

A Stabilized Physics Informed Neural Networks Method for Wave Equations

Yuling Jiao · Yuhui Liu · Jerry Zhijian Yang · Cheng Yuan

Received: date / Accepted: date

Abstract In this article, we propose a novel Stabilized Physics Informed Neural Networks method (SPINNs) for solving wave equations. In general, this method not only demonstrates theoretical convergence but also exhibits higher efficiency compared to the original PINNs. By replacing the L^2 norm with H^1 norm in the learning of initial condition and boundary condition, we theoretically proved that the error of solution can be upper bounded by the risk in SPINNs. Based on this, we decompose the error of SPINNs into approximation error, statistical error and optimization error. Furthermore, by applying the approximating theory of $ReLU^3$ networks and the learning theory on Rademacher complexity, covering number and pseudo-dimension of neural networks, we present a systematical non-asymptotic convergence analysis on our method, which shows that the error of SPINNs can be well controlled if the number of training samples, depth and width of the deep neural networks have been appropriately chosen. Two illustrative numerical examples on 1-dimensional and 2-dimensional wave equations demonstrate that SPINNs can achieve a faster and better convergence than classical PINNs method.

Keywords PINNs · $ReLU^3$ neural network · Wave equations · Error analysis

Mathematics Subject Classification (2020) 68T07 · 65M12 · 62G05

Yuling Jiao
School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
E-mail: yulingjiaomath@whu.edu.cn

Yuhui Liu
School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
E-mail: liu_yuhui@whu.edu.cn

Jerry Zhijian Yang
School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
E-mail: zjyang.math@whu.edu.cn

Cheng Yuan
School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
E-mail: yuancheng@whu.edu.cn

1 Introduction

During the past few decades, numerical methods of Partial differential equations (PDEs) have been widely studied and applied in various fields of scientific computation [1,2,3,4]. Among these, due to the central significance in solid mechanics, acoustics and electromagnetism, the numerical solution for wave equation attracts considerable attention, and a lot of work has been done to analyze the convergence rate, improve the solving efficiency and deal with practical problems such as boundary conditions. For many real problems with complex region, however, designing an efficient and accurate algorithms with practical absorbing boundary conditions is still difficult, especially for problems with irregular boundary. Furthermore, in high-dimensional case, many traditional methods may become even intractable due to the Curse of Dimensionality, which leads to an exponential increase in degree of freedom with the dimension of problem.

More recently, inspired by the great success of deep learning in fields of natural language processing and computational visions, solving PDEs with deep learning has become as a highly promising topic [5,6,7,8,9,10]. Several numerical schemes have been proposed to solve PDEs using neural networks, including the deep Ritz method (DRM) [11], physics-informed neural networks (PINNs) [12], weak adversarial neural networks(WANs) [13] and their extensions [14,15,16]. Due to the simplicity and flexibility in its formulation, PINNs turns out to be the most concerned method. In the field of wave equations, researchers have successfully applied PINNs to the modeling of scattered acoustic fields [17], including transcranial ultrasound wave [18] and seismic wave [19]. In these works, all of the authors observed an interesting phenomenon that training PINNs without any boundary constraints may lead to a solution under absorbing boundary condition. In another word, the waves obtained by PINNs without boundary loss will be naturally absorbed at the boundary. This phenomenon, in fact, greatly improves the application value of PINNs in wave simulation, especially for inverse scattering problems. On the other hand, although PINNs have been widely used in the simulation of waves, a rigorous numerical analysis of PINNs for wave equations and more efficient training strategy are still needed.

In this work, we propose the Stabilized Physics Informed Neural Networks (SPINNs) for simulation of waves. By replacing the L^2 norm in initial condition and boundary condition with H^1 norm, we obtain a stable PINNs method, in the sense that the error in solution can be upper bounded by the risk during training. It is worth mentioning that, in 2017 a similar idea called Sobolev Training has been proposed to improve the efficiency for regression [20]. Later in [21] and [22], the authors generalized this idea to the training of PINNs, with applications to heat equation, Burgers' equation, Fokker-Planck equation and elasto-plasticity models. One main difference between our model and these works is that, we still use the L^2 norm, rather than H^1 norm for the residual and initial velocity in the loss of SPINNs. This designing, as we will demonstrate, turns out to be a sufficient condition to guarantee the stability,

which also enables us to achieve a lower error with same and even less training samples. Furthermore, based on this stability, we firstly give a systematical convergence rate analysis of PINNs for wave equations. In general, our main contributions are summarized as follows:

Main contributions

- We propose a novel Stabilized Physics Informed Neural Networks method (SPINNs) for solving wave equations, for which we prove that the error in solution can be bounded by the training risk.
- We numerically show that SPINNs can achieve a faster and better convergence than original PINNs.
- We present a systematical convergence analysis of SPINNs. According to our result, once the network depth, width, as well as the number of training samples have been appropriately chosen, the error between the numerical solution u_ϕ from SPINNs and the exact solution u^* can be arbitrarily small in the H^1 norm

$$\mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \|\hat{u}_\phi - u^*\|_{H^1(\overline{\mathcal{D}_T})} \leq \varepsilon.$$

The rest of this paper is organized as follows. In Section 2, we describe the problem setting and introduce the SPINNs method. In Section 3, we study the convergence rate of the SPINNs method for solving wave equations. In Section 4, we present several numerical examples to illustrate the efficiency of SPINNs. Finally in Section 5 the main conclusion will be provided.

2 The Stabilized PINNs method

In this section, we would introduce a stabilized PINNs (SPINNs) method for solving the wave equation. For completeness, we first list the following notations of neural networks and function spaces we will use. After that, the formulation of SPINNs will be presented.

2.1 Preliminary

Let $\mathcal{D} \in \mathbb{N}$, we would call function f a neural network if it is implemented by:

$$\begin{aligned} f_0(x) &= x, \\ f_l(x) &= \rho_l(A_l f_{l-1} + b_l), \quad \text{for } l = 1, \dots, \mathcal{D} - 1, \\ f &:= f_{\mathcal{D}}(x) = A_{\mathcal{D}} f_{\mathcal{D}-1} + b_{\mathcal{D}}, \end{aligned}$$

where $A_l = (a_{ij}^{(l)}) \in \mathbb{R}^{n_l \times n_{l-1}}$, $b_l = (b_i^{(l)}) \in \mathbb{R}^{n_l}$. And $\rho_l : \mathbb{R}^{n_l} \rightarrow \mathbb{R}^{n_l}$, is the active function. The hyper-parameters \mathcal{D} and $\mathcal{W} := \max\{N_l, l = 0, \dots, \mathcal{D}\}$ are called the depth and the width of the network, respectively. Let Φ be a set of activation functions and X be a Banach space, the normed neural network function class can be defined as

$$\mathcal{N}(\mathcal{D}, \mathcal{W}, \{\|\cdot\|_X, \mathcal{B}\}, \Phi) := \{f : f \text{ is implemented by a neural network with}$$

depth \mathcal{D} and width \mathcal{W} , $\|f\|_X \leq \mathcal{B}, \rho_i^{(l)} \in \Phi$ for each i and l .

Next, we introduce several standard function spaces, including the continuous function space and Sobolev space:

$C(\Omega) := \{\text{all the continuous functions defined on } \Omega\},$

$C^s(\Omega) := \{f : \Omega \rightarrow \mathbb{R} \mid D^\alpha f \in C(\Omega), |\alpha| \leq s\},$

$C(\overline{\Omega}) := \{\text{all the continuous functions defined on } \overline{\Omega}\}, \|f\|_{C(\overline{\Omega})} := \max_{x \in \overline{\Omega}} |f(x)|,$

$C^s(\overline{\Omega}) := \{f : \Omega \rightarrow \mathbb{R} \mid D^\alpha f \in C(\overline{\Omega}), |\alpha| \leq s\}, \|f\|_{C^s(\overline{\Omega})} := \max_{x \in \overline{\Omega}, |\alpha| \leq s} |D^\alpha f(x)|,$

$L^p(\Omega) := \left\{ f : \Omega \rightarrow \mathbb{R} \mid \int_{\Omega} |f|^p dx < \infty \right\},$

$\|f\|_{L^p(\Omega)} := \left[\int_{\Omega} |f|^p dx \right]^{1/p}, \forall p \in [1, \infty),$

$L^\infty(\Omega) := \{f : \Omega \rightarrow \mathbb{R} \mid \exists C > 0 \text{ s.t. } |f| \leq C \text{ a.e.}\},$

$\|f\|_{L^\infty(\Omega)} := \inf\{C \mid |f| \leq C \text{ a.e.}\},$

$H^s(\Omega) := \{f : \Omega \rightarrow \mathbb{R} \mid D^\alpha f \in L^2(\Omega), |\alpha| \leq s\},$

$\|f\|_{H^s(\Omega)} := \left(\sum_{|\alpha| \leq s} \|D^\alpha f\|_{L^2(\Omega)}^2 \right)^{1/2}.$

2.2 Stabilized PINNs for wave equations

Considering the following wave equation:

$$\begin{cases} u_{tt} - \Delta u = f, & (x, t) \in \Omega_T, \\ u(x, 0) = \varphi(x), & x \in \Omega, \\ u_t(x, 0) = \psi(x), & x \in \Omega, \\ u(x, t) = g(x, t), & x \in \partial\Omega, t \in [0, T], \end{cases} \quad (1)$$

where $\Delta u = \sum_{i=1}^d u_{x_i x_i}$, $\Omega = (0, 1)^d$ and $\Omega_T = \Omega \times [0, T]$. Through this article, we would assume this problems defines a unique solution:

Assumption 1 Assume (1) has a unique strong solution $u^* \in C^2(\Omega_T) \cap C(\overline{\Omega_T})$.

Without loss of generality, we further assume that f, φ, ψ, g and their derivatives are L^∞ bounded by a constant κ . Denote $\mathcal{B} \triangleq \max\{2\|u^*\|_{C^2(\overline{\Omega_T})}, \kappa^4\}$. Instead of solving problem (1) by traditional numerical methods, we turn to formulate (1) as a minimization problem on $C^2(\Omega_T) \cap C(\overline{\Omega_T})$, with the loss functional \mathcal{L} being defined as:

$$\begin{aligned} \mathcal{L}(u) := & \|u_{tt}(x, t) - \Delta u(x, t) - f\|_{L^2(\Omega_T)}^2 + \|u(x, 0) - \varphi(x)\|_{H^1(\Omega)}^2 \\ & + \|u_t(x, 0) - \psi(x)\|_{L^2(\Omega)}^2 + \|u(x, t) - g(x, t)\|_{H^1(\partial\Omega_T)}^2. \end{aligned} \quad (2)$$

Remark 1 Different from the original loss in PINNs, we adopt H^1 -norm instead of L^2 -norm in the learning of initial position and boundary condition. This modification, as we will demonstrate, offers advantages in both theoretical analysis and numerical computation.

