AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(10): 19018–19025.
DOI: 10.3934/math.20221045
Received: 30 March 2022
Revised: 07 August 2022
Accepted: 08 August 2022
Published: 29 August 2022

Research article

On the uniform approximation estimation of deep ReLU networks via frequency decomposition

Liang Chen* **and Wenjun Liu**

Department of Mathematics, Jiujiang University, Jiujiang 332005, China

* **Correspondence:** Email: chenliang3@mail2.sysu.edu.cn.

Abstract: A recent line of works established the approximation complexity estimation of deep ReLU networks for the bandlimited functions in the MSE (mean square error) sense. In this note, we significantly enhance this result, that is, we estimate the approximation complexity in the L_∞ sense. The key to the proof is to establish a frequency decomposition lemma which may be of independent interest.

Keywords: bandlimited functions; fat-shattering dimension; deep ReLU networks; approximation complexity
Mathematics Subject Classification: 41A30, 41A63

1. Introduction

The approximation theory of neural networks has been widely concerned in recent years [6, 12, 18, 21, 22, 24–27, 30–35]. There are two widely accepted advantages of neural networks, one is that neural network can overcome the curse of dimensionality [1, 4, 5, 7, 9, 11, 15–17], the other is that as the number of layers increases, deep neural networks can fit high-oscillation or high-frequency functions [3, 17–19, 21, 23]. For the bandlimited functions, the approximation complexity of deep ReLU networks has been studied [3, 17] in the MSE sense. Compared with the results of the shallow neural networks [1], these results show deep ReLU networks can approximate the ultrawide bandwidth high-dimensional signals with very low complexity. In this paper, we continue this line of research, we estimate the approximation complexity of deep ReLU networks in the L_∞ sense, thus, this estimate is significantly better than the previous results [3, 17]. It worths mentioning that the approximate complexity estimations of the shallow neural networks are significantly different for different error norms [9, 13]. Similarly, for the deep ReLU networks, we can not resort to the Maurey's theorem [20] in the L_∞ case. We need to use the fat-shattering dimension theory [8] to establish a frequency decomposition lemma which is our main result.

Lemma 1.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a bandlimited function of the form*

$$f(x) = \int_{[-M,M]^d} \hat{f}(w) e^{i2\pi x \cdot w} dw,$$

and

$$\int_{[-M,M]^d} |\hat{f}(w)| dw := C_f < \infty,$$

there exist $2l$ frequencies

$$w_{1,1}, w_{1,2}, \ldots, w_{1,l}, w_{2,1}, w_{2,2}, \ldots, w_{2,l} \in [-M, M]^d,$$

and $2l$ coefficients

$$a_{1,1}, a_{1,2}, \ldots, a_{1,l}, a_{2,1}, a_{2,2}, \ldots, a_{2,l} \in \mathbb{R}$$

satisfying

$$\sum_{j=1}^{l} (|a_{1,j}| + |a_{2,j}|) \leq 2C_f$$

and

$$l \geq C(dC_f^2/\epsilon^2) \log(C_D M + \log(C_f/\epsilon)) \ln^2(C_f/\epsilon),$$

such that

$$\sup_{x \in D} \left| f(x) - \sum_{j=1}^{l} \Big(a_{1,j} \cos(2\pi w_{1,j} \cdot x) + a_{2,j} \sin(2\pi w_{2,j} \cdot x) \Big) \right| < \epsilon, \tag{1.1}$$

where C is some universal constant, $0 < \epsilon < 1/2$, D is any bounded subset of $\mathbb{R}^d$ and

$$C_D := \sup_{x \in [-1,1]^d} \sup_{y \in D} |x \cdot y|.$$

On the other hand, a *feedforward neural network* $\tilde{X}$ consists of layers of nodes, each node of the form

$$y = \sigma\Big(\sum_{k=1}^{N} w_k x_k + b\Big), \tag{1.2}$$

where $\{x_k : k = 1, 2, \ldots, N\}$ are nodes in the previous layers, and σ is a activation function. If $\sigma(x) = \max(x, 0)$, the multilayer feedforward neural network is the deep ReLU (Rectified Linear Unit) network. The complexity of deep neural networks approximating to sinusoidal functions has been discussed [3, 17, 19].

