

Convergence Rate Analysis for Deep Ritz Method

Chenguang Duan¹, Yuling Jiao^{1,2}, Yanming Lai¹, Dingwei Li¹,
Xiliang Lu^{1,2} and Jerry Zhijian Yang^{1,2,*}

¹ School of Mathematics and Statistics, Wuhan University, Wuhan 430072,
P.R. China.

² Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan 430072,
P.R. China.

Received 7 October 2021; Accepted (in revised version) 26 January 2022

Abstract. Using deep neural networks to solve PDEs has attracted a lot of attentions recently. However, why the deep learning method works is falling far behind its empirical success. In this paper, we provide a rigorous numerical analysis on deep Ritz method (DRM) [47] for second order elliptic equations with Neumann boundary conditions. We establish the first nonasymptotic convergence rate in H^1 norm for DRM using deep networks with ReLU² activation functions. In addition to providing a theoretical justification of DRM, our study also shed light on how to set the hyperparameter of depth and width to achieve the desired convergence rate in terms of number of training samples. Technically, we derive bound on the approximation error of deep ReLU² network in C^1 norm and bound on the Rademacher complexity of the non-Lipschitz composition of gradient norm and ReLU² network, both of which are of independent interest.

AMS subject classifications: 62G05, 65N12, 65N15, 68T07

Key words: Deep Ritz method, deep neural networks, convergence rate analysis.

1 Introduction

Partial differential equations (PDEs) have broad applications in physics, chemistry, biology, geology and engineering. A great deal of efforts have been devoted to studying numerical methods for solving PDEs [5, 7, 16, 22, 43]. However, it is still a challenging task to develop numerical scheme for solving PDEs in high-dimension. Due to the success of deep learning for high-dimensional data analysis in computer vision and natural

*Corresponding author. *Email addresses:* cgduan.math@whu.edu.cn (C. Duan), yulingjiaomath@whu.edu.cn (Y. Jiao), laiyanning@whu.edu.cn (Y. Lai), lidingv@whu.edu.cn (D. Li), xllv.math@whu.edu.cn (X. Lu), zjyang.math@whu.edu.cn (J. Z. Yang)

language processing, people have been paying more attention to using (deep) neural network to solve PDEs in high dimension with may be complex domain, an idea that goes back to 1990's [19, 21]. In the last few years, there are growing literatures on neural network based numerical methods for PDEs. These works can be roughly classified into two categories.

In the first category, deep neural networks are used to improve classical methods. [10] designs a neural network to estimate artificial viscosity in discontinuous Galerkin schemes, see also [6]. [32] trains a neural network serving as a troubled-cell indicator in high-resolution schemes for conservation laws. [41] proposes a universal discontinuity detector using convolution neural network and applies it in conjunction of solving non-linear conservation. [46] uses reinforcement learning to find new and potentially better data-driven solvers for conservation laws.

In the second category, deep neural networks are utilized to approximate the solution of the PDEs directly. Being benefit from the excellent approximation power of deep neural networks and SGD training, these methods have been successfully applied to solve PDEs in high-dimension. [3, 9] convert nonlinear parabolic PDEs into backward stochastic differential equations and solve them by deep neural networks, which can deal with high-dimensional problems. Methods based on the strong form of PDEs [31, 39] are also proposed. In [31], physics-informed neural networks (PINNs) use the squared residuals on the domain as the loss function and treat boundary conditions as penalty term. There are several extensions of PINNs for different types of PDEs, including fractional PINNs [30], nonlocal PINNs [29], conservative PINNs [18], eXtended PINNs [17], among others. A similar method presented in [25] proposes a residual-based adaptive refinement method to improve the training efficiency.

In contrast to minimizing squared residuals of strong form, a natural alternative approach to derive loss functions are based on the variational form of PDEs [47, 50]. Inspired by Ritz method, [47] proposes deep Ritz method (DRM) to solve variational problems arising from PDEs. The idea of Galerkin method has also been used in [50], where, they propose a deep Galerkin method (DGM) via reformulating the problem of finding the weak solution of PDEs into an operator norm minimization problem induced by the weak formulation.

1.1 Related works and contributions

Although there are great empirical achievements in recent years as mentioned above, a challenging and interesting question is that can we give rigorous analysis to guarantee their performances as people has done in the classical counterpart such as finite element method (FEM) [7] and finite difference method [22] ? Several recent efforts have been devoted to making processes along this line. [26] consider the optimization and generalization error of second-order linear PDEs with two-layer neural networks in the scenario of over-parametrization. [27, 36, 37] study the convergence of PINNs with deep neural networks. When we were about to finish our draft, we aware that [24] give an error anal-

ysis that focuses on analyzing one hidden layer shallow networks with ReLU-Cosine activation functions to solve elliptic PDEs whose solutions are restricted to spectral Barron space, see also [48] for handling general equations with solutions living in spectral Barron space via two layer ReLU^k networks. **Two important questions have not been addressed in the above mentioned related study are those: (1) What is the influence of the topological structure of the networks, say the depth and width, in the quantitative error analysis? (2) How to determine these hyper-parameters to achieve a desired convergence rate?** In this paper, we give a firm answers on these questions by studying convergence rate of the deep Ritz method to solve second order elliptic equations with Neumann boundary conditions by using ReLU² networks with arbitrary depth. As far as we know, we establish the first nonasymptotic bound on DRM. The main contributions of this paper are summarized as follows.

- We derive a bound on the approximation error of deep ReLU² network in C¹ norm, which is of independent interest, i.e., we prove that for any $u^* \in C^2(\bar{\Omega})$, there exist a ReLU² network \bar{u}_{ϕ} with depth $\mathcal{D} \leq \lceil \log_2 d \rceil + 2$, width $\mathcal{W} = C_{d, \|u^*\|_{C^2(\bar{\Omega})}} \left(\frac{1}{\epsilon}\right)^d$ such that

$$\|u^* - \bar{u}_{\phi}\|_{C^1(\bar{\Omega})} \leq \epsilon.$$

- We establish a bound on the statistical error in DRM with the tools of Pseudo dimension, especially we give an bound on

$$\mathbb{E}_{Z_i, \sigma_i, i=1, \dots, n} \left[\sup_{u_{\phi} \in \mathcal{N}^2} \frac{1}{n} \left| \sum_i \sigma_i \|\nabla u_{\phi}(Z_i)\|^2 \right| \right],$$

i.e., the Rademacher complexity of the non-Lipschitz composition of gradient norm and ReLU² network, via calculating the Pseudo dimension of networks with both ReLU and ReLU² activation functions. We believe that the technique we used here is helpful for bounding the statistical errors for other deep PDEs solvers where the Rademacher complexity of non-Lipschitz composition is hard to handle.

- Based on the above to error bounds we establish the first nonasymptotic convergence rate of deep Ritz method. We prove that if we set the depth and width in ReLU² networks to be

$$\mathcal{D} = \lceil \log_2 d \rceil + 2, \quad \mathcal{W} = C_{d, \|u^*\|_{C^2(\bar{\Omega})}} n^{\frac{d}{d+2+v}},$$

the H^1 norm error of DRM in expectation is

$$\mathcal{O}(n^{-1/(2d+4+v)}),$$

where n is the number of training samples on both the domain and the boundary, v is a positive number that can be an arbitrarily small. Our theory shed lights on choosing the topological structure of the employed networks to achieve the desired convergence rate in terms of number of training samples.

- By comparing the known results in nonparametric regression, where the optimal convergence rate in H^1 norm for estimating functions in H^2 with n paired samples is $\mathcal{O}(n^{-\frac{1}{4+d}})$ [40], we conjecture that the optimal convergence rate of DRM in H^1 norm is also $\mathcal{O}(n^{-1/(d+4)})$.

The rest of the paper are organized as follows. In Section 2, we give some preliminaries. In Section 3, we present the detail analysis on the convergence rate of DRM. We give conclusion and short discussion in Section 4.

2 Preliminaries

Consider the following elliptic equation with Neumann boundary conditions

$$\begin{cases} -\Delta u + wu = f, & \text{in } \Omega, \\ \frac{\partial u}{\partial n} = g, & \text{on } \partial\Omega, \end{cases} \quad (2.1)$$

where Ω is a bounded open subset of \mathbb{R}^d , $d > 1$, $f(x) \in L^2(\Omega)$, $w(x) \in L^\infty(\Omega)$ satisfying $w(x) \geq c_1 > 0$ a.e., and $g(s) \in L^2(\partial\Omega)$. Without loss of generality we assume $\Omega = (0,1)^d$. Define

$$\mathcal{L}(u) = \frac{1}{2} \|u\|_{H^1(\Omega)}^2 + \frac{1}{2} \|u\|_{L^2(\Omega;w)}^2 - \langle u, f \rangle_{L^2(\Omega)} - \langle Tu, g \rangle_{L^2(\partial\Omega)}, \quad (2.2)$$

where T is the trace operator.

Lemma 2.1. *The unique weak solution $u^* \in H^1(\Omega)$ of (2.1) is the unique minimizer of $\mathcal{L}(u)$ over $H^1(\Omega)$. Moreover, $u^* \in H^2(\Omega)$.*

Proof. Well known results, see for example [13]. □

A function $\mathbf{f}: \mathbb{R}^d \rightarrow \mathbb{R}^{N_L}$ implemented by a neural network is defined by

$$\begin{aligned} \mathbf{f}_0(\mathbf{x}) &= \mathbf{x}, \\ \mathbf{f}_\ell(\mathbf{x}) &= \varrho_\ell(\mathbf{A}_\ell \mathbf{f}_{\ell-1} + \mathbf{b}_\ell) \quad \text{for } \ell = 1, \dots, L-1, \\ \mathbf{f} &= \mathbf{f}_L(\mathbf{x}) := \mathbf{A}_L \mathbf{f}_{L-1} + \mathbf{b}_L, \end{aligned}$$

where $\mathbf{A}_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$, $\mathbf{b}_\ell \in \mathbb{R}^{N_\ell}$ and the activation function ϱ_ℓ is understood to act component-wise (note that here we allow different activation functions in different layers). L is called the depth of the network and $\max\{N_\ell, \ell = 0, \dots, L\}$ is called the width of the network. We will use \mathcal{L} and \mathcal{W} to denote the depth and width of neural networks \mathbf{f} , respectively. $\sum_{\ell=1}^L N_\ell$ is called number of unites of \mathbf{f} and $\phi = \{\mathbf{A}_\ell, \mathbf{b}_\ell\}_\ell$ are called the parameters. For simplicity we also use \mathbf{f}_ϕ to refer to the network. We use $\mathcal{N}_{\mathcal{D}, \mathcal{W}, \mathcal{B}}^2$ to denote the set of neural networks with depth \mathcal{D} , width \mathcal{W} , output of the function values and their square norm of gradients bounded by \mathcal{B} , activation function $\text{ReLU}^2(x) = \max\{0, x^2\}$.

Denote $\mathcal{N}_{\mathcal{D},\mathcal{W},\mathcal{B}}^{1,2}$ as the set of neural networks with depth \mathcal{D} , width \mathcal{W} , output bounded by \mathcal{B} , activation functions $\text{ReLU}(x) = \max\{0, x\}$ and $\text{ReLU}^2(x) = \max\{0, x^2\}$.