With assumption 1, we know that u^* is also the unique minimizer of loss functional \mathcal{L} such that $\mathcal{L}(u^*) = 0$. Let $|\Omega|$ and $|\partial\Omega|$ be the measure of Ω and $\partial\Omega$, namely, $|\Omega| := \int_{\Omega} 1dx$, $|\partial\Omega| := \int_{\partial\Omega} 1dx$ and $|T| := \int_0^T 1dt$, then $\mathcal{L}(u)$ can be equivalently written as

$$\begin{aligned} \mathcal{L}(u) = & |\Omega||T|\mathbb{E}_{X \in U(\Omega), T \in U([0, T])} \left(u_{tt}(X, T) - \Delta u(X, T) - f(X, T) \right)^2 \\ & + |\Omega|\mathbb{E}_{X \in U(\Omega)} \left[\left(u(X, 0) - \varphi(X) \right)^2 + \sum_{i=1}^d \left(u_{x_i}(X, 0) - \varphi_{x_i}(X) \right)^2 \right. \\ & \left. + \left(u_t(X, 0) - \psi(X) \right)^2 \right] \\ & + |\partial\Omega||T|\mathbb{E}_{Y \in U(\partial\Omega), T \in U([0, T])} \left[\left(u(Y, T) - g(Y, T) \right)^2 \right. \\ & \left. + \left(u_t(Y, T) - g_t(Y, T) \right)^2 + \sum_{i=1}^d \left(u_{x_i}(Y, T) - g_{x_i}(Y, T) \right)^2 \right] \end{aligned} \quad (3)$$

where $U(\Omega)$, $U(\partial\Omega)$, $U([0, T])$ are uniform distribution on Ω , $\partial\Omega$ and $[0, T]$, respectively. To solve minimization of $\mathcal{L}(u)$ approximately, a Monte Carlo discrete version of \mathcal{L} will be used:

$$\begin{aligned} \hat{\mathcal{L}}(u) = & \frac{|\Omega||T|}{NK} \sum_{n=1}^N \sum_{k=1}^K \left(u_{tt}(X_n, T_k) - \Delta u(X_n, T_k) - f(X_n, T_k) \right)^2 \\ & + \frac{|\Omega|}{N} \sum_{n=1}^N \left[\left(u(X_n, 0) - \varphi(X_n) \right)^2 + \sum_{i=1}^d \left(u_{x_i}(X_n, 0) - \varphi_{x_i}(X_n) \right)^2 \right. \\ & \left. + \left(u_t(X_n, 0) - \psi(X_n) \right)^2 \right] \\ & + \frac{|\partial\Omega||T|}{MK} \sum_{m=1}^M \sum_{k=1}^K \left[\left(u(Y_m, T_k) - g(Y_m, T_k) \right)^2 \right. \\ & \left. + \left(u_t(Y_m, T_k) - g_t(Y_m, T_k) \right)^2 + \sum_{i=1}^d \left(u_{x_i}(Y_m, T_k) - g_{x_i}(Y_m, T_k) \right)^2 \right] \end{aligned} \quad (4)$$

where $\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M$ and $\{T_k\}_{k=1}^K$ are independent and identically distributed random samples according to the uniform distribution $U(\Omega)$, $U(\partial\Omega)$ and $U([0, T])$, respectively. With this approximation, we would solve the original problem (1) by using the empirical risk minimization:

$$\hat{u}_\phi = \arg \min_{u_\phi \in \mathcal{P}} \hat{\mathcal{L}}(u_\phi), \quad (5)$$

where the admissible set \mathcal{P} refers to the deep neural network function class parameterized by ϕ . In this work, we will choose \mathcal{P} as the $ReLU^3$ network function space, to ensure $\mathcal{P} \subset C^2(\Omega_T) \cap C(\overline{\Omega_T})$. More precisely,

$$\mathcal{P} = \mathcal{N}(\mathcal{D}, \mathcal{W}, \{\|\cdot\|_{C^2(\overline{\Omega_T})}, \mathcal{B}\}, \{ReLU^3\}),$$

$\{\mathcal{D}, \mathcal{W}\}$ will be given later to ensure the desired accuracy. The $ReLU^3$ activation function is defined by

$$\rho(x) = \begin{cases} x^3, & x \geq 0, \\ 0, & \text{others.} \end{cases}$$

In practical, the minimizer of problem (5) is usually obtained through some optimization algorithm \mathcal{A} . We would denote the minimizer as u_{ϕ_A} .

3 Convergence analysis of SPINNs

In this section, we will present a systematical error analysis of SPINNs for wave equations. To begin with, we first review some basic notations and theorem in the PDEs theory on wave equations.

For wave equation, its total energy consists of two parts: kinetic energy U and potential energy V , both of which can be expressed by multiple integrals,

$$U = \frac{1}{2} \int_{\Omega} u_t^2 dx,$$

$$V = \frac{1}{2} \int_{\Omega} \left(\sum_{i=1}^d u_{x_i}^2 \right) dx,$$

and their sum is called energy integral, the total energy of the wave equation (1) excluding a constant factor is denoted as,

$$E(t) = \int_{\Omega} \left(u_t^2 + \sum_{i=1}^d u_{x_i}^2 \right) dx. \quad (6)$$

Theorem 1 (Energy stability) *We denote $E_0(t) := \int_{\Omega} u^2 dx$, which stands for the square norm estimation of u . We have the energy inequality as below.*

$$E(t) + E_0(t) \leq C(T)(E(0) + E_0(0) + \int_0^T \int_{\Omega} f^2 dx dt) \\ + 2 \int_0^T \int_{\partial\Omega} |u_t| \cdot \|\nabla u\| ds dt).$$

Proof See Appendix 6.1 for details.

3.1 Risk decomposition

By the definition of u^* and u_{ϕ_A} , we can decompose the risk in SPINNs as

$$\begin{aligned}\mathcal{L}(u_{\phi_A}) - \mathcal{L}(u^*) &= \mathcal{L}(u_{\phi_A}) - \hat{\mathcal{L}}(u_{\phi_A}) + \hat{\mathcal{L}}(u_{\phi_A}) - \hat{\mathcal{L}}(\hat{u}_\phi) + \hat{\mathcal{L}}(\hat{u}_\phi) \\ &\quad - \hat{\mathcal{L}}(\bar{u}) + \hat{\mathcal{L}}(\bar{u}) - \mathcal{L}(\bar{u}) + \mathcal{L}(\bar{u}) - \mathcal{L}(u^*) \\ &= \left[\hat{\mathcal{L}}(\hat{u}_\phi) - \hat{\mathcal{L}}(\bar{u}) \right] + \left[\mathcal{L}(u_{\phi_A}) - \hat{\mathcal{L}}(u_{\phi_A}) \right] + \left[\hat{\mathcal{L}}(\bar{u}) - \mathcal{L}(\bar{u}) \right] \\ &\quad + \left[\mathcal{L}(\bar{u}) - \mathcal{L}(u^*) \right] + \left[\hat{\mathcal{L}}(u_{\phi_A}) - \hat{\mathcal{L}}(\hat{u}_\phi) \right],\end{aligned}$$

where \bar{u} is an arbitrarily element in \mathcal{P} . Since $\hat{\mathcal{L}}(\hat{u}_\phi) - \hat{\mathcal{L}}(\bar{u}) \leq 0$, and u is an arbitrarily element in \mathcal{P} , we have:

$$\mathcal{L}(u_{\phi_A}) - \mathcal{L}(u^*) \leq \underbrace{2 \sup_{u \in \mathcal{P}} \left| \mathcal{L}(u) - \hat{\mathcal{L}}(u) \right|}_{\varepsilon_{sta}} + \underbrace{\inf_{u \in \mathcal{P}} \left| \mathcal{L}(u) - \mathcal{L}(u^*) \right|}_{\varepsilon_{app}} + \underbrace{\left[\hat{\mathcal{L}}(u_{\phi_A}) - \hat{\mathcal{L}}(\hat{u}_\phi) \right]}_{\varepsilon_{opt}}$$

Thus, we have decomposed the total risk into approximation error (ε_{app}), statistical error (ε_{sta}) and optimization error (ε_{opt}). While the approximation error describes the expressive power of $ReLU^3$ network, the statistical error is caused by the discretization of the Monte Carlo method and the optimization error represents performance of the solver \mathcal{A} we use. In this work, we compromise assume that the neural network can be well trained such that $\varepsilon_{opt} = 0$, and leave the optimization error as future study. In this case, it can be found that $\hat{u}_\phi = u_{\phi_A}$.

3.2 Lower bound of risk

Next, based on the energy stability of wave equations, we shall present a lower bound of risk $\mathcal{L}(\hat{u}_\phi) - \mathcal{L}(u^*)$ in SPINNs. As we will demonstrate later, the risk can be arbitrary small if the network and sample complexity have been well chosen, and thus we can assume $\mathcal{L}(\hat{u}_\phi) - \mathcal{L}(u^*) < 1$. Let $v = \hat{u}_\phi - u^*$ be the error between numerical solution and exact solution, we have

$$\begin{cases} v_{tt} - \Delta v = (\hat{u}_\phi)_{tt} - \Delta \hat{u}_\phi - f \triangleq \tilde{f}, & x \in \Omega, t \in [0, T], \\ v(x, 0) = \hat{u}_\phi(x, 0) - \varphi(x), & x \in \Omega, \\ v_t(x, 0) = (\hat{u}_\phi)_t(x, 0) - \psi(x), & x \in \Omega, \\ v(x, t) = \hat{u}_\phi(x, 0) - g(x, t), & x \in \partial\Omega, t \in [0, T], \end{cases} \quad (7)$$

and $\|v\|_{C^2(\overline{\Omega_T})} \leq \frac{3}{2}\mathcal{B}$. By applying theorem (1) to equation (7), we obtain

$$\|\hat{u}_\phi - u^*\|_{H_1}^2$$

$$\begin{aligned}
&= \int_{\Omega} \left(v_t^2 + \sum_{i=1}^d v_{x_i}^2 \right) dx + \int_{\Omega} v^2 dx \\
&\leq C(T) \left(E(0) + E_0(0) + \int_0^T \int_{\Omega} \tilde{f}^2 dx dt + 2 \int_0^T \int_{\partial\Omega} |v_t| \cdot \|\nabla v\| ds dt \right) \\
&\leq C(T) \left(\int_{\Omega} (v_t(x, 0)^2 + \sum_{i=1}^d v_{x_i}(x, 0)^2) dx + \int_{\Omega} v(x, 0)^2 dx \right. \\
&\quad \left. + \int_0^T \int_{\Omega} ((\hat{u}_{\phi})_{tt} - \Delta \hat{u}_{\phi} - f)^2 dx dt + 3\sqrt{d}\mathcal{B} \int_0^T \int_{\partial\Omega} |v_t| ds dt \right) \\
&\leq C(T) \left(\int_{\Omega} (v_t(x, 0)^2 + \sum_{i=1}^d v_{x_i}(x, 0)^2) dx + \int_{\Omega} v(x, 0)^2 dx \right. \\
&\quad \left. + \int_0^T \int_{\Omega} ((\hat{u}_{\phi})_{tt} - \Delta \hat{u}_{\phi} - f)^2 dx dt + 3\sqrt{d}\mathcal{B} |\partial\Omega_T| \left(\int_0^T \int_{\partial\Omega} v_t^2 ds dt \right)^{\frac{1}{2}} \right) \\
&\leq C(T) \left(\|(\hat{u}_{\phi})_{tt} - \Delta \hat{u}_{\phi} - f\|_{L^2(\Omega_T)}^2 + \|(\hat{u}_{\phi})_t(x, 0) - \psi(x)\|_{L^2(\Omega)}^2 \right. \\
&\quad \left. + \|\hat{u}_{\phi}(x, 0) - \varphi(x)\|_{H^1(\Omega)}^2 + 3\sqrt{d}\mathcal{B} |\partial\Omega_T| \|\hat{u}_{\phi}(x, t) - g(x, t)\|_{H^1(\partial\Omega_T)} \right) \\
&\leq C(T) \left(\mathcal{L}(\hat{u}_{\phi}) + 3\sqrt{d}\mathcal{B} |\partial\Omega_T| \mathcal{L}(\hat{u}_{\phi})^{\frac{1}{2}} \right) \\
&\leq C(T) (1 + 3\sqrt{d}\mathcal{B} |\partial\Omega_T|) \mathcal{L}(\hat{u}_{\phi})^{\frac{1}{2}} \quad (\mathcal{L}(\hat{u}_{\phi}) < 1) \\
&\leq \gamma \left(\mathcal{L}(\hat{u}_{\phi}) - \mathcal{L}(u^*) \right)^{\frac{1}{2}}, \quad (\mathcal{L}(u^*) = 0)
\end{aligned}$$

where we define $\gamma(T, d, \mathcal{B}, |\partial\Omega_T|) \triangleq C(T)(1 + 3\sqrt{d}\mathcal{B} |\partial\Omega_T|)$. Combined this lower bound with previous risk decomposition, we can arrive at:

$$\|\hat{u}_{\phi} - u^*\|_{H_1}^4 \leq \gamma^2(\varepsilon_{app} + \varepsilon_{sta}). \quad (8)$$

