Physics-informed Kernel Learning

Nathan Doumèche   NATHAN.DOUMECHE@SORBONNE-UNIVERSITE.FR
Sorbonne University, EDF R&D

Francis Bach   FRANCIS.BACH@INRIA.FR
INRIA, École Normale Supérieure, PSL University

Gérard Biau   GERARD.BIAU@SORBONNE-UNIVERSITE.FR
Sorbonne University, Institut universitaire de France

Claire Boyer   CLAIRE.BOYER@UNIVERSITE-PARIS-SACLAY.FR
Université Paris-Saclay, Institut universitaire de France

Abstract

Physics-informed machine learning typically integrates physical priors into the learning process by minimizing a loss function that includes both a data-driven term and a partial differential equation (PDE) regularization. Building on the formulation of the problem as a kernel regression task, we use Fourier methods to approximate the associated kernel, and propose a tractable estimator that minimizes the physics-informed risk function. We refer to this approach as physics-informed kernel learning (PIKL). This framework provides theoretical guarantees, enabling the quantification of the physical prior's impact on convergence speed. We demonstrate the numerical performance of the PIKL estimator through simulations, both in the context of hybrid modeling and in solving PDEs. In particular, we show that PIKL can outperform physics-informed neural networks in terms of both accuracy and computation time. Additionally, we identify cases where PIKL surpasses traditional PDE solvers, particularly in scenarios with noisy boundary conditions.

Keywords: Physics-informed machine learning, Kernel methods, Physics-informed neural networks, Rates of convergence, Physical regularization

1. Introduction

Physics-informed machine learning. Physics-informed machine learning (PIML), as described by Raissi et al. (2019), is a promising framework that combines statistical and physical principles to leverage the strengths of both fields. PIML can be applied to a variety of problems, such as solving partial differential equations (PDEs) using machine learning techniques, leveraging PDEs to accelerate the learning of unknown functions (hybrid modeling), and learning PDEs directly from data (inverse problems). For an introduction to the field and a literature review, we refer to Karniadakis et al. (2021) and Cuomo et al. (2022).

Hybrid modeling setting. We consider in this paper the classical regression model, which aims at learning the unknown function $f^\star : \mathbb{R}^d \to \mathbb{R}$ such that $Y = f^\star(X) + \varepsilon$, where $Y \in \mathbb{R}$ is the output, $X \in \Omega$ are the features with $\Omega \subseteq [-L, L]^d$ the input domain, and ε is a random noise. Using n observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, independent copies of (X, Y), the goal is to construct an estimator $\hat{f}_n$ of $f^\star$. What makes PIML special compared to other regression settings is the prior knowledge that $f^\star$ approximately follows a PDE. Therefore, we assume that $f^\star$ is

weakly differentiable up to the order $s > \frac{d}{2}$ and that there exists a known differential operator $\mathscr{D}$ such that $\mathscr{D}(f^\star) \simeq 0$. This framework typically accounts for modeling error by recognizing that $\mathscr{D}(f^\star)$ may not be exactly zero, since most PDEs in physics are derived under ideal conditions and may not hold exactly in practice. For example, if $f^\star$ is expected to satisfy the wave equation $\partial_t^2 f(x,t) \simeq \partial_x^2 f(x,t)$, we define the operator $\mathscr{D}(f)(x,t) = \partial_t^2 f(x,t) - \partial_x^2 f(x,t)$ for $(x,t) \in \Omega$.

To estimate $f^\star$, we consider the minimizer of the physics-informed empirical risk

$$R_n(f) = \frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_n \|f\|_{H^s(\Omega)}^2 + \mu_n \|\mathscr{D}(f)\|_{L^2(\Omega)}^2 \tag{1}$$

over the class $\mathscr{F} = H^s(\Omega)$ of candidate functions, where $\lambda_n > 0$ and $\mu_n \geqslant 0$ are hyperparameters that weight the relative importance of each term. Here, $H^s(\Omega)$ denotes the Sobolev space of functions with weak derivatives up to order s. The empirical risk function $R_n(f)$ is characteristic of hybrid modeling, as it is composed of:

- A data fidelity term $\frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2$, which is standard in supervised learning and measures the discrepancy between the predicted values $f(X_i)$ and the observed targets Y_i;

- A regularization term $\lambda_n \|f\|_{H^s(\Omega)}^2$, which penalizes the regularity of the estimator;

- A model error term $\mu_n \|\mathscr{D}(f)\|_{L^2(\Omega)}^2$, which measures the deviation of f from the physical prior encoded in the differential operator $\mathscr{D}$. To put it simply, the lower this term, the more closely the estimator aligns with the underlying physical principles.

Throughout the paper, we refer to $\hat{f}_n$ as the unique minimizer of the empirical risk function, i.e.,

$$\hat{f}_n = \underset{f \in H^s(\Omega)}{\operatorname{argmin}} \ R_n(f). \tag{2}$$

Algorithms to solve the PIML problem. Various algorithms have been proposed to compute the estimator $\hat{f}_n$, and physics-informed neural networks (PINNs) have emerged as a leading approach (e.g., Raissi et al., 2019; Arzani et al., 2021; Karniadakis et al., 2021; Kurz et al., 2022; Agharafeie et al., 2023; Wang et al., 2023). PINNs are usually trained by minimizing a discretized version of the risk over a class of neural networks using gradient descent strategies. Leveraging the good approximation properties of neural networks, as the size of the PINN grows, this type of estimator typically converges to the unique minimizer over the entire space $H^s(\Omega)$ (Shin et al., 2020; Doumèche et al., 2025; Mishra and Molinaro, 2023; Shin et al., 2023; Bonito et al., 2025). However, apart from the fact that optimizing PINNs by gradient descent is an art in itself, the theoretical understanding of the estimators derived through this approach is far from complete (Bonfanti et al., 2024; Rathore et al., 2024), and only a few initial studies have begun to outline their theoretical contours (Krishnapriyan et al., 2021; Wang et al., 2022a; Doumèche et al., 2025). Alternative algorithms for physics-informed learning have since been developed, primarily based on kernel methods, and are seen as promising candidates for bridging the gap between machine learning and PDEs. The connections between PDEs and kernel methods are now well established (e.g., Schaback and Wendland, 2006; Chen et al., 2021; Batlle et al., 2025). Recently, a kernel method has been adapted to perform operator learning (Nelsen and Stuart, 2024). It consists of solving a PDE using samples of the initial condition (with a purely data-driven empirical risk).

Quantifying the impact of physics. Understanding how physics can enhance learning is of critical importance to the PIML community. Arnone et al. (2022) show that for second-order elliptic PDEs in dimension $d = 2$, the PIML estimator converges at a rate of $n^{-4/5}$, outperforming the Sobolev minimax rate of $n^{-2/3}$.

Kernel formulation in the Fourier space. This work builds on the results of Doumèche et al. (2024), which shows that, in the case of hybrid modeling (potentially including a noise $\varepsilon \neq 0$ and a modeling error, i.e., $\mathscr{D}(f^\star) \neq 0$), the PIML problem (2) can be reformulated as a kernel regression task. Provided the associated kernel K is made explicit, this reformulation allows to obtain a closed-form estimator that converges at least at the Sobolev minimax rate. However, the kernel K is highly dependent on the underlying PDE, and its computation can be tedious even for simple priors, such as $\mathscr{D} = \frac{d}{dx}$ in one dimension. Thus, one of the goals of the present paper is to propose an approximation of K, making it possible to implement this kernel method in practice.

For general linear PDEs in dimension d, Doumèche et al. (2024) have adapted to PIML the notion of effective dimension, a central idea in kernel methods that quantify their convergence rate. As a result, for $d = 1$, $s = 1$, $\Omega = [-L, L]$, and $\mathscr{D} = \frac{d}{dx}$, the authors show that the L^2-error of the physics-informed kernel method is of the order of $\log(n)^2/n$ when $\mathscr{D}(f^\star) = 0$, and achieves the Sobolev minimax rate $n^{-2/3}$ otherwise. However, extending this type of results to more complex differential operators $\mathscr{D}$ remains a challenge. In this context, we show how to approximate the effective dimension, making it possible to experimentally estimate the convergence rate of a given PIML problem.

Contributions. Building on the characterization of the PIML problem as a kernel regression task, we use Fourier methods to approximate the associated kernel K and, in turn, propose a tractable estimator minimizing the physics-informed risk function. The approach involves developing the kernel K along the Fourier modes with frequencies bounded by m, and then taking m as large as possible. We refer to this approach as the physics-informed kernel learning (PIKL) method. Subsequently, for general linear operators $\mathscr{D}$, a numerical strategy is developed to estimate the effective dimension of the kernel problem, allowing for the quantification of the expected statistical convergence rate when incorporating the physics prior into the learning process. Finally, we demonstrate the numerical performance of the PIKL estimator through simulations, both in the context of hybrid modeling and in solving partial differential equations. In short, the PIKL algorithm consistently outperforms specialized PINNs from the literature, which were specifically designed for the applications under consideration.

2. The PIKL Estimator

In this section, we detail the construction of the PIKL estimator, our approximate kernel method for physics-informed learning. Throughout this paper, we assume that the differential operator $\mathscr{D}$ is linear with constant coefficients, as stated in the following assumption.

Assumption 2.1 (Linear differential operator with constant coefficients) *The differential operator $\mathscr{D} : H^s(\Omega) \to L^2(\Omega)$ is linear with constant coefficients, i.e., $\mathscr{D}(f) = \sum_{|\alpha| \leqslant s} a_\alpha \partial^\alpha f$ for some $s \in \mathbb{N}^\star$ and $a_\alpha \in \mathbb{R}$.*

We begin by observing that solving the PIML problem (2) is equivalent to performing a kernel regression task, as shown by Doumèche et al. (2024, Theorem 3.3). Thus, leveraging the extensive
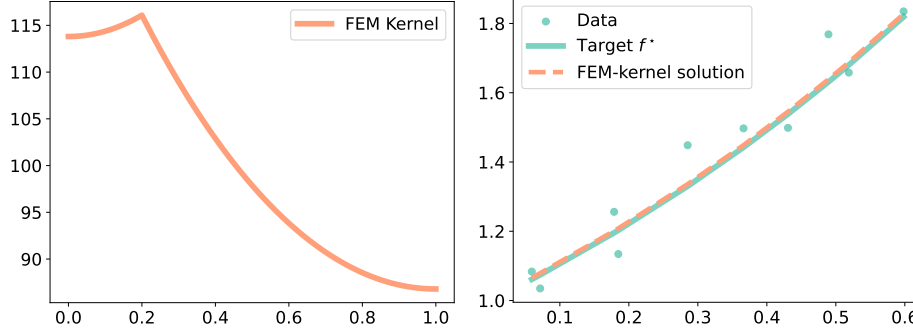
Figure 1: **Left:** Kernel function $K(0.4, \cdot)$ estimated by the FEM. **Right:** Kernel method $\hat{f}_n$ combined with the FEM.

literature on kernel methods, it follows that the estimator $\hat{f}_n$ has the closed-form expression

$$\hat{f}_n = \Big(x \mapsto (K(x, X_1), \dots, K(x, X_n))(\mathbb{K} + nI_n)^{-1}\mathbb{Y}\Big),$$

where $K : \Omega^2 \to \mathbb{R}$ is the kernel associated with the squared norm $\lambda_n\|f\|^2_{H^s(\Omega)} + \mu_n\|\mathscr{D}(f)\|^2_{L^2(\Omega)}$ of (1), and $\mathbb{K}$ is the $n \times n$ kernel matrix defined by $\mathbb{K}_{i,j} = K(X_i, X_j)$.

A finite-element-method approach. The analysis of Doumèche et al. (2024, Proposition 3.4) reveals that the kernel related to the PIML problem is uniquely characterized as the solution to a weak PDE. Indeed, for all $x \in \Omega$, the function $y \mapsto K(x, y)$ is the unique solution in $H^s(\Omega)$ to the weak formulation

$$\forall \phi \in H^s(\Omega), \quad \lambda_n \int_\Omega \Big[K(x, \cdot)\, \phi + \sum_{|\alpha|=s} \partial^\alpha K(x, \cdot)\, \partial^\alpha \phi\Big] + \mu_n \int_\Omega \mathscr{D}(K(x, \cdot))\, \mathscr{D}(\phi) = \phi(x). \quad (3)$$

(This is a consequence of Doumèche et al., 2024, Proposition 3.4, applied to the risk R_n.) A spontaneous idea is to approximate the kernel K using finite element methods (FEM). For illustrative purposes, we have applied this approach in numerical experiments with $d = 1$, $\Omega = [0, 1]$, and $\mathscr{D}(f) = \frac{d}{dx}f - f$. Figure 1 (Left) depicts the associated kernel function $K(0.4, \cdot)$ with $\lambda_n = 10^{-2}$, $\mu_n = 0$, and 100 nodes. Figure 1 (Right) shows that the PIML method (2) successfully reconstructs $f^\star(x) = \exp(x)$ using $n = 10$ data points, $\varepsilon \sim \mathcal{N}(0, 10^{-2})$, $\lambda_n = 10^{-10}$, and $\mu_n = 1000$. However, solving the weak formulation (3) in full generality is quite challenging, particularly when dealing with arbitrary domains Ω in dimension $d > 1$. In fact, FEM strategies need to be specifically tailored to the PDE and the domain in question. Additionally, standard kernel methods combined with FEM approaches come at a high computational cost, since storing the matrix $\mathbb{K}$ requires $O(n^2)$ memory. This becomes prohibitive for large amounts of data, as $n = 10^4$ already requires several gigabytes of RAM.

Fourier approximation. Our primary objective in this article is to develop a more agile, flexible, and efficient method capable of handling arbitrary domains Ω. Following Doumèche et al. (2024), our methodology first requires extending the learning problem from $\Omega \subseteq [-L, L]^d$ to the torus $[-2L, 2L]^d$. Indeed, any function of $H^s(\Omega)$ can be periodically extended into a function

of $H^s_{\text{per}}([-2L, 2L]^d)$ (see Doumèche et al., 2024, Proposition A.6 and Figure 1). The choice of $[-2L, 2L]^d$ is commented in Appendix A.2. This initial technical step allows us to use approximations with the standard Fourier basis, given for $k \in \mathbb{Z}^d$ and $x \in [-2L, 2L]^d$ by

$$\phi_k(x) = (4L)^{-d/2} e^{\frac{i\pi}{2L} \langle k, x \rangle},$$

particularly adapted to periodic functions on $[-2L, 2L]^d$. Therefore, the minimization of the risk R_n defined in (1) over $H^s(\Omega)$ can then be transferred into the minimization of the PIML risk

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

over the periodic Sobolev space $H^s_{\text{per}}([-2L, 2L]^d)$ (see Appendix B for an introduction to periodic Sobolev spaces). This results in a slightly modified kernel, determined by the RKHS norm

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

It is important to note that the estimators derived from the minimization of either R_n or $\bar{R}_n$ share the same statistical guarantees, as both kernel methods have been shown to converge to $f^\star$ at the same rate (Doumèche et al., 2024, Theorem 4.6).

To implement this kernel method, a natural approach is to expand the kernel using a truncated Fourier series, i.e., $K_m(x, y) = \sum_{\|k\|_\infty \leqslant m} a_k \phi_k(x) \phi_k(y)$, with $(\phi_k)_{\|k\|_\infty \leqslant m}$ the Fourier basis, $(a_k)_{\|k\|_\infty \leqslant m}$ the kernel coefficients in this basis, and m the order of approximation. This idea is at the core of techniques such as random Fourier features (RFF) (e.g., Rahimi and Recht, 2007; Yang et al., 2012). However, unlike RFF, the Fourier features in our problem are not random quantities, as they systematically correspond to the low-frequency modes. This low-frequency approximation is particularly well-suited to the Sobolev penalty, which more strongly regularizes high frequencies (the analogous RFF algorithm would involve sampling random frequencies k according to a density that is proportional to the Sobolev decay). In addition, and more importantly, the use of such approximations bypasses the need to discretize the domain into finite elements and requires only the knowledge of the (partial) Fourier transform of $\mathbf{1}_\Omega$, as will be explained later.

A key milestone in the development of our method is to minimize $\bar{R}_n$ not over the entire space $H^s_{\text{per}}([-2L, 2L]^d)$, but rather on the finite-dimensional subspace $H_m = \text{Span}((\phi_k)_{\|k\|_\infty \leqslant m})$. This leads to the PIKL estimator, defined by

$$\hat{f}^{\text{PIKL}} = \underset{f \in H_m}{\text{argmin}} \ \bar{R}_n(f). \tag{4}$$

This naturally transforms the PIML problem into a finite-dimensional kernel regression task, where the associated kernel K_m corresponds to a Fourier expansion of K, as will be clarified in the following paragraph. Of course, H_m provides better approximates of $H^s_{\text{per}}([-2L, 2L]^d)$ as m increases, since for any function $f \in H^s_{\text{per}}([-2L, 2L]^d)$, $\lim_{m \to \infty} \min_{g \in H_m} \|f - g\|_{H^s_{\text{per}}([-2L,2L]^d)} = 0$. Remarkably, the key advantage of using Fourier approximations in our PIKL algorithm lies in the fact that both the squared Sobolev norm $\|f\|^2_{H^s_{\text{per}}([-2L,2L]^d)}$ and the PDE penalty $\|\mathscr{D}(f)\|^2_{L^2(\Omega)}$ are bilinear functions of the Fourier coefficients of f. As shown below, these bilinear forms can be represented as closed-form matrices, easing the computation of the estimator.