Obviously,

$$\begin{aligned} \mathcal{L}(u) = & |\Omega| \mathbb{E}_{X \sim U(\Omega)} [\|\nabla u(X)\|_2^2 / 2 + w(X)u^2(X) / 2 - u(X)f(X)] \\ & - |\partial\Omega| \mathbb{E}_{Y \sim U(\partial\Omega)} [Tu(Y)g(Y)], \end{aligned}$$

where $U(\Omega), U(\partial\Omega)$ are the uniform distributions on Ω and $\partial\Omega$, respectively. The main idea of deep Ritz method (DRM) [47] is employing a $u_\phi \in \mathcal{N}^2 := \mathcal{N}_{\mathcal{D},\mathcal{W},\mathcal{B}}^2$ to approximate the minimizer u^* of \mathcal{L} , i.e., finding u_ϕ such that $\mathcal{L}(u_\phi)$ closes to $\mathcal{L}(u^*)$. To this end, by Lemma 2.1, one may consider the following empirical loss minimization problem

$$\hat{u}_\phi \in \min_{u_\phi \in \mathcal{N}^2} \hat{\mathcal{L}}(u_\phi), \tag{2.3}$$

where

$$\begin{aligned} \hat{\mathcal{L}}(u_\phi) = & \frac{|\Omega|}{N} \sum_{i=1}^N \left[\frac{\|\nabla u_\phi(X_i)\|_2^2}{2} + \frac{w(X_i)u_\phi^2(X_i)}{2} - u_\phi(X_i)f(X_i) \right] \\ & - \frac{\partial\Omega}{M} \sum_{j=1}^M [u_\phi(Y_j)g(Y_j)] \end{aligned} \tag{2.4}$$

is a discrete version of the functional $\mathcal{L}(u_\phi)$ with $\{X_i\}_{i=1}^N$ being identically and independently distributed (i.i.d.) according to $U(\Omega)$, $\{Y_j\}_{j=1}^M$ being identically and independently drawn from $U(\partial\Omega)$. Then, we call a (random) solver \mathcal{A} , say SGD, to minimize (2.3) and denote the output of \mathcal{A} , say $u_{\phi_{\mathcal{A}}} \in \mathcal{N}^2$, as the final solution.

3 Error analysis

In this section we prove the convergence rate analysis for DRM with deep ReLU² networks. The following Lemma play an important role by decoupling the total errors into three types of errors.

Lemma 3.1.

$$\begin{aligned} & \|u_{\phi_{\mathcal{A}}} - u^*\|_{H^1(\Omega)}^2 \\ \leq & \frac{2}{c_1 \wedge 1} \left[\underbrace{\frac{\|w\|_{L^\infty(\Omega)} \vee 1}{2} \cdot \inf_{\bar{u} \in \mathcal{N}^2} \|\bar{u} - u^*\|_{H^1(\Omega)}^2}_{\mathcal{E}_{app}} + 2 \underbrace{\sup_{u \in \mathcal{N}^2} |\mathcal{L}(u) - \hat{\mathcal{L}}(u)|}_{\mathcal{E}_{sta}} + \underbrace{\hat{\mathcal{L}}(u_{\phi_{\mathcal{A}}}) - \hat{\mathcal{L}}(\hat{u}_\phi)}_{\mathcal{E}_{opt}} \right]. \end{aligned}$$

Proof. For any $\bar{u} \in \mathcal{N}^2$, we have

$$\begin{aligned} & \mathcal{L}(u_{\phi_A}) - \mathcal{L}(u^*) \\ &= \mathcal{L}(u_{\phi_A}) - \widehat{\mathcal{L}}(u_{\phi_A}) + \widehat{\mathcal{L}}(u_{\phi_A}) - \widehat{\mathcal{L}}(\widehat{u}_\phi) + \widehat{\mathcal{L}}(\widehat{u}_\phi) - \widehat{\mathcal{L}}(\bar{u}) \\ & \quad + \widehat{\mathcal{L}}(\bar{u}) - \mathcal{L}(\bar{u}) + \mathcal{L}(\bar{u}) - \mathcal{L}(u^*) \\ & \leq [\mathcal{L}(\bar{u}) - \mathcal{L}(u^*)] + 2 \sup_{u \in \mathcal{N}^2} \left| \mathcal{L}(u) - \widehat{\mathcal{L}}(u) \right| + \left[\widehat{\mathcal{L}}(u_{\phi_A}) - \widehat{\mathcal{L}}(\widehat{u}_\phi) \right], \end{aligned}$$

where the last step is due to the fact that $\widehat{\mathcal{L}}(\widehat{u}_\phi) - \widehat{\mathcal{L}}(\bar{u}) \leq 0$. Since \bar{u} can be any element in \mathcal{N}^2 , we take the infimum of \bar{u} on both side of the above display,

$$\begin{aligned} \mathcal{L}(u_{\phi_A}) - \mathcal{L}(u^*) & \leq \inf_{\bar{u} \in \mathcal{N}^2} [\mathcal{L}(\bar{u}) - \mathcal{L}(u^*)] + 2 \sup_{u \in \mathcal{N}^2} \left| \mathcal{L}(u) - \widehat{\mathcal{L}}(u) \right| \\ & \quad + \left[\widehat{\mathcal{L}}(u_{\phi_A}) - \widehat{\mathcal{L}}(\widehat{u}_\phi) \right]. \end{aligned} \tag{3.1}$$

Now for any $u \in \mathcal{N}^2$, set $v = u - u^*$, then

$$\begin{aligned} & \mathcal{L}(u) = \mathcal{L}(u^* + v) \\ &= \frac{1}{2}(\nabla(u^* + v), \nabla(u^* + v))_{L^2(\Omega)} + \frac{1}{2}(u^* + v, u^* + v)_{L^2(\Omega; w)} \\ & \quad - \langle u^* + v, f \rangle_{L^2(\Omega)} - \langle Tu^* + Tv, g \rangle_{L^2(\partial\Omega)} \\ &= \frac{1}{2}(\nabla u^*, \nabla u^*)_{L^2(\Omega)} + \frac{1}{2}(u^*, u^*)_{L^2(\Omega; w)} - \langle u^*, f \rangle_{L^2(\Omega)} - \langle Tu^*, g \rangle_{L^2(\partial\Omega)} \\ & \quad + \frac{1}{2}(\nabla v, \nabla v)_{L^2(\Omega)} + \frac{1}{2}(v, v)_{L^2(\Omega; w)} \\ & \quad + \left[(\nabla u^*, \nabla v)_{L^2(\Omega)} + (u^*, v)_{L^2(\Omega; w)} - \langle v, f \rangle_{L^2(\Omega)} - \langle Tv, g \rangle_{L^2(\partial\Omega)} \right] \\ &= \mathcal{L}(u^*) + \frac{1}{2}(\nabla v, \nabla v)_{L^2(\Omega)} + \frac{1}{2}(v, v)_{L^2(\Omega; w)}, \end{aligned}$$

where the last equality is due to the fact that u^* is the weak solution of Eq. (2.1). Hence

$$\begin{aligned} \frac{c_1 \wedge 1}{2} \|v\|_{H^1(\Omega)}^2 & \leq \mathcal{L}(u) - \mathcal{L}(u^*) = \frac{1}{2}(\nabla v, \nabla v)_{L^2(\Omega)} + \frac{1}{2}(v, v)_{L^2(\Omega; w)} \\ & \leq \frac{\|w\|_{L^\infty(\Omega)} \vee 1}{2} \|v\|_{H^1(\Omega)}^2, \end{aligned}$$

that is,

$$\frac{c_1 \wedge 1}{2} \|u - u^*\|_{H^1(\Omega)}^2 \leq \mathcal{L}(u) - \mathcal{L}(u^*) \leq \frac{\|w\|_{L^\infty(\Omega)} \vee 1}{2} \|u - u^*\|_{H^1(\Omega)}^2. \tag{3.2}$$

Combining (3.1) and (3.2) yields

$$\begin{aligned} & \|u_{\phi_A} - u^*\|_{H^1(\Omega)}^2 \\ & \leq \frac{2}{c_1 \wedge 1} \left[\frac{\|w\|_{L^\infty(\Omega)} \vee 1}{2} \cdot \inf_{\bar{u} \in \mathcal{N}^2} \|\bar{u} - u^*\|_{H^1(\Omega)}^2 \right. \\ & \quad \left. + 2 \sup_{u \in \mathcal{N}^2} \left| \mathcal{L}(u) - \widehat{\mathcal{L}}(u) \right| + \left[\widehat{\mathcal{L}}(u_{\phi_A}) - \widehat{\mathcal{L}}(\widehat{u}_\phi) \right] \right]. \end{aligned}$$

This completes the proof. □

The approximation error \mathcal{E}_{app} describes the expressive power of the ReLU² networks \mathcal{N}^2 in H^1 norm, which corresponds to the approximation error in FEM known as the C ea’s lemma [7]. The statistical error \mathcal{E}_{sta} is caused by the Monte Carlo discretization of $\mathcal{L}(\cdot)$ defined in (2.2) with $\widehat{\mathcal{L}}(\cdot)$ in (2.4). While, the optimization error \mathcal{E}_{opt} indicates the performance of the solver \mathcal{A} we utilized. In contrast, this error is corresponding to the error of solving the linear systems in FEM. In this paper we focus on the first two errors, i.e, considering the scenario of perfect training with $\mathcal{E}_{opt} = 0$.

4 Approximation error

The current literature on network approximation theory are mainly focus on the $L^p, p \in [1, +\infty]$ norm for deep networks [23, 28, 33, 35, 38, 42, 49]. The approximation error of ReLU network in Sobolev norm are considered in [12, 15]. In this section we derive an upper bound on the approximation error ReLU² networks \mathcal{N}^2 in H^1 norm, which is of independent interest.

4.1 Approximation by B-splines

In this section we derive the approximation theorem of B-splines in Sobolev spaces based on average Taylor expansion and quasi-interpolation, from which we upper bound the approximation error of deep ReLU² network in the next subsection.

We denote by π_ℓ the uniform partition of $[0, 1]$:

$$\pi_\ell: F \ t_0^{(\ell)} = 0 < t_1^{(\ell)} < \dots < t_{\ell-1}^{(\ell)} < t_\ell^{(\ell)} = 1$$

with $t_i^{(\ell)} = i/\ell$ ($0 \leq i \leq \ell$). For $k \in \mathbb{N}_+$, we consider an extended partition $\widehat{\pi}_\ell$:

$$\widehat{\pi}_\ell: t_{-k+1}^{(\ell)} = \dots = t_0^{(\ell)} = 0 < t_1^{(\ell)} < \dots < t_{\ell-1}^{(\ell)} < t_\ell^{(\ell)} = \dots = t_{\ell+k-1}^{(\ell)} = 1$$

with $t_i^{(\ell)} = i/\ell$ ($0 \leq i \leq \ell$). Then the univariate B-spline of order k with respect to partition $\widehat{\pi}_\ell$ is defined by

$$N_{\ell,i}^{(k)}(x) = (-1)^k \left(t_{i+k}^{(\ell)} - t_i^{(\ell)} \right) \cdot \left[t_i^{(\ell)}, \dots, t_{i+k}^{(\ell)} \right] (x-t)_+^{k-1}, \quad x \in [0, 1], \quad i \in I, \quad (4.1)$$

where

$$I = \{-k+1, -k+2, \dots, \ell-1\}.$$

By the definition and properties of divided difference $[t_i^{(\ell)}, \dots, t_{i+k}^{(\ell)}]$ (see, for example, [34]) and some calculations, B-splines can also be equivalently written as:

$$N_{\ell,i}^{(k)}(x) = \begin{cases} \frac{\ell^{k-1}}{(k-1)!} \sum_{j=0}^k (-1)^j \binom{k}{j} \left(x - \frac{i+j}{\ell}\right)_+^{k-1}, & 0 \leq i \leq \ell-k+1, \\ \sum_{j=0}^{k-1} a_{ij} \left(x - \frac{j}{\ell}\right)_+^{k-1} + \sum_{n=1}^{k-2} b_{in} x^n + b_{i0}, & -k+1 \leq i < 0, \\ \sum_{j=\ell-k+1}^{\ell} c_{ij} \left(x - \frac{j}{\ell}\right)_+^{k-1}, & \ell-k+1 < i \leq \ell-1, \end{cases} \quad (4.2)$$

where $\{a_{ij}\}, \{b_{in}\}, \{c_{ij}\}$ are some constants. We must mention that the values of (4.1) and (4.2) at $x = 1$ is different for $\ell-k+1 < i \leq \ell-1$. In fact, for $\ell-k+1 < i \leq \ell-1$, (4.1) is discontinuous at $x = 1$ and (4.2) is its continuous modification. In the following when we mention $N_{\ell,i}^{(k)}$, we always refer to the continuous version (4.2).

The multivariate B-spline is defined by the product of univariate B-splines:

$$N_{\ell,\mathbf{i}}^{(k)}(\mathbf{x}) = \prod_{j=1}^d N_{\ell,i_j}^{(k)}(x_j), \quad \mathbf{i} \in I^d.$$

The readers are referred to [34] [8] for more discussion of B-splines.

Now we list some properties of B-splines, which are the basis of our approximation result.