3.3 Approximation error

By applying the following lemma proved in our previous work, we can get the upper bound of approximation error:

Lemma 1 $\forall \bar{u} \in C^3(\overline{\Omega_T})$ and $\varepsilon > 0$, there exist a ReLU³ network u_{ϕ} with depth $\lceil \log_2 d \rceil + 2$ and width $C(d, \|\bar{u}\|_{C^3(\overline{\Omega_T})}) (\frac{1}{\varepsilon})^{d+1}$ such that

$$\|\bar{u} - u_{\phi}\|_{C^2(\overline{\Omega_T})} \leq \varepsilon.$$

Proof A special case of Corollary 4.2 in [23].

Theorem 2 *Under the Assumption 1 and the condition that $u^* \in C^3(\overline{\Omega_T})$, for any $\varepsilon > 0$, if we choose the following neural network function class:*

$$\mathcal{P} = \mathcal{N}([\log_2 d] + 2, C(d, |\Omega|, |\partial\Omega|, |T|, \|u^*\|_{C^3(\overline{\Omega_T})})\left(\frac{1}{\varepsilon}\right)^{d+1}, \{\|\cdot\|_{C^2(\overline{\Omega_T})}, 2\|u^*\|_{C^2(\overline{\Omega_T})}\}, \{ReLU^3\}),$$

then the approximation error $\varepsilon_{app} \leq C(d, |\Omega_T|, \partial\Omega_T)\varepsilon^2$.

Proof See Appendix 6.2 for details.

3.4 Statistical error

The following theorem demonstrates that with sufficiently large sample complexity, the statistical error can be well controlled:

Theorem 3 *Let $\mathcal{D}, \mathcal{W} \in \mathbb{N}, \mathcal{B} \in \mathbb{R}^+$. For any $\varepsilon \geq 0$, if the number of samples satisfy:*

$$\begin{cases} N = C(d, |\Omega|, \mathcal{B})\mathcal{D}^4\mathcal{W}^2(\mathcal{D} + \log(\mathcal{W}))\left(\frac{1}{\varepsilon}\right)^{2+\delta}, \\ K = C(d, |T|, \mathcal{B})\mathcal{D}^2f_K(\mathcal{D}, \mathcal{W})\left(\frac{1}{\varepsilon}\right)^{k_1}, \\ M = C(d, |\partial\Omega|, \mathcal{B})f_M(\mathcal{D}, \mathcal{W})\left(\frac{1}{\varepsilon}\right)^{k_2}, \end{cases}$$

where $f_K(\mathcal{D}, \mathcal{W}) \geq 1$, $f_M(\mathcal{D}, \mathcal{W}) \geq 1$, δ is an arbitrarily small number such that

$$\begin{cases} k_1 + k_2 = 2 + \delta, \\ f_k(\mathcal{D}, \mathcal{W}) \cdot f_M(\mathcal{D}, \mathcal{W}) = \mathcal{D}^2\mathcal{W}^2(\mathcal{D} + \log(\mathcal{W})), \end{cases}$$

then we have:

$$\mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \sup_{u \in \mathcal{P}} |\mathcal{L}(u) - \hat{\mathcal{L}}(u)| \leq \varepsilon$$

Proof See Appendix 6.3 for details.

3.5 Convergence rate of SPINNs

With the preparation in last two sections on the bounds of approximation and statistical errors, we will give the main results in this section.

Theorem 4 *Under the Assumption 1 and the condition that $u^* \in C^3(\overline{\Omega_T})$. For any $\varepsilon > 0$, if we choose the parameterized neural network class*

$$\mathcal{P} = \mathcal{N}([\log_2 d] + 2, C(d, |\Omega|, |\partial\Omega|, |T|, \|u^*\|_{C^3(\overline{\Omega_T})})\left(\frac{1}{\varepsilon^2}\right)^{d+1}, \{\|\cdot\|_{C^2(\overline{\Omega_T})}, 2\|u^*\|_{C^2(\overline{\Omega_T})}\}, \{ReLU^3\})$$

and let the number of samples be

$$\begin{cases} N = C(d, |\Omega|, \mathcal{B})\left(\frac{1}{\varepsilon^4}\right)^{d+3+\delta}, \\ K = C(d, |T|, \mathcal{B})\left(\frac{1}{\varepsilon^4}\right)^{k_1+\tilde{k}_1}, \\ M = C(d, |\partial\Omega|, \mathcal{B})\left(\frac{1}{\varepsilon^4}\right)^{k_2+\tilde{k}_2}, \end{cases}$$

where $\tilde{k}_1, \tilde{k}_2 \geq 0$, δ is arbitrarily small such that

$$\begin{cases} k_1 + k_2 = 2 + \delta, \\ \tilde{k}_1 + \tilde{k}_2 = d + 1, \end{cases}$$

then we have:

$$\mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \|\hat{u}_\phi - u^*\|_{H^1(\overline{\Omega_T})} \leq \varepsilon$$

Proof By theorem 2, if we set the neural network function class as:

$$\begin{aligned} \mathcal{P} = \mathcal{N}([\log_2 d] + 2, C(d, |\Omega|, |\partial\Omega|, |T|, \|u^*\|_{C^3(\overline{\Omega_T})})\left(\frac{1}{\varepsilon^2}\right)^{d+1}, \\ \{\|\cdot\|_{C^2(\overline{\Omega_T})}, 2\|u^*\|_{C^2(\overline{\Omega_T})}, \{ReLU^3\}\}) \end{aligned} \quad (9)$$

the approximation error can be arbitrarily small:

$$\varepsilon_{app} \leq \frac{\varepsilon^4}{2\gamma^2} \quad (10)$$

Without loss of generality we assume that ε is small enough such that

$$\|\hat{u}_\phi\|_{C^2(\overline{\Omega_T})} \leq \|u^* - \hat{u}_\phi\|_{C^2(\overline{\Omega_T})} + \|u^*\|_{C^2(\overline{\Omega_T})} \leq 2\|u^*\|_{C^2(\overline{\Omega_T})}.$$

By theorem 3, when the number of samples be:

$$\begin{cases} N = C(d, |\Omega|, \mathcal{B})\left(\frac{1}{\varepsilon^4}\right)^{d+3+\delta}, \\ K = C(d, |T|, \mathcal{B})\left(\frac{1}{\varepsilon^4}\right)^{k_1+\tilde{k}_1}, \\ M = C(d, |\partial\Omega|, \mathcal{B})\left(\frac{1}{\varepsilon^4}\right)^{k_2+\tilde{k}_2}, \end{cases} \quad (11)$$

where δ is an arbitrarily positive number and

$$\begin{cases} k_1 + k_2 = 2 + \delta, \\ \tilde{k}_1 + \tilde{k}_2 = d + 1, \end{cases} \quad (12)$$

we have:

$$\mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \varepsilon_{sta} \leq \frac{\varepsilon^4}{2\gamma^2} \quad (13)$$

Combining (8), (10) and (13) together, we get the final result:

$$\begin{aligned} & \mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \|\hat{u}_\phi - u^*\|_{H^1(\overline{\Omega_T})} \\ & \leq \mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} [\gamma^2 (\mathcal{L}(\hat{u}_\phi) - \mathcal{L}(u^*))]^{1/4} \\ & \leq \varepsilon. \end{aligned} \quad (14)$$

4 Numerical Experiments

In this section, we would use SPINNs to solve wave equation in both one dimension and two dimension.

4.1 1D example

Consider the following 1D wave equation on $\Omega = [-2, 2]$ from $t = 0$ to $t = T = 8$:

$$\begin{cases} u_{tt} = u_{xx}, & x \in \Omega, \quad 0 \leq t \leq 8, \\ u(0, x) = 0, & x \in \Omega, \\ u_t(0, x) = 0, & x \in \Omega, \\ u(t, -2) = \sin(0.8\pi t), & 0 \leq t \leq 8, \\ u(t, 2) = 0, & 0 \leq t \leq 8. \end{cases} \quad (15)$$

For the training with SPINNs, we use a four-layer *ReLU*³ network with 64 neurons in each layer to approximate the solution. We choose the Adam algorithm to implement the minimization, and the initial learning rate is set as 1E-3.

As for sample complexity, we train SPINNs with 10000 interior points, 500 boundary points (250 for each end) and 250 initial points in each epoch, all of which are sampled according to a uniform distribution (see (a) in Figure 1). Further more, to obtain a better accuracy, we apply GAS method [24] as an adaptive sampling strategy. After every 250 epochs, we would adaptively add 600 inner points, 30 boundary points and 15 initial points based on a Gaussian mixture distribution. The GAS procedure will be repeated for 10 times. See [24] for more details. For evaluation, we use the central difference method with a fine mesh ($dx = 0.01$, $dt = 0.009$) to obtain a reference solution u_{cdm} (see (b) in Figure 1), with which we can calculate the following relative error by using numerical integration:

$$\text{Relative error} = \frac{\|u_\phi - u_{cdm}\|_2}{\|u_{cdm}\|_2}.$$

Figure 2 demonstrates the numerical result of PINNs and SPINNs for (15) after N_G times of adaptive sampling by GAS. As we can see, the SPINNs method converges faster than the classical PINNs, e.g., after 5 times of adaptive sampling, SPINNs have already captured all the six peaks of standing waves generated by the superposition of reflected wave and right-traveling waves. Furthermore, we present the relative error of PINNs and SPINNs in Figure 3, which shows that our method can achieve a lower relative error at the early stage of training. On the other hand, with the times of adaptive sampling increasing, the classical PINNs can also arrive at a comparable accuracy, which has also been revealed by (g) in Figure 2. These results reflect the fact that training with SPINNs can speed up the convergence in solution, especially when the number of samples is relatively small.

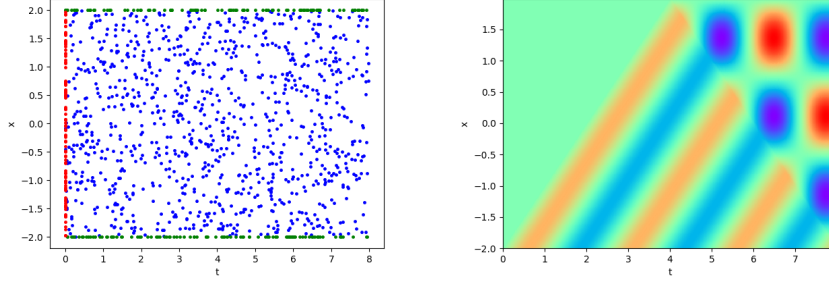


Fig. 1 Collection points (left) and reference solution (right) in $[0, T] \times \Omega$. The red, green and blue points stand for the initial points, boundary points and interior points correspondingly.

