J. Math. Study
doi: 10.4208/jms.v53n2.20.03

Vol. **53**, No. 2, pp. 159-191
June 2020

PowerNet: Efficient Representations of Polynomials and Smooth Functions by Deep Neural Networks with Rectified Power Units

Bo Li[1,2], Shanshan Tang[3] and Haijun Yu[*,1,2]

[1] *NCMIS & LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Beijing 100190, China.*
[2] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.*
[3] *China Justice Big Data Institute, Beijing 100043, China.*

Received September 6, 2019; Accepted Febuary 17, 2020;
published online April 20, 2020.

Dedicated to Professor Jie Shen on the Occasion of his 60th Birthday

Abstract. Deep neural network with rectified linear units (ReLU) is getting more and more popular recently. However, the derivatives of the function represented by a ReLU network are not continuous, which limit the usage of ReLU network to situations only when smoothness is not required. In this paper, we construct deep neural networks with rectified power units (RePU), which can give better approximations for smooth functions. Optimal algorithms are proposed to explicitly build neural networks with sparsely connected RePUs, which we call PowerNets, to represent polynomials with no approximation error. For general smooth functions, we first project the function to their polynomial approximations, then use the proposed algorithms to construct corresponding PowerNets. Thus, the error of best polynomial approximation provides an upper bound of the best RePU network approximation error. For smooth functions in higher dimensional Sobolev spaces, we use fast spectral transforms for tensor-product grid and sparse grid discretization to get polynomial approximations. Our constructive algorithms show clearly a close connection between spectral methods and deep neural networks: PowerNets with n hidden layers can exactly represent polynomials up to degree s^n, where s is the power of RePUs. The proposed PowerNets have potential applications in the situations where high-accuracy is desired or smoothness is required.

AMS subject classifications: 65M12, 65M15, 65P40

*Corresponding author. *Email addresses:* libo11713092280lsec.cc.ac.cn (B. Li), tangshanshan@lsec.cc.ac.cn (S. Tang), hyu@lsec.cc.ac.cn (H. Yu)

Key words: Deep neural network, rectified linear unit, rectified power unit, sparse grid, PowerNet.

1 Introduction

Artificial neural network (ANN) has been a hot research topic for several decades. Deep neural network (DNN), a special class of ANN with multiple hidden layers, is getting more and more popular recently. Since 2006, when efficient training methods were introduced by Hinton et al [1], DNNs have brought significant improvements in several challenging problems including image classification, speech recognition, computational chemistry and numerical solutions of high-dimensional partial differential equations, see e.g. [2–6], and references therein.

The success of ANNs relies on the fact that they have good representation power. The universal approximation property of neural networks is well-known: neural networks with one hidden layer of continuous/monotonic sigmoid activation functions are dense in continuous function space $C([0,1]^d)$ and $L^1([0,1]^d)$, see e.g. [7–9] for different proofs in different settings. Actually, for neural network with non-polynomial C^∞ activation functions, the upper bound of approximation error is of spectral type even using only one-hidden layer, i.e. error rate $\varepsilon = n^{-k/d}$ can be obtained theoretically for approximation functions in Sobolev space $W^k([-1,1]^d)$, where d is the number of dimensions, n is the number of hidden nodes in the neural network [10]. However, it is believed that one of the basic reasons behind the success of DNNs is the fact that deep neural networks have broader scopes of representation than shallow ones. Recently, several works have demonstrated or proved this in different settings. For example, by using the composition function argument, Poggio et al [11] showed that deep networks can avoid the curse of dimensionality for an important class of problems corresponding to compositional functions. In the general function approximation aspect, it has been proved by Yarotsky [12] that DNNs using rectified linear units (abbr. ReLU, a non-smooth activation function defined as $\sigma_1(x) := \max\{0,x\}$) need at most $\mathcal{O}(\varepsilon^{\frac{d}{k}}(\log|\varepsilon|+1))$ units and nonzero weights to approximation functions in Sobolev space $W^{k,\infty}([-1,1]^d)$ within ε error. This is similar to the results of shallow networks with one hidden layer of C^∞ activation units, but only optimal up to a $\mathcal{O}(\log|\varepsilon|)$ factor. Similar results for approximating functions in $W^{k,p}([-1,1]^d)$ with $p<\infty$ using ReLU DNNs are given by Petersen and Voigtlaender [13]. The significance of the works by Yarotsky [12] and Peterson and Voigtlaender [13] is that by using a very simple rectified nonlinearity, DNNs can obtain high order approximation property. It is also proved by E and Wang [14] that thin and deep ReLU networks can approximate analytic functions exponentially fast. Shallow networks do not hold such a good property. Other works show deeper ReLU DNNs have better approximation property include the work by He *et al.* [15] and the work by Opschoor *et al.* [16], which relate ReLU DNNs to finite element methods.