RKHS norm in Fourier space. Suppose that the differential operator $\mathscr{D}$ is linear with constant coefficients, i.e., it can be expressed as $\mathscr{D}(f) = \sum_{|\alpha| \leqslant s} a_\alpha \partial^\alpha f$ for some $s \in \mathbb{N}^\star$ and $a_\alpha \in \mathbb{R}$. If $f \in H_m$, then f can be rewritten in terms of its Fourier coefficients as

$$f(x) = \langle z, \Phi_m(x) \rangle_{\mathbb{C}^{(2m+1)^d}},$$

where $\langle \cdot, \cdot \rangle_{\mathbb{C}^{(2m+1)^d}}$ denotes the canonical inner product on $\mathbb{C}^{(2m+1)^d}$, z is the vector of Fourier coefficients of f, and

$$\Phi_m(x) = \left((4L)^{-d/2} e^{\frac{i\pi}{2L} \langle k,x \rangle} \right)_{\|k\|_\infty \leqslant m}.$$

According to Parseval's theorem, the L^2-norm of the derivatives of $f \in H_{\mathrm{per}}^s([-2L, 2L]^d)$ can be expressed using the Fourier coefficients of f as follows: for $r \leqslant s$ and $1 \leqslant i_1, \ldots, i_r \leqslant d$,

$$\|\partial_{i_1,\ldots,i_r}^r f\|_{L^2([-2L,2L]^d)}^2 = (2L)^{-2k} \sum_{\|j\|_\infty \leqslant m} |z_j|^2 \prod_{\ell=1}^r j_{i_\ell}^2.$$

With this notation, the Sobolev norm reads

$$\|f\|_{H_{\mathrm{per}}^s([-2L,2L]^d)}^2 = \sum_{\|j\|_\infty \leqslant m, \|k\|_\infty \leqslant m} z_j \bar{z}_k \left(1 + \left(\frac{\|k\|_2^2}{(2L)^d} \right)^s \right) \delta_{j,k},$$

and, similarly,

$$\|\mathscr{D}(f)\|_{L^2(\Omega)}^2 = \sum_{\|j\|_\infty \leqslant m, \|k\|_\infty \leqslant m} z_j \bar{z}_k \frac{P(j)\bar{P}(k)}{(4L)^d} \int_\Omega e^{\frac{i\pi}{2L} \langle k-j,x \rangle} dx,$$

where $P(k) = \sum_{|\alpha| \leqslant s} a_\alpha (\frac{-i\pi}{2L})^{|\alpha|} \prod_{\ell=1}^d (k_\ell)^{\alpha_\ell}$. Therefore, introducing M_m the $(2m+1)^d \times (2m+1)^d$ matrix with coefficients indexed by $j, k \in \{-m, \ldots, m\}^d$,

$$(M_m)_{j,k} = \lambda_n \left(1 + \left(\frac{\|k\|_2^2}{(2L)^d} \right)^s \right) \delta_{j,k} + \mu_n \frac{P(j)\bar{P}(k)}{(4L)^d} \int_\Omega e^{\frac{i\pi}{2L} \langle k-j,x \rangle} dx, \tag{5}$$

we obtain that the RKHS norm of f is expressed as a bilinear form of its Fourier coefficients z, i.e.,

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

It is important to note that M_m is Hermitian,[1] positive,[2] and definite.[3] Therefore, the spectral theorem (see Theorem B.6) ensures that M_m is invertible, and that its positive inverse square root $M_m^{-1/2}$ is unique and well-defined. We have now all the ingredients to define the PIKL algorithm.

Remark 2.2 (Linear PDEs with non-constant coefficients) *This framework could be adapted to PDEs with non-constant coefficients, i.e., to operators $\mathscr{D}(f) = \sum_{|\alpha| \leqslant s} a_\alpha \partial^\alpha f$ for some $s \in \mathbb{N}^\star$ and $a_\alpha \in C^0(\mathbb{R})$. In this case, the polynomial P in (5) should be replaced by convolutions involving the Fourier coefficients of the functions a_α.*

---

1. since $M_m^\star = \bar{M}_m^\top = M_m$.
2. since $\langle z, M_m z \rangle_{\mathbb{C}^{(2m+1)^d}} = \|f\|_{\mathrm{RKHS}}^2 \geqslant 0$.
3. since $\langle z, M_m z \rangle_{\mathbb{C}^{(2m+1)^d}} = 0$ implies $\|f\|_{H^s([-2L,2L]^d)} = 0$, i.e., $f = 0$.

Computing the PIKL estimator. For a function $f \in H_m$, one can evaluate f at x by $f(x) = \langle M_m^{1/2} z, M_m^{-1/2} \Phi_m(x) \rangle_{\mathbb{C}^{(2m+1)^d}}$. This reproducing property indicates that minimizing the risk $\bar{R}_n$ on H_m is a kernel method governed by the kernel

$$K_m(x, y) = \langle M_m^{-1/2} \Phi_m(x), M_m^{-1/2} \Phi_m(y) \rangle_{\mathbb{C}^{(2m+1)^d}}.$$

Define $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top$ and $\mathbb{K}_m \in \mathcal{M}_n(\mathbb{C})$ to be the matrix such that $(\mathbb{K}_m)_{i,j} = K_m(X_i, X_j)$ for all $1 \leqslant i, j \leqslant n$. The PIKL estimator (4), minimizer of $\bar{R}_n$ restricted to H_m, is therefore given by

$$\hat{f}^{\mathrm{PIKL}}(x) = (K_m(x, X_1), \ldots, K_m(x, X_n))(\mathbb{K}_m + n I_n)^{-1} \mathbb{Y}$$
$$= \Phi_m(x)^\star (\Phi^\star \Phi + n M_m)^{-1} \Phi^\star \mathbb{Y}, \tag{6}$$

where $\Phi = \begin{pmatrix} \Phi_m(X_1)^\star \\ \vdots \\ \Phi_m(X_n)^\star \end{pmatrix}$ is an $n \times (2m+1)^d$ matrix. The formula obtained in (6) is provided by the so-called kernel trick. This step offers a significant advantage to the PIKL estimator as it reduces the computational burden in large sample regimes: instead of storing and inverting the $n \times n$ matrix $\mathbb{K}_m + n I_n$, we only need to store and invert the $(2m+1)^d \times (2m+1)^d$ matrix $\Phi^\star \Phi + n M_m$. Moreover, the computation of $\Phi^\star \Phi$ and $\Phi^\star \mathbb{Y}$ can be performed online and in parallel as n grows. Of course, this approach is subject to the curse of dimensionality. However, it is unreasonable to try to learn more parameters than the sample complexity n. Therefore, in practice, $(2m+1)^d \ll n$, which justifies the preference of the $(2m+1)^{2d}$ storage complexity over the n^2 storage complexity of the FEM-based algorithm. In addition, similar to PINNs, the PIKL estimator has the advantage that its training phase takes longer than its evaluation at certain points. In fact, once the $(2m+1)^d$ Fourier modes of $\hat{f}^{\mathrm{PIKL}}$ (given by $(\Phi^\star \Phi + n M_m)^{-1} \Phi^\star \mathbb{Y}$) are computed, the evaluation of $\Phi_m^\star(x)$ is straightforward. This is in sharp contrast to the FEM-based strategy, which requires approximating the kernel vector $(K(x, X_1), \ldots, K(x, X_n))$ at each query point x.

We also emphasize that the PIKL predictor is characterized by low-frequency Fourier coefficients, which, in turn, enhance its interpretability. This methodology differs significantly from PINNs, which are less interpretable and rely on gradient descent for optimization (see, e.g., Wang et al., 2022a).

Remark 2.3 (PIKL vs. spectral methods) *In the PIML context, the RFF approach resembles a well-known class of powerful tools for solving PDEs, known as spectral and pseudo-spectral methods (e.g., Canuto et al., 2007). These methods solve PDEs by selecting a basis of orthogonal functions and computing the coefficients of the solution on that basis to satisfy both the boundary conditions and the PDE itself. For example, the Fourier basis $(x \mapsto \exp(\frac{i\pi}{2L} \langle k, x \rangle))_{k \in \mathbb{Z}^d}$ already used in this paper is particularly well suited for solving linear PDEs on the square domain $[-2L, 2L]^d$ with periodic boundary conditions. Spectral methods such as these have already been used in the PIML community to integrate PDEs with machine learning techniques (e.g., Meuris et al., 2023). However, the basis functions used in spectral and pseudo-spectral methods must be specifically tailored to the domain Ω, the differential operator $\mathscr{D}$, and the boundary conditions. For more information on this topic, please refer to Appendix A.1.*

Computing M_m for specific domains. Computing the matrix M_m requires the evaluation of the integrals $(j, k) \mapsto \int_\Omega e^{\frac{i\pi}{2L} \langle k-j, x \rangle} dx$. In general, these integrals can be approximated using numerical integration schemes or Monte Carlo methods. However, it is possible to provide closed-form expressions for specific domains Ω. To do so, for $d \in \mathbb{N}^\star$, $L > 0$, and $\Omega \subseteq [-L, L]^d$, we define the characteristic function F_Ω of Ω by

$$F_\Omega(k) = \frac{1}{(4L)^d} \int_\Omega e^{\frac{i\pi}{2L} \langle k, x \rangle} dx.$$

Proposition 2.4 (Closed-form characteristic functions) *The characteristic functions associated with the cube and the Euclidean ball can be analytically obtained as follows.*

- *(Cube) Let $\Omega = [-L, L]^d$. Then, for $k \in \mathbb{Z}^d$,*

$$F_\Omega(k) = \prod_{j=1}^d \frac{\sin(\pi k_j / 2)}{\pi k_j}.$$

- *(Euclidean ball) Let $d = 2$ and $\Omega = \{x \in [-L, L], \|x\|_2 \leqslant L\}$. Then, for $k \in \mathbb{Z}^d$,*

$$F_\Omega(k) = \frac{J_1(\pi \|k\|_2 / 2)}{4 \|k\|_2},$$

where J_1 is the Bessel function of the first kind of parameter 1.

This proposition, along with similar analytical results for other domains, can be found in Bracewell (2000, Table 13.4), noting that F_Ω is the Fourier transform of the indicator function $\mathbf{1}_\Omega$ and is also the characteristic function of the uniform distribution on Ω evaluated at $\frac{k}{2L}$. We can extend these computations further since, given the characteristic functions of elementary domains Ω, it is easy to compute the characteristic functions of translation, dilation, disjoint unions, and Cartesian products of such domains (see Proposition C.1 in Appendix C). For instance, it is straightforward to obtain the characteristic function of the three-dimensional cylinder $\Omega = \{x \in [-L, L], \|x\|_2 \leqslant L\} \times [-L, L]$ as

$$F_\Omega(k_1, k_2, k_3) = \frac{J_1(\pi (k_1^2 + k_2^2)^{1/2} / 2)}{4 (k_1^2 + k_2^2)^{1/2}} \times \frac{\sin(\pi k_3 / 2)}{\pi k_3}.$$

3. The PIKL Algorithm in Practice

To enhance the reproducibility of our work, we provide a `Python` package that implements the PIKL estimator, designed to handle any linear PDE prior with constant coefficients in dimensions $d = 1$ and $d = 2$. This package is available at `https://github.com/NathanDoumeche/numerical_PIML_kernel`. Note that this package implements the matrix inversion of the PIKL formula (6) by solving a linear system using the LU decomposition. Of course, any other efficient method to avoid direct matrix inversion could be used instead, such as solving a linear system with the conjugate gradient method.

Through numerical experiments, we demonstrate the performance of our approach in simulations for hybrid modeling (Subsection 3.1), and derive experimental convergence rates that quantify the benefits of incorporating PDE knowledge into a learning regression task (Subsection 3.2).

3.1 Hybrid Modeling

Perfect modeling with closed-form PDE solutions. We start by assessing the performance of the PIKL estimator in a perfect modeling situation (i.e., $\mathscr{D}(f^\star) = 0$), where the solutions of the PDE $\mathscr{D}(f) = 0$ can be decomposed on a basis $(f_k)_{k \in \mathbb{N}}$ of closed-form solution functions. In this ideal case, the spectral method suggests an alternative estimator, which involves learning the coefficients $a_k \in \mathbb{R}$ of $f^\star = \sum_{k \in \mathbb{N}} a_k f_k$ in this basis. For example, consider the one-dimensional case ($d = 1$) with domain $\Omega = [-\pi, \pi]$, and the harmonic oscillator differential prior $\mathscr{D}(f) = \frac{d^2 f}{dx^2} + \frac{df}{dx} + f$. In this case, the solutions of $\mathscr{D}(f) = 0$ are the linear combinations $f = a_1 f_1 + a_2 f_2$, where $(a_1, a_2) \in \mathbb{R}^2$, $f_1(x) = \exp(-x/2) \cos(\sqrt{3}x/2)$, and $f_2(x) = \exp(-x/2) \sin(\sqrt{3}x/2)$. Thus, the spectral method focuses on learning the vector $(a_1, a_2) \in \mathbb{R}^2$, in-

Figure 2: OLS and PIKL estimators for the harmonic oscillator with $d = 1$, sample size $n = 10$.

stead of learning the Fourier coefficients of $f^\star$, which is the approach taken by the PIKL algorithm.

A baseline that exactly leverages the particular structure of this problem, referred to as the ordinary least squares (OLS) estimator, is therefore $\hat{g}_n = \hat{a}_1 f_1 + \hat{a}_2 f_2$, where

$$(\hat{a}_1, \hat{a}_2) = \operatorname*{argmin}_{(a_1,a_2) \in \mathbb{R}^2} \frac{1}{n} \sum_{i=1}^{n} |a_1 f_1(X_i) + a_2 f_2(X_i) - Y_i|^2.$$

To compare the PIKL and OLS estimators, we generate data such that $Y = f^\star(X) + \varepsilon$, where $X \sim \mathcal{U}(\Omega)$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.5$, and the target function is $f^\star = f_1$ (corresponding to $(a_1, a_2) = (1, 0)$). We implement the PIKL algorithm with 601 Fourier modes ($m = 300$) and $s = 2$. Figure 2 shows that even with very few data points ($n = 10$) and high noise levels, both the OLS and PIKL methods effectively reconstruct $f^\star$, both incorporating physical knowledge in their own way. In Figure 3, we display the L^2-error of both estimators for different sample sizes n. The two methods have an experimental convergence rate of $n^{-1.1}$, which is consistent with the expected parametric rate of n^{-1}. This sanity check shows that under perfect modeling conditions, the PIKL estimator with $m = 300$ performs as well as the OLS estimator specifically designed to explore the space of PDE solutions.

Figure 3: L^2-error (mean $\pm$ std over 5 runs) of the OLS and PIKL estimators for the harmonic oscillator with $d = 1$, w.r.t. n in $\log_{10} - \log_{10}$ scale. The dashed lines represent adjusted linear models w.r.t. n, for both L^2-errors.

Combining the best of physics and data in imperfect modeling. In this paragraph, we deal with an imperfect modeling scenario using the heat differential operator $\mathscr{D}(f) = \partial_1 f - \partial^2_{2,2} f$ in dimension $d = 2$ over the domain $\Omega = [-\pi, \pi]^2$. The data are generated according to the model $Y = f^\star(X) + \varepsilon$, where $\|\mathscr{D}(f^\star)\|_{L^2(\Omega)} \neq 0$. We assume, however, that the PDE
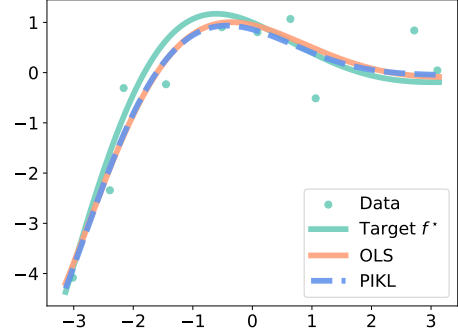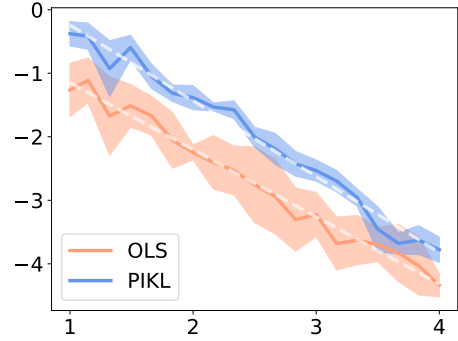
serves as a good physical prior, meaning that $\|f^\star\|^2_{L^2(\Omega)}$ is significantly larger than the modeling error $\|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)}$. The hybrid model is implemented using the PIKL estimator with parameters $s = 2$, $\lambda_n = n^{-2/3}/10$, and $\mu_n = 100/n$. These hyperparameters are selected to ensure that, when only a small amount of data is available, the model relies heavily on the PDE. Yet, as more data become available, the model can use the data to correct the modeling error. The performance of the PIKL estimator is compared with that of a purely data-driven estimator, referred to as the Sobolev estimator, and a strongly PDE-penalized estimator, referred to as the PDE estimator. The Sobolev estimator uses the same parameter $s = 2$ and $\lambda_n = n^{-2/3}/10$, but sets $\mu_n = 0$. This configuration ensures that the estimator relies entirely on the data without considering the PDE as a prior. On the other hand, the PDE estimator is configured with parameters $s = 2$, $\lambda_n = 10^{-10}$, and $\mu_n = 10^{10}$. These hyperparameters are set to ensure that the resulting PDE estimator effectively satisfies the heat equation, making it highly dependent on the physical model.

We perform an experiment where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.5$, and $f^\star(t, x) = \exp(-t)\cos(x) + 0.5\sin(2x)$. This scenario is an example of imperfect modeling, since $\|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)} = \pi > 0$. However, the heat equation serves as a strong physical prior, since $\|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)}/\|f^\star\|^2_{L^2(\Omega)} \simeq 4 \times 10^{-3}$. Figure 4 illustrates the performance of the different estimators.