4.2 2D example

Consider the following 2D wave equation on $\Omega = [-2, 2]^2$ from $t = 0$ to $t = 1$:

$$\begin{cases} u_{tt} = u_{xx} + u_{yy}, & x \in \Omega, \quad 0 \leq t \leq 1, \\ u(0, x, y) = \sin(2\pi\sqrt{x^2 + y^2}), & \sqrt{x^2 + y^2} < 1, \\ u(0, x, y) = 0, & \sqrt{x^2 + y^2} \geq 1, \\ u_t(0, x, y) = 0, & (x, y) \in \Omega, \\ u(t, x, y) = 0, & (x, y) \in \partial\Omega, 0 \leq t \leq 1, \end{cases} \quad (16)$$

For the training with SPINNs, based on the experiment, we choose a three-layer *ReLU*³ network with 512 neurons in each layer to approximate the solution. The optimization algorithm and the initial learning rate are kept as before.

As for sample complexity, we train SPINNs with 2000 interior points, 4000 boundary points (1000 for each edge) and 1000 initial points in each epoch, all of which are sampled according to a uniform distribution. For evaluation, we use the central difference method with a fine mesh ($dx = dy = 0.01$, $dt = 0.004$) to obtain a reference solution u_{cdm} , with which we can calculate the pointwise absolute error $|u_{cdm} - u_\phi|$. As we can observe from Figure 4 and Figure 5, the SPINNs method can achieve a lower pointwise error, after training for same epochs. This superiority, in fact, can be understood as that learning with derivative information improves accuracy in the fitting of initial condition.

5 Conclusion

In this work, we propose a stabilized physics informed neural networks method SPINNs for wave equations. With some numerical analysis, we rigorously prove SPINNs is a stable learning method, in which the solution error can be well

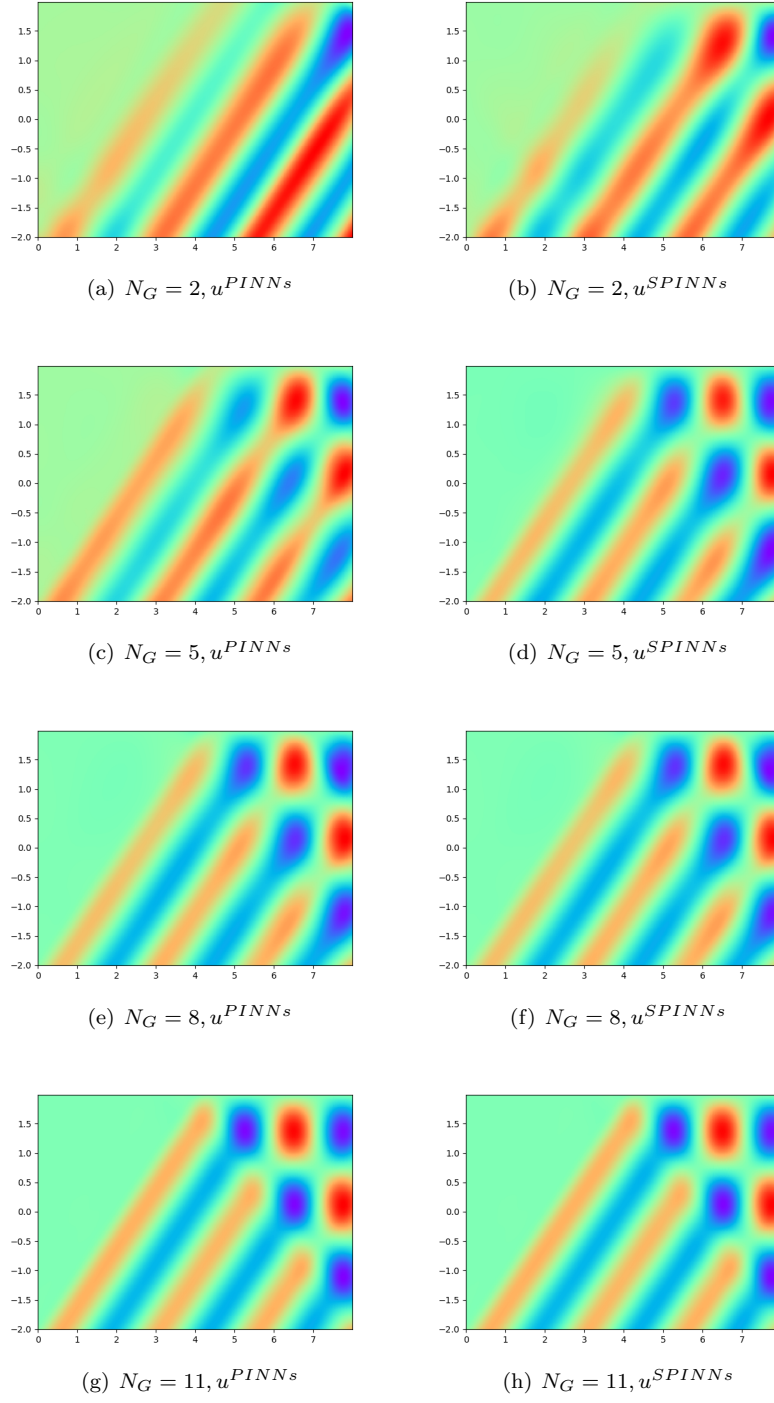


Fig. 2 The numerical result of PINNs and SPINNs for equation (15).

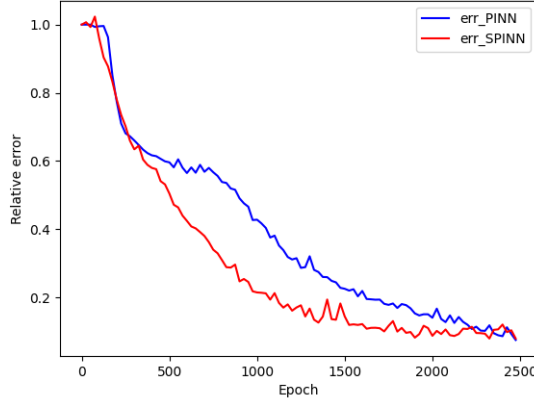


Fig. 3 The relative error of PINNs and SPINNs.

controlled by the loss term. Based on this, a non-asymptotic convergence rate of SPINNs is presented, which provide people with a solid theoretical foundation to use it. Furthermore, by applying SPINNs to the simulation of two wave propagation problems, we numerically demonstrate that SPINNs can achieve a higher training accuracy and efficiency, especially when the number of samples is limited. On the other hand, how to extend this method to more difficult situations such as high dimensional problems and how to handel the optimization error in our convergence analysis are still needed to be studied. We will leave these topics as our future research.

Acknowledgement

This work is supported by the National Key Research and Development Program of China (No.2020YFA0714200), by the National Nature Science Foundation of China (No.12371441, No.12301558, No.12125103, No.12071362), and by the Fundamental Research Funds for the Central Universities.

6 Appendix

6.1 Appendix for energy integral of wave equations

According to the Gaussian formula, we have

$$\begin{aligned}
 \int_{\Omega} \left(\sum_{i=1}^d (u_{tx_i} u_{x_i}) + u_t \left(\sum_{i=1}^d u_{x_i x_i} \right) \right) dx &= \int_{\Omega} (\nabla \cdot (u_t \nabla u)) dx \\
 &= \int_{\partial\Omega} u_t \nabla u \cdot \mathbf{n} ds, \quad (17)
 \end{aligned}$$

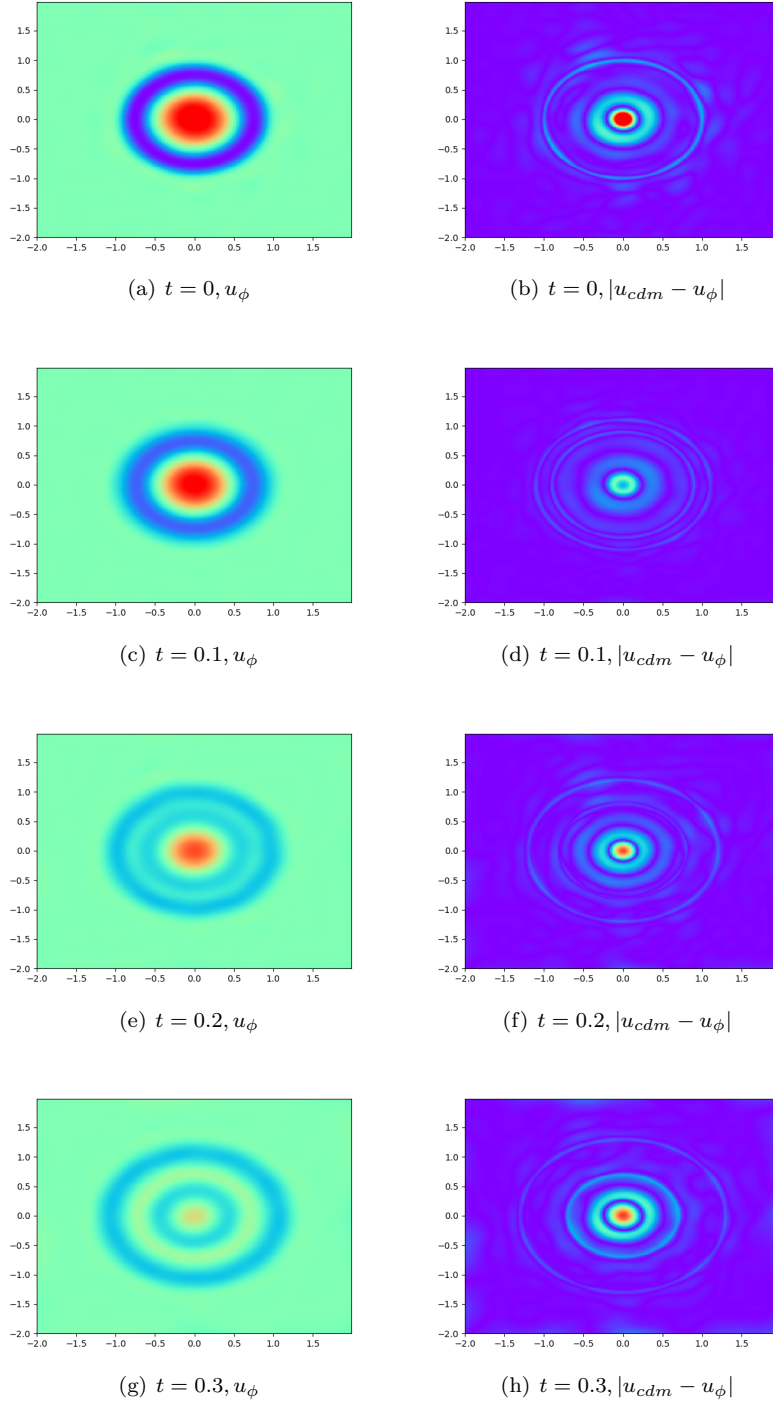


Fig. 4 The numerical result for (16) after 10000 epochs training of PINNs.

