# The impact of different methods of gradient descent on the spectral bias of physics-informed neural networks

Author:
Alexander van den Arend Schmidt
A.F.vandenArendSchmidt@student.tudelft.nl

Supervisors:
Dr. Jing Sun, Dr. Alexander Heinlein,
Dr .Tiexing Wang

Examiner:
Dr. Hayley Hung

TU Delft

## 1 Introduction

- **Physics-informed neural networks (PINNs)**

Physics-Informed Neural Networks (PINNs) are intended to solve complex problems that obey physical rules or laws but have noisy or little data. These problems are encountered in a wide range of fields including for instance bioengineering, fluid mechanics, meta-material design and high-dimensional partial differential equations (PDEs) [1]. Where a classic deep neural network uses known labeled data to calculate the current loss of a neural network, a PINN uses rules and calculates how closely the neural network adheres to the rules.

- **Spectral Bias of PINNs**

Whilst PINNs show promising results, they often fail to converge in the presence of higher frequency components [2]; a problem known as the spectral bias. Multiple studies have explored ways to overcome or minimize spectral bias specifically for PINNs [3]. This paper builds on previous studies by investigating the impact of different gradient descent methods on the spectral bias.

## 2 Research Question

***How do different methods of gradient descent impact the spectral bias of physcis-informed neural networks for partial differential equations?***

## 3 Methodology

The effect of gradient descent methods on the spectral bias of PINNs for PDEs has been investigated as follows:

**1. First, gradient descent methods to test are selected based on their expected strengths [4] regarding a PINN loss landscape [5].**

*Selected methods of gradient descent:*

- Normal Stochastic Gradient Descent (SGD)
- Stochastic Gradient Descent with Momentum (SGDM)
- Nesterov Accelerated Gradient Descent (Nesterov)
- Adagrad gradient descent (Adagrad)
- Adam Gradient Descent (ADAM)

**2. Two PDEs are selected specifically so that their frequency can be varied while keeping other parameters constant and still having an analytical solution for each frequency.**

*Selected PDEs:*

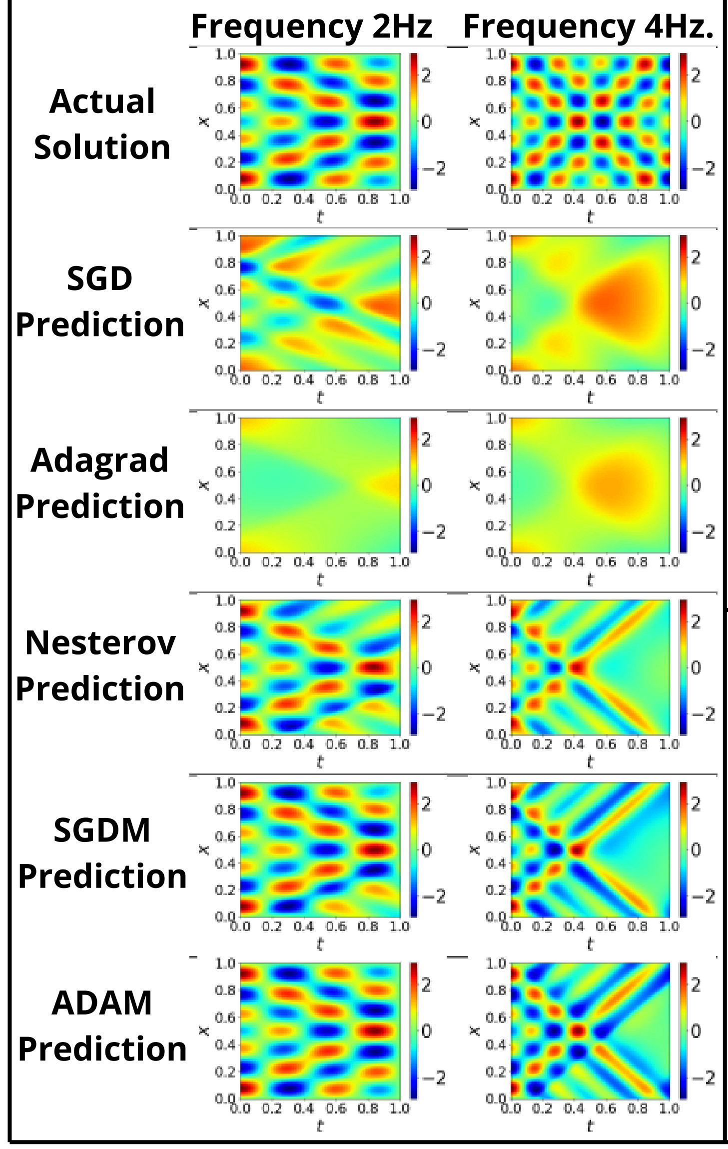
- 1D Wave PDE
- 1D Poisson PDE

**3. Finally, for all methods of gradient descent, and for all PDEs, the effect of increasing the frequency of the PDE on the convergence of the PINN will be explored.**

To do this multiple experiments have been performed. Firstly, for the 1D Wave PDE two frequnecies are chosen and the results are compared. For the 1D Poisson PDE an experiment is performed in which the frequency is increased in small steps and the corresponding loss is stored.