Proposition 4.1 (Theorem 4.9 and Corollary 4.10, [34]). (1) $N_{\ell,i}^{(k)} = 0$ for $x \notin (t_i, t_{i+k})$ and $N_{\ell,i}^{(k)} > 0$ for $x \in (t_i, t_{i+k})$.

(2) $\{N_{\ell,i}^{(k)}\}_{i \in I}$ forms a basis of space $\{f \in C^{k-2}([0,1]) : f \text{ is } k-1 \text{ degree piecewise polynomial with respect to partition } \hat{\tau}_I\}$.

Proposition 4.2 (Theorem 4.22, [34]). For $r \in \mathbb{N}_0$,

$$\left| D^r N_{\ell,i}^{(k)} \right| \leq 2^r \ell^r.$$

For $\mathbf{i} \in I^d$, define

$$\Omega_{\mathbf{i}} = \{\mathbf{x} : x_j \in [t_{i_j}, t_{i_j+k}], j = 1, \dots, d\}.$$

Proposition 4.3 (Theorem 12.5, [34]). There exists a set of linear functionals $\{\lambda_{\mathbf{i}}\}$ mapping $L^1(\Omega)$ to \mathbb{R} such that

$$\lambda_{\mathbf{i}}(N_{\ell,\mathbf{j}}^{(k)}) = \delta_{\mathbf{i},\mathbf{j}},$$

and

$$|\lambda_i(f)| \leq (2k+1)^d \mathfrak{g}^{d(k-1)} \left(\frac{k}{\ell}\right)^{-d/p} \|f\|_{L^p(\Omega_i)},$$

for any $f \in L^p(\Omega)$ with $p \in [1, \infty]$.

Now we define the quasi-interpolant of f by

$$Qf := \sum_{\mathbf{i} \in I^d} \lambda_i(f) N_{\ell, \mathbf{i}}^{(k)}, \tag{4.3}$$

and Q is called the interpolation operator.

Proposition 4.4. Let p be a multivariate polynomial with respect to x_1, \dots, x_d and be of order k with respect to each variable, then

$$Qp = p.$$

Proof. By definition, there exists $\{c_{\mathbf{ff}}\} \subset \mathbb{R}$ such that

$$p(x_1, \dots, x_d) = \sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} x_1^{\alpha_1} \dots x_d^{\alpha_d}.$$

By Proposition 4.1(2), there exists $\{e_{i,j,\alpha_j}\} \subset \mathbb{R}$ such that

$$x_j^{\alpha_j} = \sum_{i \in I} e_{i,j,\alpha_j} N_{\ell,i}^{(k)}(x_j).$$

Hence

$$\begin{aligned} p(x_1, \dots, x_d) &= \sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} \left(\prod_{j=1}^d \sum_{i \in I} e_{i,j,\alpha_j} N_{\ell,i}^{(k)}(x_j) \right) \\ &= \sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} \left(\sum_{\mathbf{i} \in I^d} e_{\mathbf{i}, \mathbf{ff}} N_{\ell, \mathbf{i}}^{(k)}(\mathbf{x}) \right) \\ &= \sum_{\mathbf{i} \in I^d} \left(\sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} e_{\mathbf{i}, \mathbf{ff}} \right) N_{\ell, \mathbf{i}}^{(k)}(\mathbf{x}). \end{aligned}$$

Then

$$\begin{aligned} Qp &= \sum_{\mathbf{i} \in I^d} \left(\sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} e_{\mathbf{i}, \mathbf{ff}} \right) Q N_{\ell, \mathbf{i}}^{(k)} = \sum_{\mathbf{i} \in I} \left(\sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} e_{\mathbf{i}, \mathbf{ff}} \right) \sum_{\mathbf{j} \in I} \lambda_{\mathbf{j}}(N_{\ell, \mathbf{i}}^{(k)}) N_{\ell, \mathbf{j}}^{(k)} \\ &= \sum_{\mathbf{i} \in I} \left(\sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} e_{\mathbf{i}, \mathbf{ff}} \right) \sum_{\mathbf{j} \in I} \delta_{\mathbf{i}, \mathbf{j}} N_{\ell, \mathbf{j}}^{(k)} = \sum_{\mathbf{i} \in I} \left(\sum_{\mathbf{ff}=(0, \dots, 0)}^{(k-1, \dots, k-1)} c_{\mathbf{ff}} e_{\mathbf{i}, \mathbf{ff}} \right) N_{\ell, \mathbf{i}}^{(k)} = p, \end{aligned}$$

where in the third step we apply Proposition 4.3. □

For functions in Sobolev spaces, the normal Taylor polynomials may not exist. However, the averaged Taylor polynomials defined in the following for such a function always exists. The readers are referred to [5] for more discussions.

Definition 4.1. Suppose $f \in W^{s,p}(\Omega)$. B is the ball centered at $x_0 \in \Omega$ with radius ρ and $B \subset \subset \Omega$. The corresponding Taylor polynomial of order $s+1$ of f averaged over B is defined as

$$Q_B^{s+1}f(\mathbf{x}) = \int_B T_{\mathbf{y}}^{s+1}f(\mathbf{x})\phi(\mathbf{y})d\mathbf{y},$$

where

$$T_{\mathbf{y}}^{s+1}f(\mathbf{x}) = \sum_{|\mathbf{ff}| < s+1} \frac{1}{\mathbf{ff}!} D^{\mathbf{ff}}f(\mathbf{y})(\mathbf{x}-\mathbf{y})^\alpha,$$

and ϕ is a cut-off function supported in B .

In order to present Bramble-Hilbert lemma, which shows the error bound between function in Sobolev spaces and its averaged Taylor polynomials, we need the concepts of star-shaped region and chunkiness parameter.

Definition 4.2. Ω is star-shaped with respect to a ball B if, for all $\mathbf{x} \in \Omega$ there holds $\overline{\text{conv}}(\{\mathbf{x}\} \cup B) \subset \Omega$.

Definition 4.3. Suppose Ω has diameter h and is star-shaped with respect to a ball B . Let $\rho_{\max} = \sup\{\rho : \Omega \text{ is star-shaped with respect to a ball of radius } \rho\}$. Then the chunkiness parameter of Ω is defined by $\gamma = \frac{h}{\rho_{\max}}$.

Proposition 4.5 (Bramble-Hilbert). Let B be a ball in Ω such that Ω is star-shaped with respect to B and such that its radius $\rho > \frac{1}{2}\rho_{\max}$. Let $Q_B^s f$ be the Taylor polynomial of order s of f averaged over B with $f \in W^{s,p}(\Omega)$ and $p \geq 1$. Then

$$|f - Q_B^s f|_{W^{r,p}(\Omega)} \leq C(s,d,\gamma)h^{s-r}|f|_{W^{s,p}(\Omega)}, \quad r=0,1,\dots,s,$$

where $h = \text{diam}(\Omega)$.

Proof. See [5]. □

Theorem 4.1. Let $f \in W^{s,p}(\Omega)$ with $p \in [1,\infty)$ and Qf be defined by (4.3) with $k \geq s$, there holds

$$|f - Qf|_{W^{r,p}(\Omega)} \leq C(k,s,r,p,d) \left(\frac{1}{\ell}\right)^{s-r} |f|_{W^{s,p}(\Omega)}, \quad r=0,1,\dots,s.$$

Proof. We only show the case $p \in [1,\infty)$ and the case $p = \infty$ can be shown similarly. Let $f \in W^{s,p}(\Omega)$ and $\mathbf{r} = (r_1, \dots, r_d)$ with $|\mathbf{r}| = r$. We first deal with the local integral. Denoting

$$I_{\mathbf{i},k} := \prod_{j=1}^d \{i_j - k + 1, i_j - k + 2, \dots, i_j + k - 1\},$$

and

$$\widetilde{\Omega}_i := \bigcup_{j \in I_{i,k}} \Omega_j.$$

By Proposition 4.5, there exists a ball $B_i \subset \widetilde{\Omega}_i$ such that

$$\int_{\Omega_i} |D^r(f - T_{B_i}^s f)|^p dx \leq C(s, p, d) \left(\frac{3k}{\ell}\right)^{(s-r)p} |f|_{W^{s,p}(\widetilde{\Omega}_i)}^p. \tag{4.4}$$

Hence

$$\begin{aligned} \int_{\Omega_i} |D^r(f - Qf)|^p dx &\leq \int_{\Omega_i} (|D^r(f - T_{B_i}^s f)| + |D^r(T_{B_i}^s f - Qf)|)^p dx \\ &\leq 2^{p-1} \int_{\Omega_i} |D^r(f - T_{B_i}^s f)|^p dx + 2^{p-1} \int_{\Omega_i} |D^r(T_{B_i}^s f - Qf)|^p dx \\ &\leq C(s, p, d) (3kh)^{(s-r)p} |f|_{W^{s,p}(\widetilde{\Omega}_i)}^p + 2^{p-1} \int_{\Omega_i} |D^r Q(T_{B_i}^s f - f)|^p dx, \end{aligned} \tag{4.5}$$

where in the final step we apply Proposition 4.4. Now we deal with the integral related to operator Q . For $x \in \Omega_i$, we have

$$\begin{aligned} |D^r Q(T_{B_i}^s f - f)| &= \left| \sum_{j \in I^d} \lambda_j (T_{B_i}^s f - f) D^r N_{l,j}^{(k)} \right| = \left| \sum_{j \in I_{i,k}} \lambda_j (T_{B_i}^s f - f) D^r N_{l,j}^{(k)} \right| \\ &= \left| \sum_{j \in I_{i,k}} \lambda_j (T_{B_i}^s f - f) \prod_{q=1}^d D^{r_q} N_{l,j_q}^{(k)} \right| \\ &\leq 2^r \ell^r \sum_{j \in I_{i,k}} |\lambda_j (T_{B_i}^s f - f)| \\ &\leq 2^r (2k+1)^d 9^{d(k-1)} k^{-d/p} \ell^{r+d/p} \sum_{j \in I_{i,k}} \|f - T_{B_i}^s f\|_{L^p(\Omega_j)} \\ &\leq 2^r (2k+1)^d 9^{d(k-1)} k^{-d/p} \ell^{r+d/p} |I_{i,k}| \|f - T_{B_i}^s f\|_{L^p(\widetilde{\Omega}_i)} \\ &\leq 2^r (4k^2 - 1)^d 9^{d(k-1)} k^{-d/p} \ell^{r+d/p} \|f - T_{B_i}^s f\|_{L^p(\widetilde{\Omega}_i)} \\ &\leq 2^r (4k^2 - 1)^d 9^{d(k-1)} k^{-d/p} \left(\frac{1}{\ell}\right)^{s-r-d/p} (3k)^s |f|_{W^{s,p}(\widetilde{\Omega}_i)}, \end{aligned}$$

where in the second step we apply Proposition 4.1(1), in the fourth step we use Proposition 4.2, in the fifth step we use Proposition 4.3 and in the last step we apply (4.4). Hence

$$\begin{aligned} \int_{\Omega_i} |D^r Q(T_{B_i}^s f - f)|^p dx &\leq C(k, s, r, p, d) \left(\frac{1}{\ell}\right)^{(s-r)p-d} |f|_{W^{s,p}(\widetilde{\Omega}_i)}^p |\Omega_i| \\ &= C(k, s, r, p, d) \left(\frac{1}{\ell}\right)^{(s-r)p} |f|_{W^{s,p}(\widetilde{\Omega}_i)}^p. \end{aligned} \tag{4.6}$$

Combining (4.5) and (4.6) and summing all the subregions, we have the global estimate:

$$\begin{aligned} \int_{\Omega} |D^r(f - Qf)|^p dx &= \sum_{\mathbf{i} \in I^d} \int_{\Omega_{\mathbf{i}}} |D^r(f - Qf)|^p dx \leq C(k, s, r, p, d) \left(\frac{1}{\ell}\right)^{(s-r)p} \sum_{\mathbf{i} \in I^d} |f|_{W^{s,p}(\tilde{\Omega}_{\mathbf{i}})}^p \\ &\leq C(k, s, r, p, d) \left(\frac{1}{\ell}\right)^{(s-r)p} k^d |f|_{W^{s,p}(\Omega)}^p \\ &= C(k, s, r, p, d) \left(\frac{1}{\ell}\right)^{(s-r)p} |f|_{W^{s,p}(\Omega)}^p. \end{aligned}$$

Hence

$$|f - Qf|_{W^{r,p}(\Omega)}^p = \sum_{|\mathbf{r}|=r} \int_{\Omega} |D^{\mathbf{r}}(f - Qf)|^p dx \leq C(k, s, r, p, d) \left(\frac{1}{\ell}\right)^{(s-r)p} |f|_{W^{s,p}(\Omega)}^p.$$

This completes the proof. □

When $f \in C^r(\Omega)$, from the case $p = \infty$ in Theorem 4.1 we immediately obtain

Corollary 4.1. *Let $f \in C^s(\Omega)$ with $s \geq 0$ and Qf be defined by (4.3) with $k \geq s$, there holds*

$$\|f - Qf\|_{C^r(\Omega)} \leq C(k, s, r, d) \left(\frac{1}{\ell}\right)^{s-r} \|f\|_{C^s(\Omega)}, \quad r = 0, 1, \dots, s.$$

4.2 Approximation by neural network

Now we bound the approximation error of deep ReLU² networks.