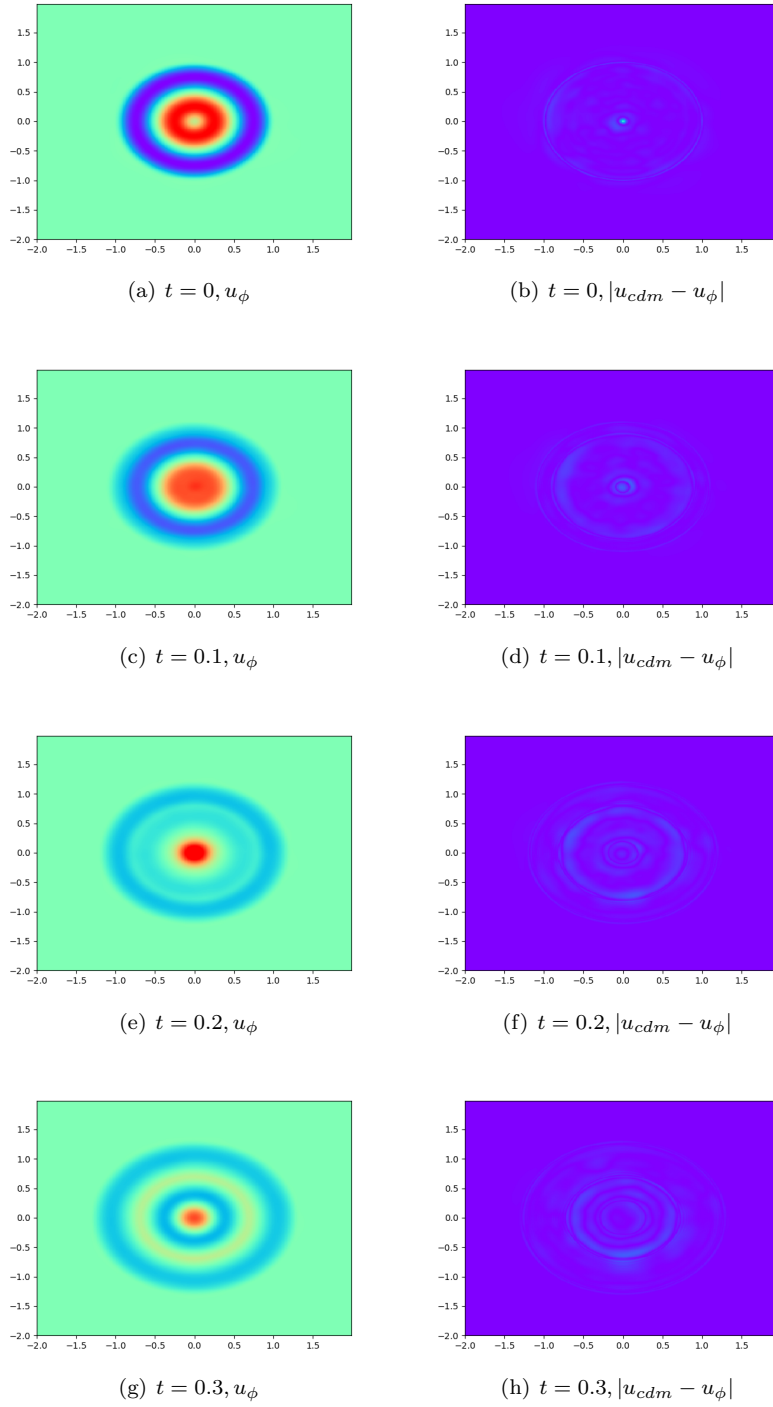


Fig. 5 The numerical result for (16) after 10000 epochs training of SPINNs.

where \mathbf{n} is the unit outer normal vector. Combine (6) and (17), we have

$$\begin{aligned}
\frac{dE(t)}{dt} &= 2 \int_{\Omega} \left(u_t u_{tt} + \sum_{i=1}^d u_{x_i} u_{x_i t} \right) dx \\
&= 2 \int_{\Omega} \left(u_t u_{tt} - u_t \sum_{i=1}^d u_{x_i x_i} \right) dx + 2 \int_{\partial\Omega} u_t \nabla u \cdot \mathbf{n} ds \\
&= 2 \int_{\Omega} u_t f dx + 2 \int_{\partial\Omega} u_t \nabla u \cdot \mathbf{n} ds \\
&\leq \int_{\Omega} (u_t^2 + f^2) dx + 2 \int_{\partial\Omega} |u_t| \cdot \|\nabla u\| ds.
\end{aligned} \tag{18}$$

Multiply both sides of the inequality (18) by e^{-t} ,

$$\frac{d(e^{-t}E(t))}{dt} \leq e^{-t} \left(\int_{\Omega} f^2 dx + 2 \int_{\partial\Omega} |u_t| \cdot \|\nabla u\| ds \right). \tag{19}$$

Then, integrating the equation (19) from 0 to t ,

$$E(t) \leq e^t \left(E(0) + \int_0^t e^{-\tau} \int_{\Omega} f^2 dx d\tau + 2 \int_0^t e^{-\tau} \int_{\partial\Omega} |u_t| \cdot \|\nabla u\| ds d\tau \right)$$

For any $t \in [0, T]$,

$$E(t) \leq C_1 \left(E(0) + \int_0^T \int_{\Omega} f^2 dx dt + 2 \int_0^T \int_{\partial\Omega} |u_t| \cdot \|\nabla u\| ds dt \right), \tag{20}$$

C_1 is a constant that is only related to T . Further we have

$$\frac{dE_0(t)}{dt} = \int_{\Omega} 2u u_t dx \leq \int_{\Omega} u^2 dx + \int_{\Omega} u_t^2 dx \leq E_0(t) + E(t) \tag{21}$$

Multiply both sides of the above equation (21) by e^{-t} ,

$$\frac{d}{dt}(e^{-t}E_0(t)) \leq e^{-t}E(t) \tag{22}$$

Integrating the equation (22) from 0 to t ,

$$E_0(t) \leq e^t E_0(0) + e^t \int_0^t e^{-\tau} E(\tau) d\tau \tag{23}$$

For any $t \in [0, T]$,

$$E_0(t) \leq C_2 (E_0(0) + E(t)), \tag{24}$$

C_2 is a constant that is only related to T . Combine (20) and (24), we have,

$$E(t) + E_0(t) \leq C \left(E(0) + E_0(0) + \int_0^T \int_{\Omega} f^2 dx dt \right. \tag{25}$$

$$\left. + 2 \int_0^T \int_{\partial\Omega} |u_t| \cdot \|\nabla u\| ds dt \right). \tag{26}$$

C is a constant that is only related to T .

6.2 Appendix for approximation error

[Proof of theorem 2] According to lemma 1, we know that for $u^* \in C^3(\overline{\Omega_T})$, $\varepsilon > 0$, there exist a $ReLU^3$ network with depth $\lceil \log_2 d \rceil + 2$ and width $C(d, \|u^*\|_{C^3(\overline{\Omega_T})})(\frac{1}{\varepsilon})^{d+1}$, such that $\|v(x, t)\|_{C^2(\overline{\Omega_T})} = \|u^* - \hat{u}_\phi\|_{C^2(\overline{\Omega_T})} \leq \varepsilon$. Hence,

$$\begin{aligned} \varepsilon_{app} &\leq |\mathcal{L}(\hat{u}_\phi) - \mathcal{L}(u^*)| \\ &= \|(\hat{u}_\phi)_{tt}(x, t) - \Delta \hat{u}_\phi(x, t) - f\|_{L^2(\Omega_T)}^2 + \|\hat{u}_\phi(x, 0) - \varphi(x)\|_{H^1(\Omega)}^2 \\ &\quad + \|(\hat{u}_\phi)_t(x, 0) - \psi(x)\|_{L^2(\Omega)}^2 + \|\hat{u}_\phi(x, t) - g(x, t)\|_{H^1(\partial\Omega_T)}^2 \\ &\leq \|v_{tt}\|_{L^2(\Omega_T)}^2 + \|\Delta v\|_{L^2(\Omega_T)}^2 + \|v(x, 0)\|_{H^1(\Omega)}^2 + \|v_t(x, 0)\|_{L^2(\Omega)}^2 + \|v\|_{H^1(\partial\Omega_T)}^2 \\ &\leq C(d, |\Omega_T|, |\partial\Omega_T|) \cdot \varepsilon^2. \end{aligned}$$

6.3 Appendix for statistical error

We will give the precise computation on the upper bounds of statistical error in this section. To begin with, we first introduce several basic concepts and results in learning theory.

Definition 1 (Rademacher complexity) The Rademacher complexity of a set $A \subseteq \mathbb{R}^N$ is defined by

$$\mathfrak{R}(A) = \mathbb{E}_{\{\sigma_k\}_{k=1}^N} \left[\sup_{a \in A} \frac{1}{N} \sum_{k=1}^N \sigma_k a_k \right],$$

where $\{\sigma_k\}_{k=1}^N$ are N i.i.d Rademacher variables with $p(\sigma_k = 1) = p(\sigma_k = -1) = \frac{1}{2}$.

Let Ω be a set and \mathcal{F} be a function class which maps Ω to \mathbb{R} . Let P be a probability distribution over Ω and $\{X_k\}_{k=1}^N$ be i.i.d. samples from P . The Rademacher complexity of \mathcal{F} associated with distribution P and sample size N is defined by

$$\mathfrak{R}_{P,N}(\mathcal{F}) = \mathbb{E}_{\{X_k, \sigma_k\}_{k=1}^N} \left[\sup_{u \in \mathcal{F}} \frac{1}{N} \sum_{k=1}^N \sigma_k u(X_k) \right].$$

Lemma 2 Let Ω be a set and P be a probability distribution over Ω . Let $N \in \mathbb{N}$. Assume that $\omega : \Omega \rightarrow \mathbb{R}$ and $|\omega(x)| \leq \mathcal{B}$ for all $x \in \Omega$, then for any function class \mathcal{F} mapping Ω to \mathbb{R} , there holds

$$\mathfrak{R}_{P,N}(\omega(x)\mathcal{F}) \leq \mathcal{B} \mathfrak{R}_{P,N}(\mathcal{F}),$$

where $\omega(x)\mathcal{F} := \{\bar{u} : \bar{u}(x) = \omega(x)u(x), u \in \mathcal{F}\}$.

Proof See [25] for the proof.

Definition 2 (Covering number) Suppose that $W \subset \mathbb{R}$. For any $\varepsilon > 0$, let $V \subset \mathbb{R}^n$ be an ε -cover of W with respect to the distance d_∞ , that is, for any $\omega \in W$, there exists a $v \in V$ such that $d_\infty(\omega, v) < \varepsilon$, where d_∞ is defined by $d_\infty(u, v) := \max_{1 \leq i \leq n} |u_i - v_i|$. The covering number $\mathcal{C}(\varepsilon, W, d_\infty)$ is defined to be the minimum cardinality among all ε -cover of W with respect to the distance d_∞ .

