u^* \in L^2(P_X)$.
If $K(x,y)$ is **universal**, i.e. if the hypothesis space $\mathcal{H}$ is rich enough
Then we have the asymptotic convergence

$$L(\widehat{u}) = \|\widehat{u} - u^*\|^2_{L^2(P_X)} \xrightarrow[n\to\infty]{} 0 \quad a.s.$$

If further assumptions on K, P_x and $u^* \implies$ rates
e.g. effective dimension, source condition...

Can we adapt this to physics-informed problems?

Physics-informed setting

- $x_1, \ldots, x_n \in \Omega$ sampled according to P_X;
- $y_i = u^*(x_i)$.

Università di Genova | MaLGa

Physics-informed setting

- $x_1, \ldots, x_n \in \Omega$ sampled according to P_X;
- $y_i = u^*(x_i)$.

- $z_1, \ldots, z_m \in \Omega$ sampled according to P_Z;
- $d_j = f(z_j) = \mathcal{D}u^*(z_j)$.

PDE reproducing property

Consider a **linear** differential operator $\mathcal{D}u(x) = \sum_{|\alpha| \leq s} c_\alpha(x) \partial_\alpha u(x)$.
We then have the PDE reproducing property:

For all $x \in \Omega$, there exists $J_x^{\mathcal{D}} \in \mathcal{H}$ such that

$$\mathcal{D}u(x) = \langle u, J_x^{\mathcal{D}} \rangle_{\mathcal{H}} \qquad \forall u \in \mathcal{H}.$$

[Kimeldorf and Wahba ’70, Fasshauer ’96, Wendland ’04]

The physics-informed empirical risk

Pick a kernel K defined over $\overline{\Omega}$, with its RKHS $\mathcal{H}$. Define the physics-informed loss, for $u \in \mathcal{H}$,

$$\widehat{L}_\lambda(u) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(u(x_i) - y_i)^2}_{L_{data}} + \underbrace{\frac{1}{m}\sum_{j=1}^{m}(\mathcal{D}u(z_j) - d_j)^2}_{L_{physics}} + \underbrace{\lambda\|u\|_{\mathcal{H}}^2}_{\text{regularization}} \quad . \tag{1}$$

- Minimizer $\hat{u} \in \mathcal{H}$.
- Can we have convergence guarantees when $n, m \to \infty$?
- What is the impact of the physics-informed term?

PIML with kernels

Representer theorem

The minimizer of the physics-informed loss has the form:

$$\widehat{u}(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) + \sum_{j=1}^{m} \beta_j \mathcal{D}_{z_j} K(z_j, x)$$

This leads to a larger, block-structured kernel matrix $\mathbf{K} \in \mathbb{R}^{(n+m)\times(n+m)}$.

Closed form solution

The vectors of coefficients $\boldsymbol{\alpha} \in \mathbb{R}^n$ and $\boldsymbol{\beta} \in \mathbb{R}^m$ are given by

$$(\boldsymbol{\alpha}, \boldsymbol{\beta})^T = (\mathbf{K} + \lambda \mathbf{I})^{-1} \begin{pmatrix} y \\ d \end{pmatrix}$$

The kernel (block) matrix

$$\mathbf{K} = \begin{pmatrix} \frac{1}{n}B_{00} & \frac{1}{n}B_{01} \\ \frac{1}{m}B_{10} & \frac{1}{m}B_{11} \end{pmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}, \tag{2}$$

where the blocks are

$$\begin{aligned}
&B_{00} \in \mathbb{R}^{n\times n} && (B_{00})_{i,i'} = K(x_i, x_{i'}) && \forall i, i' \in [\![1, n]\!] \\
&B_{01} \in \mathbb{R}^{n\times m} && (B_{01})_{i,j'} = \mathcal{D}_2 K(x_i, z_{j'}) && \forall i \in [\![1,\ n]\!], \forall j' \in [\![1, m]\!] \\
&B_{10} \in \mathbb{R}^{m\times n} && (B_{10})_{j,i'} = \mathcal{D}_1 K(z_j, x_i) && \forall j \in [\![1, m]\!],\ \forall i' \in [\![1, n]\!] \\
&B_{11} \in \mathbb{R}^{m\times m} && (B_{11})_{j,j'} = \mathcal{D}_1\mathcal{D}_2 K(z_j, z_{j'}) && \forall j, j' \in [\![1, m]\!].
\end{aligned}$$

Theoretical guarantees

Assume $u^* \in H^s(\Omega)$. If $K(x, y)$ is C^s-universal (stronger than classical universality), we have