Theorem 4.2. *Assume that $u^* \in C^2(\bar{\Omega})$. For any $\epsilon > 0$, there exists a ReLU² network u with depth $\lceil \log_2 d \rceil + 2$ and width $C_{d, \|u^*\|_{C^2(\Omega)}} \left(\frac{1}{\epsilon}\right)^d$ such that*

$$\|u^* - u\|_{C^1(\bar{\Omega})} \leq \epsilon.$$

Proof. By Theorem 4.1 and Corollary 4.1, we know that there exists $\{c_{\mathbf{i}}\}_{\mathbf{i} \in I^d} \subset \mathbb{R}$ such that

$$\left\| u^* - \sum_{\mathbf{i} \in I^d} c_{\mathbf{i}} N_{\ell, \mathbf{i}}^{(3)} \right\|_{C^1(\bar{\Omega})} \leq C(d) \|u^*\|_{C^2(\bar{\Omega})} \cdot \frac{1}{\ell}. \tag{4.7}$$

Now we show that $\{N_{\ell, \mathbf{i}}^{(3)}\}$ can be implemented by \mathcal{N}^2 without any error. Denote

$$\sigma(x) = \begin{cases} x^2, & x \geq 0, \\ 0, & \text{else,} \end{cases}$$

as the activation function in ReLU² network. By definition of $N_{l,i}^{(3)}(x)$ in (4.2), it's clear that $N_{l,i}^{(3)}(x)$ can be implemented by ReLU² network without any error with depth 2 and width 7. On the other hand ReLU² network can also realize multiplication without any error. In fact, for any $x, y \in \mathbb{R}$,

$$xy = \frac{1}{4}[(x+y)^2 - (x-y)^2] = \frac{1}{4}[\sigma(x+y) + \sigma(-x-y) - \sigma(x-y) - \sigma(y-x)].$$

Hence multivariate B-spline of order 3 can be implemented by ReLU² network exactly with depth $\lceil \log_2 d \rceil + 2$ and width $7d$.

Hence by (4.7) we conclude that there exists a ReLU² network u with depth $\lceil \log_2 d \rceil + 2$ and width $C_d \|u^*\|_{C^2(\bar{\Omega})} \left(\frac{1}{\epsilon}\right)^d$ such that $\|u^* - u\|_{C^1(\bar{\Omega})} \leq \epsilon$. □

5 Statistical error

In this section, we bound the statistical error

$$\mathcal{E}_{sta} = 2 \sup_{u \in \mathcal{N}^2} |\mathcal{L}(u) - \widehat{\mathcal{L}}(u)|.$$

Lemma 5.1.

$$\sup_{u \in \mathcal{N}^2} |\mathcal{L}(u) - \widehat{\mathcal{L}}(u)| \leq \sum_{j=1}^4 \sup_{u \in \mathcal{N}^2} |\mathcal{L}_j(u) - \widehat{\mathcal{L}}_j(u)|,$$

where

$$\mathcal{L}_1(u) = |\Omega| \mathbb{E}_{X \sim U(\Omega)} [w(X)u^2(X)/2], \quad \widehat{\mathcal{L}}_1(u) = \frac{|\Omega|}{N} \sum_{i=1}^N \left[\frac{w(X_i)u^2(X_i)}{2} \right],$$

$$\mathcal{L}_2(u) = |\Omega| \mathbb{E}_{X \sim U(\Omega)} [u(X)f(X)], \quad \widehat{\mathcal{L}}_2(u) = \frac{|\Omega|}{N} \sum_{i=1}^N [u(X_i)f(X_i)],$$

$$\mathcal{L}_3(u) = |\partial\Omega| \mathbb{E}_{Y \sim U(\partial\Omega)} [Tu(Y)g(Y)], \quad \widehat{\mathcal{L}}_3(u) = \frac{|\partial\Omega|}{M} \sum_{j=1}^M [u(Y_j)g(Y_j)],$$

$$\mathcal{L}_4(u) = |\Omega| \mathbb{E}_{X \sim U(\Omega)} [\|\nabla u(X)\|_2^2/2], \quad \widehat{\mathcal{L}}_4(u) = \frac{|\Omega|}{N} \sum_{i=1}^N \left[\frac{\|\nabla u(X_i)\|_2^2}{2} \right].$$

Proof. This lemma holds by the direct consequence of triangle inequality. □

By Lemma 5.1, we have to bound the maximum value of four random processes indexed by $u \in \mathcal{N}^2$. To this end, we recall tools in empirical process [20, 44]. Denote

$$\mathcal{B} = \max\{\|\bar{u}_{\bar{\phi}}\|_{L^\infty(\Omega)}, \|\|\nabla \bar{u}_{\bar{\phi}}\|_2^2\|_{L^\infty(\Omega)}\}, \tag{5.1}$$

where \bar{u}_ϕ is the best approximation of u^* in Theorem 4.2. Let f, g, w be bounded, say by some constant c_3 , i.e, we assume that

$$\|f\|_{L^\infty(\Omega)} \vee \|w\|_{L^\infty(\Omega)} \vee \|g\|_{L^\infty(\partial\Omega)} \vee \mathcal{B} \leq c_3 < \infty.$$

We use μ to denote $U(\Omega)(U(\partial\Omega))$. Given $n = N(M)$ i.i.d samples $\mathbf{Z}_n = \{Z_i\}_{i=1}^n$ from μ , with $Z_i = X_i(Y_i) \sim \mu$, we need the following Rademacher complexity to measure the capacity of the given function class \mathcal{N} restricted on n random samples \mathbf{Z}_n .

Definition 5.1. The Rademacher complexity of a set $A \subseteq \mathbb{R}^n$ is defined as

$$\mathfrak{R}(A) = \mathbb{E}_{\mathbf{Z}_n, \Sigma_n} \left[\sup_{a \in A} \frac{1}{n} \left| \sum_i \sigma_i a_i \right| \right],$$

where, $\Sigma_n = \{\sigma_i\}_{i=1}^n$ are n i.i.d Rademacher variables with $\mathbb{P}(\sigma_i = 1) = \mathbb{P}(\sigma_i = -1) = \frac{1}{2}$. The Rademacher complexity of function class \mathcal{N} associate with random sample \mathbf{Z}_n is defined as

$$\mathfrak{R}(\mathcal{N}) = \mathbb{E}_{\mathbf{Z}_n, \Sigma_n} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_i \sigma_i u(Z_i) \right| \right].$$

Lemma 5.2. Let $\Psi_1(x, y) = \frac{w(x)y^2}{2} : \mathbb{R}^d \times \mathbb{R}, |y| \leq c_3, \Psi_2(x, y) = f(x)y : \mathbb{R}^d \times \mathbb{R}, |y| \leq c_3, \Psi_3(x, y) = g(x)y : \mathbb{R}^d \times \mathbb{R}, |y| \leq c_3$. Then $\Psi_1(x, y), \Psi_2(x, y)$ and $\Psi_3(x, y)$ are c_3^2, c_3 and c_3 -Lipschitz continuous on y for all x and $\Psi_j(x, 0) = 0, j = 1, 2, 3$.

Proof. We give the proof for Ψ_1 and omit the details for Ψ_2, Ψ_3 since they can be shown similarly. For arbitrary y_1, y_2 with $|y_i| \leq c_3, i = 1, 2$,

$$\begin{aligned} |\Psi_1(x, y_1) - \Psi_1(x, y_2)| &= \left| \frac{w(x)y_1^2}{2} - \frac{w(x)y_2^2}{2} \right| \\ &= \frac{|w(x)(y_1 + y_2)|}{2} |y_1 - y_2| \leq c_3^2 |y_1(x) - y_2|. \end{aligned}$$

This completes the proof. □

By Corollary 3.17 in [20] and Lemma 5.2, we have the following Lipschitz contraction results on Rademacher complexity.

Lemma 5.3. Let $\mathcal{N} = \{u(x) : \|u\|_{L^\infty(\Omega)} \leq c_3\}$. Define

$$\Psi_j \circ \mathcal{N} = \{\text{composition of } \Psi_j \text{ and } \mathcal{N} : x \rightarrow \Psi_j(x, u(x)) : u \in \mathcal{N}\}, \quad j = 1, 2, 3.$$

Then $\mathfrak{R}(\Psi_1 \circ \mathcal{N}) \leq c_3^2 \mathfrak{R}(\mathcal{N}), \mathfrak{R}(\Psi_i \circ \mathcal{N}) \leq c_3 \mathfrak{R}(\mathcal{N}), i = 2, 3$.

The following symmetrization result shows that the Rademacher complexity $\mathfrak{R}(\Psi_j \circ \mathcal{N}^2)$ gives upper bound on $\sup_{u \in \mathcal{N}^2} |\mathcal{L}_j(u) - \widehat{\mathcal{L}}_j(u)|, j = 1, 2, 3$.

Lemma 5.4. $\mathbb{E}_{\mathbf{Z}_n} [\sup_{u \in \mathcal{N}^2} |\mathcal{L}_j(u) - \widehat{\mathcal{L}}_j(u)|] \leq 2\mathfrak{R}(\Psi_j \circ \mathcal{N}^2), j = 1, 2, 3.$

Proof. We give the proof for $j = 1$ and omit the proof for Ψ_2 and Ψ_3 since they can be shown similarly. Let $\tilde{\mathbf{Z}}_n = \{\tilde{Z}_i\}_{i=1}^n$ be an i.i.d ghost sample from μ and $\tilde{\mathbf{Z}}_n$ is independent of \mathbf{Z}_n .

$$\begin{aligned} & \mathbb{E}_{\mathbf{Z}_n} \left[\sup_{u \in \mathcal{N}^2} |\mathcal{L}_1(u) - \widehat{\mathcal{L}}_1(u)| \right] \\ &= \frac{|\Omega|}{n} \mathbb{E}_{\mathbf{Z}_n} \left[\sup_{u \in \mathcal{N}^2} \left| \sum_{i=1}^n \left(\mathbb{E}_{\tilde{\mathbf{Z}}_n} \left[\frac{w(\tilde{Z}_i)u^2(\tilde{Z}_i)}{2} \right] - \frac{w(Z_i)u^2(Z_i)}{2} \right) \right| \right] \\ &\leq \frac{|\Omega|}{n} \mathbb{E}_{\mathbf{Z}_n, \tilde{\mathbf{Z}}_n} \left[\sup_{u \in \mathcal{N}^2} \left| \sum_{i=1}^n \left(\frac{w(\tilde{Z}_i)u^2(\tilde{Z}_i)}{2} - \frac{w(Z_i)u^2(Z_i)}{2} \right) \right| \right] \\ &= \frac{|\Omega|}{n} \mathbb{E}_{\mathbf{Z}_n, \tilde{\mathbf{Z}}_n, \Sigma_n} \left[\sup_{u \in \mathcal{N}^2} \left| \sum_{i=1}^n \sigma_i \left(\frac{w(\tilde{Z}_i)u^2(\tilde{Z}_i)}{2} - \frac{w(Z_i)u^2(Z_i)}{2} \right) \right| \right] \\ &\leq \frac{|\Omega|}{n} \mathbb{E}_{\mathbf{Z}_n, \Sigma} \left[\sup_{u \in \mathcal{N}^2} \left| \sum_{i=1}^n \frac{w(Z_i)u^2(Z_i)}{2} \right| \right] + \frac{|\Omega|}{n} \mathbb{E}_{\tilde{\mathbf{Z}}_n, \Sigma} \left[\sup_{u \in \mathcal{N}^2} \left| \sum_{i=1}^n \frac{w(\tilde{Z}_i)u^2(\tilde{Z}_i)}{2} \right| \right] \\ &= 2|\Omega|\mathfrak{R}(\Psi_1 \circ \mathcal{N}^2), \end{aligned}$$

where the first inequality follows from the Jensen’s inequality, and the second equality holds since both $\sigma_i(\frac{w(\tilde{Z}_i)u^2(\tilde{Z}_i)}{2} - \frac{w(Z_i)u^2(Z_i)}{2})$ and $(\frac{w(\tilde{Z}_i)u^2(\tilde{Z}_i)}{2} - \frac{w(Z_i)u^2(Z_i)}{2})$ are governed by the same law, and the last equality holds since the distribution of the two terms are the same. \square

Next we give an upper bound of $\mathfrak{R}(\mathcal{N}^2)$ in terms of the covering number of \mathcal{N}^2 by using the Dudley’s entropy formula [11].