Clearly, the PDE estimator outperforms the Sobolev estimator when the data set is small ($n \leqslant 10^2$). As expected, the performance of the Sobolev estimator improves as the sample size increases ($n \geqslant 10^3$), but it remains consistently inferior to that of the PIKL. When only a small amount of data is available, the PDE provides significant benefits, and the L^2-error decreases at the super-parametric rate of n^{-2} for both the PIKL and the PDE estimators. However, in the context of imperfect modeling, the PDE estimator cannot overcome the PDE error, resulting in no further improvement beyond $n \geqslant 100$. In addition, when a large amount of data is available, the data become more reliable than the PDE. In this case, the errors for both the PIKL and the Sobolev estimators decrease at the Sobolev minimax rate of $n^{-2/3}$. Overall, the PIKL estimator successfully combines the strengths of both approaches, using the PDE when data is scarce and relying more on data when it becomes abundant.
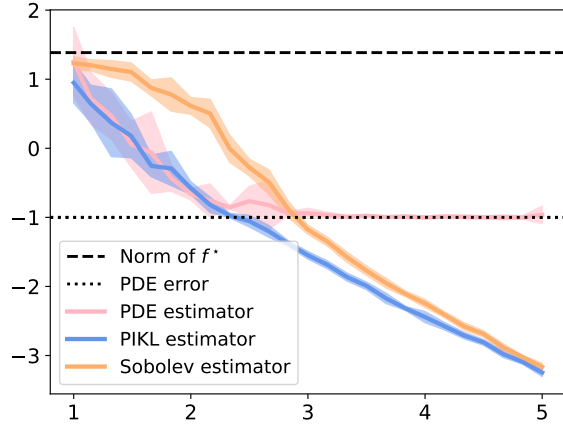

Figure 4: L^2-error (mean $\pm$ std over 5 runs) of the PDE, PIKL, and Sobolev estimators for imperfect modeling with the heat equation, as a function of n in $\log_{10} - \log_{10}$ scale. The PDE error is the L^2-norm between $f^\star$ and the PDE solution that is closest to $f^\star$.

3.2 Measuring the Impact of Physics with the Effective Dimension

The important question of measuring the impact of the differential operator $\mathscr{D}$ on the convergence rate of the PIML estimator has not yet found a clear answer in the literature. In this subsection, we propose an approach to experimentally compare the PIKL convergence rate to the Sobolev minimax rate in $H^s(\Omega)$, which is $n^{-2s/(2s+d)}$ (e.g., Tsybakov, 2009, Theorem 2.1).

Theoretical backbone. According to Doumèche et al. (2024, Theorem 4.3), if X has a bounded density and the noise ε is sub-Gamma with parameters (σ, M), the L^2-error of both estimators (2) and (4) satisfies

$$
\mathbb{E}\int_\Omega |\hat{f}_n - f^\star|^2 d\mathbb{P}_X
$$

$$
\leqslant C_4 \log^2(n)\Big(\lambda_n\|f^\star\|^2_{H^s(\Omega)} + \mu_n\|\mathscr{D}(f^\star)\|^2_{L^2(\Omega)} + \frac{M^2}{n^2\lambda_n} + \frac{\sigma^2 \mathscr{N}(\lambda_n, \mu_n)}{n}\Big), \qquad (7)
$$

where $\mathbb{P}_X$ is the distribution of X. The quantity $\mathscr{N}(\lambda_n, \mu_n)$ on the right-hand side of (7) is referred to as the effective dimension. Given the integral kernel operator L_K, defined by $L_K : f \in L^2(\mathbb{P}_X) \mapsto (x \mapsto \int_\Omega K(x,y)f(y)d\mathbb{P}_X(y)) \in L^2(\mathbb{P}_X)$, the effective dimension is the trace of the operator $(L_K + \mathrm{Id})^{-1}L_K$ (see, e.g., Caponnetto and Vito, 2007). Since λ_n and μ_n can be freely chosen by the practitioner, the effective dimension $\mathscr{N}(\lambda_n, \mu_n)$ becomes a key consideration that help quantify the impact of the physics on the learning problem. Unfortunately, bounding $\mathscr{N}(\lambda_n, \mu_n)$ is not trivial. Doumèche et al. (2024) have shown that, whenever $\frac{d\mathbb{P}_X}{dx} \leqslant \kappa$ with $\kappa \geq 1$,

$$
\mathscr{N}(\lambda_n, \mu_n) \leqslant \sum_{\lambda \in \sigma(C\mathscr{O}_n C)} \frac{1}{1 + (\kappa\lambda)^{-1}} \leqslant \kappa \sum_{\lambda \in \sigma(C\mathscr{O}_n C)} \frac{1}{1 + \lambda^{-1}},
$$

where $\mathscr{O}_n$ is the operator $\mathscr{O}_n = \lim_{m\to\infty} M_m^{-1}$ (where the limit is taken in the sense of the operator norm—see Definition B.3) and C is the operator $C(f) = 1_\Omega f$ (that is, $C(f)(x) = f(x)$ if $x \in \Omega$, and $C(f)(x) = 0$ otherwise). Therefore, a natural idea to assess the effective dimension is to replace $C\mathscr{O}_n C$ by $C_m M_m^{-1} C_m$, where $C_m : H_m \to H_m$ is defined by

$$
\forall j, k \in \{-m, \dots, m\}^d, \quad (C_m)_{j,k} = \frac{1}{(4L)^d}\int_\Omega e^{\frac{i\pi}{2L}\langle k-j, x\rangle} dx.
$$

The following theorem shows that this is a sound strategy, in the sense that computing the effective dimension using the eigenvalues of $C_m M_m^{-1} C_m$ becomes increasingly accurate as m grows.

Theorem 3.1 (Convergence of the effective dimension)

(i) One has

$$
\lim_{m\to\infty} \sum_{\lambda \in \sigma(C_m M_m^{-1} C_m)} \frac{1}{1 + \lambda^{-1}} = \sum_{\lambda \in \sigma(C\mathscr{O}_n C)} \frac{1}{1 + \lambda^{-1}}.
$$

(ii) Let $\sigma_k^\downarrow(C_m M_m^{-1} C_m)$ be the k-th highest eigenvalue of $C_m M_m^{-1} C_m$. The spectrum of the matrix $C_m M_m^{-1} C_m$ converges to the spectrum of $C\mathscr{O}_n C$ in the following sense:

$$
\forall k \in \mathbb{N}^\star, \quad \lim_{m\to\infty} \sigma_k^\downarrow(C_m M_m^{-1} C_m) = \sigma_k^\downarrow(C\mathscr{O}_n C).
$$

The provided `Python` package[4,5] includes numerical approximations of the effective dimension in dimensions $d = 1$ and $d = 2$ for any linear operator $\mathscr{D}$ with constant coefficients, when

---

4. https://github.com/NathanDoumeche/numerical_PIML_kernel
5. Also on `pip` at https://pypi.org/project/pikernel

Ω is either a cube or a Euclidean ball. The code is designed to run on both CPU and GPU. The convergence of the effective dimension as m grows is studied in greater detail in Appendix D.2.

Comparison to the closed-form case. We start by assessing the quality of the approximation encapsulated in Theorem 3.1 in a scenario where the eigenvalues can be theoretically bounded. When $d = 1$, $s = 1$, $\mathscr{D} = \frac{d}{dx}$, and $\Omega = [-\pi, \pi]$, one has (Doumèche et al., 2024, Proposition 5.2)

$$\frac{4}{(\lambda_n + \mu_n)(k+4)^2} \leqslant \sigma_k^\downarrow(C\mathscr{O}_n C) \leqslant \frac{4}{(\lambda_n + \mu_n)(k-2)^2}.$$

This shows that $\log \sigma_k^\downarrow(C\mathscr{O}_n C) \sim_{k \to \infty} -2\log(k)$. Figure 5 (Left) represents the eigenvalues of $C_m M_m^{-1} C_m$ in decreasing order, for increasing values of m, with $\lambda_n = 0.01$ and $\mu_n = 1$. For any fixed m, two distinct regimes can be clearly distinguished: initially, the eigenvalues decrease linearly on a $\log - \log$ scale and align with the theoretical values of $-2\log(k)$. Afterward, the eigenvalues suddenly drop to zero. As m increases, the spectrum progressively approaches the theoretical bound.

In Appendix D.2, we show that $m = 10^2$ Fourier modes are sufficient to accurately approximate the effective dimension when $n \leqslant 10^4$. It is evident from Figure 5 (Right) that the effective dimension exhibits a sub-linear behavior in the $\log - \log$ scale, experimentally confirming the findings of Doumèche et al. (2024), which show that $\mathscr{N}(\frac{\log(n)}{n}, \frac{1}{\log(n)}) = o_{n \to \infty}(n^\gamma)$ for all $\gamma > 0$. So, plugging this into (7) with $\lambda_n = n^{-1}\log(n)$ and $\mu_n = \log(n)^{-1}$ leads to

$$\mathbb{E}\int_{[-L,L]} |\hat{f}_n - f^\star|^2 d\mathbb{P}_X = (\|f^\star\|_{H^1(\Omega)}^2 + \sigma^2 + M^2)O_n(n^{-1}\log^3(n))$$

when $\mathscr{D}(f^\star) = 0$, i.e., when the modeling is perfect. The Sobolev minimax rate on $H^1(\Omega)$ is $n^{-2/3}$, whereas the experimental bound in this context gives a rate of n^{-1}. This indicates that when the target $f^\star$ satisfies the underlying PDE, the gain in terms of speed from incorporating the physics into the learning problem is $n^{-1/3}$.

Harmonic oscillator equation. Here, we follow up on the example of Subsection 3.1, as presented in Figures 2 and 3. Thus, we set $d = 1$, $s = 2$, $\mathscr{D}(u) = \frac{d^2}{dx^2}u + \frac{d}{dx}u + u$, and $\Omega = [-\pi, \pi]$. Recall that in this perfect modeling experiment, we observed a parametric convergence rate of n^{-1}, which is not surprising since the regression problem essentially involves learning the two parameters a_1 and a_2. Figure 6 (Left) shows the eigenvalues of $C_m M_M^{-1} C_m$, while Figure 6 (Right) shows the effective dimension as a function of n. Similarly to the previous closed-form case, we observe that $\mathscr{N}(\frac{\log(n)}{n}, \frac{1}{\log(n)}) = o_{n \to \infty}(n^\gamma)$ for all $\gamma > 0$. The same argument as in the paragraph above shows that this results in a parametric convergence rate, provided $\mathscr{D}(f^\star) = 0$.

Heat equation on the disk. Let us now consider the one-dimensional heat equation $\mathscr{D} = \frac{\partial}{\partial x} - \frac{\partial^2}{\partial y^2}$, with $d = 2$, $s = 2$, and the disk $\Omega = \{x \in \mathbb{R}^2, \|x\|_2 \leq \pi\}$. Since the heat equation is known to have C^∞ solutions with bounded energy (see, e.g., Evans, 2010, Chapter 2.3, Theorem 8), we expect the convergence rate to match that of $H^\infty(\Omega)$, which corresponds to the parametric rate of n^{-1}. Once again, we observe $\mathscr{N}(\frac{\log(n)}{n}, \frac{1}{\log(n)}) = o_{n \to \infty}(n^\gamma)$ for all $\gamma > 0$, and thus an improvement over the $n^{-2/3}$-Sobolev minimax rate on $H^2(\Omega)$ when $\mathscr{D}(f^\star) = 0$.

Quantifying the impact of physics. The three examples above show how incorporating physics can enhance the learning process by reducing the effective dimension, leading to a faster convergence
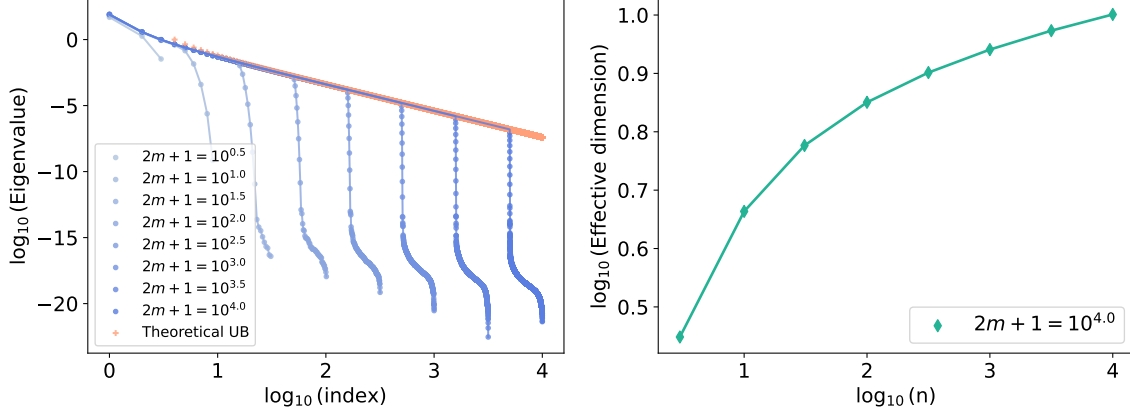
Figure 5: The case of $\mathscr{D} = \frac{d}{dx}$. **Left:** Spectrum of $C_m M_m^{-1} C_m$. **Right:** Estimation of the effective dimension $n \mapsto \mathscr{N}\left(\frac{\log(n)}{n}, \frac{1}{\log(n)}\right)$.
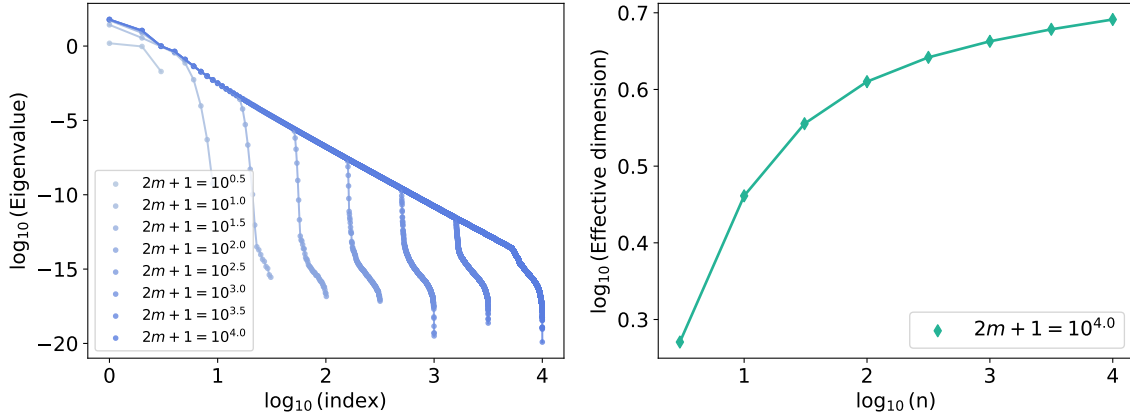

Figure 6: Harmonic oscillator. **Left:** Spectrum of $C_m M_m^{-1} C_m$. **Right:** Estimation of the effective dimension $n \mapsto \mathscr{N}\left(\frac{\log(n)}{n}, \frac{1}{\log(n)}\right)$.
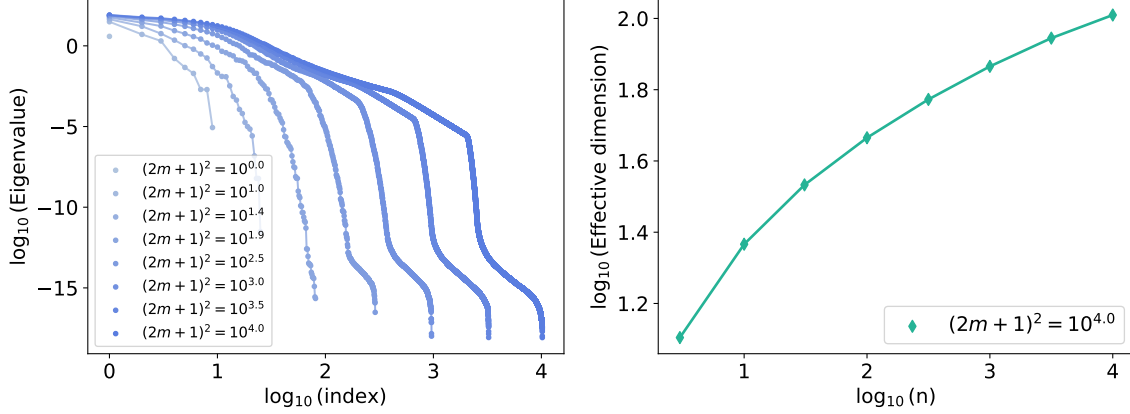
Figure 7: Heat equation. **Left:** Spectrum of $C_m M_m^{-1} C_m$. **Right:** Estimation of the effective dimension $n \mapsto \mathcal{N}\big(\frac{\log(n)}{n}, \frac{1}{\log(n)}\big)$.

rate. In all cases, the rate becomes parametric due to the PDE, achieving the fastest possible speed, as predicted by the central limit theorem. Our package can be directly applied to any linear PDE with constant coefficients to compute the effective convergence rate given a scaling of λ_n and μ_n. By identifying the optimal convergence rate, this approach can assist in determining the best parameters λ_n and μ_n for use in other PIML techniques, such as PINNs.

4. PDE Solving: Mitigating the Difficulties of PINNs with PIKL

It turns out that our PIKL algorithm can be effectively used as a PDE solver. In this scenario, there is no noise (i.e., $\varepsilon = 0$), no modeling error (i.e., $\mathscr{D}(f^\star) = 0$), and the data consist of samples of boundary and initial conditions, as is typical for PINNs. Assume for example that the objective is to solve the Laplacian equation $\Delta(f^\star) = 0$ on a domain $\Omega \subseteq [-1, 1]^2$ with the Dirichlet boundary condition $f^\star|_{\partial\Omega} = g$, where g is a known function. Then this problem can be addressed by implementing the PIKL estimator, which minimizes the risk $\bar{R}_n(f) = \frac{1}{n}\sum_{i=1}^n |f(X_i) - Y_i|^2 + \lambda_n \|f\|_{H^2_{\text{per}}([-1,1]^2)}^2 + \mu_n \|\Delta(f)\|_{L^2(\Omega)}^2$, where the X_i are uniformly sampled on $\partial\Omega$ and $Y_i = g(X_i)$. Of course, this example focuses on Dirichlet boundary conditions, but PIKL is a highly flexible framework that can incorporate a wide variety of boundary conditions, such as periodic and Neumann boundary conditions, as the next two examples will illustrate.