Lemma 1.2. *([3]) There exists a deep ReLU network $\tilde{X}_\epsilon$ with $O(\log(C_D M) \log_2^2(1/\epsilon))$ layers and $O(\log C_D M \log_2^3(1/\epsilon))$ nodes such that*

$$|\tilde{X}_\epsilon(t) - \cos(2\pi w \cdot t)| \leq \epsilon \quad \text{for} \quad w \in [-M, M]^d, \quad t \in D. \tag{1.3}$$

Based on Lemma 1.1 and Lemma 1.2, the following main theorem can be obtained.

Theorem 1.3. *Let $f : D \to \mathbb{R}$ be a bandlimited function of the form*

$$f(x) = \int_{[-M,M]^d} \hat{f}(w) e^{i2\pi x\cdot w} dw,$$

and

$$C_f := \int_{[-M,M]^d} |\hat{f}(w)| dw < \infty,$$

there exists a deep ReLU network $\tilde{X}_\epsilon$ with

$$O(\log(C_D M)\log_2^2(1/\epsilon))$$

layers and

$$O((\log(C_D M)\log_2^3(1/\epsilon) d C_f^2/\epsilon^2)\log_2(C_D M + \log(C_f/\epsilon))\ln^2(C_f/\epsilon))$$

nodes such that

$$\sup_{x\in D}\left| f(x) - \tilde{X}_\epsilon \right| < \epsilon. \tag{1.4}$$

2. Proof of the Lemma 1.1

For any complex value $z \in \mathbb{C}$, $\Re z$ and $\Im z$ denote the real part and the imaginary part of z, respectively. We denote by C_j, $j = 1, 2, \ldots$ some universal constants.

To prove Lemma 1.1, some notations and lemmas are needed. Given a class $\mathcal{H}$ of real-valued functions defined on X and a real number $\gamma \geq 0$. The *fat-shattering dimension* of $\mathcal{H}$, denoted $\text{fat}_H(\gamma)$, is the largest integer m satisfying the following property: There exist $V = \{x_1, x_2, \ldots, x_m\}$ is a subset of the domain X, and the real numbers $r_1, r_2, \ldots, r_m$ such that for all $b \in \{0, 1\}^m$ there is a function f_b in $\mathcal{H}$ with $f_b(x_i) \geq r_i + \gamma$ if $b_i = 1$ and $f_b(x_i) < r_i - \gamma$ if $b_i = 0$ for $1 \leq i \leq m$. Fat-shattering dimension is a generalization of Vapnik-Chervonenkis dimension [29] which is often used to estimate approximation complexity [10, 14, 35].

Lemma 2.1. *([2]) Let $\mathcal{H}$ be a class of functions from some domain X into* $[0, 1]$. *Then for all distributions $\mathbb{P}$ over X and for all* $0 < \epsilon, \delta < 1/2$

$$\text{Pr}_{x_1,x_2,\ldots,x_l}\left[\sup_{f\in\mathcal{H}}\left|\frac{1}{m}\sum_{i=1}^{m} f(x_i) - \text{E}_{x\sim\mathbb{P}}[f(x)]\right| \geq \epsilon\right] \leq \delta$$

where $\{x_i\}_{i=1}^l$ are l independent draws from $\mathbb{P}$, $\text{Pr}_{x_1,x_2,\ldots,x_l}$ denotes the joint probability distribution function for random variables $x_1, x_2, \ldots, x_l$ and

$$l = O\left(\frac{1}{\epsilon^2}\left(\text{fat}_H(\epsilon/5)\ln^2\frac{1}{\epsilon} + \ln\frac{1}{\delta}\right)\right).$$

Lemma 2.2. *([28]) Let $p_1, p_2, \ldots, p_l$ be d-variable real polynomials of degree at most q. If $l \geq d$, then the number of vectors*

$$(\text{sgn}(p_1(x)), \text{sgn}(p_2(x)), \ldots, \text{sgn}(p_l(x)))$$

is at most $(4elq/d)^d$, *where* $\text{sgn}(x) := \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases}$

By the Taylor expansion, the sines and cosines can be approximation by polynomials.

