Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2021-0195

Vol. **31**, No. 4, pp. 1020-1048
April 2022

Convergence Rate Analysis for Deep Ritz Method

Chenguang Duan[1], Yuling Jiao[1,2], Yanming Lai[1], Dingwei Li[1], 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 ReLU^2 activation functions. In addition to providing a theoretical justification of DRM, our study also shed light 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 bound on the approximation error of deep ReLU^2 network in C^1 norm and bound on the Rademacher complexity of the non-Lipschitz composition of gradient norm and ReLU^2 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), `laiyanming@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)

 ©2022 Global-Science Press

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 nonlinear 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-