Definition 5.2. Suppose that $W \subset \mathbb{R}^n$. For any $\epsilon > 0$, let $V \subset \mathbb{R}^n$ be a ϵ -cover of W with respect to the distance d_∞ , that is, for any $w \in W$, there exists a $v \in V$ such that $d_\infty(w, v) < \epsilon$, where d_∞ is defined by

$$d_\infty(u, v) := \|u - v\|_\infty.$$

The covering number $\mathcal{C}(\epsilon, W, d_\infty)$ is defined to be the minimum cardinality among all ϵ -cover of W with respect to the distance d_∞ .

Definition 5.3. Suppose that \mathcal{N} is a class of functions from Ω to \mathbb{R} . Given n sample $\mathbf{Z}_n = (Z_1, Z_2, \dots, Z_n) \in \Omega^n$, $\mathcal{N}|_{\mathbf{Z}_n} \subset \mathbb{R}^n$ is defined by

$$\mathcal{N}|_{\mathbf{Z}_n} = \{(u(Z_1), u(Z_2), \dots, u(Z_n)) : u \in \mathcal{N}\}.$$

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

$$\mathcal{C}_\infty(\epsilon, \mathcal{N}, n) = \max_{\mathbf{Z}_n \in \Omega^n} \mathcal{C}(\epsilon, \mathcal{N}|_{\mathbf{Z}_n}, d_\infty).$$

We need the Massart’s finite class lemma [4] for preparation.

Lemma 5.5. For any finite set $V \in \mathbb{R}^n$ with diameter $D = \sup_{v \in V} \|v\|_2$, then

$$\mathbb{E}_{\Sigma_n} \left[\sup_{v \in V} \frac{1}{n} \left| \sum_i \sigma_i v_i \right| \right] \leq \frac{D}{n} \sqrt{2 \log(2|V|)}.$$

Lemma 5.6. Assume $0 \in \mathcal{N}$ and the diameter of \mathcal{N} is less than \mathcal{B} , i.e., $\|u\|_{L^\infty(\Omega)} \leq \mathcal{B}, \forall u \in \mathcal{N}$. Then

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

Proof. The proof follows from the chaining method [44]. By definition

$$\mathfrak{R}(\mathcal{N}) = \mathfrak{R}(\mathcal{N} | \mathbf{Z}_n) = \mathbb{E}_{\mathbf{Z}_n} \left[\mathbb{E}_{\Sigma} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_i \sigma_i u(Z_i) \right| \middle| \mathbf{Z}_n \right] \right].$$

With the help of Lemma 5.5, it suffice to show

$$\mathbb{E}_{\Sigma} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_i \sigma_i u(Z_i) \right| \right] \leq \inf_{0 < \delta < \mathcal{B}} \left(4\delta + \frac{12}{\sqrt{n}} \int_\delta^{\mathcal{B}} \sqrt{2 \log \mathcal{C}(\epsilon, \mathcal{N}^2, n)} d\epsilon \right)$$

by conditioning on \mathbf{Z}_n . Given an positive integer K , let $\epsilon_k = 2^{-k+1} \mathcal{B}, k = 1, \dots, K$. Let C_k be a cover of $\mathcal{N} | \mathbf{Z}_n \subseteq \mathbb{R}^n$ whose covering number is denoted as $\mathcal{C}(\epsilon_k, \mathcal{N} | \mathbf{Z}_n, d_\infty)$. Then, by definition, $\forall u \in \mathcal{N}$, there $\exists c^k \in C_k$ such that

$$d_\infty(u | \mathbf{Z}_n, c^k) = \max\{|u(Z_i) - c_i^k|, i = 1, \dots, n\} \leq \epsilon_k, \quad k = 1, \dots, K.$$

Moreover, we denote the best approximate element of u in C_k with respect to d_∞ as $c^k(u)$. Then,

$$\begin{aligned} & \mathbb{E}_{\Sigma} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i u(Z_i) \right| \right] \\ &= \mathbb{E}_{\Sigma} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i (u(Z_i) - c_i^K(u)) + \sum_{j=1}^{K-1} \sum_{i=1}^n \sigma_i (c_i^j(u) - c_i^{j+1}(u)) + \sum_{i=1}^n \sigma_i c_i^1(u) \right| \right] \\ &\leq \mathbb{E}_{\Sigma} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i (u(Z_i) - c_i^K(u)) \right| \right] + \sum_{j=1}^{K-1} \mathbb{E}_{\Sigma} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i (c_i^j(u) - c_i^{j+1}(u)) \right| \right] \\ &+ \mathbb{E}_{\Sigma} \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i c_i^1(u) \right| \right]. \end{aligned}$$

Since $0 \in \mathcal{N}$, and the diameter of \mathcal{N} is smaller than \mathcal{B} , we can choose $C_1 = \{0\}$ such that the third term in the above display vanishes. By Hölder’s inequality, we deduce that the

first term can be bounded by ϵ_K as follows.

$$\begin{aligned} & \mathbb{E}_\Sigma \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i(u(Z_i) - c_i^K(u)) \right| \right] \\ & \leq \mathbb{E}_\Sigma \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left(\sum_{i=1}^n |\sigma_i| \right) \left(\sum_{i=1}^n \max_{i=1, \dots, n} \{ |u(Z_i) - c_i^K(u)| \} \right) \right] \\ & \leq \epsilon_K. \end{aligned}$$

Let $V_j = \{c^j(u) - c^{j+1}(u) : u \in \mathcal{N}\}$. Then by definition, the number of elements in V_j and C_j satisfying

$$|V_j| \leq |C_j| |C_{j+1}| \leq |C_{j+1}|^2.$$

And the diameter of V_j denoted as D_j can be bounded as

$$\begin{aligned} D_j &= \sup_{v \in V_j} \|v\|_2 \leq \sqrt{n} \sup_{u \in \mathcal{N}} \|c^j(u) - c^{j+1}(u)\|_\infty \\ &\leq \sqrt{n} \sup_{u \in \mathcal{N}} \|c^j(u) - u\|_\infty + \|u - c^{j+1}(u)\|_\infty \\ &\leq \sqrt{n}(\epsilon_j + \epsilon_{j+1}) \\ &\leq 3\sqrt{n}\epsilon_{j+1}. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{E}_\Sigma \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_{j=1}^{K-1} \sum_{i=1}^n \sigma_i(c_i^j(u) - c_i^{j+1}(u)) \right| \right] &\leq \sum_{j=1}^{K-1} \mathbb{E}_\Sigma \left[\sup_{v \in V_j} \frac{1}{n} \left| \sum_{i=1}^n \sigma_i v_j \right| \right] \\ &\leq \sum_{j=1}^{K-1} \frac{D_j}{n} \sqrt{2 \log(2|V_j|)} \\ &\leq \sum_{j=1}^{K-1} \frac{6\epsilon_{j+1}}{\sqrt{n}} \sqrt{\log(2|C_{j+1}|)}, \end{aligned}$$

where we use triangle inequality in the first inequality, and use Lemma 5.5 in the second inequality. Putting all the above estimates together, we get

$$\begin{aligned} \mathbb{E}_\Sigma \left[\sup_{u \in \mathcal{N}} \frac{1}{n} \left| \sum_i \sigma_i u(Z_i) \right| \right] &\leq \epsilon_K + \sum_{j=1}^{K-1} \frac{6\epsilon_{j+1}}{\sqrt{n}} \sqrt{\log(2|C_{j+1}|)} \\ &\leq \epsilon_K + \sum_{j=1}^K \frac{12(\epsilon_j - \epsilon_{j+1})}{\sqrt{n}} \sqrt{\log(2\mathcal{C}(\epsilon_j, \mathcal{N}, n))} \\ &\leq \epsilon_K + \frac{12}{\sqrt{n}} \int_{\epsilon_{K+1}}^B \sqrt{\log(2\mathcal{C}(\epsilon, \mathcal{N}, n))} d\epsilon \\ &\leq \inf_{0 < \delta < B} \left(4\delta + \frac{12}{\sqrt{n}} \int_\delta^B \sqrt{\log(2\mathcal{C}(\epsilon, \mathcal{N}, n))} d\epsilon \right), \end{aligned}$$

where last inequality holds since for $0 < \delta < \mathcal{B}$, we can choose K to be the largest integer such that $\epsilon_{K+1} > \delta$, at this time $\epsilon_K \leq 4\epsilon_{K+2} \leq 4\delta$. \square

Now we turn to handle the most difficult term $\sup_{u \in \mathcal{N}^2} |\mathcal{L}_4(u) - \widehat{\mathcal{L}}_4(u)|$, where we need to bound the Rademacher complexity of the non-Lipschitz composition of gradient norm and ReLU² network. We believe that the technique we used here is helpful for bounding the statistical errors for other deep PDEs solvers where the main difficulties is bounding the Rademacher complexity of non-Lipschitz composition induced by the gradient operator.

Lemma 5.7.

$$\begin{aligned} & \mathbb{E}_{\mathbf{Z}_n} \left[\sup_{u \in \mathcal{N}^2} |\mathcal{L}_4(u) - \widehat{\mathcal{L}}_4(u)| \right] \\ & \leq \mathbb{E}_{\mathbf{Z}_n, \Sigma_n} \left[\sup_{u \in \mathcal{N}^2} \frac{1}{n} \left| \sum_i \sigma_i \|\nabla u(Z_i)\|^2 \right| \right] \end{aligned} \tag{5.2}$$

$$\leq \mathfrak{R}(\mathcal{N}^{1,2}) = \mathbb{E}_{\mathbf{Z}_n, \Sigma_n} \left[\sup_{u \in \mathcal{N}^{1,2}} \frac{1}{n} \left| \sum_i \sigma_i u(Z_i) \right| \right] \tag{5.3}$$

$$\leq \inf_{0 < \delta < \mathcal{B}} \left(4\delta + \frac{12}{\sqrt{n}} \int_\delta^{\mathcal{B}} \sqrt{\log(2\mathcal{C}(\epsilon, \mathcal{N}^{1,2}, n))} d\epsilon \right). \tag{5.4}$$

Proof. The proof of (5.2) is based on the symmetrization method used in the proof of Lemma 5.4, we omit the detail here. The proof of (5.3) is a direct consequence of the following claim. Claim: Let u be a function implemented by a ReLU² network with depth \mathcal{D} and width \mathcal{W} . Then $\|\nabla u\|_2^2$ can be implemented by a ReLU-ReLU² network with depth $\mathcal{D}+3$ and width $d(\mathcal{D}+2)\mathcal{W}$.

Denote ReLU and ReLU² as σ_1 and σ_2 , respectively. As long as we show that each partial derivative $D_i u$ ($i = 1, 2, \dots, d$) can be implemented by a ReLU-ReLU² network respectively, we can easily obtain the network we desire, since, $\|\nabla u\|_2^2 = \sum_{i=1}^d |D_i u|^2$ and the square function can be implemented by $x^2 = \sigma_2(x) + \sigma_2(-x)$.