Definition 3 (Uniform covering number) Suppose that \mathcal{F} is a class of functions from Ω to \mathbb{R} . Given n sample $Z_n = (Z_1, Z_2, \dots, Z_n) \in \Omega^n$, $\mathcal{F}|_{Z_n} \subset \mathbb{R}^n$ is defined by

$$\mathcal{F}|_{Z_n} \subset \mathbb{R}^n = \{(u(Z_1), u(Z_2), \dots, u(Z_n)) : u \in \mathcal{F}\}.$$

The uniform covering number $\mathcal{C}_\infty(\varepsilon, \mathcal{F}, n)$ is defined by

$$\mathcal{C}_\infty(\varepsilon, \mathcal{F}, n) = \max_{Z_n \in \Omega^n} \mathcal{C}(\varepsilon, \mathcal{F}|_{Z_n}, d_\infty).$$

Lemma 3 Let Ω be a set and P be a probability distribution over Ω . Let $N \in \mathbb{N}_{\geq 1}$, and \mathcal{F} be a class of functions from Ω to \mathbb{R} such that $0 \in \mathcal{F}$ and the diameter of \mathcal{F} is less than \mathcal{B} , i.e., $\|u\|_{L^\infty(\Omega)} \leq \mathcal{B}, \forall u \in \mathcal{F}$. Then

$$\mathfrak{R}_{P,N}(\mathcal{F}) \leq \inf_{0 < \delta < \mathcal{B}} \left(4\delta + \frac{12}{\sqrt{N}} \int_\delta^{\mathcal{B}} \sqrt{\log(2\mathcal{C}_\infty(\varepsilon, \mathcal{F}, N))} d\varepsilon \right).$$

Proof This proof is base on the chaining method, see [26].

Definition 4 (Pseudo-dimension) Let \mathcal{F} be a class of functions from X to \mathbb{R} . Suppose that $S = \{x_1, x_2, \dots, x_n\} \subset X$. We say that S is pseudo-shattered by \mathcal{F} if there exists y_1, y_2, \dots, y_n such that for any $b \in \{0, 1\}^n$, there exists a $u \in \mathcal{F}$ satisfying

$$\text{sign}(u(x_i) - y_i) = b_i, i = 1, 2, \dots, n,$$

and we say that $\{y_i\}_{i=1}^n$ witnesses the shattering. The pseudo-dimension of \mathcal{F} , denoted as $Pdim(\mathcal{F})$, is defined to be the maximum cardinality among all sets pseudo-shattered by \mathcal{F} .

Lemma 4 (Theorem 12.2 in [27]) Let \mathcal{F} be a class of real functions from a domain X to the bounded interval $[0, \mathcal{B}]$. Let $\varepsilon > 0$. Then

$$\mathcal{C}_\infty(\varepsilon, \mathcal{F}, n) \leq \sum_{i=1}^{Pdim(\mathcal{F})} \binom{n}{i} \left(\frac{\mathcal{B}}{\varepsilon}\right)^i,$$

which is less than $(\frac{en\mathcal{B}}{\varepsilon \cdot Pdim(\mathcal{F})})^{Pdim(\mathcal{F})}$ for $n \geq Pdim(\mathcal{F})$.

Next, to obtain the upper bound, we would decompose the statistical error into 24 terms by using triangle inequality:

$$\begin{aligned} & \mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \sup_{u \in \mathcal{P}} |\mathcal{L}(u) - \hat{\mathcal{L}}(u)| \leq \\ & \sum_{j=1}^{24} \mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \sup_{u \in \mathcal{P}} |\mathcal{L}_j(u) - \hat{\mathcal{L}}_j(u)| \end{aligned}$$

where

$$\begin{aligned} \mathcal{L}_1 &= |\Omega| |T| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} (u_{tt}(X, T))^2, \\ \mathcal{L}_2 &= |\Omega| |T| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} \left(\sum_{i=1}^d u_{x_i x_i}(X, T) \right)^2, \\ \mathcal{L}_3 &= |\Omega| |T| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} (f(X, T))^2, \\ \mathcal{L}_4 &= -2|\Omega| |T| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} \left(\sum_{i=1}^d u_{tt}(X, T) u_{x_i x_i}(X, T) \right), \\ \mathcal{L}_5 &= -2|\Omega| |T| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} (u_{tt}(X, T) f(X, T)), \\ \mathcal{L}_6 &= 2|\Omega| |T| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} \left(\sum_{i=1}^d u_{x_i x_i}(X, T) f(X, T) \right), \\ \mathcal{L}_7 &= |\Omega| \mathbb{E}_{X \sim U(\Omega)} (u(X, 0))^2, \\ \mathcal{L}_8 &= |\Omega| \mathbb{E}_{X \sim U(\Omega)} (\varphi(X))^2, \\ \mathcal{L}_9 &= -2|\Omega| \mathbb{E}_{X \sim U(\Omega)} (u(X, 0) \varphi(X)), \\ \mathcal{L}_{10} &= |\Omega| \mathbb{E}_{X \sim U(\Omega)} \left(\sum_{i=1}^d u_{x_i}(X, 0)^2 \right), \\ \mathcal{L}_{11} &= |\Omega| \mathbb{E}_{X \sim U(\Omega)} \left(\sum_{i=1}^d \varphi_{x_i}(X)^2 \right), \\ \mathcal{L}_{12} &= -2|\Omega| \mathbb{E}_{X \sim U(\Omega)} \left(\sum_{i=1}^d u_{x_i}(X, 0) \varphi_{x_i}(X) \right), \\ \mathcal{L}_{13} &= |\Omega| \mathbb{E}_{X \sim U(\Omega)} (u_t(X, 0))^2, \\ \mathcal{L}_{14} &= |\Omega| \mathbb{E}_{X \sim U(\Omega)} (\psi(X))^2, \\ \mathcal{L}_{15} &= -2|\Omega| \mathbb{E}_{X \sim U(\Omega)} (u_t(X, 0) \psi(X)), \\ \mathcal{L}_{16} &= |\partial\Omega| |T| \mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} (u(Y, T))^2, \\ \mathcal{L}_{17} &= |\partial\Omega| |T| \mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} (g(Y, T))^2, \\ \mathcal{L}_{18} &= -2|\partial\Omega| |T| \mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} (u(Y, T) g(Y, T)), \\ \mathcal{L}_{19} &= |\partial\Omega| |T| \mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} (u_t(Y, T))^2, \\ \mathcal{L}_{20} &= |\partial\Omega| |T| \mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} (g_t(Y, T))^2, \end{aligned}$$

$$\begin{aligned}
\mathcal{L}_{21} &= -2|\partial\Omega||T|\mathbb{E}_{Y\sim U(\partial\Omega), T\sim U([0,T])}(u_t(Y,T)g_t(Y,T)), \\
\mathcal{L}_{22} &= |\partial\Omega||T|\mathbb{E}_{Y\sim U(\partial\Omega), T\sim U([0,T])}\sum_{i=1}^d(u_{x_i}(Y,T)^2), \\
\mathcal{L}_{23} &= |\partial\Omega||T|\mathbb{E}_{Y\sim U(\partial\Omega), T\sim U([0,T])}\sum_{i=1}^d(g_{x_i}(Y,T)^2), \\
\mathcal{L}_{24} &= -2|\partial\Omega||T|\mathbb{E}_{Y\sim U(\partial\Omega), T\sim U([0,T])}\sum_{i=1}^d(u_{x_i}x_i(Y,T)g_{x_i}(X,T)),
\end{aligned}$$

and $\hat{\mathcal{L}}_j(u)$ is the empirical version of $\mathcal{L}_j(u)$. The following lemma states that each of these 24 terms can be controlled by the corresponding Rademacher complexity.

Lemma 5 *Let $\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K$ be i.i.d samples from $U(\Omega), U(\partial\Omega), U([0, T])$, then we have*

$$\mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \sup_{u \in \mathcal{P}} \left| \mathcal{L}_j(u) - \hat{\mathcal{L}}_j(u) \right| \leq C(d, \mathcal{B}) \mathfrak{R}_{\mathcal{U}, N}(\mathcal{F}_j)$$

for $j = 1, 2, \dots, 24$, where:

$$\begin{aligned}
\mathcal{F}_1 &= \{\pm f : \Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x, t) = u_{tt}(x, t)^2\}, \\
\mathcal{F}_2 &= \{\pm f : \Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ } 1 \leq i \leq j \leq d \text{ s.t. } f(x, t) = u_{x_i x_i}(x, t) u_{x_j x_j}(x, t)\}, \\
\mathcal{F}_4 &= \{\pm f : \Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ } 1 \leq i \leq d \text{ s.t. } f(x, t) = u_{tt}(x, t) u_{x_i x_i}(x, t)\}, \\
\mathcal{F}_5 &= \{\pm f : \Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x, t) = u_{tt}(x, t)\}, \\
\mathcal{F}_6 &= \{\pm f : \Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ } 1 \leq i \leq d \text{ s.t. } f(x, t) = u_{x_i x_i}(x, t)\}, \\
\mathcal{F}_7 &= \{\pm f : \Omega \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x) = u(x, 0)^2\}, \\
\mathcal{F}_9 &= \{\pm f : \Omega \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x) = u(x, 0)\}, \\
\mathcal{F}_{10} &= \{\pm f : \Omega \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ } 1 \leq i \leq j \leq d \text{ s.t. } f(x) = u_{x_i}(x, 0) u_{x_j}(x, 0)\}, \\
\mathcal{F}_{12} &= \{\pm f : \Omega \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ } 1 \leq i \leq d \text{ s.t. } f(x, t) = u_{x_i}(x, 0)\}, \\
\mathcal{F}_{13} &= \{\pm f : \Omega \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x) = u_t(x, 0)^2\}, \\
\mathcal{F}_{15} &= \{\pm f : \Omega \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x) = u_t(x, 0)\}, \\
\mathcal{F}_{16} &= \{\pm f : \partial\Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x, t) = u(x, t)^2|_{\partial\Omega}\}, \\
\mathcal{F}_{18} &= \{\pm f : \partial\Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x, t) = u(x, t)|_{\partial\Omega}\}, \\
\mathcal{F}_{19} &= \{\pm f : \partial\Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x, t) = u_t(x, t)^2|_{\partial\Omega}\}, \\
\mathcal{F}_{21} &= \{\pm f : \partial\Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ s.t. } f(x, t) = u_t(x, t)|_{\partial\Omega}\}, \\
\mathcal{F}_{22} &= \{\pm f : \partial\Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ } 1 \leq i \leq j \leq d \text{ s.t. } f(x, t) = u_{x_i}(x, t) u_{x_j}(x, t)|_{\partial\Omega}\}, \\
\mathcal{F}_{24} &= \{\pm f : \partial\Omega_T \rightarrow \mathbb{R} \mid \exists u \in \mathcal{P} \text{ } 1 \leq i \leq d \text{ s.t. } f(x, t) = u_{x_i}(x, t)|_{\partial\Omega}\}.
\end{aligned}$$

Proof The proof is based on the symmetrization technique, see **lemma 4.3** in [25] for more details.