Comparison with PINNs for the convection equation. To begin, we compare the performance of our PIKL algorithm with the PINN approach developed by Krishnapriyan et al. (2021) for solving the one-dimensional convection equation $\mathscr{D}(f) = \partial_t f + \beta \partial_x f$ on the domain $\Omega = [0, 1] \times [0, 2\pi]$. The problem is subject to the following periodic boundary conditions:

$$\begin{cases} \forall x \in [0, 2\pi], & f(0, x) = \sin(x), \\ \forall t \in [0, 1], & f(t, 0) = f(t, 2\pi) = 0. \end{cases}$$

The solution of this PDE is given by $f^\star(t, x) = \sin(x - \beta t)$. Krishnapriyan et al. (2021) show that for high values of β, PINNs struggle to solve the PDE effectively. To address this challenge, we

train our PIML kernel method using $n = 100$ data points and 1681 Fourier modes (i.e., $m = 20$). The training data set $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ is constructed such that $X_i = (0, 2\pi U_i)$ and $Y_i = \sin(U_i)$, where $(U_i)_{1 \leqslant i \leqslant n}$ are i.i.d. uniform random variables. To enforce the periodic boundary conditions, we center Ω at $\tilde{\Omega} = \Omega - (0.5, \pi)$, extend it to $[-1, 1] \times [-\pi, \pi]$, and consider $\tilde{H}_m = \text{Span}((t, x) \mapsto e^{i(\frac{\pi}{2} k_1 t + k_2 x)})_{\|k\|_\infty \leqslant m}$. Noting that for all $(j_1, k_1), (j_2, k_2) \in \mathbb{Z}^2$,

$$\int_{[-1,1] \times [-\pi, \pi]} e^{i(\frac{\pi}{2}(k_1 - j_1)t + (k_2 - j_2)x)} dx = \frac{\sin(\pi(k_1 - j_1)/2)}{\pi} \delta_{k_2, j_2},$$

we let the matrix $(M_m)_{j,k}$ be as follows:

$$(M_m)_{j,k} = \lambda_n \Big(1 + \frac{\|k\|_2^2}{(2L)^2}\Big) \delta_{j,k} + \mu_n \frac{P(j)\bar{P}(k)}{(4L)^2} \frac{\sin(\pi(k_1 - j_1)/2)}{\pi} \delta_{k_2, j_2},$$

where P is the polynomial associated with the operator $\mathscr{D}$. Notice that, although $f^\star$ is a sinusoidal function, the frequency vector of $f^\star$ is $(-\beta, 1)$, which does not belong to $\frac{\pi}{2}\mathbb{Z} \oplus \mathbb{Z}$. As a result, $f^\star$ does not lie in $\tilde{H}_m$ for any m.

Table 1 compares the performance of various PIML methods using a sample of $n = 100$ initial condition points. The performance of an estimator $\hat{f}_n$ on a test set (Test) is evaluated based on the L^2 relative error $(\sum_{x \in \text{Test}} \|\hat{f}_n(x) - f^\star(x)\|_2^2 / \sum_{y \in \text{Test}} \|f^\star(y)\|_2^2)^{1/2}$. Standard deviations are computed across 10 trials. The results show that the PIML kernel estimator clearly outperforms PINNs in terms of accuracy.

Table 1: L^2 **relative error of the kernel method in solving the advection equation.**

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

$^\diamond$ *Krishnapriyan et al. (2021, Table 1)*

Comparison with PINNs for the 1d-wave equation. The performance of the PIKL algorithm is compared to the PINN methodology of Wang et al. (2022a, Section 7.3) for solving the one-dimensional wave equation $\mathscr{D}(f) = \partial_{t,t}^2 f - 4\partial_{x,x}^2 f$ on the square domain $[0, 1]^2$, with the following boundary conditions:

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

The solution of the PDE is $f^\star(t, x) = \sin(\pi x)\cos(2\pi t) + \sin(4\pi x)\cos(8\pi t)/2$. This solution serves as an interesting benchmark since $f^\star$ exhibits significant variations, with $\|\partial_t f^\star\|_2^2 / \|f^\star\|_2^2 = 16\pi^2$ (Figure 8, Left). Meanwhile, PINNs are known to have a spectral bias toward low frequencies (e.g., Deshpande et al., 2022; Wang et al., 2022b). The optimization of the PINNs in Wang et al. (2022b) is carried out using stochastic gradient descent with $80,000$ steps, each drawing 300 points at random, resulting in a sample size of $n = 2.4 \times 10^6$. The architecture of the PINNs

these authors employ is a dense neural network with $\tanh$ activation functions and layers of sizes $(2, 500, 500, 500, 1)$, resulting in $m = (2 \times 500 + 500) + 2 \times (500 \times 500 + 500) + (500 \times 1 + 1) = 503,001$ parameters. The training time for Vanilla PINNs is 7 minutes on an Nvidia L4 GPU (24 GB of RAM, 30.3 teraFLOPs for Float32). We obtain an L^2 relative error of 4.21×10^{-1}, which is consistent with the results of Wang et al. (2022a), who report a L^2 relative error of 4.52×10^{-1}. Figure 8 (Middle) shows the Vanilla PINNs.

We train our PIKL method using $n = 10^5$ data points and 1681 Fourier modes (i.e., $m = 20$). Let $(U_i)_{1 \leqslant i \leqslant n}$ be i.i.d. random variables uniformly distributed on $[0, 1]$. The training data set $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ is constructed such that

- if $1 \leqslant i \leqslant \lfloor n/4 \rfloor$, then $X_i = (0, U_i)$ and $Y_i = \sin(\pi U_i) + \sin(4\pi U_i)/2$,

- if $\lfloor n/4 \rfloor + 1 \leqslant i \leqslant 2\lfloor n/4 \rfloor$, then $X_i = (U_i, 0)$ and $Y_i = 0$,

- if $2\lfloor n/4 \rfloor + 1 \leqslant i \leqslant 3\lfloor n/4 \rfloor$, then $X_i = (U_i, 1)$ and $Y_i = 0$,

- if $3\lfloor n/4 \rfloor + 1 \leqslant i \leqslant n$, then $X_i = (1/n, U_i)$ and

$$Y_i = f(0, U_i) + \frac{1}{2n^2}\partial_{t,t}^2 f(0, U_i) = f(0, U_i) + \frac{2}{n^2}\partial_{x,x}^2 f(0, U_i)$$
$$= \left(1 - \frac{2\pi^2}{n^2}\right)\sin(\pi U_i) + \left(\frac{1}{2} - \frac{16\pi^2}{n^2}\right)\sin(4\pi U_i).$$

The final requirement enforces the initial condition $\partial_t f = 0$ in a manner similar to that of a second-order numerical scheme.

Table 2 compares the performance of the PINN approach from Wang et al. (2022a) with the PIKL estimator. Across 10 trials, the PIKL method achieves an L^2 relative error of $(8.70 \pm 0.08) \times 10^{-4}$, which is 50% better than the performance of the PINNs. This demonstrates that the kernel approach is more accurate, requiring fewer data points and parameters than the PINNs. The training time for the PIKL estimator is 6 seconds on an Nvidia L4 GPU. Thus, the PIKL estimator can be computed 70 times faster than the Vanilla PINNs. Figure 8 (Right) shows the PIKL estimator. Note that in this case, the solution $f^\star$ can be represented by a sum of complex exponential functions ($f^\star \in H_{16}$), which could have biased the result in favor of the PIKL estimator by canceling its approximation error. However, the results remain unchanged when altering the frequencies in H_m (e.g., taking $L = 0.55$ in Equation 5 instead of $L = 0.5$ yields an L^2 relative error of $(9.6 \pm 0.3) \times 10^{-4}$).

Table 2: **Performance of PINN/PIKL methods for solving the wave equation on $\Omega = [0, 1]^2$**

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

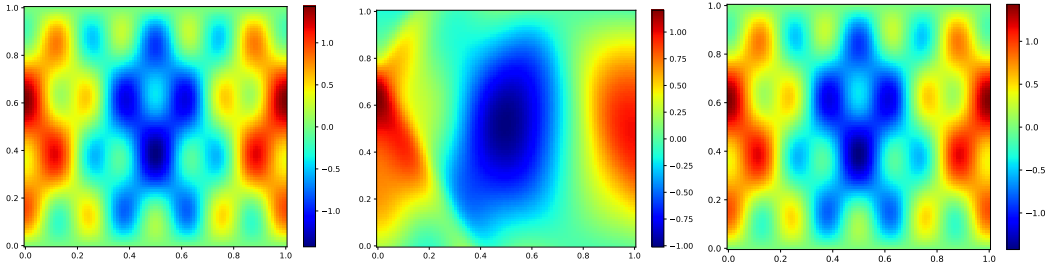$^\diamond$ *Wang et al. (2022a, Figure 6)*

Figure 8: **Left:** ground truth solution $f^\star$ to the wave equation (taken from Wang et al., 2022a, Figure 6). **Middle:** Vanilla PINNs from Wang et al. (2022a). **Right:** PIKL estimator.

5. PDE Solving with Noisy Boundary Conditions

5.1 Wave Equation in Dimension 2

Comparison with traditional PDE solvers. PIML is a promising framework for solving PDEs, particularly due to its adaptability to domains Ω with complex geometries, where most traditional PDE solvers tend to be highly domain-dependent. However, its comparative performance against traditional PDE solvers remains unclear in scenarios where both approaches can be easily implemented. The meta-analysis by McGreivy and Hakim (2024) indicates that, in some cases, PINNs may be faster than traditional PDE solvers, although they are often less accurate. In our study, solving the wave equation on a simple square domain represents a setting where traditional numerical methods are straightforward to implement and are known to perform well. Table 3 summarizes the performance of classical techniques, including the explicit Euler, Runge-Kutta 4 (RK4), and Crank-Nicolson (CN) schemes (see Appendix D.3 for a brief presentation of these methods). These methods clearly outperform both PINNs and the PIKL algorithm, even with fewer data points.

Table 3: **Performance of traditional PDE solvers for the wave equation on $\Omega = [0, 1]^2$.**

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

Noisy boundary conditions. However, a more relevant setting for comparing the performance of these methods arises when noise is introduced into the boundary conditions. This situation is common, for instance, when the initial condition of the wave is measured by a noisy sensor. Such a setting aligns with hybrid modeling, where $\varepsilon \neq 0$, but there is no modeling error (i.e., $\mathscr{D}(f^\star) = 0$). Table 4 compares the performance of all methods with Gaussian noise of variance of 10^{-2}. In this case, the PIKL estimator outperforms all other approaches.

Such PDEs with noisy boundary conditions are special cases of the hybrid modeling framework, where the data is located on the boundary of the domain. This situation arises, for example, in Cai et al. (2021) which models the temperature in the core of a nuclear reactor.

Table 4: **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}	$\mathbf{1.87 \times 10^{-2}}$
Training data (n)	2.4×10^6	4×10^4	4×10^4	4×10^4	4×10^4
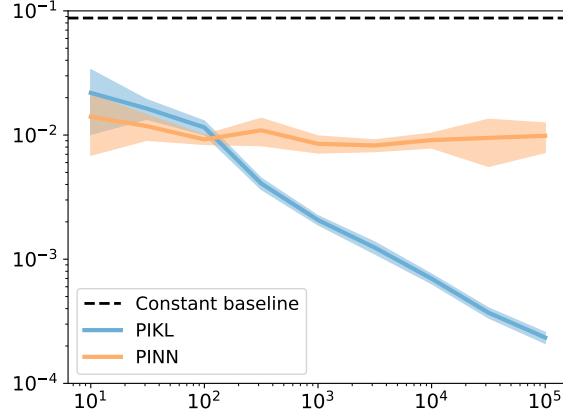
Figure 9: L^2 relative error of the models for the 4-dimensional heat equation with noisy boundary conditions as a function of the number of training points. Standard deviations are estimated over 5 runs.

5.2 Heat Equation in Dimension 4

Still in the context of noisy boundary conditions, we study the feasibility and limitations of the PIKL estimator in higher dimensions. To this aim, we consider the task of learning a solution to the heat equation in dimension 4 with noisy boundary conditions. In this setting, the goal is to learn

$$f^\star(x_1, x_2, x_3, x_4) = \exp(-3x_1/\pi^2) \cos(x_2/\pi) \cos(x_3/\pi) \cos(x_4/\pi)$$

on $\Omega = [-0.5, 0.5]^4$ given n i.i.d. observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ such that $Y = f^\star(X) + \mathcal{N}(0, 10^{-2})$, and X is sampled on $(\{0\} \times [-0.5, 0.5]^3) \cup ([-0.5, 0.5] \times \partial[-0.5, 0.5]^3)$ to respectively encode the initial and boundary conditions. The function $f^\star$ is the unique solution to the heat equation $(\partial_1 - \partial_{2,2}^2 - \partial_{3,3}^2 - \partial_{4,4}^2) f^\star = 0$ satisfying this set of initial and boundary conditions (see, e.g., Evans, 2010, Chapter 2.3, Theorem 5). The function $f^\star$ is flat, and close to the constant function equal to 1. We compare the performance of our PIKL estimator with a PINN. Here, the PIKL estimator is computed with $m = 3$, leading to 2401 Fourier modes. The PINN is a fully-connected neural network with three hidden layers of size 10, using $\tanh$ as activation function, and optimized on 2×10^5 collocation points by 2000 gradient descent steps. In this high-dimensional setting, Figure 9 shows that the PIKL estimator clearly outperforms the PINN both in terms of performance. Note that both methods outperform the constant model equal to $\int_\Omega f^\star$ (dotted line). Moreover, the PIKL estimator is more than 100 times faster than the PINN. The experimental convergence rate of the PIKL estimator is $n^{-0.53}$, which matches the parametric rate of $n^{-1/2}$.

6. Conclusion and Future Directions

In this article, we developed an efficient algorithm to solve the PIML hybrid problem (2). The PIKL estimator can be computed exactly through matrix inversion and possesses strong theoretical properties. Specifically, we demonstrated how to estimate its convergence rate based on the PDE prior $\mathscr{D}$. Moreover, through various examples, we showed that it outperforms PINNs in terms of performance, stability, and training time in certain PDE-solving tasks where PINNs struggle to escape local minima during optimization. Future work could focus on comparing PIKL with the implementation of RFF and exploring its performance against PINNs in the case of PDEs with non-constant coefficients. Another avenue for future research is to assess the effectiveness of the kernel approach compared to traditional PDE solvers, as discussed in Section 5.

Extension to nonlinear PDEs. Extending the PIKL framework to accommodate nonlinear PDEs is an important and interesting direction for future research. From a PDE theory point of view, it is expected to be harder to deal with nonlinear PDEs, since nonlinear differential operators are challenging compared to linear operators. Even in dimension $d = 1$, the solution of the ODE $y' = y^2$ with initial condition $y(0) = y_0$ is $y(t) = (1/y_0 - t)^{-1}$, which explodes at $t = y_0$. Note that the domain of the solution y is given by the set $\{t \leq y_0\}$, which intricately depends on the initial condition y_0. This prevents us from using a systematic methodology to solve this problem, and would require to design specific algorithms tailored to the condition-dependent geometry of the domain Ω.

Comparison with neural networks. In our experiments, physics-informed kernel methods systematically outperformed neural networks, both in terms of performance and running time. However, neural networks have the great property to be able to adapt to low-dimensional problems, resulting in quicker convergence rate (under the hypothesis of an exact optimization) contrary to kernel methods (see, e.g., Bach, 2024, Section 2.6). Thus, there are two scenarios in which PINNs could be preferable to PIKL. First, implementing PIKL for nonlinear PDEs is not straightforward, whereas PINNs handle them with ease. Second, for high-dimensional problems with an intrinsically low-dimensional structure—such as coupled differential equations commonly found in earth sciences or biology—the computational cost of PIKL becomes prohibitively high.

Acknowledgments

Appendix A. Comments on the PIKL Estimator

A.1 Spectral Methods and PIKL

The Fourier approximation on which the PIKL algorithm relies resembles usual spectral methods. Spectral methods are a class of numerical techniques used to solve PDEs by representing the solution as a sum of basis functions, typically trigonometric (Fourier series) or polynomial (Chebyshev or Legendre polynomials). These methods are particularly powerful for problems with smooth solutions and periodic or well-behaved boundary conditions (e.g., Canuto et al., 2007). However the basis functions used in spectral and pseudo-spectral methods must be specifically tailored to the domain Ω, the differential operator $\mathscr{D}$, and the boundary conditions. This customization ensures that the method effectively captures the characteristics of the problem being solved. For example, the Fourier basis is unable to accurately reconstruct non-periodic functions on a square domain, leading to the Gibbs phenomenon at points of periodic discontinuity. A natural solution to this problem is to extend the solution of the PDE from the domain Ω to a simpler domain that admits a known spectral basis (e.g., Matthysen and Huybrechs, 2016, for Fourier basis extension). If the solution of the PDE on Ω can be extended to a solution of the same PDE on the extended domain, it becomes possible to apply a spectral method directly to the extended domain (e.g., Badea and Daripa, 2001; Lui, 2009). However, the PDE must satisfy certain regularity conditions (e.g., ellipticity), and there must be a method to implement the boundary conditions on $\partial\Omega$ instead of on the boundary of the extended domain.

In this article, we take a slightly different approach. Although we extend $\Omega \subseteq [-L, L]^d$ to $[-2L, 2L]^d$, we impose the PDE only on Ω and not on the entire extended domain $[-2L, 2L]^d$. Also, unlike spectral methods, we do not require that $\mathscr{D}(\hat{f}^{\text{PIKL}}) = 0$. Instead, to ensure that the problem is well-posed, we regularize the PIML problem using the Sobolev norm of the periodic extension. This Tikhonov regularization is a conventional approach in kernel learning and is known to resemble spectral methods because it acts as a low-pass filter (see, e.g., Caponnetto and Vito, 2007). However, given a kernel, it is non-trivial to identify the basis of orthogonal functions that diagonalize it. The main contribution of this article is to establish an explicit connection between the Fourier basis and the PIML kernel, leading to surprisingly simple formulas for the kernel matrix M_m.

A.2 Choice of the Extended Domain

Embedding Ω in a toroidal structure is necessary to consider periodic functions, which is a requirement for our Fourier expansion. Let $L_{\mathbb{T}}$ define the length of the torus, i.e., such that an extension to $[-L_{\mathbb{T}}, L_{\mathbb{T}}]$ is considered. Since we assume that $\Omega \subseteq [-L, L]^d$, we necessarily have $L_{\mathbb{T}} > L$. This condition being satisfied, any value of $L_{\mathbb{T}} > L$ would be admissible for the PIKL framework.