Now we show that for any $i = 1, 2, \dots, d$, $D_i u$ can be implemented by a ReLU-ReLU² network. We deal with the first two layers in details since there are a little bit difference for the first two layer and apply induction for layers $k \geq 3$. For the first layer, since $\sigma_2'(x) = 2\sigma_1(x)$, we have for any $q = 1, 2, \dots, n_1$

$$D_i u_q^{(1)} = D_i \sigma_2 \left(\sum_{j=1}^d a_{qj}^{(1)} x_j + b_q^{(1)} \right) = 2\sigma_1 \left(\sum_{j=1}^d a_{qj}^{(1)} x_j + b_q^{(1)} \right) \cdot a_{qi}^{(1)}.$$

Hence $D_i u_q^{(1)}$ can be implemented by a ReLU-ReLU² network with depth 2 and width \mathcal{W} . For the second layer,

$$D_i u_q^{(2)} = D_i \sigma_2 \left(\sum_{j=1}^{n_1} a_{qj}^{(2)} u_j^{(1)} + b_q^{(2)} \right) = 2\sigma_1 \left(\sum_{j=1}^{n_1} a_{qj}^{(2)} u_j^{(1)} + b_q^{(2)} \right) \cdot \sum_{j=1}^{n_1} a_{qj}^{(2)} D_i u_j^{(1)}.$$

Since $\sigma_1\left(\sum_{j=1}^{n_1} a_{qj}^{(2)} u_j^{(1)} + b_q^{(2)}\right)$ and $\sum_{j=1}^{n_1} a_{qj}^{(2)} D_i u_j^{(1)}$ can be implemented by two ReLU-ReLU² subnetworks, respectively, and the multiplication can also be implemented by

$$\begin{aligned} x \cdot y &= \frac{1}{4} [(x+y)^2 - (x-y)^2] \\ &= \frac{1}{4} [\sigma_2(x+y) + \sigma_2(-x-y) - \sigma_2(x-y) - \sigma_2(-x+y)], \end{aligned}$$

we conclude that $D_i u_q^{(2)}$ can be implemented by a ReLU-ReLU² network. We have

$$\mathcal{D}\left(\sigma_1\left(\sum_{j=1}^{n_1} a_{qj}^{(2)} u_j^{(1)} + b_q^{(2)}\right)\right) = 2, \quad \mathcal{W}\left(\sigma_1\left(\sum_{j=1}^{n_1} a_{qj}^{(2)} u_j^{(1)} + b_q^{(2)}\right)\right) \leq \mathcal{W}$$

and

$$\mathcal{D}\left(\sum_{j=1}^{n_1} a_{qj}^{(2)} D_i u_j^{(1)}\right) = 2, \quad \mathcal{W}\left(\sum_{j=1}^{n_1} a_{qj}^{(2)} D_i u_j^{(1)}\right) \leq \mathcal{W}.$$

Thus $\mathcal{D}(D_i u_q^{(2)}) = 3$, $\mathcal{W}(D_i u_q^{(2)}) \leq \max\{2\mathcal{W}, 4\}$.

Now we apply induction for layers $k \geq 3$. For the third layer,

$$D_i u_q^{(3)} = D_i \sigma_2\left(\sum_{j=1}^{n_2} a_{qj}^{(3)} u_j^{(2)} + b_q^{(3)}\right) = 2\sigma_1\left(\sum_{j=1}^{n_2} a_{qj}^{(3)} u_j^{(2)} + b_q^{(3)}\right) \cdot \sum_{j=1}^{n_2} a_{qj}^{(3)} D_i u_j^{(2)}.$$

Since $\mathcal{D}(\sigma_1(\sum_{j=1}^{n_2} a_{qj}^{(3)} u_j^{(2)} + b_q^{(3)})) = 3$, $\mathcal{W}(\sigma_1(\sum_{j=1}^{n_2} a_{qj}^{(3)} u_j^{(2)} + b_q^{(3)})) \leq \mathcal{W}$ and

$$\mathcal{D}\left(\sum_{j=1}^{n_2} a_{qj}^{(3)} D_i u_j^{(2)}\right) = 4, \quad \mathcal{W}\left(\sum_{j=1}^{n_2} a_{qj}^{(3)} D_i u_j^{(2)}\right) \leq \max\{2\mathcal{W}, 4\mathcal{W}\} = 4\mathcal{W},$$

we conclude that $D_i u_q^{(3)}$ can be implemented by a ReLU-ReLU² network and $\mathcal{D}(D_i u_q^{(3)}) = 5$, $\mathcal{W}(D_i u_q^{(3)}) \leq \max\{5\mathcal{W}, 4\} = 5\mathcal{W}$.

We assume that $D_i u_q^{(k)}$ ($q=1, 2, \dots, n_k$) can be implemented by a ReLU-ReLU² network and $\mathcal{D}(D_i u_q^{(k)}) = k+2$, $\mathcal{W}(D_i u_q^{(k)}) \leq (k+2)\mathcal{W}$. For the $(k+1)$ -th layer,

$$\begin{aligned} D_i u_q^{(k+1)} &= D_i \sigma_2\left(\sum_{j=1}^{n_k} a_{qj}^{(k+1)} u_j^{(k)} + b_q^{(k+1)}\right) \\ &= 2\sigma_1\left(\sum_{j=1}^{n_k} a_{qj}^{(k+1)} u_j^{(k)} + b_q^{(k+1)}\right) \cdot \sum_{j=1}^{n_k} a_{qj}^{(k+1)} D_i u_j^{(k)}. \end{aligned}$$

Since

$$\mathcal{D}\left(\sigma_1\left(\sum_{j=1}^{n_k} a_{qj}^{(k+1)} u_j^{(k)} + b_q^{(k+1)}\right)\right) = k+1, \quad \mathcal{W}\left(\sigma_1\left(\sum_{j=1}^{n_k} a_{qj}^{(k+1)} u_j^{(k)} + b_q^{(k+1)}\right)\right) \leq \mathcal{W}$$

and

$$\mathcal{D}\left(\sum_{j=1}^{n_k} a_{qj}^{(k+1)} D_i u_j^{(k)}\right) = k+2, \quad \mathcal{W}\left(\sum_{j=1}^{n_k} a_{qj}^{(k+1)} D_i u_j^{(k)}\right) \leq \max\{(k+2)\mathcal{W}, 4\mathcal{W}\} = (k+2)\mathcal{W},$$

we conclude that $D_i u_q^{(k+1)}$ can be implemented by a ReLU-ReLU² network and $\mathcal{D}(D_i u_q^{(k+1)}) = k+3, \mathcal{W}(D_i u_q^{(k+1)}) \leq \max\{(k+3)\mathcal{W}, 4\} = (k+3)\mathcal{W}$.

Hence we derive that $D_i u = D_i u_1^{\mathcal{D}}$ can be implemented by a ReLU-ReLU² network and $\mathcal{D}(D_i u) = \mathcal{D}+2, \mathcal{W}(D_i u) \leq (\mathcal{D}+2)\mathcal{W}$. Finally we obtain that $\mathcal{D}(\|\nabla u\|^2) = \mathcal{D}+3, \mathcal{W}(\|\nabla u\|^2) \leq d(\mathcal{D}+2)\mathcal{W}$.

The proof of (5.4) follows from Lemma 5.6. □

By Lemma 5.4, Lemma 5.3, Lemma 5.6 and Lemma 5.7, we have to find upper bounds for the converging numbers $\mathcal{C}(\epsilon, \mathcal{N}^2, n)$ and $\mathcal{C}(\epsilon, \mathcal{N}^{1,2}, n)$ used in the Dudley’s entropy formula. To this end, we need the VC-dimension [45] and Pseudo-dimension [1].

Definition 5.4. Let \mathcal{N} be a set of functions from $X = \Omega(\partial\Omega)$ to $\{0,1\}$. Suppose that $S = \{x_1, x_2, \dots, x_n\} \subset X$. We say that S is shattered by \mathcal{N} if for any $b \in \{0,1\}^n$, there exists a $u \in \mathcal{N}$ satisfying

$$u(x_i) = b_i, \quad i = 1, 2, \dots, n.$$

Definition 5.5. The VC-dimension of \mathcal{N} , denoted as $\text{VCdim}(\mathcal{N})$, is defined to be the maximum cardinality among all sets shattered by \mathcal{N} .

VC-dimension reflects the capability of a class of functions to perform binary classification of points. The larger VC-dimension is, the stronger the capability to perform binary classification is. For more discussion of VC-dimension, readers are referred to [1].

For real-valued functions, we can generalize the concept of VC-dimension into pseudo-dimension [1].

Definition 5.6. Let \mathcal{N} be a set 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{N} if there exists y_1, y_2, \dots, y_n such that for any $b \in \{0,1\}^n$, there exists a $u \in \mathcal{N}$ satisfying

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

and we say that $\{y_i\}_{i=1}^n$ witnesses the shattering.

Definition 5.7. The pseudo-dimension of \mathcal{N} , denoted as $\text{Pdim}(\mathcal{N})$, is defined to be the maximum cardinality among all sets pseudo-shattered by \mathcal{N} .

The following proposition showing a relation between uniform covering number and pseudo-dimension.

Proposition 5.1 (Theorem 12.2, [1]). Let \mathcal{N} be a set of real functions from a domain X to the bounded interval $[0, \mathcal{B}]$. Let $\epsilon > 0$. Then

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

which is less than $\left(\frac{\epsilon n \mathcal{B}}{\epsilon \cdot \text{Pdim}(\mathcal{N})}\right)^{\text{Pdim}(\mathcal{N})}$ for $n \geq \text{Pdim}(\mathcal{N})$.

We now present the bound of pseudo-dimension for the $\mathcal{N}^{1,2}$, the class of network functions with ReLU and ReLU² activation functions. We first need a lemma stated below.

Lemma 5.8. Let p_1, \dots, p_m be polynomials with n variables of degree at most d . If $n \leq m$, then

$$|\{(\text{sign}(p_1(x)), \dots, \text{sign}(p_m(x))) : x \in \mathbb{R}^n\}| \leq 2 \left(\frac{2emd}{n}\right)^n.$$

Proof. See Theorem 8.3 in [1]. □

Theorem 5.1. Let

$$\begin{aligned} \mathcal{N} := \{ & u \in [0, 1]^d : u \text{ can be implemented by a neural network} \\ & \text{with depth no more than } \mathcal{D} \text{ and width no more than } \mathcal{W}, \\ & \text{and activation function in each unit be the ReLU or the ReLU}^2 \}. \end{aligned}$$

Then

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

Proof. The argument is follows from the proof of Theorem 6 in [2]. The result stated here is somewhat stronger than Theorem 6 in [2] since $\text{VCdim}(\text{sign}(\mathcal{N})) \leq \text{Pdim}(\mathcal{N})$.

We consider a new set of functions:

$$\widetilde{\mathcal{N}} = \{\widetilde{u}(x, y) = \text{sign}(u(x) - y) : u \in \mathcal{H}\}.$$

It is clear that $\text{Pdim}(\mathcal{N}) \leq \text{VCdim}(\widetilde{\mathcal{N}})$. We now bound the VC-dimension of $\widetilde{\mathcal{N}}$. Denoting \mathcal{M} as the total number of parameters(weights and biases) in the neural network implementing functions in \mathcal{N} , in our case we want to derive the uniform bound for

$$K_{\{x_i\}, \{y_i\}}(m) := |\{(\text{sign}(u(x_1, a) - y_1), \dots, \text{sign}(u(x_m, a) - y_m)) : a \in \mathbb{R}^{\mathcal{M}}\}|$$

over all $\{x_i\}_{i=1}^m \subset X$ and $\{y_i\}_{i=1}^m \subset \mathbb{R}$. Actually the maximum of $K_{\{x_i\}, \{y_i\}}(m)$ over all $\{x_i\}_{i=1}^m \subset X$ and $\{y_i\}_{i=1}^m \subset \mathbb{R}$ is the growth function $\mathcal{G}_{\widetilde{\mathcal{N}}}(m)$. In order to apply Lemma 5.8, we partition the parameter space $\mathbb{R}^{\mathcal{M}}$ into several subsets to ensure that in each subset $u(x_i, a) - y_i$ is a polynomial with respect to a without any breakpoints. In fact, our

partition is exactly the same as the partition in [2]. Denote the partition as $\{P_1, P_2, \dots, P_N\}$ with some integer N satisfying

$$N \leq \prod_{i=1}^{D-1} 2 \left(\frac{2emk_i(1+(i-1)2^{i-1})}{\mathcal{M}_i} \right)^{\mathcal{M}_i}, \tag{5.5}$$

where k_i and \mathcal{M}_i denotes the number of units at the i th layer and the total number of parameters at the inputs to units in all the layers up to layer i of the neural network implementing functions in \mathcal{N} , respectively. See [2] for the construction of the partition. Obviously we have

$$K_{\{x_i\},\{y_i\}}(m) \leq \sum_{i=1}^N |\{(\text{sign}(u(x_1,a) - y_1), \dots, \text{sign}(u(x_m,a) - y_m)) : a \in P_i)\}|. \tag{5.6}$$

