

Enhancing Training of Deep Neural Networks Using Multilevel and Domain Decomposition Strategies

Abstract

The training of deep neural networks (DNNs) is traditionally accomplished using stochastic gradient descent or its variants. While these methods have demonstrated certain robustness and accuracy, their convergence speed deteriorates for large-scale, highly ill-conditioned, and stiff problems, such as ones arising in scientific machine learning applications. Consequently, there is a growing interest in adopting more sophisticated training strategies that can not only accelerate convergence but may also provide parallelism, convergence control, and automatic selection of certain hyper-parameters.

In this talk, we propose to enhance the training of DNNs by leveraging nonlinear multilevel and domain decomposition strategies. We will discuss how to construct a multilevel hierarchy and how to decompose the parameters of the network by exploring the structure of the DNN architecture, properties of the loss function, and characteristics of the dataset. Furthermore, the dependency on a large number of hyper-parameters will be reduced by employing a trust-region globalization strategy. The effectiveness of the proposed training strategies will be demonstrated through a series of numerical experiments from the field of image classification and physics-informed neural networks.

References

- [1] A. Kopaničáková, H. Kothari, G. Karniadakis and R. Krause. Enhancing training of physics-informed neural networks using domain-decomposition based preconditioning strategies. Under review, 2023.
- [2] S. Gratton, A. Kopaničáková, and Ph. Toint. Multilevel Objective-Function-Free Optimization with an Application to Neural Networks Training. SIAM, Journal on Optimization (Accepted), 2023.
- [3] A. Kopaničáková. On the use of hybrid coarse-level models in multilevel minimization methods. Domain Decomposition Methods in Science and Engineering XXVII (Accepted), 2023.
- [4] A. Kopaničáková, and R. Krause. Globally Convergent Multilevel Training of Deep Residual Networks. SIAM Journal on Scientific Computing, 2022.