However, the choice of $L_{\mathbb{T}}$ may impact the algorithmic performances. In particular, the basis functions $\Phi_k(x) = \exp(i\frac{\pi}{L_{\mathbb{T}}}\langle k, x\rangle)$ depend on $L_{\mathbb{T}}$. With a fixed number m of Fourier modes, we see that for all k such that $\|k\|_\infty \leq m$, and for all $x \in \Omega$, $\lim_{L_{\mathbb{T}} \to \infty} \Phi_k(x) = 1$. This means that, in the case of an excessively large torus (when $L_{\mathbb{T}} \to \infty$ grows), our PIKL estimator will only be able to learn constant functions. Therefore, $L_{\mathbb{T}}$ should not be too large relative to L.

Letting $L_{\mathbb{T}}$ be close to L, i.e., $L_{\mathbb{T}} = (1 + \varepsilon)L$ is not a good idea. To see this, consider the case where $d = 1$ and $\Omega = [-1, 1]$. Then, if ε is small, the periodic extension of $f^\star$ to $[-1 - \varepsilon, 1 + \varepsilon]$ will admit an exploding Sobolev norm. Indeed, for any periodic function $f \in H^1_{\text{per}}([-1-\varepsilon, 1+\varepsilon])$,

we have

$$\|f\|^2_{H^1_{\mathrm{per}}([-1-\varepsilon,1+\varepsilon])} = \int_{[-1-\varepsilon,1+\varepsilon]} f^2 + (f')^2 \geq \int_{[-1-\varepsilon,1+\varepsilon]} (f')^2 \geq \int_{[-\varepsilon,\varepsilon]} (f'(x+1+\varepsilon))^2 dx.$$

Then, the Cauchy-Schwarz inequality states that

$$\int_{[-\varepsilon,\varepsilon]} (f'(x+1+\varepsilon))^2 dx \int_{[-\varepsilon,\varepsilon]} 1 dx \geq \left(\int_{[-\varepsilon,\varepsilon]} f'(x+1+\varepsilon) dx\right)^2,$$

meaning that

$$\int_{[-\varepsilon,\varepsilon]} (f'(x+1))^2 dx \geq \varepsilon^{-1} [f(1) - f(-1)]^2.$$

All in all, denoting by $E(f^\star)$ the extension of $f^\star$ to $[-1-\varepsilon, 1+\varepsilon]$, we deduce that

$$\|E(f^\star)\|^2_{H^1_{\mathrm{per}}([-1-\varepsilon,1+\varepsilon])} \geq \varepsilon^{-1}(f^\star(1) - f^\star(-1))^2.$$

This computation shows that taking the extension torus to be $[-1-\varepsilon, 1+\varepsilon]$ results in the Sobolev norm $\|f^\star\|^2_{H^1_{\mathrm{per}}([-1-\varepsilon,1+\varepsilon])}$ introducing a bias towards functions satisfying $f^\star(1) = f^\star(-1)$, i.e., favoring periodic functions on Ω.

The choice of an optimal constant $L_{\mathbb{T}}$ depending on L is non-trivial, and closely relates to computing the constant in the Sobolev embedding $H^s(\Omega) \to H^s_{\mathrm{per}}([-L_{\mathbb{T}}, L_{\mathbb{T}}])$, which is known to be a difficult problem in functional analysis. Given that $L_{\mathbb{T}} = 2L$ strikes a balance by avoiding both pathological behaviors discussed earlier, we have adopted this choice throughout the paper.

A.3 Reproducing Property

Here, we formally prove that both properties

(i) $f(x) = \langle M_m^{1/2} z, M_m^{-1/2} \Phi_m(x) \rangle_{\mathbb{C}^{(2m+1)^d}}$, and

(ii) $\|f\|^2_{\mathrm{RKHS}} = \langle z, M_m z \rangle_{\mathbb{C}^{(2m+1)^d}} = \|M_m^{1/2} z\|^2_2$

are sufficient to show that minimizing $\bar{R}_n$ is a kernel method. From (i) and (ii), we deduce that the feature map is $x \mapsto M_m^{-1/2} \Phi_m(x)$. The kernel is thus necessarily given by $K(x,y) = \langle M_m^{-1/2} \Phi_m(x), M_m^{-1/2} \Phi_m(y) \rangle_{\mathbb{C}^{(2d+1)^d}}$. From (ii), we deduce that the RKHS inner product is given by $\langle z, \tilde{z} \rangle_{\mathrm{RKHS}} = \langle M_m^{1/2} z, M_m^{1/2} \tilde{z} \rangle_{\mathbb{C}^{(2d+1)^d}}$, so that $\|z\|^2_{\mathrm{RKHS}} = \|M_m^{1/2} z\|^2_2$.

The reproducing property is then a consequence of (i) and (ii). Indeed, let $x \in \Omega$. We know that $\langle f, K(x,\cdot) \rangle_{\mathrm{RKHS}} = \langle M_m^{1/2} z, M_m^{1/2} \tilde{z} \rangle_{\mathbb{C}^{(2d+1)^d}}$, where z is the Fourier vector of f and $\tilde{z}$ the Fourier vector of $K(x,\cdot)$. Since the kernel K results from

$$K(x,y) = \langle M_m^{-1/2} \Phi_m(x), M_m^{-1/2} \Phi_m(y) \rangle_{\mathbb{C}^{(2d+1)^d}} = \langle M_m^{-1} \Phi_m(x), \Phi_m(y) \rangle_{\mathbb{C}^{(2d+1)^d}},$$

we know that $\tilde{z} = M_m^{-1} \Phi_m(x)$. This means that

$$\langle f, K(x,\cdot) \rangle_{\mathrm{RKHS}} = \langle M_m^{1/2} z, M_m^{1/2} M_m^{-1} \Phi_m(x) \rangle_{\mathbb{C}^{(2d+1)^d}} = \langle z, \Phi_m(x) \rangle_{\mathbb{C}^{(2d+1)^d}} = f(x),$$

which is the reproducing property.

Appendix B. Fundamentals of Functional Analysis on Complex Hilbert Spaces

Let $L > 0$ and $d \in \mathbb{N}^\star$. We define $L^2([-2L, 2L]^d, \mathbb{C})$ as the space of complex-valued functions f on the hypercube $[-2L, 2L]^d$ such that $\int_{[-2L,2L]^d} |f|^2 < \infty$. The real part of f is denoted by $\Re(f)$, and the imaginary part by $\Im(f)$, such that $f = \Re(f) + i\Im(f)$. Throughout the appendix, for the sake of clarity, we use the dot symbol $\cdot$ to represent functions. For example, $\|\cdot\|$ denotes the function $x \mapsto \|x\|$, and $\langle \cdot, \cdot \rangle$ stands for the function $(x, y) \mapsto \langle x, y \rangle$.

Definition B.1 (L^2-space and $\|\cdot\|_2$-norm) *The separable Hilbert space $L^2([-2L, 2L]^d, \mathbb{C})$ is associated with the inner product $\langle f, g \rangle = \int_{[-2L,2L]^d} f\bar{g}$ and the norm $\|f\|_2^2 = \int_{[-2L,2L]^d} |f|^2$.*

Let $s \in \mathbb{N}$.

Definition B.2 (Periodic Sobolev spaces) *The periodic Sobolev space $H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{R})$ is the space of real functions $f(x) = \sum_{k \in \mathbb{Z}^d} z_k \exp(\frac{i\pi}{2L} \langle k, x \rangle)$ such that the Fourier coefficients z_k satisfy $\sum_k |z_k|^2 (1 + \|k\|^{2s}) < \infty$. The corresponding complex periodic Sobolev space is defined by*

$$H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{C}) = H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{R}) \oplus i\, H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{R}).$$

It is the space of complex-valued functions $f(x) = \sum_{k \in \mathbb{Z}^d} z_k \exp(\frac{i\pi}{2L} \langle k, x \rangle)$ such that $\sum_k |z_k|^2 (1 + \|k\|^{2s}) < \infty$.

We recall that, given two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, an operator is a linear function from $\mathcal{H}_1$ to $\mathcal{H}_2$.

Definition B.3 (Operator norm) *(e.g., Brezis, 2010, Section 2.6) Let $\mathscr{O} : L^2([-2L, 2L]^d, \mathbb{C}) \to L^2([-2L, 2L]^d, \mathbb{C})$ be an operator. Its operator norm $|||\mathscr{O}|||_2$ is defined by*

$$|||\mathscr{O}|||_2 = \sup_{\substack{g \in L^2([-2L,2L]^d, \mathbb{C}) \\ \|g\|_2 = 1}} \|\mathscr{O}g\|_2 = \sup_{\substack{g \in L^2([-2L,2L]^d, \mathbb{C}) \\ g \neq 0}} \|g\|_2^{-1} \|\mathscr{O}g\|_2.$$

The operator norm is sub-multiplicative, i.e., $|||\mathscr{O}_1 \circ \mathscr{O}_2|||_2 \leqslant |||\mathscr{O}_1|||_2 \times |||\mathscr{O}_2|||_2$.

Definition B.4 (Adjoint) *Let $(\mathcal{H}, \langle \cdot, \cdot \rangle_\mathcal{H})$ be an Hilbert space and $\mathscr{O} : \mathcal{H} \to \mathcal{H}$ be an operator. The adjoint $\mathscr{O}^\star$ of $\mathscr{O}$ is the unique operator such that $\forall f, g \in \mathcal{H}$, $\langle f, \mathscr{O}g \rangle_\mathcal{H} = \langle \mathscr{O}^\star f, g \rangle_\mathcal{H}$.*

If $\mathcal{H} = \mathbb{R}^d$ with the canonical scalar product, then $\mathscr{O}^\star$ is the $d \times d$ matrix $\mathscr{O}^\star = \mathscr{O}^T$. If $\mathcal{H} = \mathbb{C}^d$ with the canonical sesquilinear inner product, then $\mathscr{O}^\star$ is the $d \times d$ matrix $\mathscr{O}^\star = \bar{\mathscr{O}}^T$.

Definition B.5 (Hermitian operator) *Let $\mathcal{H}$ be an Hilbert space and $\mathscr{O} : \mathcal{H} \to \mathcal{H}$ be an operator. The operator $\mathscr{O}$ is said to be Hermitian if $\mathscr{O} = \mathscr{O}^\star$.*

Theorem B.6 (Spectral theorem) *(e.g. Rudin, 1991, Theorems 12.29 and 12.30) Let $\mathscr{O}$ be a positive Hermitian compact operator. Then $\mathscr{O}$ is diagonalizable on an Hilbert basis with positive eigenvalues that tend to zero. We denote its eigenvalues, ordered in decreasing order, by $\sigma(\mathscr{O}) = (\sigma_k^\downarrow(\mathscr{O}))_{k \in \mathbb{N}^\star}$.*

We emphasize that, given an invertible positive self-adjoint compact operator $\mathscr{O}$ and its inverse $\mathscr{O}^{-1}$, the eigenvalues of $\mathscr{O}^{-1}$ can also be ordered in increasing order, i.e.,

$$\sigma(\mathscr{O}^{-1}) = (\sigma_k^\uparrow(\mathscr{O}^{-1}))_{k \in \mathbb{N}^\star} = (\sigma_k^\downarrow(\mathscr{O})^{-1})_{k \in \mathbb{N}^\star}. \tag{8}$$

Theorem B.7 (Courant-Fischer minmax theorem) *(Brezis, 2010, Problem 37) Let $\mathscr{O} : \mathcal{H} \to \mathcal{H}$ be a positive Hermitian compact operator. Then*

$$\sigma_k^{\downarrow}(\mathscr{O}) = \max_{\substack{H \subseteq \mathcal{H} \\ \dim H = k}} \min_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, \mathscr{O}g \rangle_{\mathcal{H}}.$$

If $\mathscr{O}$ is injective, then

$$\sigma_k^{\uparrow}(\mathscr{O}^{-1}) = \min_{\substack{H \subseteq \mathscr{O}(\mathcal{H}) \\ \dim H = k}} \max_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, \mathscr{O}^{-1}g \rangle_{\mathcal{H}}.$$

Interestingly, if $\mathscr{O}$ is a positive Hermitian compact operator, then Theorem B.7 shows that $\|\|\mathscr{O}\|\|_2$ equals its largest eigenvalue.

Definition B.8 (Orthogonal projection on H_m) *We let $\Pi_m : H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{C}) \to H_m$ be the orthogonal projection with respect to $\langle \cdot, \cdot \rangle$, i.e., for all $f \in H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{C})$,*

$$\Pi_m f(y) = \sum_{\|k\|_{\infty} \leqslant m} \Big(\frac{1}{(4L)^d} \int_{[-2L, 2L]^d} \exp\Big(-\frac{i\pi}{2L}\langle k, x\rangle\Big) f(x)dx \Big) \exp\Big(\frac{i\pi}{2L}\langle k, y\rangle\Big).$$

Note that Π_m is Hermitian and that, for all $f \in L^2([-2L, 2L]^d, \mathbb{C})$, $\lim_{m \to \infty} \|f - \Pi_m f\|_2 = 0$.

Appendix C. Theoretical Results for PIKL

C.1 Detailed Computation of the Fourier Expansion of the Differential Penalty

The formula relating $\|\mathscr{D}(f)\|_{L^2(\Omega)}$ to the Fourier coefficients of f follows from (i) expanding f in the Fourier basis, (ii) leveraging the linearity of $\mathscr{D}$, (iii) applying the property $\mathscr{D}(x \mapsto e^{-\frac{i\pi}{2L}\langle k, x\rangle}) = (x \mapsto P(k)e^{-\frac{i\pi}{2L}\langle k, x\rangle})$, and (iv) recognizing that $|z|^2 = z\bar{z}$. From these steps, we deduce that

$$\|\mathscr{D}(f)\|_{L^2(\Omega)}^2 = \int_{\Omega} |\mathscr{D}(f)(x)|^2 dx$$

$$= \int_{\Omega} \Big| \sum_{\|j\|_{\infty} \leqslant m} z_j \frac{P(j)}{(4L)^{d/2}} e^{-\frac{i\pi}{2L}\langle j, x\rangle} \Big|^2 dx,$$

$$= \int_{\Omega} \Big(\sum_{\|j\|_{\infty} \leqslant m} z_j \frac{P(j)}{(4L)^{d/2}} e^{-\frac{i\pi}{2L}\langle j, x\rangle} \Big) \Big(\sum_{\|k\|_{\infty} \leqslant m} \bar{z}_k \frac{\bar{P}(k)}{(4L)^{d/2}} e^{\frac{i\pi}{2L}\langle k, x\rangle} dx \Big) dx$$

$$= \sum_{\|j\|_{\infty} \leqslant m, \|k\|_{\infty} \leqslant m} z_j \bar{z}_k \frac{P(j)\bar{P}(k)}{(4L)^d} \int_{\Omega} e^{\frac{i\pi}{2L}\langle k-j, x\rangle} dx.$$

C.2 Proof of Proposition 2.4

We have

$$\frac{1}{(4L)^d} \int_{[-L,L]^d} e^{\frac{i\pi}{2L}\langle k,x\rangle} dx = \frac{1}{(4L)^d} \prod_{j=1}^d \int_{[-L,L]} e^{\frac{i\pi}{2L}k_j x} dx = \prod_{j=1}^d \left[\frac{1}{2i\pi} e^{\frac{i\pi}{2L}k_j x}\right]_{x=-L}^L$$

$$= \prod_{j=1}^d \frac{e^{\frac{i\pi}{2}k_j} - e^{-\frac{i\pi}{2}k_j}}{2i\pi k_j} = \prod_{j=1}^d \frac{\sin(\frac{\pi}{2}k_j)}{\pi k_j}.$$

The characteristic function of the Euclidean ball is computed in Bracewell (2000, Table 13.4).

C.3 Operations on Characteristic Functions

Proposition C.1 (Operations on characteristic functions) *Consider $d \in \mathbb{N}^\star$, $L > 0$, and $\Omega \subseteq [-L, L]^d$.*

- *Let $a \in [-1, 1]$. Then $a \cdot \Omega \subseteq [-L, L]^d$ and*

$$F_{a\cdot\Omega}(k) = |a|^d \times F_\Omega(a \cdot k).$$

- *Let $\tilde{\Omega} \subseteq [-L, L]^d$ be a domain such that $\Omega \cap \tilde{\Omega} = \emptyset$. Then $\Omega \sqcup \tilde{\Omega} \subseteq [-L, L]^d$ and*

$$F_{\Omega \sqcup \tilde{\Omega}}(k) = F_\Omega(k) + F_{\tilde{\Omega}}(k).$$

- *Assume that $\Omega \subseteq [-L/2, L/2]^d$, and let $z \in \mathbb{R}^d$ be such that $\|z\|_\infty < L/2$. Then $\Omega + z \subseteq [-L, L]^d$ and*

$$F_{\Omega+z}(k) = F_\Omega(k) \times \exp\left(\frac{i\pi}{2L}\langle k, z\rangle\right).$$

- *Assume that $\Omega = \Omega_1 \times \Omega_2$, where $\Omega_1 \subseteq [-L, L]^{d_1}$, $\Omega_2 \subseteq [-L, L]^{d_2}$, and $d_1 + d_2 = d$. Then*

$$F_\Omega(k) = F_{\Omega_1}(k_1, \ldots, k_{d_1}) \times F_{\Omega_2}(k_{d_1+1}, \ldots, k_d).$$

C.4 Operator Extensions

Definition C.2 (Projection on Ω) *The projection $C : L^2([-2L, 2L]^d, \mathbb{C}) \to L^2([-2L, 2L]^d, \mathbb{C})$ on Ω is defined by $C(f) = 1_\Omega f$.*

Definition C.3 (Operator extensions) *The operators $C_m : H_m \to H_m$, $M_m : H_m \to H_m$, and $M_m^{-1} : H_m \to H_m$ can be extended to $L^2([-2L, 2L]^d, \mathbb{C})$ by $C_m = \Pi_m C_m \Pi_m$, $M_m = \Pi_m M_m \Pi_m$, and $M_m^{-1} = \Pi_m M_m^{-1} \Pi_m$.*

From now on, we consider the extensions of these operators, allowing us to express equivalently

$$|||M_m^{-1}|||_2 = \sup_{\substack{g \in H_m \\ \|g\|_2 = 1}} \|M_m^{-1}g\|_2 = \sup_{\substack{g \in L^2([-2L,2L]^d) \\ \|g\|_2 = 1}} \|M_m^{-1}g\|_2.$$

It is important to note that the extended operator M_m^{-1} is no longer the inverse of the extended operator M_m.