## 4 Results

***Figure 1:*** Heatmaps of 1D Wave PDE: Solution (Top) and predictions below.




- In Figure 1 it can be seen that the higher frequency (4Hz) is predicted less accurately than the lower frequency (2Hz) for all gradient descent methods. Adagrad does not converge at all even for the lower frequency.
- Overall in Figure 1 gradient descent methods that use momentum (Nesterov, SGDM and ADAM) achieve a (visually) better solution than those that do not.
- Within the methods with momentum, Nesterov performs worst, and ADAM performs best.
- In Figure 2 an L2 residual loss of 1 means that the PINN did not converge.
- The frequency (a) at which the gradient descent method start to fail can therefore be seen in Figure 2 where the loss increases from 0 to 1.
- The first methods to fail are the methods without momentum (SGD and Adagrad), and the gradient descent methods that use momentum (Nesterov, SGDM and ADAM) perform well up to higher frequencies.
- Also, in Figure 2 we see that ADAM fails last when increasing the frequency and performs best at high frequencies.

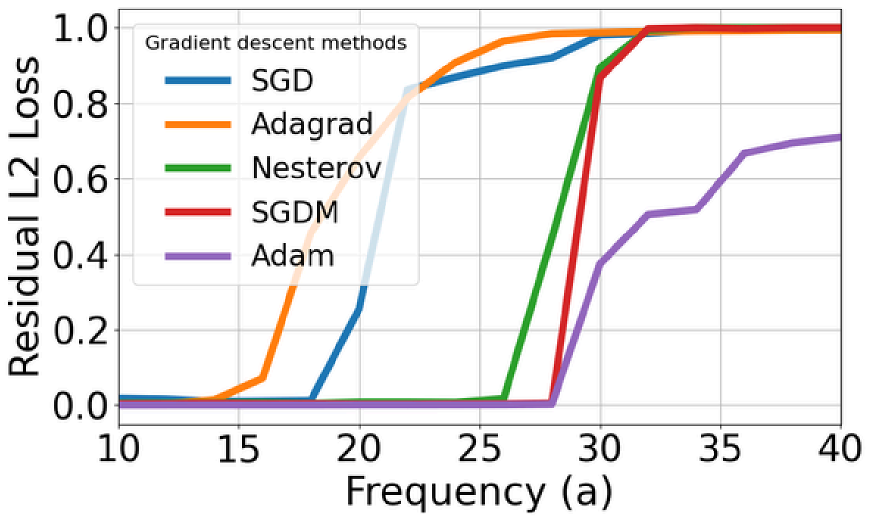



***Figure 2:*** Residual L2 losses for different gradient descent methods and varying frequency of the 1D Poisson PDE.

## 5 Conclusion

- The method of **gradient descent has a significant impact** on the **spectral bias of PINNs.**
- **Momentum** seems to be **the most important component** of a gradient descent optimizer.
- **Within** the methods with **momentum**, **Nesterov performs worst**, and **ADAM performs best.**
- Also, **ADAM**'s performance **deteriorates** more **gradually**, still providing an approximation of the target function at the higher frequencies rather than complete failure.

## References

[1] Yuyao Chen, Lu Lu, George Em Karniadakis, and Luca Dal Negro. Physics-informed neural networks for inverse problems in nano-optics and metamaterials. Optics express, 28(8):11618–11633, 2020.
[2] Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why pinns fail to train: A neural tangent kernel perspective. Journal of Computational Physics, 449, 2022.
[3] Xintao Chai, Wenjun Cao, and Jianhui Li. Overcoming the spectral bias problem of physics-informed neural networks in solving the frequency-domain acoustic wave equation. IEEE, 2024.
[4] Sebastian Ruder. An overview of gradient descent optimization algorithms. arXiv, 2016.
[5] Pratik Rathore, Weimu Lei, and Zachary Frangella. Challenges in training pinns: A loss landscape perspective. arXiv, 2024.