Note that $u(x_i,a) - y_i$ is a polynomial with respect to a with degree the same as the degree of $u(x_i,a)$, which is equal to $1 + (\mathcal{D} - 1)2^{\mathcal{D}-1}$ as shown in [2]. Hence by Lemma 5.8, we have

$$\begin{aligned} & |\{(\text{sign}(u(x_1,a) - y_1), \dots, \text{sign}(u(x_m,a) - y_m)) : a \in P_i)\}| \\ & \leq 2 \left(\frac{2em(1 + (\mathcal{D} - 1)2^{\mathcal{D}-1})}{\mathcal{M}_D} \right)^{\mathcal{M}_D}. \end{aligned} \tag{5.7}$$

Combining (5.5), (5.6), (5.7) yields

$$K_{\{x_i\},\{y_i\}}(m) \leq \prod_{i=1}^D 2 \left(\frac{2emk_i(1+(i-1)2^{i-1})}{\mathcal{M}_i} \right)^{\mathcal{M}_i}.$$

We then have

$$\mathcal{G}_{\widehat{\mathcal{N}}}(m) \leq \prod_{i=1}^D 2 \left(\frac{2emk_i(1+(i-1)2^{i-1})}{\mathcal{M}_i} \right)^{\mathcal{M}_i},$$

since the maximum of $K_{\{x_i\},\{y_i\}}(m)$ over all $\{x_i\}_{i=1}^m \subset X$ and $\{y_i\}_{i=1}^m \subset \mathbb{R}$ is the growth function $\mathcal{G}_{\widehat{\mathcal{N}}}(m)$. Some algebras as that of the proof of Theorem 6 in [2], we obtain

$$\text{Pdim}(\mathcal{N}) \leq \mathcal{O}(\mathcal{D}^2\mathcal{W}^2 \log \mathcal{U} + \mathcal{D}^3\mathcal{W}^2) = \mathcal{O}(\mathcal{D}^2\mathcal{W}^2(\mathcal{D} + \log \mathcal{W})),$$

where \mathcal{U} refers to the number of units of the neural network implementing functions in \mathcal{N} . □

5.1 Main results

With the above preparation we present the main results of this paper in the scenario $\mathcal{E}_{opt} = 0$.

Theorem 5.2. Let u^* be the solution of (2.1) with bounded f, g, w and assume that $u^* \in C^2(\bar{\Omega})$. Let \hat{u}_ϕ be the minimizer of deep Ritz method defined in (2.3) with $n = N(M)$ random samples. If we set the network parameters depth and width in the ReLU² network $\mathcal{N}_{\mathcal{D}, \mathcal{W}, \mathcal{B}}^2$ as

$$\mathcal{D} = \lceil \log_2 d \rceil + 2, \quad \mathcal{W} = C_{d, \|u^*\|_{C^2(\bar{\Omega})}} n^{\frac{d}{d+2+\nu}}.$$

Then,

$$\mathbb{E}_{\mathbf{X}, \mathbf{Y}} [\|\hat{u}_\phi - u^*\|_{H^1(\Omega)}^2] \leq C_{\|u^*\|_{C^2(\bar{\Omega})}, c_1, c_2, c_3, d} \mathcal{O}(N^{-1/(d+2+\nu)} + M^{-1/(d+2+\nu)}),$$

where $\nu > 0$ but can be arbitrarily small.

Proof. In order to apply Lemma 5.6, we need to bound the term

$$\begin{aligned} & \frac{1}{\sqrt{n}} \int_\delta^{\mathcal{B}} \sqrt{\log(2\mathcal{C}(\epsilon, \mathcal{N}, n))} d\epsilon \\ & \leq \frac{\mathcal{B}}{\sqrt{n}} + \frac{1}{\sqrt{n}} \int_\delta^{\mathcal{B}} \sqrt{\log\left(\frac{en\mathcal{B}}{\epsilon \cdot \text{Pdim}(\mathcal{N})}\right)^{\text{Pdim}(\mathcal{N})}} d\epsilon \\ & \leq \frac{\mathcal{B}}{\sqrt{n}} + \left(\frac{\text{Pdim}(\mathcal{N})}{n}\right)^{1/2} \int_\delta^{\mathcal{B}} \sqrt{\log\left(\frac{en\mathcal{B}}{\epsilon \cdot \text{Pdim}(\mathcal{N})}\right)} d\epsilon, \end{aligned}$$

where in the first inequality we use Proposition 5.1. Now we calculate the integral. Set

$$t = \sqrt{\log\left(\frac{en\mathcal{B}}{\epsilon \cdot \text{Pdim}(\mathcal{N})}\right)}$$

then $\epsilon = \frac{en\mathcal{B}}{\text{Pdim}(\mathcal{N})} \cdot e^{-t^2}$. Denote

$$t_1 = \sqrt{\log\left(\frac{en\mathcal{B}}{\mathcal{B} \cdot \text{Pdim}(\mathcal{N})}\right)}, \quad t_2 = \sqrt{\log\left(\frac{en\mathcal{B}}{\delta \cdot \text{Pdim}(\mathcal{N})}\right)}.$$

And

$$\begin{aligned} & \int_\delta^{\mathcal{B}} \sqrt{\log\left(\frac{en\mathcal{B}}{\epsilon \cdot \text{Pdim}(\mathcal{N})}\right)} d\epsilon \\ & = \frac{2en\mathcal{B}}{\text{Pdim}(\mathcal{N})} \int_{t_1}^{t_2} t^2 e^{-t^2} dt \\ & = \frac{2en\mathcal{B}}{\text{Pdim}(\mathcal{N})} \int_{t_1}^{t_2} t \left(\frac{-e^{-t^2}}{2}\right)' dt \end{aligned}$$

$$\begin{aligned} &= \frac{en\mathcal{B}}{\text{Pdim}(\mathcal{N})} \left[t_1 e^{-t_1^2} - t_2 e^{-t_2^2} + \int_{t_1}^{t_2} e^{-t^2} dt \right] \\ &\leq \frac{en\mathcal{B}}{\text{Pdim}(\mathcal{N})} \left[t_1 e^{-t_1^2} - t_2 e^{-t_2^2} + (t_2 - t_1) e^{-t_1^2} \right] \\ &\leq \frac{en\mathcal{B}}{\text{Pdim}(\mathcal{N})} \cdot t_2 e^{-t_1^2} = \mathcal{B} \sqrt{\log \left(\frac{en\mathcal{B}}{\delta \cdot \text{Pdim}(\mathcal{N})} \right)}. \end{aligned}$$

Choosing $\delta = \mathcal{B} \left(\frac{\text{Pdim}(\mathcal{N})}{n} \right)^{1/2} \leq \mathcal{B}$, by Lemma 5.6 and the above display, we get for both

$$\mathcal{N} = \mathcal{N}^2 \quad \text{and} \quad \mathcal{N} = \mathcal{N}^{1,2},$$

there holds

$$\begin{aligned} &\mathfrak{R}(\mathcal{N}) \\ &\leq 4\delta + \frac{12}{\sqrt{n}} \int_{\delta}^{\mathcal{B}} \sqrt{\log(2\mathcal{C}(\epsilon, \mathcal{N}, n))} d\epsilon \\ &\leq 4\delta + \frac{12\mathcal{B}}{\sqrt{n}} + 12\mathcal{B} \left(\frac{\text{Pdim}(\mathcal{N})}{n} \right)^{1/2} \sqrt{\log \left(\frac{en\mathcal{B}}{\delta \cdot \text{Pdim}(\mathcal{N})} \right)} \\ &\leq 28\sqrt{\frac{3}{2}}\mathcal{B} \left(\frac{\text{Pdim}(\mathcal{N})}{n} \right)^{1/2} \sqrt{\log \left(\frac{en}{\text{Pdim}(\mathcal{N})} \right)}. \end{aligned} \tag{5.8}$$

Then by Lemmas 5.1, 5.3, 5.4, 5.6, 5.7 and Eq. (5.8), we have

$$\begin{aligned} \mathcal{E}_{sta} &= 2 \sup_{u \in \mathcal{N}^2} |\mathcal{L}(u) - \widehat{\mathcal{L}}(u)| \\ &\leq 2(c_3 + c_3 + c_3^2)\mathfrak{R}(\mathcal{N}^2) + 2\mathfrak{R}(\mathcal{N}^{1,2}) \\ &\leq 56\sqrt{\frac{3}{2}}(2c_3 + c_3^2)\mathcal{B} \left(\frac{\text{Pdim}(\mathcal{N}^2)}{n} \right)^{1/2} \sqrt{\log \left(\frac{en}{\text{Pdim}(\mathcal{N}^2)} \right)} \\ &\quad + 56\sqrt{\frac{3}{2}}\mathcal{B} \left(\frac{\text{Pdim}(\mathcal{N}^{1,2})}{n} \right)^{1/2} \sqrt{\log \left(\frac{en}{\text{Pdim}(\mathcal{N}^{1,2})} \right)}. \end{aligned}$$

Since $\text{Pdim}(\mathcal{N}^2) \leq \text{Pdim}(\mathcal{N}^{1,2})$, plugging the upper bound of $\text{Pdim}(\mathcal{N}^{1,2})$ derived in Theorem 5.1 into the above display and using the relationship of depth and width between \mathcal{N}^2 and $\mathcal{N}^{1,2}$ proved in Lemma 5.7, we get

$$\mathcal{E}_{sta} \leq C_{\mathcal{B},c_3} \left[d(\mathcal{D}+3)(\mathcal{D}+2)\mathcal{W} \sqrt{\frac{\mathcal{D}+3+\log(d(\mathcal{D}+2)\mathcal{W})}{n}} \right]^{1-\nu}, \tag{5.9}$$

where $\nu > 0$ can be arbitrarily small. Combing (5.9) with the approximation error in Theorem 4.2 and taking $\epsilon^2 = C_{\mathcal{B},c_1,c_2,c_3} C \left(\frac{1}{n}\right)^{\frac{1}{d+2+\nu}}$, we get that

$$\begin{aligned} & \mathbb{E}_{\mathbf{X},\mathbf{Y}}[\|\widehat{u}_\phi - u^*\|_{H^1(\Omega)}^2] \\ & \leq \frac{2}{c_1 \wedge 1} \left[C_{\mathcal{B},c_3} \left(d(\mathcal{D}+3)(\mathcal{D}+2)\mathcal{W} \sqrt{\frac{\mathcal{D}+3+\log(d(\mathcal{D}+2)\mathcal{W})}{n}} \right)^{1-\mu} + \frac{c_3+1}{2}\epsilon^2 \right] \\ & \leq \frac{2}{c_1 \wedge 1} \left[C_{\mathcal{B},c_3} \left(d(\lceil \log d \rceil + 6)(\lceil \log d \rceil + 5) C_{d,\|u^*\|_{C^2(\bar{\Omega})}} \left(\frac{1}{\epsilon}\right)^d \cdot \right. \right. \\ & \quad \left. \left. \sqrt{\frac{\lceil \log d \rceil + 6 + \log \left(d(\lceil \log d \rceil + 5) \cdot C_{d,\|u^*\|_{C^2(\bar{\Omega})}} \left(\frac{1}{\epsilon}\right)^d \right)}{n}} \right)^{1-\mu} + \frac{c_3+1}{2}\epsilon^2 \right] \\ & \leq C_{\|u^*\|_{C^2(\bar{\Omega})},c_1,c_2,c_3,d} \mathcal{O}\left(n^{-1/(d+2+\nu)}\right). \end{aligned}$$

This completes the proof. □

Remark 5.1. Deep Ritz method is actually a kinds of deep nonparametric estimation method where we estimate functions from random samples. The benefit is that we can use DRM to handle PDEs in high dimension since only a small bach of samples is needed during SGD training. In contrast, we have form the loading matrix and vector explicitly in FEM. However, what we have to pay is that the convergence rate as illustrate as follows. Let $N = N = \mathcal{O}\left(\frac{1}{h^d}\right)$, where h is the size of the mesh in FEM. From Theorem 5.2, we get

$$\mathbb{E}_{\mathbf{X},\mathbf{Y}}[\|\widehat{u}_\phi - u^*\|_{H^1(\Omega)}^2] = \mathcal{O}(h^{\frac{d}{d+2+\nu}}).$$

In the case $d=2$, by Markov’s inequality and the above display, we get with high probability,

$$\|\widehat{u}_\phi - u^*\|_{H^1(\Omega)} \leq \mathcal{O}(h^{\frac{1}{4+\nu}}).$$

Comparing the well known results of FEM where the convergence rate in H^1 norm is $\mathcal{O}(h)$, the rate proved here for the DRM method is far from satisfactory.

