Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2024-0096

Vol. **36**, No. 4, pp. 1090-1112
October 2024

Loss Jump During Loss Switch in Solving PDEs with Neural Networks

Zhiwei Wang[1], Lulu Zhang[1], Zhongwang Zhang[1] and Zhi-Qin John Xu[1,*]

[1] *Institute of Natural Sciences, School of Mathematical Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, P.R. China.*

Received 20 April 2024; Accepted (in revised version) 22 August 2024

Abstract. Using neural networks to solve partial differential equations (PDEs) is gaining popularity as an alternative approach in the scientific computing community. Neural networks can integrate different types of information into the loss function. These include observation data, governing equations, and variational forms, etc. These loss functions can be broadly categorized into two types: observation data loss directly constrains and measures the model output, while other loss functions indirectly model the performance of the network, which can be classified as model loss. However, this alternative approach lacks a thorough understanding of its underlying mechanisms, including theoretical foundations and rigorous characterization of various phenomena. This work focuses on investigating how different loss functions impact the training of neural networks for solving PDEs. We discover a stable loss-jump phenomenon: when switching the loss function from the data loss to the model loss, which includes different orders of derivative information, the neural network solution significantly deviates from the exact solution immediately. Further experiments reveal that this phenomenon arises from the different frequency preferences of neural networks under different loss functions. We theoretically analyze the frequency preference of neural networks under model loss. This loss-jump phenomenon provides a valuable perspective for examining the underlying mechanisms of neural networks in solving PDEs.

AMS subject classifications: 68T15, 68Q01

Key words: Loss jump, frequency bias, neural network, loss switch.

1 Introduction

The use of neural networks for solving partial differential equations (PDEs) has emerged as a promising alternative to traditional numerical methods in the scientific computing

*Corresponding author. *Email addresses:* victorywzw@sjtu.edu.cn (Z. Wang), zhangl9661@sjtu.edu.cn (L. Zhang), 0123zzw666@sjtu.edu.cn (Z. Zhang), xuzhiqin@sjtu.edu.cn (Z.-Q. J. Xu)

 ©2024 Global-Science Press

community. By incorporating various types of information into the loss function, such as observation data, governing equations, and variational forms, neural networks offer a flexible and powerful framework for approximating the solution of PDEs. These loss functions can be broadly classified into two categories: **data loss**, which directly constrains and measures the model output using observation data, and **model loss**, which indirectly models the performance of the network using equations and variational forms.

Despite the growing interest in this approach, a comprehensive understanding of the underlying mechanisms governing the behavior of neural networks in solving PDEs is still lacking. While several works have explored the capabilities and limitations of physics-informed learning [5, 6, 10, 12, 13, 23, 33] and the challenges in training physics-informed neural networks (PINNs) [8, 25, 29, 33], the impact of different loss functions on the training dynamics and convergence properties of neural networks remains an open question.

Recent studies have shown that the derivatives of the target functions in the loss function play a crucial role in the convergence of frequencies [18, 32, 33]. A key observation is that neural networks often exhibit a frequency principle, learning from low to high frequencies [24, 32–34]. This phenomenon has inspired a series of theoretical works aimed at understanding the convergence properties of neural networks [1, 2, 4, 19, 20].

Moreover, the development of deep learning theory and algorithms has greatly benefited from the accurate description of stable phenomena. For instance, it has been observed that heavily over-parameterized neural networks usually do not overfit [3, 35], neurons in the same layer tend to condense in the same direction [21, 37, 38], and stochastic gradient descent or dropout tends to find flat minima [7, 14, 26, 31, 37, 39]. Additionally, a series of multiscale neural networks have been developed for solving differential equations [11, 16, 17, 28, 30, 36] and fitting functions [22, 27].

Motivated by these findings, we aim to investigate the impact of different loss functions on the training dynamics and convergence properties of neural networks for solving PDEs. We focus on the interplay between data loss and model loss, which incorporate different orders of derivative information. We discover a stable **loss-jump phenomenon**: when switching the loss function from the data loss to the model loss, which includes different orders of derivative information, the neural network solution significantly deviates from the exact solution immediately. Additionally, the fitted curve often exhibits an overall shift relative to the target function.

From a convergence perspective, there is a simple analysis for the sudden jump in the loss function: L^p convergence of u does not imply the L^p convergence of ∇u and in general $\mathcal{L}u$ where $\mathcal{L}$ is any differential operator. However, this analysis does not explain why the fitted curve undergoes an overall shift relative to the target function rather than making minor adjustments to the details of the original fitted curve. We note that the overall shift is often caused by changes in the low-frequency components of the function. Therefore, it is reasonable and valuable to explain and analyze this phenomenon from the perspective of frequency preference.

In this work, we analyze the training process and the dynamics induced by different