Lemma 2.3. *Let* $0 < \epsilon < 1$, $a > 0$ *and* $f(x) = \sin(x)$ *or* $\cos(x)$. *Then there exist* $C_1 > 0$ *and a polynomial* Q *of order less than or equal to* $C_1(a + \log(1/\epsilon))$ *such that* $\sup_{x\in[-a,a]} |f(x) - Q(x)| \leq \epsilon$.

Proof. The Taylor expansion for an n times differentiable function f at 0 with the Lagrange remainder is given by

$$f(x) = \sum_{j=1}^{n-1} \frac{f^{(j)}(0)}{j!} x^j + \frac{f^{(n)}(x_0)}{n!} x^n, \tag{2.1}$$

where x_0 is between 0 and x, $f^{(j)}$ denotes the derivative of order j, and the factorial $n! = n(n-1)\cdots 1$. Applying the Stirling formula to $n!$, we have

$$n! \geq \sqrt{2\pi n}\Big(\frac{n}{e}\Big)^n > \Big(\frac{n}{e}\Big)^n, \ n \in \mathbb{N}. \tag{2.2}$$

Note that $|\sin^{(n)}(x)| \leq 1$ and $|\cos^{(n)}(x)| \leq 1$ for all $x \in \mathbb{R}$. Thus, by (2.1) and (2.2), we have for all $n \geq \max\{2ea, \frac{1}{\log 2}\log(\frac{1}{\epsilon})\}$

$$\sup_{x\in[-a,a]} \Big|\frac{f^{(n)}(x_0)}{n!}(x)^n\Big| \leq \frac{a^n}{n!} < \Big(\frac{ea}{n}\Big)^n < \frac{1}{2^n} < \epsilon.$$

The proof is hence complete. □

Proof of Lemma 1.1. Since f is a real-valued function, we have for all $x \in \mathbb{R}^d$

$$\begin{aligned} f(x) &= \int_{[-M,M]^d} \hat{f}(w) e^{i2\pi w\cdot x} dw \\ &= \int_{[-M,M]^d} (\Re \hat{f}(w) + i\Im \hat{f}(w))(\cos(2\pi w\cdot x) + i\sin(2\pi w \cdot x))dw \\ &= \int_{[-M,M]^d} \cos(2\pi w\cdot x)\Re\hat{f}(w) - \sin(2\pi w\cdot x))\Im\hat{f}(w) dw \\ &=: f_1 - f_2 - f_3 + f_4, \end{aligned} \tag{2.3}$$

where

$$f_1(x) := \int_{[-M,M]^d} \cos(2\pi w\cdot x)\max(0, \Re\widehat{f}(w))dw,$$

$$f_2(x) := \int_{[-M,M]^d} \cos(2\pi w\cdot x)\max(-\Re\widehat{f}(w), 0)dw,$$

$$f_3(x) := \int_{[-M,M]^d} \sin(2\pi w\cdot x))\max(0, \Im\widehat{f}(w))dw,$$

$$f_4(x) := \int_{[-M,M]^d} \sin(2\pi w\cdot x))\max(-\Im\widehat{f}(w), 0)dw.$$

Let $F_1 := \max(0, \Re\widehat{f})$. Without loss of generality, we assume that

$$\int_{[-M,M]^d} F_1(w)dw > 0.$$

We only discuss the approximation for the first term

$$f_1(x) = \int_{[-M,M]^d} \cos(2\pi w \cdot x) F_1(w) dw,\ x \in \mathbb{R}^d.$$