Proposition

There exists a choice of $\lambda_{n,m}$ such that almost surely, when $n, m \to +\infty$, we have

$$\|\widehat{u} - u^*\|^2_{L^2(P_X)} \xrightarrow[n \to \infty]{} 0$$

and

$$\|\mathcal{D}\widehat{u} - \mathcal{D}u^*\|^2_{L^2(P_Z)} \xrightarrow[n \to \infty]{} 0.$$

Theoretical guarantees

Assume $u^* \in H^s(\Omega)$. If $K(x,y)$ is C^s-universal (stronger than classical universality), we have

Proposition

There exists a choice of $\lambda_{n,m}$ such that almost surely, when $n, m \to +\infty$, we have

$$\|\widehat{u} - u^*\|^2_{L^2(P_X)} \quad \xrightarrow[n\to\infty]{} \quad 0$$

and

$$\|\mathcal{D}\widehat{u} - \mathcal{D}u^*\|^2_{L^2(P_Z)} \quad \xrightarrow[n\to\infty]{} \quad 0.$$

Convergence to the target and PDE consistency!

Better convergence: an example

Assume $\mathcal{D}$ is elliptic of order 2 and P_X and P_Z are equivalent to the Lebesgue measure on Ω.

Corollary

There exists a choice of $\lambda_{n,m}$ such that almost surely, when $n, m \to +\infty$, for any open V such that $\overline{V} \subset \Omega$, we have

$$\widehat{u}_{\lambda_{n,m}} \xrightarrow{H^2(V)} u^*.$$

The PDE information allows to get a stronger convergence!

Some experiments

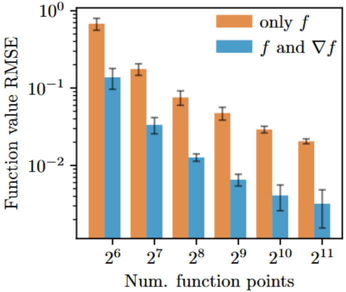


Figure 1: Improved learning accuracy with derivative data. A dataset of function values (in red) is augmented by gradient (in blue) or laplacian (in green) information to improve accuracy.

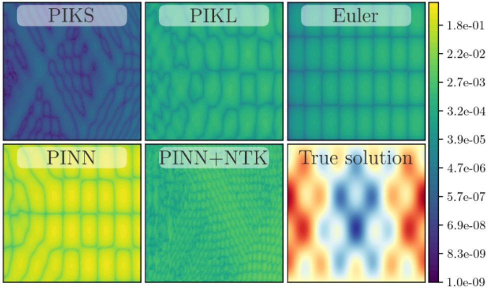


Figure 2: Comparison of errors on 1D wave equation. The color-scale is logarithmic and at this scale of errors (between 10^{-3} and 10^{-7}), only the standard PINN produces a solution which is qualitatively incorrect.

Some experiments

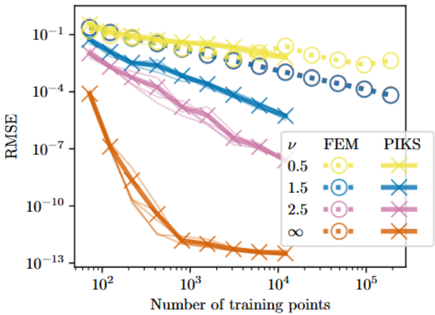


Figure 3: FEM vs. PIKS on noiseless data of different smoothness. All FEM results apart from $\nu = 0.5$ overlap at the top of the plot. The PIKS results with $\nu = \infty$ plateau due to numerical accuracy.

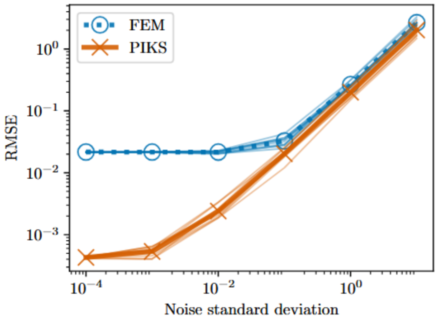


Figure 4: FEM vs. PIKS with increasing noise on the boundary conditions. 500 data-points were used; the true function comes from a Matérn 3/2 kernel.

Conclusion

- For linear PDEs, it is possible to extend KRR to physics-informed problems
- Theoretical guarantees under universality assumption
- Can work in some experimental settings where PINNs struggle

Some questions:

- Are these methods scalable?
- Convergence rates, scaling laws...
- What about nonlinear PDEs?

Conclusion

- For linear PDEs, it is possible to extend KRR to physics-informed problems
- Theoretical guarantees under universality assumption
- Can work in some experimental settings where PINNs struggle

Some questions:

- Are these methods scalable?
- Convergence rates, scaling laws...
- What about nonlinear PDEs?

Thank you for your attention!