In the literature of nonparametric regression, where functions living in certain function classes are estimated from n paired random samples, the best convergence rate in H^1 norm for estimating functions in H^2 is $\mathcal{O}(n^{-\frac{1}{4+d}})$ [40]. Obviously, the nonparametric learning task in DRM is not easier than nonparametric regressions. What we proved here is $\mathcal{O}(n^{-\frac{1}{4+2d+\nu}})$ in H^1 norm with ν can be arbitrary small. **We conjecture that the best convergence rate in H^1 norm of DRM with both n training samples in both the domain and boundary is also $\mathcal{O}(n^{-\frac{1}{4+d}})$.** Upon this conjecture, we get the optimal convergence rate of DRM for $d=2$ is $\|\widehat{u}_\phi - u^*\|_{H^1(\Omega)} \leq \mathcal{O}(h^{\frac{1}{3}})$ with high probability.

The convergence rate of DRM proved here suffers the curse of dimensionality. One possible direction to improve the convergence rate and reduce the curse of dimensionality is considering solutions of PDE (2.1) with higher regularity. For example, if we assume $f \in H^s(\Omega)$, $s \geq 1$, deep Ritz method using deep neural network with ReLU^{s+2} activation functions will reduce the curse since the higher regularity assumption can improve both the approximation and statistical error. We leave the detail of this idea in a following up work.

6 Conclusion and extension

In this paper, we provide a rigorous numerical analysis on deep Ritz method (DRM) [47] for second order elliptic equations with Neumann boundary conditions. We establish the first nonasymptotic convergence rate in H^1 norm on DRM for general deep networks with ReLU^2 activation functions. In addition to provide theoretical justification of DRM, our study also provides guidance on how to set the hyper-parameter of depth and width to achieve the desired convergence rate in terms of number of training samples. Technically, we derive bounds on the approximation error of deep ReLU^2 network in C^1 norm and on the Rademacher complexity of the non-Lipschitz composition of gradient norm and ReLU^2 network.

There are several directions for our future exploration. First, it is easy to extend the current analysis for general second order elliptic equations with variational form under Dirichlet or Robin boundary conditions. Second, the approximation and statistical error bounds deriving here can be used for studying the nonasymptotic convergence rate for residual based method (PINNS). Studying deep DGM by combing current analysis with the tools for analyzing GAN [14] is also of immense interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2020YFA0714200), the Science and Technology Major Project of Hubei Province under Grant 2021AAA010, the National Science Foundation of China (Nos. 12125103, 12071362, 11871474, 11871385), by the Natural Science Foundation of Hubei Province (No. 2019CFA007), and by the research fund of KLATASDSMOE. The numerical simulations in this work have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

References

- [1] MARTIN ANTHONY AND PETER L. BARTLETT, *Neural Network Learning: Theoretical Foundations*, Cambridge University Press, 2009.

- [2] PETER L. BARTLETT, NICK HARVEY, CHRISTOPHER LIAW, AND ABBAS MEHRABIAN, *Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks*, *J. Mach. Learn. Res.*, 20 (2019), pp. 1–17.
- [3] CHRISTIAN BECK, WEINAN E, AND ARNULF JENTZEN, *Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations*, *Journal of Nonlinear Science*, 29 (2019), pp. 1563–1619.
- [4] STÉPHANE BOUCHERON, GÁBOR LUGOSI, AND PASCAL MASSART, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013.
- [5] SUSANNE BRENNER AND RIDGWAY SCOTT, *The Mathematical Theory of Finite Element Methods*, vol. 15, Springer Science & Business Media, 2007.
- [6] FAN CHEN, JIANGUO HUANG, CHUNMEI WANG, AND HAIZHAO YANG, *Friedrichs learning: Weak solutions of partial differential equations via deep learning*, arXiv preprint arXiv:2012.08023, (2020).
- [7] PHILIPPE G. CIARLET, *The Finite Element Method for Elliptic Problems*, SIAM, 2002.
- [8] CARL DE BOOR AND CARL DE BOOR, *A Practical Guide to Splines*, vol. 27, Springer-Verlag New York, 1978.
- [9] A. DEBUSSCHE, M. FUHRMAN, AND G. TESSITORE, *Optimal control of a stochastic heat equation with boundary-noise and boundary-control*, *ESAIM: Control, Optimisation and Calculus of Variations*, 13 (2007), pp. 178–205.
- [10] NICCOL DISCACCIATI, JAN S. HESTHAVEN, AND DEEP RAY, *Controlling oscillations in high-order discontinuous Galerkin schemes using artificial viscosity tuned by neural networks*, *Journal of Computational Physics*, 409 (2020), p. 109304.
- [11] R.M DUDLEY, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, *Journal of Functional Analysis*, 1 (1967), pp. 290–330.
- [12] DINH DŨNG AND VAN KIEN NGUYEN, *Sparse-grid sampling recovery and deep ReLU neural networks in high-dimensional approximation*, arXiv preprint arXiv:2007.08729, (2020).
- [13] LAWRENCE C. EVANS, *Partial differential equations*, Graduate Studies in Mathematics, 19 (1998).
- [14] IAN GOODFELLOW, JEAN POUGET-ABADIE, MEHDI MIRZA, BING XU, DAVID WARDEFARLEY, SHERJIL OZAIR, AARON COURVILLE, AND YOSHUA BENGIO, *Generative adversarial nets*, in *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., 2014, pp. 2672–2680.
- [15] INGO GÜHRING, GITTA KUTYNIOK, AND PHILIPP PETERSEN, *Error bounds for approximations with deep ReLU neural networks in w s , p norms*, *Analysis and Applications*, 18 (2020), pp. 803–859.
- [16] T. J. HUGHES, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Courier Corporation, 2012.
- [17] AMEYA D. JAGTAP AND GEORGE EM KARNIADAKIS, *Extended physics-informed neural networks (xPINNs): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations*, *Communications in Computational Physics*, 28 (2020), pp. 2002–2041.
- [18] 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 (2020), p. 113028.
- [19] 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

- (1998), pp. 987–1000.
- [20] MICHEL LEDOUX AND MICHEL TALAGRAND, *Probability in Banach Spaces: Isoperimetry and Processes*, Springer Science & Business Media, 2013.
 - [21] HYUK LEE AND IN SEOK KANG, *Neural algorithm for solving differential equations*, Journal of Computational Physics, 91 (1990), pp. 110–131.
 - [22] RANDALL J. LEVEQUE, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, SIAM, 2007.
 - [23] BO LI, SHANSHAN TANG, AND HAIJUN YU, *Better approximations of high dimensional smooth functions by deep neural networks with rectified power units*, arXiv preprint arXiv:1903.05858, (2019).
 - [24] JIANFENG LU, YULONG LU, AND MIN WANG, *A priori generalization analysis of the deep Ritz method for solving high dimensional elliptic equations*, arXiv preprint arXiv:2101.01708, (2021).
 - [25] LU LU, XUHUI MENG, ZHIPING MAO, AND GEORGE E. KARNIADAKIS, *DeepXDE: A deep learning library for solving differential equations*, CoRR, abs/1907.04502 (2019).
 - [26] TAO LUO AND HAIZHAO YANG, *Two-layer neural networks for partial differential equations: Optimization and generalization theory*, arXiv preprint arXiv:2006.15733, (2020).
 - [27] SIDDHARTHA MISHRA AND ROBERTO MOLINARO, *Estimates on the generalization error of physics informed neural networks (PINNs) for approximating PDEs*, arXiv preprint arXiv:2006.16144, (2020).
 - [28] HADRIEN MONTANELLI AND QIANG DU, *New error bounds for deep ReLU networks using sparse grids*, SIAM Journal on Mathematics of Data Science, 1 (2019), pp. 78–92.
 - [29] 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 (2020), p. 109760.
 - [30] GUOFEI PANG, LU LU, AND GEORGE EM KARNIADAKIS, *fPINNs: Fractional physics-informed neural networks*, SIAM Journal on Scientific Computing, 41 (2019), pp. A2603–A2626.
 - [31] 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 (2019), pp. 686–707.
 - [32] DEEP RAY AND JAN S. HESTHAVEN, *An artificial neural network as a troubled-cell indicator*, Journal of Computational Physics, 367 (2018), pp. 166–191.
 - [33] JOHANNES SCHMIDT-HIEBER, *Deep ReLU network approximation of functions on a manifold*, arXiv preprint arXiv:1908.00695, (2019).
 - [34] LARRY SCHUMAKER, *Spline Functions: Basic Theory*, Cambridge University Press, 2007.
 - [35] ZUOWEI SHEN, HAIZHAO YANG, AND SHIJUN ZHANG, *Deep network approximation characterized by number of neurons*, arXiv preprint arXiv:1906.05497, (2019).
 - [36] YEONJONG SHIN, JEROME DARBON, AND GEORGE EM KARNIADAKIS, *On the convergence and generalization of physics informed neural networks*, arXiv preprint arXiv:2004.01806, (2020).
 - [37] YEONJONG SHIN, ZHONGQIANG ZHANG, AND GEORGE EM KARNIADAKIS, *Error estimates of residual minimization using neural networks for linear PDEs*, arXiv preprint arXiv:2010.08019, (2020).
 - [38] JONATHAN W SIEGEL AND JINCHAO XU, *Approximation rates for neural networks with general activation functions*, Neural Networks, 128 (2020), pp. 313–321.
 - [39] JUSTIN A. SIRIGNANO AND K. SPILIOPOULOS, *DGM: A deep learning algorithm for solving partial differential equations*, Journal of Computational Physics, 375 (2018), pp. 1339–1364.
 - [40] CHARLES J STONE, *Optimal global rates of convergence for nonparametric regression*, The Annals of Statistics, (1982), pp. 1040–1053.

- [41] ZHENG SUN, SHUYI WANG, LO-BIN CHANG, YULONG XING, AND DONGBIN XIU, *Convolution neural network shock detector for numerical solution of conservation laws*, *Communications in Computational Physics*, 28 (2020), pp. 2075–2108.
- [42] TAIJI SUZUKI, *Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: Optimal rate and curse of dimensionality*, arXiv preprint arXiv:1810.08033, (2018).
- [43] J.W. THOMAS, *Numerical Partial Differential Equations: Finite Difference Methods*, vol. 22, Springer Science & Business Media, 2013.
- [44] AAD W. VAN DER VAART AND JON A. WELLNER, *Weak convergence*, in *Weak Convergence and Empirical Processes*, Springer, 1996.
- [45] VLADIMIR N VAPNIK AND A YA CHERVONENKIS, *On the uniform convergence of relative frequencies of events to their probabilities*, in *Measures of Complexity*, Springer, 2015, pp. 11–30.
- [46] YUFEI WANG, ZIJU SHEN, ZICHAO LONG, AND BIN DONG, *Learning to discretize: Solving 1D scalar conservation laws via deep reinforcement learning*, *Communications in Computational Physics*, 28 (2020), pp. 2158–2179.
- [47] 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 (2017), pp. 1–12.
- [48] JINCHAO XU, *The finite neuron method and convergence analysis*, arXiv preprint arXiv:2010.01458, (2020).
- [49] DMITRY YAROTSKY, *Error bounds for approximations with deep ReLU networks*, *Neural Networks*, 94 (2017), pp. 103–114.
- [50] YAOHUA ZANG, GANG BAO, XIAOJING YE, AND HAOMIN ZHOU, *Weak adversarial networks for high-dimensional partial differential equations*, *Journal of Computational Physics*, 411 (2020), p. 109409.