The other three terms f_2, f_3, f_4 in (2.3) can be handled in a similar manner. Observe that

$$|w \cdot x| \le C_D M \text{ for all } w \in [-M, M]^d \text{ and } x \in D.$$

By Lemma 2.3, there exists $C_2 > 0$ and a polynomial Q of degree $C_2(C_D M + \log(1/\epsilon))$ such that

$$\Big|\cos(2\pi w \cdot x) - Q(2\pi w \cdot x)\Big| < \epsilon \text{ for all } w \in [-M, M]^d \text{ and } x \in D. \tag{2.4}$$

Let

$$\tilde{f}(x) := \int_{[-M,M]^d} Q(2\pi w \cdot x) F_1(w) dw,\ x \in \mathbb{R}^d.$$

Then, we get

$$|\tilde{f}(x) - f_1(x)| < \epsilon \int_{[-M,M]^d} F_1(w) dw, \quad x \in D. \tag{2.5}$$

Set

$$G := \{g_x(w) = \cos(2\pi x \cdot w) : x \in D, w \in [-M, M]^d\}$$

and

$$G_\epsilon = \{g_x(w) = Q(2\pi x \cdot w) : x \in D, w \in [-M, M]^d\}.$$

By Lemma 2.2, the number of vectors

$$(\text{sgn}(Q(2\pi w_1 \cdot x)), \text{sgn}(Q(2\pi w_2 \cdot x)), \ldots, \text{sgn}(Q(2\pi w_l \cdot x))$$

is at most $(4elC_2(C_D M + \log(1/\epsilon))/d)^d$. Thus,

$$2^{fat_{G_\epsilon}(0)} \le (4e fat_{G_\epsilon}(0) C_2(C_D M + \log(1/\epsilon)/d)^d,$$

that is

$$fat_{G_\epsilon}(0) \le d \log_2 (4eC_2(C_D M + \log(1/\epsilon)) + d \log_2 \Big(\frac{fat_{G_\epsilon}(0)}{d}\Big).$$

Note that

$$d \log_2 \Big(\frac{fat_{G_\epsilon}(0)}{d}\Big) \le \frac{1}{2} fat_{G_\epsilon}(0) \quad \text{for} \quad fat_{G_\epsilon}(0) > 4d,$$

then

$$\text{fat}_G \Big(\epsilon \int_{[-M,M]^d} F_1(w) dw\Big) \le \text{fat}_{G_\epsilon}(0) \le 2d \log_2 (8eC_2(C_D M + \log(1/\epsilon))) + 4d.$$

By Lemma 2.1, choosing $\delta < 1/4$ and

$$l \gtrsim \frac{1}{\epsilon^2}\Big(d \log_2(C_D M + \log(1/\epsilon)) \ln^2 \frac{1}{\epsilon} + \ln \frac{1}{\delta}\Big),$$

we have

$$Pr_{w_1,w_2,\dots,w_l} \left(\sup_{x\in D,} \left| \frac{\int_{[-M,M]^d} F_1(w)dw}{l} \sum_{j=1}^{l} \cos(w_j \cdot x) - f_1(x) \right| > \epsilon \int_{[-M,M]^d} F_1(w)dw \right) \le \delta < 1,$$

where $\{w_i\}_{i=1}^{l}$ obeys the probability distribution generated by the density function

$$\frac{F_1(w)}{\int_{[-M,M]^d} F_1(w)dw}.$$

Therefore, there exist $w_1, w_2, \dots, w_l$ such that

$$\left| \frac{\int_{[-M,M]^d} F_1(w)dw}{l} \sum_{j=1}^{l} \cos(w_j \cdot x) - f_1(x) \right| \le O(\epsilon \int_{[-M,M]^d} F_1(w)dw). \tag{2.6}$$

The approximation for f_2 is similar to f_1 given as in (2.6). Hence, we can obtain an approximation for f_3 or f_4 by the linear combination of sines. The sum of $2l$ coefficients in (1.1) is bounded by

$$\begin{aligned} \sum_{j=1}^{l}(|a_{1,j}| + |a_{2,j}|) \quad &\le \int_{[-M,M]^d} (\max(0, \Re \widehat{f}(w)) + \max(0, -\Re \widehat{f}(w)) \\ &+ \max(0, \Im \widehat{f}(w)) + \max(0, -\Im \widehat{f}(w)))dw \le 2C_f. \end{aligned} \tag{2.7}$$

The proof is hence complete.