Proposition C.4 (Compact operator extension) *Let $\mathcal{O}$ be a positive Hermitian compact operator on $L^2([-2L, 2L]^d, \mathbb{R})$. Then its unique extension $\tilde{\mathcal{O}}$ to $L^2([-2L, 2L]^d, \mathbb{C})$ is a positive Hermitian compact operator with the same real eigenfunctions and positive eigenvalues.*

Proof Since $\tilde{\mathcal{O}}$ is $\mathbb{C}$-linear, we necessarily have $\tilde{\mathcal{O}}(f) = \mathcal{O}(\Re(f)) + i\mathcal{O}(\Im(f))$. Therefore, the extension is unique. Since $\mathcal{O}$ is compact, $\tilde{\mathcal{O}}$ is also compact. According to Theorem B.6, the operator $\mathcal{O}$ is diagonalizable in a Hermitian basis $(f_k)_{k \in \mathbb{N}^\star}$. Thus, for all $f \in L^2([-2L, 2L], \mathbb{R})$,

$$\mathcal{O}(f) = \sum_{k \in \mathbb{N}^\star} \sigma_k^\downarrow(\mathcal{O}) \langle f, f_k \rangle_{L^2([-2L,2L],\mathbb{R})} f_k.$$

Thus, for all $f \in L^2([-2L, 2L], \mathbb{C})$

$$\tilde{\mathcal{O}}(f) = \sum_{k \in \mathbb{N}^\star} \sigma_k^\downarrow(\mathcal{O}) \left(\langle \Re(f), f_k \rangle_{L^2([-2L,2L],\mathbb{R})} + i \langle \Im(f), f_k \rangle_{L^2([-2L,2L],\mathbb{R})} \right) f_k$$

$$= \sum_{k \in \mathbb{N}^\star} \sigma_k^\downarrow(\mathcal{O}) \langle f, f_k \rangle_{L^2([-2L,2L],\mathbb{C})} f_k.$$

This formula shows that $\tilde{\mathcal{O}}$ is Hermitian and diagonalizable with the same real eigenfunctions and positive eigenvalues as $\mathcal{O}$. ∎

Recall that $\mathcal{O}_n$ is the operator $\mathcal{O}_n = \lim_{m \to \infty} M_m^{-1}$, where the limit is taken in the sense of the operator norm (see Proposition B.2 Doumèche et al., 2024).

Definition C.5 (Operator M) *Proposition C.4 shows that the operator $\mathcal{O}_n$ can be extended to $L^2([-2L, 2L], \mathbb{C})$. We denote the extension of $\mathcal{O}_n$ by M^{-1}.*

The uniqueness of the extension in Proposition C.4 implies that the extension of the operator $C\mathcal{O}_n C : L^2([-2L, 2L]^d) \to L^2([-2L, 2L]^d)$ to $\mathbb{C}$ is indeed $CM^{-1}C : L^2([-2L, 2L]^d, \mathbb{C}) \to L^2([-2L, 2L]^d, \mathbb{C})$. Proposition C.4 shows that $C\mathcal{O}_n C$ has the same eigenvalues as $CM^{-1}C$.

C.5 Convergence of M_m^{-1}

Lemma C.6 (Bounding the spectrum of M_m^{-1}) *Let $m \in \mathbb{N}^\star$. Then, for all $k \in \mathbb{N}^\star$,*

$$\sigma_k^\downarrow(M_m^{-1}) \leqslant \sigma_k^\downarrow(M^{-1}).$$

Proof Let $f \in H_m$. Then,

$$\langle f, M_m f \rangle = \langle f, M f \rangle = \lambda_n \|f\|^2_{H_{\mathrm{per}}^s([-2L,2L]^d)} + \mu_n \|\mathscr{D}(f)\|^2_{L^2(\Omega)}.$$

Thus, using Theorem B.7, we deduce that

$$\begin{aligned}
\sigma_k^\uparrow(M_m) &= \min_{\substack{H \subseteq H_m \\ \dim H = k}} \max_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, M_m g \rangle \\
&= \min_{\substack{H \subseteq H_m \\ \dim H = k}} \max_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, M g \rangle \\
&\geqslant \min_{\substack{H \subseteq H_{\mathrm{per}}^s([-2L,2L]^d) \\ \dim H = k}} \max_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, M g \rangle \\
&= \sigma_k^\uparrow(M).
\end{aligned}$$

From (8), we deduce that $\sigma_k^{\downarrow}(M_m^{-1}) \leqslant \sigma_k^{\downarrow}(M^{-1})$. $\blacksquare$

Lemma C.7 (Spectral convergence of M_m) *Let $m \in \mathbb{N}^{\star}$. Then, for all $k \in \mathbb{N}^{\star}$, one has*

$$\lim_{m \to \infty} \sigma_k^{\downarrow}(M_m^{-1}) = \sigma_k^{\downarrow}(M^{-1}).$$

Proof By continuity of the RKHS norm $f \mapsto \langle f, Mf \rangle$ on $H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{C})$ (Doumèche et al., 2024, Proposition B.1), we know that, for all function $f \in H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{C})$, the quantity $\lambda_n \|\Pi_m(f)\|_{H_{\mathrm{per}}^s([-2L,2L]^d,\mathbb{C})}^2 + \mu_n \|\mathscr{D}(\Pi_m(f))\|_{L^2(\Omega,\mathbb{C})}^2$ converges to $\lambda_n \|f\|_{H_{\mathrm{per}}^s([-2L,2L]^d,\mathbb{C})}^2 + \mu_n \|\mathscr{D}(f)\|_{L^2(\Omega,\mathbb{C})}^2$ as m goes to the infinity. Thus,

$$\forall f \in H_{\mathrm{per}}^s([-2L, 2L]^d, \mathbb{C}), \quad \lim_{m \to \infty} \langle f, (M - M_m)f \rangle = 0.$$

Next, consider $f_1, \ldots, f_k$ to be the eigenfunctions of M associated with the ordered eigenvalue $\sigma_1^{\uparrow}(M), \ldots, \sigma_k^{\uparrow}(M)$. Since, for any $1 \leqslant j, \ell \leqslant k$, we have that $\lim_{m \to \infty} \langle f_j + f_\ell, M_m(f_j + f_\ell) \rangle = \langle f_j + f_\ell, M(f_j + f_\ell) \rangle$ and $\lim_{m \to \infty} \langle f_j + f_\ell, M_m(f_j + f_\ell) \rangle = \lim_{m \to \infty} \langle f_j, M_m f_j \rangle + \lim_{m \to \infty} \langle f_\ell, M_m f_\ell \rangle + 2 \lim_{m \to \infty} \Re(\langle f_j, M_m f_\ell \rangle)$, we deduce that $\lim_{m \to \infty} \Re(\langle f_j, M_m f_\ell \rangle) = \Re(\langle f_j, M f_\ell \rangle)$. Using the same argument by developing $\langle f_j + if_\ell, M_m(f_j + if_\ell) \rangle$ shows that $\lim_{m \to \infty} \Im(\langle f_j, M_m f_\ell \rangle) = \Im(\langle f_j, M f_\ell \rangle)$. Overall,

$$\forall 1 \leqslant j, \ell \leqslant k, \quad \lim_{m \to \infty} \langle f_j, M_m f_\ell \rangle = \langle f_j, M f_\ell \rangle. \tag{9}$$

Now, observe that

$$(g \in \mathrm{Span}(f_1, \ldots, f_k) \text{ and } \|g\|_2 = 1) \Leftrightarrow (\exists (a_1, \ldots, a_k) \in \mathbb{C}^k, \ g = \sum_{j=1}^k a_j f_j \text{ and } \sum_{j=1}^k |a_j|^2 = 1).$$

Thus,

$$\max_{\substack{g \in \mathrm{Span}(f_1, \ldots, f_k) \\ \|g\|_2 = 1}} |\langle g, M_m g \rangle - \langle g, M g \rangle| \leqslant \max_{\|a\|_2 = 1} \sum_{i,j=1}^k |a_i a_j| |\langle f_i, (M_m - M) f_j \rangle|$$

$$\leqslant k \max_{1 \leqslant i,j \leqslant k} |\langle f_i, (M_m - M) f_j \rangle|$$

$$\xrightarrow{m \to \infty} 0 \quad \text{according to (9).}$$

So,

$$\lim_{m \to \infty} \max_{\substack{g \in \mathrm{Span}(\Pi_m(f_1), \ldots, \Pi_m(f_k)) \\ \|g\|_2 = 1}} \langle g, M_m g \rangle = \lim_{m \to \infty} \max_{\substack{g \in \mathrm{Span}(f_1, \ldots, f_k) \\ \|g\|_2 = 1}} \langle g, M_m g \rangle$$

$$= \max_{\substack{g \in \mathrm{Span}(f_1, \ldots, f_k) \\ \|g\|_2 = 1}} \langle g, M g \rangle$$

$$= \sigma_k^{\uparrow}(M). \tag{10}$$

Note that $\mathrm{Span}(\Pi_m(f_1), \ldots, \Pi_m(f_k)) \subseteq H_m$. Moreover, for m large enough, we have that $\dim \mathrm{Span}(\Pi_m(f_1), \ldots, \Pi_m(f_k)) = k$. Therefore, according to Theorem B.7,

$$\sigma_k^\uparrow(M_m) = \min_{\substack{H \subseteq H_m \\ \dim H = k}} \max_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, M_m g \rangle \leqslant \max_{\substack{g \in \mathrm{Span}(\Pi_m(f_1), \ldots, \Pi_m(f_k)) \\ \|g\|_2 = 1}} \langle g, M_m g \rangle.$$

Combining this inequality with identity (10) shows that $\limsup_{m \to \infty} \sigma_k^\uparrow(M_m) \leqslant \sigma_k^\uparrow(M)$. Equivalently,

$$\liminf_{m \to \infty} \sigma_k^\downarrow(M_m^{-1}) \geqslant \sigma_k^\downarrow(M^{-1}).$$

Finally, by Lemma C.6, we have $\sigma_k^\downarrow(M_m^{-1}) \leqslant \sigma_k^\downarrow(M^{-1})$. We conclude that $\lim_{m \to \infty} \sigma_k^\downarrow(M_m^{-1}) = \sigma_k^\downarrow(M^{-1})$. ∎

Lemma C.8 (Eigenfunctions convergence) *Let $(f_{j,m})_{j \in \mathbb{N}^\star}$ be the eigenvectors of M_m associated with the eigenvalues $(\sigma_j^\uparrow(M_m))_{j \in \mathbb{N}^\star}$. Let $E_j = \ker(M - \sigma_j^\uparrow(M)\mathrm{Id})$. Then*

$$\forall j \in \mathbb{N}^\star, \quad \lim_{m \to \infty} \min_{y \in E_j} \|f_{j,m} - y\|_2 = 0.$$

Proof Let $(f_j)_{j \in \mathbb{N}^\star}$ be the eigenvectors of M associated with the eigenvalues $(\sigma_j^\uparrow(M))_{j \in \mathbb{N}^\star}$.

The proof proceeds by contradiction. Assume that the lemma is false, and consider the minimum integer $p \in \mathbb{N}^\star$ such that $\limsup_{m \to \infty} \min_{y \in E_p} \|f_{p,m} - y\|_2 > 0$. Let $k_1 < p$ be the largest integer such that $\sigma_{k_1}^\uparrow(M_m) < \sigma_p^\uparrow(M_m)$, and let $k_2 > p$ be the smallest integer such that $\sigma_{k_2}^\uparrow(M_m) > \sigma_p^\uparrow(M_m)$. Observe that $E_p = \mathrm{Span}(f_{k_1+1}, \ldots, f_{k_2-1})$.

Let $k_1 < j < k_2$. We know that $\langle f_j, M_m f_j \rangle = \sum_{\ell \in \mathbb{N}^\star} \sigma_j^\uparrow(M_m) |\langle f_{\ell,m}, f_j \rangle|^2$ from diagonalizing M_m. The minimality assumption on j ensures that, for all $\ell \leqslant k_1$, $\lim_{m \to \infty} \min_{y \in E_\ell} \|f_{\ell,m} - y\|_2 = 0$. Since $E_j \subseteq \mathrm{Span}(f_1, \ldots, f_{k_1})$, we deduce that $\forall \ell \leqslant k_1, \lim_{m \to \infty} \langle f_{\ell,m}, f_j \rangle = 0$. Thus,

$$\sum_{\ell \geqslant k_1} |\langle f_{\ell,m}, f_j \rangle|^2 = 1 + o_{m \to \infty}(1), \tag{11}$$

and

$$\sum_{k_1 \leqslant \ell < k_2} \sigma_j^\uparrow(M_m) |\langle f_{\ell,m}, f_j \rangle|^2 = o_{m \to \infty}(1). \tag{12}$$

By Lemma C.7, using $|\langle f_{\ell,m}, f_j \rangle|^2 \leqslant 1$, we have

$$\sum_{k_1 \leqslant \ell \leqslant k_2} \sigma_j^\uparrow(M_m) |\langle f_{\ell,m}, f_j \rangle|^2 = \sigma_j^\uparrow(M) \sum_{k_1 \leqslant \ell < k_2} |\langle f_{\ell,m}, f_j \rangle|^2 + o_{m \to \infty}(1). \tag{13}$$

Combining (12) and (13), we deduce that

$$\langle f_j, M_m f_j \rangle = o_{m \to \infty}(1) + \sigma_j^\uparrow(M) \sum_{k_1 \leqslant \ell < k_2} |\langle f_{\ell,m}, f_j \rangle|^2 + \sum_{\ell \geqslant k_2} \sigma_j^\uparrow(M_m) |\langle f_{\ell,m}, f_j \rangle|^2.$$

27

Moreover, identity (9) ensures that $\langle f_j, M_m f_j \rangle = \sigma_j^{\uparrow}(M) + o_{m \to \infty}(1)$. Thus,

$$\sigma_j^{\uparrow}(M)(1 - \sum_{k_1 \leqslant \ell < k_2} |\langle f_{\ell,m}, f_j \rangle|^2) = o_{m \to \infty}(1) + \sum_{\ell \geqslant k_2} \sigma_j^{\uparrow}(M_m) |\langle f_{\ell,m}, f_j \rangle|^2.$$

However, according to Lemma C.7, there is $\varepsilon > 0$ such that, for m large enough,

$$\forall \ell \geqslant k_2, \quad \sigma_j^{\uparrow}(M_m) \geqslant \sigma_k^{\uparrow}(M) + \varepsilon.$$

Hence,

$$\sigma_j^{\uparrow}(M)(1 - \sum_{k_1 \leqslant \ell < k_2} |\langle f_{\ell,m}, f_j \rangle|^2) \geqslant o_{m \to \infty}(1) + (\sigma_j^{\uparrow}(M) + \varepsilon) \sum_{\ell \geqslant k_2} |\langle f_{\ell,m}, f_j \rangle|^2.$$

Combining this inequality with (11), this means that

$$0 \geqslant o_{m \to \infty}(1) + \varepsilon \sum_{\ell \geqslant k_2} |\langle f_{\ell,m}, f_j \rangle|^2.$$

Thus, $\lim_{m \to \infty} \sum_{\ell \geqslant k_2} |\langle f_{\ell,m}, f_j \rangle|^2 = 0$ and $\lim_{m \to \infty} \sum_{k_1 \leqslant \ell < k_2} |\langle f_{\ell,m}, f_j \rangle|^2 = 1$.

We deduce that, for all $k_1 < j < k_2$, $\lim_{m \to \infty} \min_{y \in \mathrm{Span}(f_{k_1+1,m}, \ldots, f_{k_2-1,m})} \|f_j - y\|_2 = 0$. By symmetry of the ℓ^2-distance between two spaces of the same dimension $k_2 - k_1 - 1$, for all $k_1 < j < k_2$, $\lim_{m \to \infty} \min_{y \in \mathrm{Span}(f_{k_1+1}, \ldots, f_{k_2-1})} \|f_{j,m} - y\|_2 = 0$. This contradicts the fact that $\limsup_{m \to \infty} \min_{y \in E_p} \|f_{p,m} - y\|_2 > 0$. ∎

Lemma C.9 (Convergence of M_m^{-1}) *One has*

$$\lim_{m \to \infty} |||M^{-1} - M_m^{-1}|||_2 = 0.$$

Proof Let $(f_{j,m})_{j \in \mathbb{N}^\star}$ be the eigenvectors of M_m, each associated with the corresponding eigenvalues $(\sigma_j^{\uparrow}(M_m))_{j \in \mathbb{N}^\star}$. Let $(f_j)_{j \in \mathbb{N}^\star}$ be the eigenvectors of M, each associated with the eigenvalues $(\sigma_j^{\uparrow}(M))_{j \in \mathbb{N}^\star}$. By Lemma C.6, $\sigma_j^{\downarrow}(M_m^{-1}) \leqslant \sigma_j^{\downarrow}(M^{-1})$; by Lemma C.7, $\lim_{m \to \infty} \sigma_j^{\downarrow}(M_m^{-1}) = \sigma_j^{\downarrow}(M^{-1})$; and by Lemma C.8 $\lim_{m \to \infty} \min_{y \in E_j} \|f_{j,m} - y\|_2 = 0$.

