Journal of Computational Mathematics
Vol.43, No.6, 2025, 1488–1511.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2406-m2023-0143

A NEURON-WISE SUBSPACE CORRECTION METHOD FOR THE FINITE NEURON METHOD*

Jongho Park[1)], Jinchao Xu and Xiaofeng Xu
Applied Mathematics and Computational Sciences Program, CEMSE, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia
Emails: jongho.park@kaust.edu.sa, jinchao.xu@kaust.edu.sa, xiaofeng.xu@kaust.edu.sa

Abstract

In this paper, we propose a novel algorithm called neuron-wise parallel subspace correction method for the finite neuron method that approximates numerical solutions of partial differential equations (PDEs) using neural network functions. Despite extremely extensive research activities in applying neural networks for numerical PDEs, there is still a serious lack of effective training algorithms that can achieve adequate accuracy, even for one-dimensional problems. Based on recent results on the spectral properties of linear layers and analysis for single neuron problems, we develop a special type of subspace correction method that optimizes the linear layer and each neuron in the nonlinear layer separately. An optimal preconditioner that resolves the ill-conditioning of the linear layer is presented for one-dimensional problems, so that the linear layer is trained in a uniform number of iterations with respect to the number of neurons. In each single neuron problem, a local minimum is found by a superlinearly convergent algorithm. Numerical experiments on function approximation problems and PDEs demonstrate better performance of the proposed method than other gradient-based methods.

Mathematics subject classification: 65D15, 65N22, 65N30, 65N55, 68T07.
Key words: Finite neuron method, Subspace correction method, Training algorithm, Preconditioner, Function approximation, Partial differential equation.

1. Introduction

Neural networks, thanks to the universal approximation property [9,26], are promising tools for numerical solutions of partial differential equations. Moreover, it was shown in [30] that the approximation properties of neural networks have higher asymptotic approximation rates than that of traditional numerical methods such as the finite element method. Such powerful approximation properties, however, can hardly be observed in numerical experiments in simple tasks of function approximation despite extensive research on numerical solutions of PDEs in recent years, e.g., physics-informed neural networks [27], the deep Ritz method [11], and the finite neuron method [40]. Even in one dimension, using gradient-based methods to training a shallow neural network does not produce accurate solutions with a substantial number of iterations. This poor convergence behavior was analyzed rigorously in [15] and is due to the ill-conditioning of the problem. Therefore, applying neural networks to solutions of PDEs must require novel training algorithms, different from the conventional ones for regression, image classification, and pattern recognition tasks. In this viewpoint, there are many works

* Received July 12, 2023 / Revised version received January 25, 2024 / Accepted June 4, 2024 / Published online August 29, 2024 /
1) Corresponding author

on designing and analyzing training algorithms for neural networks to try to speed up the convergence or narrow the gap between the theoretical optimum and training results. For instance, a hybrid least squares/gradient descent method was proposed in [10] from an adaptive basis perspective. The active neuron least squares method in [2] was designed to avoid plateau phenomena that slow down the gradient dynamics of training ReLU shallow neural networks [1, 22]. In addition, as a completely different approach, the orthogonal greedy algorithm was shown to achieve an optimal convergence rate [31, 32].

The aim of this paper is to develop a novel training algorithm using several recent theoretical results on training of neural networks. A surprising recent result on training of neural networks shows that optimizing the linear layer parameters in a neural network is one bottleneck that leads to a large number of iterations of a gradient-based method. More precisely, it was proven in [15] that optimizing the linear layer parameters in a ReLU shallow neural network requires solving a very ill-conditioned linear problem in general. This work motivates us to separately design efficient solvers for the outer linear layer and the inner nonlinear layer, respectively, and train them alternately.

Meanwhile, some recent works suggest that learning a single neuron may be a more hopeful task than learning a nonlinear layer with multiple neurons. In [33, 35, 43], convergence analyses of gradient methods for the single neuron problem with ReLU activation were presented under various assumptions on input distributions. In [37], the case of a single ReLU neuron with bias was analyzed. All of these results show that the global convergence of gradient methods for the single ReLU neuron problem can be attained under certain conditions.

Inspired by the above results, we consider training of the finite neuron method in one dimension as an example, and use the well-known framework of subspace correction [39] to combine insights from the spectral properties of linear layers [15] and landscape analysis for single neuron problems [37]. Subspace correction methods provide a unified framework to design and analyze many modern iterative numerical methods such as block coordinate descent, multigrid, and domain decomposition methods. Mathematical theory of subspace correction methods for convex optimization problems is established in [23, 34], and successful applications of it to various nonlinear optimization problems in engineering fields can be found in, e.g., [4, 19]. In particular, there have been successful applications of block coordinate descent methods [38] for training of neural networks [44, 45]. Therefore, we expect that the idea of the subspace correction method is also suitable for training of neural networks for the finite neuron method.

We propose a new training algorithm called neuron-wise parallel subspace correction method (NPSC), which is a special type of subspace correction method for the finite neuron method [40]. The proposed method utilizes a space decomposition for the linear layer and each individual neuron. In the first step of each epoch of the NPSC, the linear layer is fully trained by solving a linear system using an iterative solver. We prove, both theoretically and numerically, that we can design an optimal preconditioner for the linear layer for one-dimensional problems, based on the relation between ReLU neural networks and linear finite elements investigated in [12, 15]. In the second step, we train each single neuron in parallel, taking advantages of better convergence properties of learning a single neuron and a superlinearly convergent algorithm [20]. Finally, an update for the parameters in the nonlinear layer is computed by assembling the corrections obtained in the local problems for each neuron. Due to the intrinsic parallel structure, NPSC is suitable for parallel computation on distributed memory computers. We present applications of NPSC to various function approximation problems and PDEs, and numerically verify that it outperforms conventional training algorithms.