Lemma 6 Let $\Phi = \{ReLU, ReLU^2, ReLU^3\}$. There holds

$$\begin{aligned}
\mathcal{F}_1 &\subset \mathcal{N}_1 := \mathcal{N}(\mathcal{D} + 5, (\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_2 &\subset \mathcal{N}_2 := \mathcal{N}(\mathcal{D} + 5, 2(\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_4 &\subset \mathcal{N}_4 := \mathcal{N}(\mathcal{D} + 5, 2(\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_5 &\subset \mathcal{N}_5 := \mathcal{N}(\mathcal{D} + 4, (\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi), \\
\mathcal{F}_6 &\subset \mathcal{N}_6 := \mathcal{N}(\mathcal{D} + 4, (\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi), \\
\mathcal{F}_7 &\subset \mathcal{N}_7 := \mathcal{N}(\mathcal{D} + 1), \mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_9 &\subset \mathcal{N}_9 := \mathcal{N}(\mathcal{D}, \mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi), \\
\mathcal{F}_{10} &\subset \mathcal{N}_{10} := \mathcal{N}(\mathcal{D} + 2, 2(\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_{12} &\subset \mathcal{N}_{12} := \mathcal{N}(\mathcal{D} + 2, (\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi), \\
\mathcal{F}_{13} &\subset \mathcal{N}_{13} := \mathcal{N}(\mathcal{D} + 3, (\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_{15} &\subset \mathcal{N}_{15} := \mathcal{N}(\mathcal{D} + 2, (\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi), \\
\mathcal{F}_{16} &\subset \mathcal{N}_{16} := \mathcal{N}(\mathcal{D} + 1, \mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_{18} &\subset \mathcal{N}_{18} := \mathcal{N}(\mathcal{D}, \mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi), \\
\mathcal{F}_{19} &\subset \mathcal{N}_{19} := \mathcal{N}(\mathcal{D} + 3, (\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_{21} &\subset \mathcal{N}_{21} := \mathcal{N}(\mathcal{D} + 2, (\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi), \\
\mathcal{F}_{22} &\subset \mathcal{N}_{22} := \mathcal{N}(\mathcal{D} + 3, 2(\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi), \\
\mathcal{F}_{24} &\subset \mathcal{N}_{24} := \mathcal{N}(\mathcal{D} + 2, (\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi).
\end{aligned}$$

Proof The proof is an application of **proposition 4.2** in [25]. Take \mathcal{F}_1 as an example, since $u \in \mathcal{P}$, we have $u_t \in \mathcal{N}(\mathcal{D} + 2, (\mathcal{D} + 2)\mathcal{W}, \{\|\cdot\|_{C^1(\overline{\Omega_T})}, \mathcal{B}\}, \Phi)$ and $u_{tt} \in \mathcal{N}(\mathcal{D} + 4, (\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}\}, \Phi)$. Notice the square operation can be implemented as $x^2 = ReLU^2(x) + ReLU^2(-x)$, thus we get that $u_{tt}^2 \in \mathcal{N}(\mathcal{D} + 5, (\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W}, \{\|\cdot\|_{C(\overline{\Omega_T})}, \mathcal{B}^2\}, \Phi)$.

Lemma 7 (Proposition 4.3 in [25]) For any $\mathcal{D}, \mathcal{W} \in \mathbb{N}$,

$$Pdim(\mathcal{N}(\mathcal{D}, \mathcal{W}, \{ReLU, ReLU^2, ReLU^3\})) = \mathcal{O}(\mathcal{D}^2 \mathcal{W}^2 (\mathcal{D} + \log \mathcal{W})).$$

Now we are ready to prove Theorem 3 on the statistical error.

Proof (The proof of Theorem 3) According to lemma 3 and lemma 6,

• For $i = 1, 2, 4, 5, 6$, when the sample numbers $n = NK > Pdim(\mathcal{F}_i)$, we have

$$\mathfrak{R}_{P(\mathcal{X}), NK}(\mathcal{F}_i) \leq \inf_{0 < \delta < \mathcal{B}_i} \left(4\delta + \frac{12}{\sqrt{NK}} \int_{\delta}^{\mathcal{B}_i} \sqrt{\log(2C_{\infty}(\varepsilon, \mathcal{F}_i, N))} d\varepsilon \right)$$

$$\begin{aligned}
&\leq \inf_{0 < \delta < \mathcal{B}_i} \left(4\delta + \frac{12}{\sqrt{NK}} \int_{\delta}^{\mathcal{B}_i} \sqrt{\log \left(2 \left(\frac{eNK\mathcal{B}_i}{\varepsilon \cdot Pdim(\mathcal{F}_i)} \right)^{Pdim(\mathcal{F}_i)} \right)} d\varepsilon \right) \\
&\leq \inf_{0 < \delta < \mathcal{B}_i} \left(4\delta + \frac{12\mathcal{B}_i}{\sqrt{NK}} + \frac{12}{\sqrt{NK}} \int_{\delta}^{\mathcal{B}_i} \sqrt{Pdim(\mathcal{F}_i) \log \left(\frac{eNK\mathcal{B}_i}{\varepsilon \cdot Pdim(\mathcal{F}_i)} \right)} d\varepsilon \right)
\end{aligned} \tag{27}$$

Let $t = \sqrt{\log \left(\frac{eNK\mathcal{B}_i}{\varepsilon \cdot Pdim(\mathcal{F}_i)} \right)}$, then $\varepsilon = \frac{eNK\mathcal{B}_i}{Pdim(\mathcal{F}_i)} e^{-t^2}$. Denoting:

$$t_1 = \sqrt{\log \left(\frac{eNK\mathcal{B}_i}{\mathcal{B}_i \cdot Pdim(\mathcal{F}_i)} \right)}, t_2 = \sqrt{\log \left(\frac{eNK\mathcal{B}_i}{\delta \cdot Pdim(\mathcal{F}_i)} \right)}$$

we have:

$$\begin{aligned}
&\int_{\delta}^{\mathcal{B}_i} \sqrt{\log \left(\frac{eNK\mathcal{B}_i}{\varepsilon \cdot Pdim(\mathcal{F}_i)} \right)} d\varepsilon \\
&= \frac{2eNK\mathcal{B}_i}{Pdim(\mathcal{F}_i)} \int_{t_1}^{t_2} t^2 e^{-t^2} dt \\
&= \frac{2eNK\mathcal{B}_i}{Pdim(\mathcal{F}_i)} \int_{t_1}^{t_2} t \left(\frac{-e^{-t^2}}{2} \right)' dt \\
&= \frac{eNK\mathcal{B}_i}{Pdim(\mathcal{F}_i)} \left[t_1 e^{-t_1^2} - t_2 e^{-t_2^2} + \int_{t_1}^{t_2} e^{-t^2} dt \right] \\
&\leq \frac{eNK\mathcal{B}_i}{Pdim(\mathcal{F}_i)} \left[t_1 e^{-t_1^2} - t_2 e^{-t_2^2} + (t_2 - t_1) e^{-t_1^2} \right] \\
&\leq \frac{eNK\mathcal{B}_i}{Pdim(\mathcal{F}_i)} t_2 e^{-t_1^2} \\
&\leq \mathcal{B}_i \sqrt{\log \left(\frac{eNK\mathcal{B}_i}{\delta \cdot Pdim(\mathcal{F}_i)} \right)}
\end{aligned} \tag{28}$$

Substitute (28) into (27) and choose $\delta = \mathcal{B}_i \left(\frac{Pdim(\mathcal{F}_i)}{NK} \right)^{\frac{1}{2}} \leq \mathcal{B}_i$, we have:

$$\begin{aligned}
\mathfrak{R}_{P(\mathcal{X}), NK}(\mathcal{F}_i) &\leq \inf_{0 < \delta < \mathcal{B}_i} \left(4\delta + \frac{12\mathcal{B}_i}{\sqrt{NK}} + \frac{12}{\sqrt{NK}} \int_{\delta}^{\mathcal{B}_i} \sqrt{Pdim(\mathcal{F}_i) \log \left(\frac{eNK\mathcal{B}_i}{\varepsilon \cdot Pdim(\mathcal{F}_i)} \right)} d\varepsilon \right) \\
&\leq \inf_{0 < \delta < \mathcal{B}_i} \left(4\delta + \frac{12\mathcal{B}_i}{\sqrt{NK}} + \frac{12\mathcal{B}_i \sqrt{Pdim(\mathcal{F}_i)}}{\sqrt{NK}} \sqrt{\log \left(\frac{eNK\mathcal{B}_i}{\delta \cdot Pdim(\mathcal{F}_i)} \right)} \right) \\
&\leq 28\sqrt{\frac{3}{2}} \mathcal{B}_i \left(\frac{Pdim(\mathcal{F}_i)}{NK} \right)^{\frac{1}{2}} \sqrt{\log \left(\frac{eNK}{Pdim(\mathcal{F}_i)} \right)}
\end{aligned}$$

$$\begin{aligned}
&\leq 28\sqrt{\frac{3}{2}}\mathcal{B}_i\left(\frac{Pdim(\mathcal{N}_i)}{NK}\right)^{\frac{1}{2}}\sqrt{\log\left(\frac{eNK}{Pdim(\mathcal{N}_i)}\right)} \\
&\leq 28\sqrt{\frac{3}{2}}\max\{\mathcal{B}, \mathcal{B}^2\}\left(\frac{\mathcal{H}_1}{NK}\right)^{\frac{1}{2}}\sqrt{\log\left(\frac{eNK}{\mathcal{H}_1}\right)} \quad (29)
\end{aligned}$$

where

$$\mathcal{H}_1 = C_1(\mathcal{D} + 2)^2(\mathcal{D} + 4)^2(\mathcal{D} + 5)^2\mathcal{W}^2[(\mathcal{D} + 5) + \log(2(\mathcal{D} + 2)(\mathcal{D} + 4)\mathcal{W})].$$

The last step above is due to lemma 7.

• For $i = 7, 9, 10, 12, 13, 15$, when the sample numbers $n = N > Pdim(\mathcal{F}_i)$, we can similarly prove that

$$\mathfrak{R}_{\mathcal{U}(\Omega), N}(\mathcal{F}_i) \leq 28\sqrt{\frac{3}{2}}\max\{\mathcal{B}, \mathcal{B}^2\}\left(\frac{\mathcal{H}_2}{N}\right)^{\frac{1}{2}}\sqrt{\log\left(\frac{eN}{\mathcal{H}_2}\right)} \quad (30)$$

where

$$\mathcal{H}_2 = C_2(\mathcal{D} + 2)^2(\mathcal{D} + 3)^2\mathcal{W}^2[(\mathcal{D} + 3) + \log(2(\mathcal{D} + 2)\mathcal{W})].$$

• For $i = 16, 18, 19, 21, 22, 24$, when the sample numbers $n = MK > Pdim(\mathcal{F}_i)$, we have

$$\mathfrak{R}_{\mathcal{U}(\partial\Omega_T), MK}(\mathcal{F}_i) \leq 28\sqrt{\frac{3}{2}}\max\{\mathcal{B}, \mathcal{B}^2\}\left(\frac{\mathcal{H}_3}{MK}\right)^{\frac{1}{2}}\sqrt{\log\left(\frac{eMK}{\mathcal{H}_3}\right)} \quad (31)$$

where

$$\mathcal{H}_3 = C_3(\mathcal{D} + 2)^2(\mathcal{D} + 3)^2\mathcal{W}^2[(\mathcal{D} + 3) + \log(2(\mathcal{D} + 2)\mathcal{W})].$$