Notice that $M_m^{-1} = \sum_{\ell \in \mathbb{N}} \sigma_j^{\downarrow}(M_m^{-1}) \langle f_{j,m}, \cdot \rangle f_{j,m}$ and that $M^{-1} = \sum_{\ell \in \mathbb{N}} \sigma_j^{\downarrow}(M^{-1}) \langle f_j, \cdot \rangle f_j$. Let $g \in L^2([-2L, 2L]^d, \mathbb{C})$ be such that $\|g\|_2 = 1$. Then,

$$\begin{aligned}
\|(M^{-1} - M_m^{-1})g\|_2 &\leqslant \sum_{j \in \mathbb{N}^\star} \|\sigma_j^{\downarrow}(M_m^{-1}) \langle f_{j,m}, g \rangle f_{j,m} - \sigma_j^{\downarrow}(M^{-1}) \langle f_j, g \rangle f_j\|_2 \\
&\leqslant \sum_{j \in \mathbb{N}^\star} (\sigma_j^{\downarrow}(M_m^{-1}) - \sigma_j^{\downarrow}(M^{-1})) \|\langle f_{j,m}, g \rangle f_{j,m}\|_2 \\
&\quad + \sum_{j \in \mathbb{N}^\star} \sigma_j^{\downarrow}(M^{-1}) \|\langle f_{j,m} - f_j, g \rangle f_{j,m}\|_2 \\
&\quad + \sum_{j \in \mathbb{N}^\star} \sigma_j^{\downarrow}(M^{-1}) \|\langle f_j, g \rangle (f_{j,m} - f_j)\|_2.
\end{aligned}$$

28

Since $\|f_{j,m}\|_2 = \|f_j\|_2 = 1$, it follows that $|\langle f_{j,m}, g \rangle| \leqslant 1$. Additionally, by the Cauchy-Schwarz inequality, $|\langle f_j, g \rangle| \leqslant 1$ and $|\langle f_{j,m} - f_j, g \rangle| \leqslant \|(f_{j,m} - f_j)\|_2$. Thus, the above inequality can be simplified as

$$\|(M^{-1} - M_m^{-1})g\|_2 \leqslant \sum_{j \in \mathbb{N}^\star} (\sigma_j^\downarrow(M_m^{-1}) - \sigma_j^\downarrow(M^{-1}) + 2\sigma_j^\downarrow(M^{-1})\|f_{j,m} - f_j\|_2).$$

Thus,

$$|||M^{-1} - M_m^{-1}|||_2 \leqslant \sum_{j \in \mathbb{N}^\star} (\sigma_j^\downarrow(M_m^{-1}) - \sigma_j^\downarrow(M^{-1}) + 2\sigma_j^\downarrow(M^{-1})\|f_{j,m} - f_j\|_2).$$

Clearly, since $|\sigma_j^\downarrow(M_m^{-1}) - \sigma_j^\downarrow(M^{-1})| \leqslant 2\sigma_j^\downarrow(M^{-1})$ and $\|f_{j,m} - f_j\|_2 \leqslant 2$,

$$|\sigma_j^\downarrow(M_m^{-1}) - \sigma_j^\downarrow(M^{-1}) + 2\sigma_j^\downarrow(M^{-1})\|f_{j,m} - f_j\|_2| \leqslant 4\sigma_j^\downarrow(M^{-1}).$$

Moreover, $\sum_{j \in \mathbb{N}^\star} \sigma_j^\downarrow(M^{-1}) < \infty$ (Doumèche et al., 2024, Proposition B.6). Thus, since we have that $\lim_{m \to \infty} |\sigma_j^\downarrow(M_m^{-1}) - \sigma_j^\downarrow(M^{-1})| = \lim_{m \to \infty} \|(f_{j,m} - f_j)\|_2 = 0$, we conclude with the dominated convergence theorem that $\lim_{m \to \infty} |||M^{-1} - M_m^{-1}|||_2 = 0$, as desired. $\blacksquare$

C.6 Operator Norms of C_m and C

Lemma C.10 *One has $|||C|||_2 \leqslant 1$ and $|||C_m|||_2 \leqslant 1$, for all $m \in \mathbb{N}^\star$.*

Proof Let $g \in L^2([-2L, 2L]^d, \mathbb{C})$. Then, by definition, $Cg = 1_\Omega g$, and $\|Cg\|_2 = \|1_\Omega g\|_2 \leqslant \|g\|_2$. Therefore, $|||C|||_2 \leqslant 1$.

Let $m \in \mathbb{N}^\star$. Then, since $C_m : L^2([-2L, 2L]^d, \mathbb{C}) \to L^2([-2L, 2L]^d, \mathbb{C})$ is a positive Hermitian compact operator, Theorem B.7 states that

$$\sigma_1^\downarrow(C_m) = \max_{\substack{h \in L^2([-2L,2L]^d, \mathbb{C}) \\ \|h\|_2 = 1}} \langle h, C_m h \rangle = \max_{\substack{h \in L^2([-2L,2L]^d, \mathbb{C}) \\ \|h\|_2 = 1}} \|\Pi_m h\|_2^2 \leqslant \max_{\substack{h \in L^2([-2L,2L]^d, \mathbb{C}) \\ \|h\|_2 = 1}} \|h\|_2^2 = 1.$$

Since $\sigma_1^\downarrow(C_m^2) = \sigma_1^\downarrow(C_m)^2 \leqslant 1$, we deduce that

$$1 \geqslant \sigma_1^\downarrow(C_m)^2 = \max_{\substack{h \in L^2([-2L,2L]^d, \mathbb{C}) \\ \|h\|_2 = 1}} \langle h, C_m^2 h \rangle = \max_{\substack{h \in L^2([-2L,2L]^d, \mathbb{C}) \\ \|h\|_2 = 1}} \|C_m h\|_2^2.$$

This shows that $|||C_m|||_2 \leqslant 1$. $\blacksquare$

C.7 Proof of Theorem 3.1

Note that if H is a linear subspace of H_m, then $C_m H$ is also a subspace of H_m, and $\dim H \geqslant \dim C_m H$. Therefore,

$$
\begin{aligned}
\sigma_k^\downarrow(C_m M_m^{-1} C_m) &= \max_{\substack{H \subseteq H_m \\ \dim H = k}} \min_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, C_m M_m^{-1} C_m g \rangle \\
&= \max_{\substack{H \subseteq H_m \\ \dim H = k}} \min_{\substack{g \in H \\ \|g\|_2 = 1}} \langle (C_m g), M_m^{-1}(C_m g) \rangle \\
&\leqslant \max_{\substack{H \subseteq H_m \\ \dim H = k}} \min_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, M_m^{-1} g \rangle \\
&= \sigma_k^\downarrow(M_m^{-1}) \\
&\leqslant \sigma_k^\downarrow(M^{-1}).
\end{aligned}
$$

Moreover, according to Lemma C.9, one has $|||M_m^{-1} - M^{-1}|||_2 \to 0$. Thus, $\sup_{\|g\|_2 = 1} \|(M_m^{-1} - M^{-1})(g)\|_2 \to 0$. Using Lemma C.10, we see that

$$
\begin{aligned}
&\|C_m M_m^{-1} C_m g - C M^{-1} C g\|_2 \\
&\leqslant \|(C_m - C) M^{-1} C g\|_2 + \|C_m (M_m^{-1} C_m - M^{-1} C) g\|_2 \\
&\leqslant |||(C_m - C) M^{-1}|||_2 + \|(M_m^{-1} C_m - M^{-1} C) g\|_2 \\
&\leqslant |||(C_m - C) M^{-1}|||_2 + \|(M_m^{-1} - M^{-1}) C_m g\|_2 + \|M^{-1}(C_m - C) g\|_2 \\
&\leqslant |||(C_m - C) M^{-1}|||_2 + |||M_m^{-1} - M^{-1}|||_2 + |||M^{-1}(C_m - C)|||_2.
\end{aligned}
$$

Thus,

$$
|||C_m M_m^{-1} C_m - C M^{-1} C|||_2 \leqslant |||(C_m - C) M^{-1}|||_2 + |||M_m^{-1} - M^{-1}|||_2 + |||M^{-1}(C_m - C)|||_2.
$$

By diagonalizing M^{-1} and using the facts that $\sum_{\ell \in \mathbb{N}^\star} \sigma_\ell^\downarrow(M^{-1}) < \infty$ and that $\lim_{m \to \infty} \|(C_m - C) f\|_2 = 0 \; \forall f \in L^2([-2L, 2L]^d)$, it is easy to see that

$$
\lim_{m \to \infty} |||(C_m - C) M^{-1}|||_2 = \lim_{m \to \infty} |||M^{-1}(C_m - C)|||_2 = 0.
$$

Applying Lemma C.9, we deduce that

$$
\lim_{m \to \infty} |||C_m M_m^{-1} C_m - C M^{-1} C|||_2 = 0.
$$

But, by Theorem B.7,

$$
\sigma_k^\downarrow(C_m M_m^{-1} C_m) = \max_{\substack{H \subseteq L^2([-2L, 2L]^d, \mathbb{C}) \\ \dim H = k}} \min_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, C_m M_m^{-1} C_m g \rangle,
$$

and

$$
\sigma_k^\downarrow(C M^{-1} C) = \max_{\substack{H \subseteq L^2([-2L, 2L]^d, \mathbb{C}) \\ \dim H = k}} \min_{\substack{g \in H \\ \|g\|_2 = 1}} \langle g, C M^{-1} C g \rangle.
$$

Clearly, for all $g \in L^2([-2L, 2L]^d, \mathbb{C})$,

$$|\langle g, CM^{-1}Cg \rangle - \langle g, C_m M_m^{-1} C_m g \rangle| = |\langle g, (CM^{-1}C - C_m M_m^{-1} C_m)g \rangle|$$
$$\leqslant |||C_m M_m^{-1} C_m - CM^{-1}C|||_2.$$

Therefore,

$$|\sigma_k^{\downarrow}(C_m M_m^{-1} C_m) - \sigma_k^{\downarrow}(CM^{-1}C)| \leqslant |||C_m M_m^{-1} C_m - CM^{-1}C|||_2,$$

and, in turn, $\lim_{m \to \infty} \sigma_k^{\downarrow}(C_m M_m^{-1} C_m) = \sigma_k^{\downarrow}(CM^{-1}C)$.

To conclude the proof, observe that, on the one hand,

$$\frac{1}{1 + \sigma_k^{\downarrow}(C_m M_m^{-1} C_m)^{-1}} = \frac{\sigma_k^{\downarrow}(C_m M_m^{-1} C_m)}{1 + \sigma_k^{\downarrow}(C_m M_m^{-1} C_m)} \leqslant \sigma_k^{\downarrow}(C_m M_m^{-1} C_m) \leqslant \sigma_k^{\downarrow}(M^{-1}),$$

with $\sum_{k \in \mathbb{N}^\star} \sigma_k^{\downarrow}(M^{-1}) < \infty$. On the other hand,

$$\lim_{m \to \infty} \frac{1}{1 + \sigma_k^{\downarrow}(C_m M_m^{-1} C_m)^{-1}} = \frac{1}{1 + \sigma_k^{\downarrow}(CM^{-1}C)^{-1}},$$

by continuity on $\mathbb{R}^+$ of the function $x \mapsto \frac{x}{1+x}$. Thus, applying the dominated convergence theorem, we are led to

$$\lim_{m \to \infty} \sum_{\lambda \in \sigma(C_m M_m^{-1} C_m)} \frac{1}{1 + \lambda^{-1}} = \sum_{\lambda \in \sigma(CM^{-1}C)} \frac{1}{1 + \lambda^{-1}}.$$

Appendix D. Experiments

D.1 Numerical Precision

Enabling high numerical precision is crucial for efficient kernel inversion. Setting the default precision to *Float32* and *Complex64* can lead to significant numerical errors when approximating the kernel. For example, consider the harmonic oscillator case with $d = 1$, $s = 2$, and the operator $\mathscr{D} = \frac{d^2}{dx^2} u + \frac{d}{dx} u + u$, with $\lambda_n = 0.01$ and $\mu_n = 1$. Figure 10 (left) shows the spectrum of $C_m M_m^{-1} C_m$ using *Float32* precision, while Figure 10 (right) shows the same spectrum with *Float64* precision. It is evident that with *Float32*, the diagonalization results in lower eigenvalues compared to *Float64*. In more physical terms, some energy of the matrix is lost when the last digits of the matrix coefficients are ignored. This leads to a problematic interpretation of the situation, as the *Float64* estimation of the eigenvalues shows a clear convergence of the spectrum of $C_m M_m^{-1} C_m$, whereas the *Float32* estimation appears to indicate divergence.

D.2 Convergence of the Effective Dimension Approximation

The PIKL algorithm relies on a Fourier approximation of the PIML kernel, as developed in Section 2. The precision of this approximation is determined by the number m of Fourier modes used to compute the kernel. However, determining an *optimal* value of m for a specific regression problem is challenging. There is a trade-off between the accuracy of the kernel estimation, which improves with higher values of m, and the computational complexity of the algorithm, which also
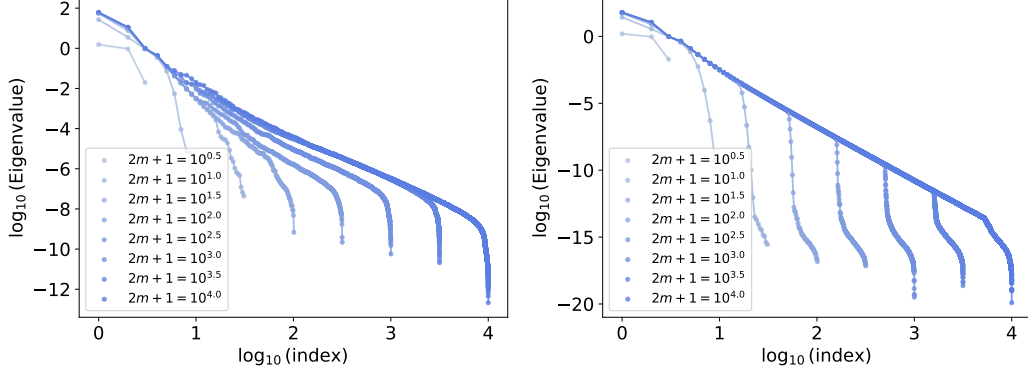
Figure 10: Spectrum of $C_m M_m^{-1} C_m$. **Left:** *Float32* precision. **Right:** *Float64* precision.
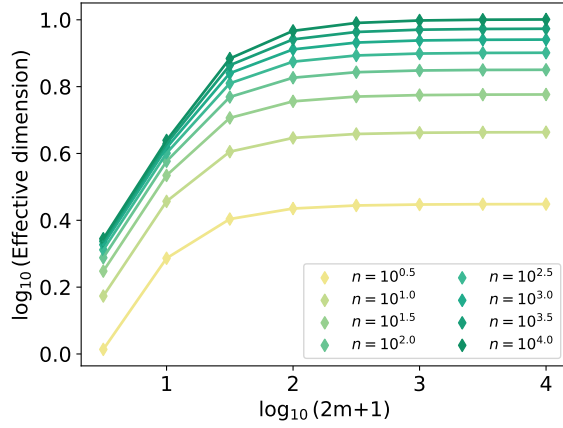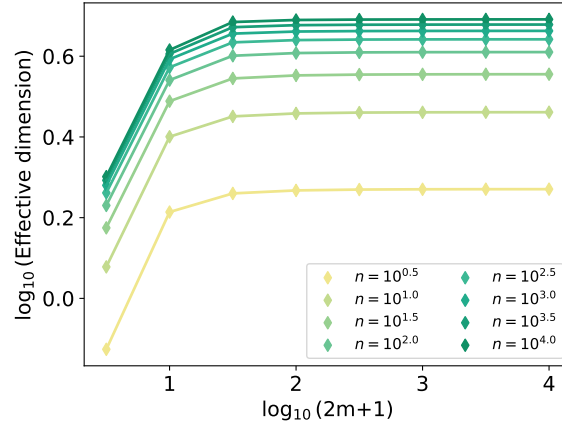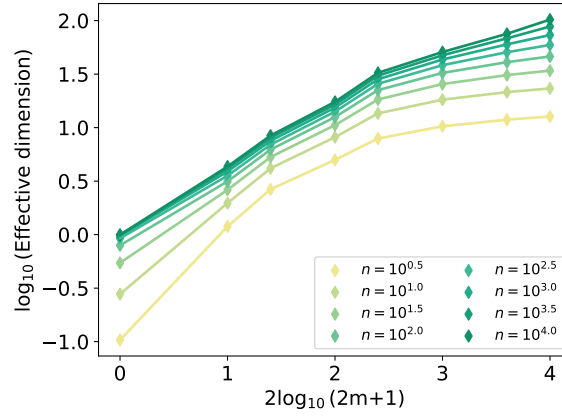

Figure 11: Convergence of the effective dimension as m grows for $\mathscr{D} = \frac{d}{dx}$.

increases with m. There is a trade-off between the accuracy of the kernel approximation, which improves with higher values of m, and the computational complexity of the algorithm, which also increases with m.

An interesting tool to leverage here is the effective dimension, as it captures the underlying degrees of freedom of the PIML problem and, consequently, the precision of the method. Theorem 3.1 states that the estimation of the effective dimension on H_m converges to the effective dimension on $H^s(\Omega)$ as m increases to infinity. Therefore, the smallest value $m^\star$ at which the effective dimension stabilizes is a strong candidate for balancing accuracy and computational complexity.

Figures 11, 12, and 13 illustrate the convergence of the effective dimension estimation, using the eigenvalues of $C_m M_M^{-1} C_m$, as m increases, for different values of n. These figures provide insights into the PIK algorithm. As expected, the Fourier approximations converge more slowly as the dimension d increases. Specifically, Figures 11 and 12 show that in dimension $d = 1$, $m^\star \simeq 10^2$ for $n \leqslant 10^4$, while Figure 13 indicates that in dimension $d = 2$, $m^\star \simeq 10^2$ for $n \leqslant 10^3$.

Figure 12: Convergence of the effective dimension as m grows for the harmonic oscillator

Figure 13: Convergence of the effective dimension as m grows for the heat equation on the disk.

