Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Deep ReLU neural network expression for elliptic multiscale problems

V.H. Hoang and Ch. Schwab

Research Report No. 2020-24
April 2020

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding: The Singapore MOE Tier 2 grant MOE2017-T2-144

DEEP RELU NEURAL NETWORK EXPRESSION FOR ELLIPTIC MULTISCALE PROBLEMS

VIET HA HOANG[1] AND CHRISTOPH SCHWAB[2]

Abstract. We analyze expression rates of deep ReLU neural network (DNN) approximations for several solution families of two-scale, linear, second order elliptic boundary value problems with either locally periodic or quasi-periodic setting. We prove that DNNs can approximate the multiscale solution families with error $\delta > 0$ in the norm of the Sobolev space H^1 at an NN expression rate which is essentially independent of the scale parameter ε.

1. Introduction

After fundamental advances in classification related tasks in data science and in data intensive applications, recent years have seen a brisk advance of deep neural networks (DNNs) for the numerical solution of partial differential equations. There has been a broad research effort to harness the approximation capabilities of DNNs for the numerical approximation of solutions of PDEs

Many of the publications and results in recent years have been algorithmic in nature, i.e., algorithms were proposed which offered capability of approximation of solutions of PDEs. We mention only [20, 23, 25] for so-called "physics-informed" DNNs which use, in some form, explicitly first principles of physical processes (e.g. conservation of mass, momentum, energy) modelled by the PDE under consideration. Additional applications to computational uncertainty quantification of PDEs by means of DNNs have also been proposed in recent years; let us mention only [27, 26] and the references there, and [24] for theory. It has been found in these references that depth of NNs is very beneficial for enhanced approximation properties of the DNNs.

These computational developments have been paralleled, with slight delay, by theoretical insights which explain the at times competitive approximation properties, or "*expressive power*" of DNNs, i.e., the ability of the DNN to represent a PDE solution to, say, accuracy $\delta > 0$ in a physically meaningful norm (which is not, in general, the mean square or pointwise error, but rather an "energy norm" induced by the physical properties of the PDE of interest).

Expression rate analysis of ReLU DNNs for various function systems have been given in [21]. These results emphasize quantative expression rate bounds which are more refined, in a sense, than the early results on universality of DNNs from the 90ies; these results are surveyed in [22] and the references there. Expression

VHH is supported by The Singapore MOE Tier 2 grant MOE2017-T2-144. This work was initiated during a visit of VHH to the Institute for Mathematical Research (FIM) ETH Zürich in October 2019. He thanks the Institute for the hospitality and the excellent working conditions.

rate analysis of DNNs for data-to-QoI maps in PDE constrained Bayesian inverse problems has recently been provided in [12].

1.1. **Previous Results.** Numerous references have appeared in recent years which propose *algorithms for the numerical approximation of PDE solution families* by DNNs. We mention only [27, 25, 24, 26] and the references there for DNN based PDE solvers, and [4, 23], and the references there. In particular, in these references DNNs were proposed as a PDE solution discretization, and the NN parameters in the training process were obtained from minimizing a physical variational principle, such as, e.g., the potential energy in the system described by a particular NN configuration. Without aiming at a complete list, we mention recent works which address the *mathematical analysis of expressive power of DNNs*; we focus on references with particular relevance to the presently developed results. Representative papers of this direction of thought are [27, 25, 24, 26] and the references there for DNN based PDE solvers, and [4, 23, 2].

Also DNN approximation theory has advanced in recent years. After fundamental results on DNN universality in 90ies (see [22] and the references there) in recent years, more refined results on DNN expressive power rates for specific classes of inputs have been developed. We mention only [5, 17, 19, 6, 30]. In these works, the DNN approximation of certain elementary functions, most notably univariate polynomials, has been analyzed. One remarkable result in [17, 30] is that *ReLU DNNs are able to represent polynomials of high degree at an exponential rate* in terms of the DNN size, despite their realizations being continuous, piecewise affine functions.

Viceversa, fixed order, piecewise polynomials such as splines and continuous, piecewise affine "Courant" Finite Element functions are *equivalent* to ReLU DNNs (see, e.g., [11] [19] for the (elementary) details of the argument).

To establish DNN expression rates for linear elliptic partial differential equations (PDEs for short) with multiple scales that are explicit in the scale parameter in the PDE and in the expression error is the purpose of the present paper.