• For $i = 3$, we have,

$$\begin{aligned}
&\mathbb{E}_{\{X_n\}_{n=1}^N, \{T_k\}_{k=1}^K} \left| \mathcal{L}_3 - \hat{\mathcal{L}}_3 \right| \\
&= |\Omega||T| \mathbb{E}_{\{X_n\}_{n=1}^N, \{T_k\}_{k=1}^K} \left| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} f(X, T)^2 - \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K f(X_n, T_k)^2 \right| \\
&\leq |\Omega||T| \sqrt{\mathbb{E}_{\{X_n\}_{n=1}^N, \{T_k\}_{k=1}^K} \left| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} f(X, T)^2 - \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K f(X_n, T_k)^2 \right|^2} \\
&= \frac{|\Omega||T|}{NK} \sqrt{\mathbb{E}_{\{X_n\}_{n=1}^N, \{T_k\}_{k=1}^K} \sum_{n=1}^N \sum_{k=1}^K \left| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} f(X, T)^2 - f(X_n, T_k)^2 \right|^2} \\
&= \frac{|\Omega||T|}{NK} \sqrt{NK \cdot \mathbb{E}_{X_1 \sim U(\Omega), T_1 \sim U([0, T])} \left| \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} f(X, T)^2 - f(X_1, T_1)^2 \right|^2} \\
&= \frac{|\Omega||T|}{NK} \sqrt{NK} \sigma(f(X, T)^2)
\end{aligned}$$

$$= |\Omega||T| \frac{\sigma(f(X, T)^2)}{\sqrt{NK}}$$

where $\sigma(f(X, T))$ is the standard deviation of $f(X, T)$. With the bound of f , we can further obtain,

$$\begin{aligned} \sigma^2(f(X, T)^2) &= \mathbb{E}((f(X, T)^2)^2) - (\mathbb{E}(f(X, T)^2))^2 \\ &\leq \frac{1}{|\Omega||T|} \left(\int_{\Omega_T} (f(X, T)^2)^2 dX dT \right) \\ &\leq \frac{1}{|\Omega||T|} \left(\int_{\Omega_T} \kappa^4 dX dT \right) \\ &\leq \frac{1}{|\Omega||T|} \cdot |\Omega||T| \mathcal{B} \\ &= \mathcal{B} \end{aligned}$$

then we have,

$$\mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} \left| \mathcal{L}_3 - \hat{\mathcal{L}}_3 \right| \leq |\Omega||T| \sqrt{\frac{\mathcal{B}}{NK}}.$$

- Similarly, for $i = 8, 14, 17, 20$,

$$\begin{aligned} \mathbb{E}_{X \sim U(\Omega), T \sim U([0, T])} \left| \mathcal{L}_8 - \hat{\mathcal{L}}_8 \right| &\leq |\Omega| \sqrt{\frac{\mathcal{B}}{N}}, \\ \mathbb{E}_{X \sim U(\Omega)} \left| \mathcal{L}_{14} - \hat{\mathcal{L}}_{14} \right| &\leq |\Omega| \sqrt{\frac{\mathcal{B}}{N}}, \\ \mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} \left| \mathcal{L}_{17} - \hat{\mathcal{L}}_{17} \right| &\leq |\partial\Omega||T| \sqrt{\frac{\mathcal{B}}{MK}}, \\ \mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} \left| \mathcal{L}_{20} - \hat{\mathcal{L}}_{20} \right| &\leq |\partial\Omega||T| \sqrt{\frac{\mathcal{B}}{MK}}. \end{aligned}$$

- For $i = 11$,

$$\begin{aligned} \sigma^2\left(\sum_{i=1}^d \varphi_{x_i}(x)^2\right) &= \mathbb{E}\left(\left(\sum_{i=1}^d \varphi_{x_i}(x)^2\right)^2\right) - \left(\mathbb{E}\left(\sum_{i=1}^d \varphi_{x_i}(x)^2\right)\right)^2 \\ &\leq \frac{1}{|\Omega|} \left[\int_{\Omega} \left(\sum_{i=1}^d \varphi_{x_i}(x)^2\right)^2 dx \right] \\ &\leq \frac{1}{|\Omega|} \left[\int_{\Omega} (d\kappa^2)^2 dx \right] \\ &\leq d^2 \mathcal{B} \end{aligned}$$

then we have,

$$\mathbb{E}_{X \sim U(\Omega)} \left| \mathcal{L}_{11} - \hat{\mathcal{L}}_{11} \right| \leq d|\Omega| \sqrt{\frac{\mathcal{B}}{N}}.$$

- Similarly, for $i = 23$,

$$\mathbb{E}_{Y \sim U(\partial\Omega), T \sim U([0, T])} \left| \mathcal{L}_{23} - \hat{\mathcal{L}}_{23} \right| \leq d|\partial\Omega||T| \sqrt{\frac{\mathcal{B}}{MK}}.$$

Hence, we have,

$$\begin{aligned} & \mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \sup_{u \in \mathcal{P}} |\mathcal{L}(u) - \hat{\mathcal{L}}(u)| \\ & \leq \sum_{j=1}^{24} \mathbb{E}_{\{X, Y, T\}} \sup_{u \in \mathcal{P}} |\mathcal{L}_j(u) - \hat{\mathcal{L}}_j(u)| \\ & \leq 28 \sqrt{\frac{3}{2}} \max\{\mathcal{B}, \mathcal{B}^2\} \left(5|\Omega||T|C_1 \left(\frac{\mathcal{H}_1}{NK} \right)^{\frac{1}{2}} \sqrt{\log \left(\frac{eNK}{\mathcal{H}_1} \right)} + \right. \\ & \quad \left. 6|\Omega|C_2 \left(\frac{\mathcal{H}_2}{N} \right)^{\frac{1}{2}} \sqrt{\log \left(\frac{eN}{\mathcal{H}_2} \right)} + 6|\partial\Omega||T|C_3 \left(\frac{\mathcal{H}_3}{MK} \right)^{\frac{1}{2}} \sqrt{\log \left(\frac{eMK}{\mathcal{H}_3} \right)} \right) \\ & \quad \left(|\Omega||T| \sqrt{\frac{\mathcal{B}}{NK}} + (2+d)|\Omega| \sqrt{\frac{\mathcal{B}}{N}} + (2+d)|\partial\Omega||T| \sqrt{\frac{\mathcal{B}}{MK}} \right), \end{aligned}$$

where C_1, C_2, C_3 are constants associated with dimensionality d and bound \mathcal{B} . Hence, for any $\varepsilon \geq 0$, if the number of samples satisfies:

$$\begin{cases} N &= C(d, |\Omega|, \mathcal{B}) \mathcal{D}^4 \mathcal{W}^2 (\mathcal{D} + \log(\mathcal{W})) \left(\frac{1}{\varepsilon} \right)^{2+\delta}, \\ K &= C(d, |T|, \mathcal{B}) \mathcal{D}^2 f_K(\mathcal{D}, \mathcal{W}) \left(\frac{1}{\varepsilon} \right)^{k_1}, \\ M &= C(d, |\partial\Omega|, \mathcal{B}) f_M(\mathcal{D}, \mathcal{W}) \left(\frac{1}{\varepsilon} \right)^{k_2}, \end{cases}$$

where:

$$\begin{cases} k_1 + k_2 = 2 + \delta, \\ f_k(\mathcal{D}, \mathcal{W}) \cdot f_M(\mathcal{D}, \mathcal{W}) = \mathcal{D}^2 \mathcal{W}^2 (\mathcal{D} + \log(\mathcal{W})). \end{cases}$$

with restriction $f_K(\mathcal{D}, \mathcal{W}) \geq 1$, $f_M(\mathcal{D}, \mathcal{W}) \geq 1$, and δ is arbitrarily small. Then we have:

$$\mathbb{E}_{\{X_n\}_{n=1}^N, \{Y_m\}_{m=1}^M, \{T_k\}_{k=1}^K} \sup_{u \in \mathcal{P}} |\mathcal{L}(u) - \hat{\mathcal{L}}(u)| \leq \varepsilon.$$

References

1. Susanne Brenner and Ridgway Scott. *The mathematical theory of finite element methods*, volume 15. Springer Science & Business Media, 2007.
2. Philippe G Ciarlet. *The finite element method for elliptic problems*. SIAM, 2002.
3. A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*, volume 23. Springer Science & Business Media, 2008.
4. J.W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*, volume 22. Springer Science & Business Media, 2013.
5. Isaac E Lagaris, Aristidis Likas, and Dimitrios I Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks*, 9(5):987–1000, 1998.
6. Cosmin Anitescu, Elena Atroshchenko, Naif Alajlan, and Timon Rabczuk. Artificial neural network methods for the solution of second order boundary value problems. *Computers, Materials & Continua*, 59(1):345–359, 2019.
7. Julius Berner, Markus Dablander, and Philipp Grohs. Numerically solving parametric families of high-dimensional kolmogorov partial differential equations via deep learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 16615–16627. Curran Associates, Inc., 2020.
8. Jiequn Han, Arnulf Jentzen, and E Weinan. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.
9. Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. Deepxde: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021.
10. Justin Sirignano and Konstantinos Spiliopoulos. Dgm: A deep learning algorithm for solving partial differential equations. *Journal of computational physics*, 375:1339–1364, 2018.
11. E. Weinan and Ting Yu. The deep ritz method: A deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2017.
12. Maziar Raissi, Paris Perdikaris, and George E 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.
13. Yaohua Zang, Gang Bao, Xiaojing Ye, and Haomin Zhou. Weak adversarial networks for high-dimensional partial differential equations. *Journal of Computational Physics*, 411:109409, 2020.
14. Ameya D Jagtap, Ehsan Kharazmi, and George Em Karniadakis. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 365:113028, 2020.
15. G. Pang, M. D’Elia, M. Parks, and G.E. Karniadakis. npinns: Nonlocal physics-informed neural networks for a parametrized nonlocal universal laplacian operator. algorithms and applications. *Journal of Computational Physics*, 422:109760, 2020.
16. Guofei Pang, Lu Lu, and George Em Karniadakis. fpinns: Fractional physics-informed neural networks. *SIAM Journal on Scientific Computing*, 41(4):A2603–A2626, 2019.
17. Hao Wang, Jian Li, Linfeng Wang, Lin Liang, Zhoumo Zeng, and Yang Liu. On acoustic fields of complex scatters based on physics-informed neural networks. *Ultrasonics*, 128:106872, 2023.
18. Linfeng Wang, Hao Wang, Lin Liang, Jian Li, Zhoumo Zeng, and Yang Liu. Physics-informed neural networks for transcranial ultrasound wave propagation. *Ultrasonics*, 132:107026, 2023.
19. Yi Ding, Su Chen, Xiaojun Li, Suyang Wang, Shaokai Luan, and Hao Sun. Self-adaptive physics-driven deep learning for seismic wave modeling in complex topography. *Engineering Applications of Artificial Intelligence*, 123:106425, 2023.
20. Wojciech M Czarnecki, Simon Osindero, Max Jaderberg, Grzegorz Swirszcz, and Razvan Pascanu. Sobolev training for neural networks. *Advances in neural information processing systems*, 30, 2017.

21. Hwijae Son, Jin Woo Jang, Woo Jin Han, and Hyung Ju Hwang. Sobolev training for physics informed neural networks. *Communications in Mathematical Sciences*, 21:1679–1705, 2013.
22. Nikolaos N Vlassis and WaiChing Sun. Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening. *Computer Methods in Applied Mechanics and Engineering*, 377:113695, 2021.
23. Yuling Jiao, Xiliang Lu, Jerry Zhijian Yang, Cheng Yuan, and Pingwen Zhang. Improved analysis of pinns: Alleviate the cod for compositional solutions. *Annals of Applied Mathematics*, 39(3):239–263, 2023.
24. Yuling Jiao, Di Li, Xiliang Lu, Jerry Zhijian Yang, and Cheng Yuan. Gas: A gaussian mixture distribution-based adaptive sampling method for pinns. *arXiv preprint arXiv:2303.15849*, 2023.
25. Yuling Jiao, Yanming Lai, Dingwei Li, Xiliang Lu, Fengru Wang, Jerry Zhijian Yang, et al. A rate of convergence of physics informed neural networks for the linear second order elliptic pdes. *Communications in Computational Physics*, 31(4):1272–1295, 2022.
26. Aad W Van Der Vaart, Adrianus Willem van der Vaart, Aad van der Vaart, and Jon Wellner. *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media, 1996.
27. Martin Anthony and Peter L. Bartlett. Neural network learning: Theoretical foundations. *Ai Magazine*, 22(2):99–100, 1999.