Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2023-0253

Vol. **35**, No. 3, pp. 761-815
March 2024

Generalization Error in the Deep Ritz Method with Smooth Activation Functions

Janne Siipola*

Department of Mathematics and Statistics, University of Helsinki, Finland.

Received 27 September 2023; Accepted (in revised version) 12 March 2024

Abstract. Deep Ritz method is a deep learning paradigm to solve partial differential equations. In this article we study the generalization error of the Deep Ritz method. We focus on the quintessential problem which is the Poisson's equation. We show that generalization error of the Deep Ritz method converges to zero with rate $\frac{C}{\sqrt{n}}$, and we discuss about the constant C. Results are obtained for shallow and residual neural networks with smooth activation functions.

AMS subject classifications: 35A35, 65C05, 92B20, 68T10

Key words: Deep learning, Deep Ritz method, Poisson's equation, residual neural networks, shallow neural networks, generalization.

1 Introduction

We are studying the Deep Ritz method [1], in particular the problem of approximating the solution of the Poisson's equation in a ball with deep neural networks [2, 3]; the results given in this article can easily be generalized for more complex domains, what is important is that the domain is bounded. Let us give a few definitions and remarks to frame what we are talking about here.

1.1 Deep Ritz Method

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with a smooth boundary, where $d \in \mathbb{N}$ is the dimension. Let's consider the Poisson's equation

$$-\Delta u = f,$$

*Corresponding author. *Email address:* `janne.siipola@helsinki.fi` (J. Siipola)

where f is some function, possibly $f \in L^2(\Omega)$, sometimes referred as the source term. We can consider this equation with Dirichlet boundary condition as

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{1.1}$$

The Poisson's equation can also be stated in a variational form as

$$\min_{u \in H_0^1(\mathbb{R}^d)} \int_\Omega \frac{1}{2}|\nabla u|^2 - fu, \tag{1.2}$$

meaning that a minimizer of the integral in (1.2), among a set of admissible functions $H_0^1(\mathbb{R}^d)$, is the solution of the Poisson's equation. Here a natural choice for the set of admissible functions is the Sobolev space $H_0^1(\Omega)$. This variational form is what was proposed for a loss function in [1], known as the Deep Ritz method (DRM). However, in DRM we cannot minimize the integral over some Sobolev space – the set of admissible functions will be some set of neural networks.

When the set of admissible functions is some set of neural networks, then the zero boundary condition might not be met, and one must add a penalty term in order to force the solution to meet the boundary condition. Therefore the loss function for the Deep Ritz method will be

$$\min_{u \in \mathcal{H}} \int_\Omega \left(\frac{1}{2}|\nabla u|^2 - fu\right) + \lambda \int_{\partial\Omega} u^2, \tag{1.3}$$

where the penalty term $\lambda > 0$ is a constant. The minimal solution for (1.3) must take into account the boundary values as well, and so this is the actual theoretical loss function for the DRM. The latter variational form is related to another boundary value problem for the Poisson's equation, namely the Robin boundary value problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u + \frac{1}{2\lambda}\partial_\nu u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1.4}$$

where $\partial_\nu u$ is the derivative of u into direction of the normal of the boundary $\partial\Omega$. These two boundary value problems are connected with the parameter λ, and it is a known result that if we let $\lambda \to \infty$ then the solution of the Robin problem converges to the solution of the Dirichlet problem [4].

The theoretical loss function (1.3) can be represented as

$$\begin{aligned} \mathcal{L}_\lambda(u) = |\Omega| \mathbb{E}_{X \sim U(\Omega)} \left[\frac{\|\nabla u(X)\|_2^2}{2} - u(X) f(X) \right] \\ + \frac{\lambda}{2} |\partial\Omega| \mathbb{E}_{Y \sim U(\partial\Omega)} [T(u(Y)^2)], \end{aligned} \tag{1.5}$$