1.2. **Contributions.** In the present paper, we make the following contributions. For a classical, linear elliptic 2-scale homogenization problem, with periodic dependence of the diffusion coefficient on the fast variable, we prove rates of expression for DNN approximation of the parametric family $\{u_\varepsilon : 0 < \varepsilon \leq 1\} \subset H^1_0(\mathrm{D})$ of solutions. This regularity is optimal under the assumption that the forcing f of the elliptic PDE in D is in $H^{-1}(\mathrm{D})$ and the coefficient function is Lipschitz. More restrictive structural hypotheses on the data f and a will, of course, imply corresponding stronger expression rate results, and we shall explore some of these here as well.

To prove DNN expression rate bounds for solutions of the two-scale PDE, we construct DNN approximations of the solution in two steps: first, using (classical) 2-scale asymptotic expansions with explicit Fourier series expansion of the 2-scale corrector with respect to the fast scale and, subsequently, by ReLU DNN reapproximation of the Fourier expansion. Here, a key observation is the moderate increase of DNN depth with respect to the Fourier index k and the scale parameters ε, which was pointed out in [10]. These results are limited to homogenization problems with two separated scales in coefficients. In the final section, we prove several generalizations. In particular, we indicate in one space dimension that DNNs are capable of similar approximation rates for considerably more general coefficients; among these are almostperiodic and certain classes of fractal coefficient functions.

1.3. **Structure of this paper.** The structure of the present paper is as follows. In Section 2 we present formulation of the homogenization problem. In particular, we define the function space setting, and recapitulate known results on existence, uniqueness and on regularity of the two scale family of solutions. Furthermore, we indicate results and error bounds on the 2-scale asymptotics of the solution family $\{u_\varepsilon : 0 < \varepsilon \leq 1\}$ which will be required subsequently for the derivation and proof of the DNN expression rate bounds. In Section 3 we present basic notation and terminology of the DNNs which we shall consider for the solution approximation. In particular, we focus on so-called deep ReLU NNs. Section 4 has our main results: ReLU DNN expression rate bounds for the solution family $\{u^\varepsilon : 0 < \varepsilon \leq 1/2\} \subset H_0^1(\mathrm{D})$ of the two-scale problem (2.1) - (2.2), for $f \in H^{-1}(\mathrm{D})$ which are explicit in the $H^1(\mathrm{D})$-accuracy of the NN and in the scale parameter ε. Section 5 discusses various conclusions and generaliations. In particular, also an extension of the results to problems with a continuum of scales, i.e., without scale separation.

Throughout the paper, by #, we denote Banach spaces of periodic function over the period cube Y.

2. Two-scale Model Elliptic Homogenization Problem

Let $\mathrm{D} \subset \mathbb{R}^d$ be a bounded polytope of dimension $d \geq 1$ with plane faces so that $\partial\mathrm{D}$ is Lipschitz and denote by $\mathrm{Y} = (0,1)^d$ the cell in $\mathbb{R}^d$. We also assume that $\mathrm{diam}(\mathrm{D}) = O(1)$ in order to render the notion of two-scale problem with nondimensional small scale parameter $0 < \varepsilon \leq 1/2$ meaningful[1].

Assume given a diffusion coefficient $a \in C^0(\overline{\mathrm{D} \times \mathrm{Y}}; \mathbb{R}^{d\times d}_{sym})$ (subsequently, somewhat stronger smoothness requirements will be imposed in order to obtain approximation rate bounds for DNNs).

Assume in particular that for every $x \in \overline{\mathrm{D}}$, the map $[\mathrm{Y} \ni y \mapsto a(x,y)]$ is Y-periodic, and that a is elliptic: there exist constants $0 < a_{min} \leq a_{max} < \infty$ such that

$$\forall x \in \mathrm{D}, y \in \mathrm{Y},\ \forall \xi \in \mathbb{R}^d : \quad a_{min}|\xi|^2 \leq \xi^\top a(x,y)\xi \leq a_{max}|\xi|^2\ , \tag{2.1}$$

where $|\cdot|$ denotes the Euclidean norm of a vector ξ in $\mathbb{R}^d$.