3. Conclusions

In this paper we prove an interesting frequency decomposition lemma, as a corollary, we establish the uniform approximation estimation of bandlimited functions via deep ReLU networks. We will discuss the learnability of neural networks for bandlimied functions in the following work. These theoretical results will deepen the understanding of machine learning methods in signal processing.

Acknowledgements

The authors are thankful to their Universities for their support. This work was supported by the National Natural Science Foundation of China (No. 11661065).

Conflict of interest

The authors declare that they have no competing interests.

References

1. A. R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, *IEEE T. Inform. Theory*, **39** (1993), 930–945. https://doi.org/10.1109/18.256500
2. P. L. Bartlett, P. M. Long, More theorems about scalesensitive dimensions and learning, *Proceedings of the Eighth Annual Conference on Computational Learning Theory*, 1995, 392–401.

3. L. Chen, C. Wu, A note on the expressive power of deep rectified linear unit networks in high-dimensional spaces, *Math. Methods Appl. Sci.*, **42** (2019), 3400–3404. https://doi.org/10.1002/mma.5575
4. N. Cohen, O. Sharir, A. Shashua, On the expressive power of deep learning: A tensor analysis, *29th Annual Conference on Learning Theory*, 2016, 698–728.
5. R. Eldan, O. Shamir, The power of depth for feedforward neural networks, *29th Annual Conference on Learning Theory*, 2016, 907–940.
6. I. Gühring, G. Kutyniok, P. Petersen, Error bounds for approximations with deep ReLU neural networks in $W^{s,p}$ norms, *Anal. Appl.*, **18** (2020), 803–859. https://doi.org/10.1142/S0219530519410021
7. A. Jentzen, D. Salimova, T. Welti, A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients, *arXiv*, 2018. https://arxiv.org/abs/1809.07321
8. M. J. Kearns, R. E. Schapire, Efficient distribution-free learning of probabilistic concepts, *J. Comput. Syst. Sci.*, **48** (1994), 464–497. https://doi.org/10.1016/S0022-0000(05)80062-5
9. J. M. Klusowski, A. R. Barron, Approximation by combinations of ReLU and squared ReLU ridge functions with ℓ^1 and ℓ^0 controls, *IEEE T. Inform. Theory*, **64** (2018), 7649–7656. https://doi.org/10.1109/TIT.2018.2874447
10. V. N. Konovalov, On the orders of nonlinear approximations for classes of functions of given form, *Math. Notes*, **78** (2005), 88–104. https://doi.org/10.1007/s11006-005-0102-3
11. H. Lee, R. Ge, T. Ma, A. Risteski, S. Arora, On the ability of neural nets to express distributions, *Proceedings of the 2017 Conference on Learning Theory*, **65** (2017), 1271–1296.
12. S. Liang, R. Srikant, Why deep neural networks for function approximation? *arXiv*, 2016. https://arxiv.org/abs/1610.04161
13. Y. Makovoz, Uniform approximation by neural networks, *J. Approx. Theory*, **95** (1998), 215–228. https://doi.org/10.1006/jath.1997.3217
14. V. Maiorov, J. Ratsaby, On the degree of approximation by manifolds of finite pseudo-dimension, *Constr. Approx.*, **15** (1999), 291–300. https://doi.org/10.1007/s003659900108
15. H. N. Mhaskar, T. Poggio Deep vs. shallow networks: An approximation theory perspective, *Anal. Appl.*, **14** (2016), 829–848. https://doi.org/10.1142/S0219530516400042
16. H. Montanelli, Q. Du, New error bounds for deep ReLU networks using sparse grids, *SIAM J. Math. Data Sci.*, **1** (2019), 78–92. https://doi.org/10.1137/18M1189336
17. H. Montanelli, H. Yang, Q. Du, Deep ReLU networks overcome the curse of dimensionality for bandlimited functions, *arXiv*, 2019. https://arxiv.org/abs/1903.00735
18. G. F. Montúfar, R. Pascanu, K. Cho, Y. Bengio, On the number of linear regions of deep neural networks, *Adv. Neural Inform. Proc. Syst.*, 2014, 2924–2932.
19. D. Perekrestenko, P. Grohs, D. Elbrachter, H. Bölcskei, The universal approximation power of finite-width deep ReLU networks, *arXiv*, 2018. https://arxiv.org/abs/1806.01528
20. G. Pisier, Remarques sur un résultat non publié de B. Maurey, *Séminaire Anal. Fonct. (Polytechnique)*, 1981, 1–12.
21. M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, J. Sohl-Dickstein On the expressive power of deep neural networks, *Proceedings of the 34th International Conference on Machine Learning*, **70** (2017), 2847–2854.

