Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2023-0058

Vol. **34**, No. 4, pp. 869-906
October 2023

Learning Specialized Activation Functions for Physics-Informed Neural Networks

Honghui Wang[1], Lu Lu[2], Shiji Song[1] and Gao Huang[1,*]

[1] *Department of Automation, Tsinghua University, Beijing 100084, P.R. China.*
[2] *Department of Statistics and Data Science, Yale University, New Haven, CT 06511, USA.*

Received 28 February 2023; Accepted (in revised version) 9 August 2023

Abstract. Physics-informed neural networks (PINNs) are known to suffer from optimization difficulty. In this work, we reveal the connection between the optimization difficulty of PINNs and activation functions. Specifically, we show that PINNs exhibit high sensitivity to activation functions when solving PDEs with distinct properties. Existing works usually choose activation functions by inefficient trial-and-error. To avoid the inefficient manual selection and to alleviate the optimization difficulty of PINNs, we introduce adaptive activation functions to search for the optimal function when solving different problems. We compare different adaptive activation functions and discuss their limitations in the context of PINNs. Furthermore, we propose to tailor the idea of learning combinations of candidate activation functions to the PINNs optimization, which has a higher requirement for the smoothness and diversity on learned functions. This is achieved by removing activation functions which cannot provide higher-order derivatives from the candidate set and incorporating elementary functions with different properties according to our prior knowledge about the PDE at hand. We further enhance the search space with adaptive slopes. The proposed adaptive activation function can be used to solve different PDE systems in an interpretable way. Its effectiveness is demonstrated on a series of benchmarks. Code is available at https://github.com/LeapLabTHU/AdaAFforPINNs.

AMS subject classifications: 65M99, 68T07

Key words: Partial differential equations, deep learning, adaptive activation functions, physics-informed neural networks.

1 Introduction

Recent years have witnessed the remarkable progress of physics-informed neural networks (PINNs) on modeling the dynamics of physics systems [1–6]. The underlying

*Corresponding author. *Email addresses:* wanghh20@mails.tsinghua.edu.cn (H. Wang), lu.lu@yale.edu (L. Lu), shijis@mail.tsinghua.edu.cn (S. Song), gaohuang@tsinghua.edu.cn (G. Huang)

physics laws, usually presented as ordinary and partial differential equations (ODEs and PDEs), are embedded as soft constraints to guide the learning process of deep neural networks [1, 7, 8]. Despite the effectiveness, the introduction of PDE-based loss function makes the optimization more ill-conditioned [9, 10]. Efforts have been made to alleviate this problem from the aspects of loss weight balancing [9, 11], loss function design [12–14], adaptive collocation point sampling [15–18], domain decomposition [19–21], and curriculum learning [10]. However, since most previous works adopt the standard fully-connected networks, the relationship between network architectures and the optimization difficulty of PINNs is less explored, regardless of the fact that the advancement of network architectures is the key to the success on deep learning in computer vision [22, 23] and natural language processing [24].

At the heart of network architectures lie the activation functions, which play a significant role in the expressiveness and optimization of models. Recent works [25–29] have observed that the choice of activation functions affects the learning of continuous signal representations. For example, the hyperbolic tangent function is shown to suffer from numerical instability when simulating vortex induced vibrations, while a PINN with sinusoidal function can be optimized smoothly [25]. Another important observation is that the optimal activation function depends on the problem at hand. While the Rectified Linear Unit [30–33] is widely adopted in most computer vision and natural language processing tasks [34], there is no such default choice of activation functions for PINNs when applied to physical systems with distinct properties. In fact, PINNs show great sensitivity to activation functions. These observations reveal the possibility and necessity of reducing the training difficulty of PINNs by selecting an appropriate activation function.

The various characteristics of different PDE systems make the choice of activation functions a critical aspect in PINNs. The common practice to find the optimal activation functions is by trial-and-error, which requires extensive computational resources and human knowledge. As an alternative, the adaptive activation functions are designed to find specialized activation functions for different architectures and tasks. In this work, we aim to explore different adaptive activation functions and discuss their limitations in the context of PINNs. The major difference of existing methods lies in the search space. The piece-wise linear function is adopted as the universal function approximator in some methods, such as APL [35], RePLU [36] and PWLU [37]. The formulation of SLAF [38] is based on Taylor approximation with polynomial basis. PAU [39] leverages Padé approximation to form its search space. ACON [40] is proposed as a smooth approximator to the general Maxout family activation functions [41]. Built upon previous works [42–44] to learn activation functions as linear combinations of candidate activation functions, the Adaptive Blending Units (ABU) [45] explore different normalization methods for the combination coefficients, which correspond to different restrictions imposed on the search space. In the literature of PINNs, the recent work [46] proposes the Kronecker neural network (KNN) based on adaptive linear combinations of candidate activation functions and apples it to solve PDEs. However, only one type of adaptive activation function is discussed in their work under the context of PINNs, and the designs