Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.OA-2024-0026

Vol. **17**, No. 4, pp. 996-1017
November 2024

Error Analysis for Empirical Risk Minimization Over Clipped ReLU Networks in Solving Linear Kolmogorov Partial Differential Equations

Jichang Xiao* and Xiaoqun Wang

Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China

Received 9 March 2024; Accepted (in revised version) 12 June 2024

Abstract. Deep learning algorithms have been successfully applied to numerically solve linear Kolmogorov partial differential equations (PDEs). A recent research shows that if the initial functions are bounded, the empirical risk minimization (ERM) over clipped ReLU networks generalizes well for solving the linear Kolmogorov PDE. In this paper, we propose to use a truncation technique to extend the generalization results for polynomially growing initial functions. Specifically, we prove that under an assumption, the sample size required to achieve an generalization error within ε with a confidence level ϱ grows polynomially in the size of the clipped neural networks and $(\varepsilon^{-1}, \varrho^{-1})$, which means that the curse of dimensionality is broken. Moreover, we verify that the required assumptions hold for Black-Scholes PDEs and heat equations which are two important cases of linear Kolmogorov PDEs. For the approximation error, under certain assumptions, we establish approximation results for clipped ReLU neural networks when approximating the solution of Kolmogorov PDEs. Consequently, we establish that the ERM over artificial neural networks indeed overcomes the curse of dimensionality for a larger class of linear Kolmogorov PDEs.

AMS subject classifications: 60H30, 65C30, 62M45, 68T07

Key words: Linear Kolmogorov PDE, curse of dimensionality, empirical risk minimization, generalization error.

1. Introduction

Partial differential equations (PDEs) have been widely used in modeling problems in physics, finance and engineering. Traditional numerical methods for solving PDEs

*Corresponding author. *Email addresses:* `xiaojc19@mails.tsinghua.edu.cn` (J. Xiao), `wangxiaoqun@mail.tsinghua.edu.cn` (X. Wang)

such as finite differences [34] and finite elements [7] suffer from the curse of dimensionality, which implies that the computational cost grows exponentially as the dimension increases. Recently, many deep learning-based algorithms have been proposed for different classes of PDEs, see [3, 11, 18, 28–30, 33].

In this paper, we focus on the numerical approximation of linear Kolmogorov PDEs, with a special attention to heat equations and the Black-Scholes PDEs. Beck *et al.* [3] first proposed a deep learning algorithm to approximate the solution over a full hypercube by reformulating the approximation problem into an optimization problem. Berner *et al.* [5] further developed the deep learning algorithm to solve parametric Kolmogorov PDEs. Richter *et al.* [30] introduced different formulations of the risk functional to enhance the robustness of the deep learning algorithm. The numerical results in [3, 5, 30] demonstrated the effectiveness of the deep learning algorithms even in high dimensions. These findings suggest that deep learning algorithms for solving linear Kolmogorov PDEs do not suffer from the curse of dimensionality.

One theoretical explanation for this phenomenon is that the solutions of Kolmogorov PDEs can be approximated by deep artificial neural networks in which the number of parameters grows polynomially in both the prescribed accuracy $\varepsilon \in (0,1)$ and the dimension d, see [12, 15, 20, 21]. Besides the analysis of the approximation error for deep artificial neural networks, Berner *et al.* [6] proved that under suitable conditions, the generalization error resulted from empirical risk minimization (ERM) also breaks the curse of dimensionality. From the perspective of learning theory, the generalization results in [6] shows the agnostic probably approximately correct (PAC) learnability of the clipped ReLU neural networks for the quadratic loss function and the region $[u,v]^d \times [-D,D]$.

Since the Hoeffding's inequality is commonly used to establish the PAC-type inequality, the generalization results in [6] are established only for linear Kolmogorov PDEs with bounded initial functions. There are many linear Kolmogorov PDEs in which the initial functions are in fact unbounded, for example, the Black-Scholes PDEs in the problems of the option pricing of basket call and call on max options. In this paper, we prove that for linear Kolmogorov PDEs with unbounded initial function, the generalization error still does not suffer from the curse of dimensionality if some assumptions are satisfied.

Under an assumption regarding the tail probability, we extend the generalization results in [6] for a broader class of linear Kolmogorov PDEs, where the initial functions grow polynomially instead of being bounded. We can prove that the required assumption holds for some important equations such as Black-Scholes PDEs and heat equations. In fact, the extended generalization results hold for the general regression problem under Setting 2.1. The truncation technique makes it possible to establish the PAC bound of the generalization error when the empirical loss is unbounded. Specifically, the idea is to truncate both the true loss and empirical loss, then we can apply the Hoeffding's inequality to the truncated parts which are now bounded. By carefully choosing an appropriate truncation threshold, we can establish generalization error bound for ERM over the clipped neural networks. As for the approximation error, we