CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2020-0004

Vol. **2**, No. 3, pp. 532-550
September 2021

Enhanced Expressive Power and Fast Training of Neural Networks by Random Projections

Jian-Feng Cai[1,*], Dong Li[2], Jiaze Sun[3] and Ke Wang[1]

[1] *Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, P.R. China.*
[2] *SUSTech International Center for Mathematics and Department of Mathematics, Southern University of Science and Technology, Shenzhen 518055, China.*
[3] *Department of Electrical and Electronic Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.*

Received 28 September 2020; Accepted 17 March 2021

Abstract. Random projections are able to perform dimension reduction efficiently for datasets with nonlinear low-dimensional structures. One well-known example is that random matrices embed sparse vectors into a low-dimensional subspace nearly isometrically, known as the restricted isometric property in compressed sensing. In this paper, we explore some applications of random projections in deep neural networks. We provide the expressive power of fully connected neural networks when the input data are sparse vectors or form a low-dimensional smooth manifold. We prove that the number of neurons required for approximating a Lipschitz function with a prescribed precision depends on the sparsity or the dimension of the manifold and weakly on the dimension of the input vector. The key in our proof is that random projections embed stably the set of sparse vectors or a low-dimensional smooth manifold into a low-dimensional subspace. Based on this fact, we also propose some new neural network models, where at each layer the input is first projected onto a low-dimensional subspace by a random projection and then the standard linear connection and non-linear activation are applied. In this way, the number of parameters in neural networks is significantly reduced, and therefore the training of neural networks can be accelerated without too much performance loss.

AMS subject classifications: 68T05, 41A30

Key words: Neural network, approximation, random projection.

---

*Corresponding author. *Email addresses:* `jfcai@ust.hk` (J.-F. Cai), `mpdongli@gmail.com` (D. Li), `j.sun19@imperial.ac.uk` (J. Sun), `kewang@ust.hk` (K. Wang)

1 Introduction

Over the past few years, learning via multiple-layer neural network has been widely studied and has achieved unprecedented success. It has many important applications in image recognition, speech recognition, and natural language processing.

One of the fundamental theoretical question in deep learning is the expressive power of a neural network, which describes its ability to approximate functions. The celebrated universal approximation theorem, which was proved by Cybenko [10], Hornick [17] et al., Funahashi [15] and Barron [3], states that sufficiently large shallow (that is, depth-2 or equivalently, one hidden layer) neural networks can approximate any continuous function on a bounded domain to arbitrary accuracy. But, with a fixed approximation accuracy, the required size of such networks can be exponentially increasing with respect to the dimension. Indeed, Eldan and Shamir [11] proved that there is a continuous function expressed by a small depth-3 feedforward neural networks which cannot be approximated by any shallow network to more than a certain constant accuracy, unless its width grows exponentially in the dimension. This shows the power of depth for feedforward neural network. Lu-Pu-Wang-Hu-Wang [19] studied the expressive power of neural networks from the width point of view. They shown that there exists a class of width-$\mathcal{O}(k^2)$ shallow ReLU network that cannot be approximated by any width-$\mathcal{O}(k^{1.5})$ and depth-k neural network.

However, the data input in the real world applications are usually structured. For example, images modelled as piecewise smooth functions can have sparse representations under certain orthonormal bases or frames [21]. This means that the intrinsic dimension of the input data is significantly smaller than the ambient space dimension. This fact is often ignored in aforementioned classical approximation results. The expressive power of a neural network may be improved by exploring the structure of the input data. In this direction, Shaham-Cloninger-Coifman [31] studied approximations of functions on a smooth k-dimensional submanifold embedded in $\mathbb{R}^d$. They constructed a depth-4 network and controlled the error of its approximation, where the size of their network depends on k but just weakly on d. Chui-Lin-Zhou [9] studied the expressive power of neural networks in the regression setting when the samples are located approximately on some unknown manifold. They showed that the error of the approximation of their trained depth-3 neural network to the regression function depends on the number of samples, and the dimension of the manifold instead of the ambient dimension.

In this paper, we consider a different approach to analyze the theoretical performance of neural networks with structured input data. Based on our analysis, we propose a new architecture of neural networks, for which the training can be significantly accelerated compared to conventional fully connected or convolutional neural networks. Our main idea is to use linear random projections developed in compressed sensing [14].

For simplicity, we assume that the input data are sparse vectors, namely, k-sparse vectors in $\mathbb{R}^d$. Using the theory of compressed sensing, one can construct a random projection onto an $\mathcal{O}(k\log(d/k))$-dimensional space that satisfies the so-called restricted