D.3 Numerical Schemes

We detail below the numerical schemes used as benchmarks in Section 4 for solving the wave equation. All these numerical schemes are constructed by discretizing the domain $\Omega = [0,1]^2$ into the grid $(\ell_1^{-1}\mathbb{Z}/\ell_1\mathbb{Z}) \times (\ell_2^{-1}\mathbb{Z}/\ell_2\mathbb{Z})$. The initial and boundary conditions are then enforced on $n = 2\ell_1 + \ell_2$ points, with the approximation $\hat{f}_n$ defined accordingly as

- for all $0 \leqslant \ell \leqslant \ell_2$, $\hat{f}_n(0, \ell/\ell_2) = \sin(\pi\ell/\ell_2) + \sin(4\pi\ell/\ell_2)/2$,

- for all $0 \leqslant \ell \leqslant \ell_1$, $\hat{f}_n(\ell/\ell_1, 0) = 0$,

- for all $0 \leqslant \ell \leqslant \ell_1$, $\hat{f}_n(\ell/\ell_1, 1) = 0$.

Let the discrete Laplacian $\Delta_{(\ell_1,\ell_2)}$ be defined for all $(a, b) \in \mathbb{Z}/\ell_1\mathbb{Z} \times \mathbb{Z}/\ell_2\mathbb{Z}$ by

$$(\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, b/\ell_2) = \ell_2^2(\hat{f}_n(a/\ell_1, (b+1)/\ell_2) - 2\hat{f}_n(a/\ell_1, b/\ell_2) + \hat{f}_n(a/\ell_1, (b-1)/\ell_2)).$$

If $f^\star \in C^2([0,1]^2)$, its Taylor expansion leads to $(\Delta_{(\ell_1,\ell_2)}f^\star)(a/\ell_1, b/\ell_2) = \partial^2_{x,x}f^\star(a/\ell_1, b/\ell_2) + o_{\ell_2 \to 0}(1)$. Similarly, let the second-order time partial derivative operative $\partial^2_{t,t,(\ell_1,\ell_2)}$ be defined for all $(a, b) \in \mathbb{Z}/\ell_1\mathbb{Z} \times \mathbb{Z}/\ell_2\mathbb{Z}$ by

$$(\partial^2_{t,t,(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, b/\ell_2) = \ell_2^2(\hat{f}_n((a+1)/\ell_1, b/\ell_2) - 2\hat{f}_n(a/\ell_1, b/\ell_2) + \hat{f}_n((a-1)/\ell_1, b/\ell_2)).$$

Euler explicit. The Euler explicit scheme is initialized using the Taylor expansion $f(t, x) = f(0, x) + t\partial_t f(0, x) + t^2\partial^2_{t,t}f(0, x)/2 + o_{t \to 0}(t^2)$. With the initial condition $\partial_t f(0, x) = 0$ and the wave equation $\partial^2_{t,t}f(0, x) = 4\partial^2_{x,x}f(0, x)$, this simplifies to $f(t, x) = f(0, x) + 2t^2\partial^2_{x,x}f(0, x) + o_{t \to 0}(t^2)$. This leads to the initialization

$$\forall 0 \leqslant b \leqslant \ell_2, \quad \hat{f}_n(1/\ell_1, b/\ell_2) = \hat{f}_n(0, b/\ell_2) + 2\ell_1^{-2}(\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(0, b/\ell_2).$$

The wave equation $\partial^2_{t,t}f^\star = 4\partial^2_{x,x}f^\star$ can then be discretized as $\partial^2_{t,t,(\ell_1,\ell_2)}\hat{f}_n = 4\Delta_{(\ell_1,\ell_2)}\hat{f}_n$. This leads to the explicit Euler recursive formula

$$\hat{f}_n((a+1)/\ell_1, b/\ell_2) = 2\hat{f}_n(a/\ell_1, b/\ell_2) - \hat{f}_n((a-1)/\ell_1, b/\ell_2) + 4\ell_1^{-2}(\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, b/\ell_2).$$

This formula allows to compute $\hat{f}_n((a+1)/\ell_1, \cdot)$ given the values of $\hat{f}_n(0, \cdot), \ldots, \hat{f}_n(a/\ell_1, \cdot)$.

Runge-Kutta 4. The RK4 scheme is a numerical scheme applied on both $f^\star$ and its derivative $\partial_t f^\star$. Here, $\hat{g}_n$ represents the approximation of $\partial_t f^\star$. The initial condition $\partial_t f(0, \cdot) = 0$ translates into

$$\forall 0 \leqslant b \leqslant \ell_2, \quad \hat{g}_n(0, b/\ell_2) = 0.$$

To infer $\hat{f}_n((a+1)/\ell_1, \cdot)$ and $\hat{g}_n((a+1)/\ell_1, \cdot)$ given the values of $\hat{f}_n(0, \cdot), \ldots, \hat{f}_n(a/\ell_1, \cdot)$ and $\hat{g}_n(0, \cdot), \ldots, \hat{g}_n(a/\ell_1, \cdot)$, the RK4 scheme introduces intermediate estimates as follows:

- $\tilde{f}_1 = \hat{g}_n(a/\ell_1, \cdot)/\ell_1$,

- $\tilde{g}_1 = 4(\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, \cdot)/\ell_1$,

- $\tilde{f}_2 = (\hat{g}_n(a/\ell_1, \cdot) + 0.5\tilde{f}_1)/\ell_1$,

- $\tilde{g}_2 = 4(0.5\tilde{g}_1 + (\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, \cdot))/\ell_1,$

- $\tilde{f}_3 = (\hat{g}_n(a/\ell_1, \cdot) + 0.5\tilde{f}_2)/\ell_1,$

- $\tilde{g}_3 = 4(0.5\tilde{g}_2 + (\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, \cdot))/\ell_1,$

- $\tilde{f}_4 = (\hat{g}_n(a/\ell_1, \cdot) + \tilde{f}_3)/\ell_1,$

- $\tilde{g}_4 = 4(\tilde{g}_3 + (\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, \cdot))/\ell_1,$

- $\hat{f}_n((a+1)/\ell_1, \cdot) = \hat{f}_n(a/\ell_1, \cdot) + (\tilde{f}_1 + 2\tilde{f}_2 + 2\tilde{f}_3 + \tilde{f}_4)/6,$

- $\hat{g}_n((a+1)/\ell_1, \cdot) = \hat{g}_n(a/\ell_1, \cdot) + (\tilde{g}_1 + 2\tilde{g}_2 + 2\tilde{g}_3 + \tilde{g}_4)/6.$

Similarly to the Euler explicit scheme, the RK4 relies on a recursive formulas to compute $\hat{f}_n$.

Crank-Nicolson. The CN scheme is an implicit scheme defined as follows. Similar to the Euler explicit scheme, the $\partial_t f(0, \cdot) = 0$ initial condition is implemented as

$$\forall 0 \leqslant \ell \leqslant \ell_2, \quad \hat{f}_n(1/\ell_1, \ell/\ell_2) = \hat{f}_n(0, \ell/\ell_2) + 2\ell_1^{-2}(\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(0, \ell/\ell_2).$$

Then, the recursive formula of this scheme takes the form

$$\partial_{t,t,(\ell_1,\ell_2)}^2 \hat{f}_n(a/\ell_1, \cdot) = 2((\Delta_{(\ell_1,\ell_2)}\hat{f}_n)(a/\ell_1, \cdot) + (\Delta_{(\ell_1,\ell_2)}\hat{f}_n)((a+1)/\ell_1, \cdot)).$$

This leads to the recursion

$$\hat{f}_n((a+1)/\ell_1, \cdot) = (\mathrm{Id} + 2\ell_2^2\ell_1^{-2}\Delta)^{-1}(3\hat{f}_n(a/\ell_1, \cdot) - \hat{f}_n((a-1)/\ell_1, \cdot)) - \hat{f}_n((a-1)/\ell_1, \cdot),$$

where $\Delta = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}$ is the discrete Laplacian matrix.

D.4 PINN Training

Figures 14 and 15 illustrate the performance of the PINNs during training while solving the 1d wave equation with noisy boundary conditions.
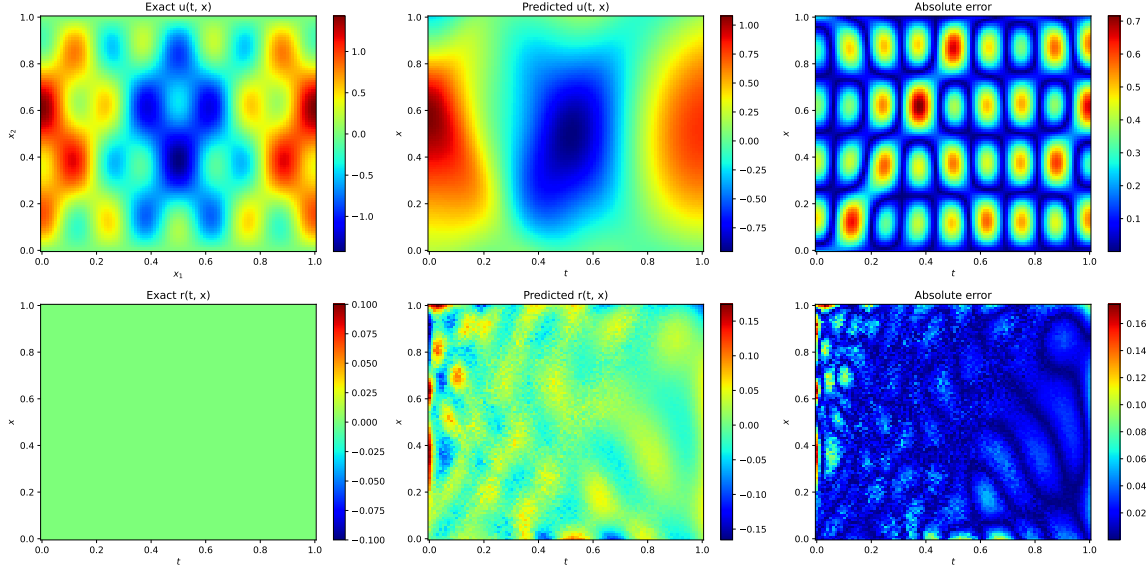
Figure 14: **Left**: Ground truth. **Middle**: PINN estimator. **Right**: Error = PINN - Ground truth.
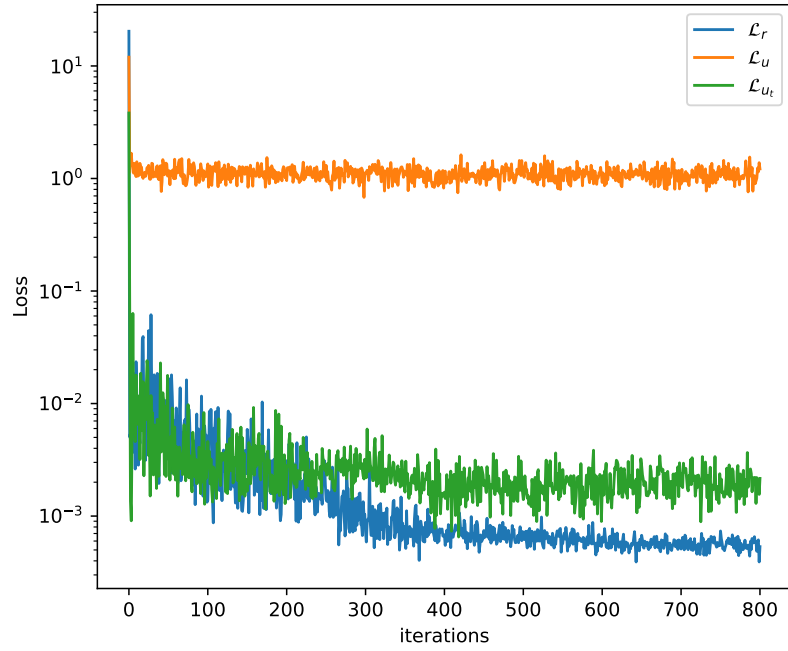

Figure 15: PINN training with noisy boundary conditions.

References

R. Agharafeie, J. R. C. Ramos, J. M. Mendes, and R. Oliveira. From shallow to deep bioprocess hybrid modeling: Advances and future perspectives. Fermentation, 9, 2023.

E. Arnone, A. Kneip, F. Nobile, and L. M. Sangalli. Some first results on the consistency of spatial regression with partial differential equation regularization. Statistica Sinica, 32:209–238, 2022.

A. Arzani, J.-X. Wang, and R. M. D'Souza. Uncovering near-wall blood flow from sparse data with physics-informed neural networks. Physics of Fluids, 33:071905, 2021.

F. Bach. Learning Theory from First Principles. MIT Press, 2024.

L. Badea and P. Daripa. On a boundary control approach to domain embedding methods. SIAM Journal on Control and Optimization, 40:421–449, 2001.

P. Batlle, Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart. Error analysis of kernel/GP methods for nonlinear and parametric PDEs. Journal of Computational Physics, 520:113488, 2025.

A. Bonfanti, G. Bruno, and C. Cipriani. The challenges of the nonlinear regime for physics-informed neural networks. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, Advances in Neural Information Processing Systems, volume 37, pages 41852–41881. Curran Associates, Inc., 2024.

A. Bonito, R. DeVore, G. Petrova, and J. W. Siegel. Convergence and error control of consistent PINNs for elliptic PDEs. IMA Journal of Numerical Analysis, draf008, 2025.

R. Bracewell. The Fourier Transform and its Applications. Electrical Engineering series. McGraw-Hill International Editions, Boston, 3 edition, 2000.

H. Brezis. Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, New York, 2010.

S. Cai, Z. Wang, S. Wang, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks for heat transfer problems. Journal of Heat Transfer, 143:060801, 2021.

C. Canuto, A. Quarteroni, M. Y. Hussaini, and T. A. Zang. Spectral Methods. Scientific Computation. Springer, Berlin, 1 edition, 2007.

A. Caponnetto and E. D. Vito. Optimal rates for the regularized least-squares algorithm. Foundations of Computational Mathematics, 7:331–368, 2007.

Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. Journal of Computational Physics, 447:110668, 2021.

S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli. Scientific machine learning through physics-informed neural networks: Where we are and what's next. Journal of Scientific Computing, 92:88, 2022.

M. Deshpande, S. Agarwal, and A. K. Bhattacharya. Investigations on convergence behaviour of physics informed neural networks across spectral ranges and derivative orders. In 2022 IEEE Symposium Series on Computational Intelligence (SSCI), pages 1172–1179, 2022.

N. Doumèche, F. Bach, G. Biau, and C. Boyer. Physics-informed machine learning as a kernel method. In S. Agrawal and A. Roth, editors, Proceedings of Thirty Seventh Conference on Learning Theory, volume 247 of Proceedings of Machine Learning Research, pages 1399–1450. PMLR, 2024.

N. Doumèche, G. Biau, and C. Boyer. Convergence and error analysis of PINNs. Bernoulli, 31: 2127–2151, 2025.

L. Evans. Partial Differential Equations, volume 19 of Graduate Studies in Mathematics. American Mathematical Society, Providence, 2nd edition, 2010.

G. E. Karniadakis, I. G. Kevrekidis, L. Lu, S. Wang, P. Perdikaris, and L. Yang. Physics-informed machine learning. Nature Reviews Physics, 3:422–440, 2021.

A. S. Krishnapriyan, A. Gholami, S. Zhe, R. M. Kirby, and M. W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, Advances in Neural Information Processing Systems, volume 34, pages 26548–26560. Curran Associates, Inc., 2021.

S. Kurz, H. De Gersem, A. Galetzka, A. Klaedtke, M. Liebsch, D. Loukrezis, S. Russenschuck, and M. Schmidt. Hybrid modeling: Towards the next level of scientific computing in engineering. Journal of Mathematics in Industry, 12, 2022.

S. Lui. Spectral domain embedding for elliptic PDEs in complex domains. Journal of Computational and Applied Mathematics, 225:541–557, 2009.

R. Matthysen and D. Huybrechs. Fast algorithms for the computation of Fourier extensions of arbitrary length. SIAM Journal on Scientific Computing, 38:A899–A922, 2016.

N. McGreivy and A. Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. Nature Machine Intelligence, 6:1256––1269, 2024.

B. Meuris, S. Qadeer, and P. Stinis. Machine-learning-based spectral methods for partial differential equations. Scientific Reports, 13:1739, 2023.

S. Mishra and R. Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. IMA Journal of Numerical Analysis, 43:1–43, 2023.

N. H. Nelsen and A. M. Stuart. Operator learning using random features: A tool for scientific computing. SIAM Review, 66:535–571, 2024.

A. Rahimi and B. Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems, volume 20. Curran Associates, Inc., 2007.

M. Raissi, P. Perdikaris, and G. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378:686–707, 2019.

PHYSICS-INFORMED KERNEL LEARNING

P. Rathore, W. Lei, Z. Frangella, L. Lu, and M. Udell. Challenges in training PINNs: A loss landscape perspective. In Proceedings of the 41st International Conference on Machine Learning, volume 235, pages 42159–42191. PMLR, 2024.

W. Rudin. Functional Analysis. International series in Pure and Applied Mathematics. McGraw-Hill, New-York, 2 edition, 1991.

R. Schaback and H. Wendland. Kernel techniques: From machine learning to meshless methods. Acta Numerica, 15:543–639, 2006.

Y. Shin, J. Darbon, and G. E. Karniadakis. On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type PDEs. Communications in Computational Physics, 28:2042–2074, 2020.

Y. Shin, Z. Zhang, and G. E. Karniadakis. Error estimates of residual minimization using neural networks for linear PDEs. Journal of Machine Learning for Modeling and Computing, 4:73–101, 2023.

A. Tsybakov. Introduction to Nonparametric Estimation. Springer, New York, 2009.

S. Wang, X. Yu, and P. Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. Journal of Computational Physics, 449:110768, 2022a.

Y. Wang, C.-Y. Lai, J. Gómez-Serrano, and T. Buckmaster. Asymptotic self-similar blow-up profile for three-dimensional axisymmetric euler equations using neural networks. Phys. Rev. Lett., 130: 244002, 2023.

Z. Wang, W. Xing, R. Kirby, and S. Zhe. Physics informed deep kernel learning. In Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, volume 151, pages 1206–1218. PMLR, 2022b.

T. Yang, Y.-F. Li, M. Mahdavi, R. Jin, and Z.-H. Zhou. Nyström method vs random Fourier features: A theoretical and empirical comparison. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, Advances in Neural Information Processing Systems, volume 25. Curran Associates, Inc., 2012.