22. U. Shaham, A. Cloninger, R. R. Coifman, Provable approximation properties for deep neural networks, *Appl. Comput. Harmon. Anal.*, **44** (2018), 537–557. https://doi.org/10.1016/j.acha.2016.04.003
23. T. Serra, C. Tjandraatmadja, S. Ramalingam, Bounding and counting linear regions of deep neural networks, *Proceedings of the 35th International Conference on Machine Learning*, 2018, 4558–4566.
24. Z. Shen, H. Yang, S. Zhang, Nonlinear approximation via compositions, *Neural Networks*, **119** (2019), 74–84. https://doi.org/10.1016/j.neunet.2019.07.011
25. Z. Shen, H. Yang, S. Zhang, Deep network approximation with discrepancy being reciprocal of width to power of depth, *arXiv*, 2020.
26. Z. Shen, H. Yang, S. Zhang, Neural network approximation: Three hidden layers are enough, *Neural Networks*, **141** (2021), 160–173. https://doi.org/10.1016/j.neunet.2021.04.011
27. M. Telgarsky, Representation benefits of deep feedforward networ, *arXiv*, 2015. https://arxiv.org/abs/1509.08101
28. H. E. Warren, Lower bounds for approximation by nonlinear manifolds, *Trans. Amer. Math. Soc.*, **133** (1968), 167–178. https://doi.org/10.1090/S0002-9947-1968-0226281-1
29. V. N. Vapnik, A. Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, In: V. Vovk, H. Papadopoulos, A. Gammerman, *Measures of complexity*, Springer, 1971, 11–30. https://doi.org/10.1007/978-3-319-21852-6_3
30. D. Yarotsky, Error bounds for approximations with deep ReLU networks, *Neural Networks*, **94** (2017), 103–114. https://doi.org/10.1016/j.neunet.2017.07.002
31. D. X. Zhou, Universality of deep convolutional neural networks, *Appl. Comput. Harmon. Anal.*, **48** (2020), 787–794. https://doi.org/10.1016/j.acha.2019.06.004
32. Y. Xu, H. Zhang, Convergence of deep convolutional neural networks, *Neural Networks*, **153** (2022), 553–563. https://doi.org/10.1016/j.neunet.2022.06.031
33. R. DeVore, H. Boris, P. Guergana, Neural network approximation, *Acta Numer.*, **30** (2021), 327–444. https://doi.org/10.1017/S0962492921000052
34. I. Gühring, R. Mones, Approximation rates for neural networks with encodable weights in smoothness spaces, *Neural Networks*, **134** (2021), 107–130. https://doi.org/10.1016/j.neunet.2020.11.010
35. D. Dũng, V. K. Nguyen, Deep ReLU neural networks in high-dimensional approximation, *Neural Networks*, **142** (2021), 619–635. https://doi.org/10.1016/j.neunet.2021.07.027