Let $\varepsilon > 0$ be a small quantity that represents the microscopic scale of the physical problem. We denote the two scale coefficient $a^\varepsilon : \mathrm{D} \to \mathbb{R}^{d\times d}_{sym}$ as

$$a^\varepsilon(x) = a(x, \frac{x}{\varepsilon}).$$

We denote $V = H_0^1(\mathrm{D})$ and assume given $f \in V' = H^{-1}(\mathrm{D})$, where duality is taken w.r. to the pivot space $L^2(\mathrm{D})$. We consider the elliptic problem

$$-\nabla \cdot (a^\varepsilon \nabla u^\varepsilon) = f \quad \text{in} \quad \mathrm{D} \tag{2.2}$$

with homogeneous Dirichlet boundary condition $u^\varepsilon = 0$ on $\partial\mathrm{D}$. The variational form of (2.2) reads: given $f \in V'$ and $0 < \varepsilon \leq 1/2$, find $u^\varepsilon \in V$ such that for every $v \in V$ holds

$$(a(\cdot, \frac{\cdot}{\varepsilon})\nabla u^\varepsilon, \nabla v) = (f, v)\ . \tag{2.3}$$

[1]The assumption $0 < \varepsilon \leq 1/2$ is to render $|\log \varepsilon|$ uniformly bounded from 0

Here, $(\cdot,\cdot)$ signifies the $L^2(\mathrm{D})$-innerproduct, extended in the right hand side of (2.3) in the usual way to a duality pairing between V' and V by density and continuity. The ellipticity (2.1) and the assumption $f \in V'$ imply that for every $0 < \varepsilon < 1$ the weak problem (2.3) admits a unique solution $u^\varepsilon \in V$. We are interested in the expression of the solution set $\{u^\varepsilon(x) : x \in \mathrm{D}, 0 < \varepsilon \le 1/2\}$ by DNNs.

We summarize some well known results in the theory of homogenization for (2.2). Formally, the usual two-scale asymptotic expansion (e.g. [3, 1, 16])

$$u^\varepsilon(x) = u_0(x, \frac{x}{\varepsilon}) + \varepsilon u_1(x, \frac{x}{\varepsilon}) + \dots, \tag{2.4}$$

where $u_i(x, y)$ are Y-periodic with respect to y implies that u_0 does not depend on y. Furthermore, the first order "corrector" term u_1 can be written in terms of u_0 and of the solution of a parametric family of so-called "unit-cell problems" (assuming sufficient regularity of this solution w.r. to the slow variable x).

Specifically, for $i = 1, \dots, d$, and for every $x \in \mathrm{D}$, let $N^i(x, \cdot)$ denote a Y-periodic function of $y \in \mathrm{Y}$ which is the solution of the parametric (with parameter $x \in \mathrm{D}$) family of unit cell problems:

$$-\nabla_y \cdot (a(x, y)(e^i + \nabla_y N^i(x, y))) = 0 \ \text{ in } \ \mathrm{Y}\ . \tag{2.5}$$

Here e^i is the ith unit vector in the standard basis of $\mathbb{R}^d$. It is known (e.g. [3, 1, 16]) that the so-called corrector term in (2.4) has separated form:

$$u_1(x, y) = \sum_{i=1}^{d} \frac{\partial u_0}{\partial x_i}(x) N^i(x, y). \tag{2.6}$$

When u_0 and N^i are sufficiently smooth functions of their arguments, we have the following approximations for u^ε.

Proposition 2.1. *Assume that $u_0 \in H^2(\mathrm{D})$ and that $N^i \in C^1(\bar{\mathrm{D}}, C^1_{per}(\bar{\mathrm{Y}}) \bigcap H^2_{per}(\mathrm{Y}))$. Then*

$$\left\| \nabla u^\varepsilon - [\nabla u_0(\cdot) + \nabla_y u_1(\cdot, \frac{\cdot}{\varepsilon})] \right\|_{L^2(\mathrm{D})} \le c\varepsilon^{1/2} \tag{2.7}$$

where the positive constant c does not depend on ε.

Remark 2.2. *Estimate* (2.7) *has been established under stronger regularity conditions in the literature. For example, under the assumption $u_0 \in H^2(\mathrm{D}) \cap W^{1,\infty}(\mathrm{D})$ and $u_0 \in C^2(\bar{\mathrm{D}})$ in* [3, pg.66] *and* [16, pg. 28]*, respectively. However,* (2.7) *holds under the weaker regularity assumption $u_0 \in H^2(\mathrm{D})$ (see, e.g,* [15, Prop. 5.1]*).*

3. Deep ReLU neural networks

We adopt the notations in the recent publications [10, 21] and the references there.

We denote the ReLU activation $\sigma(x) = \max\{x, 0\}$. A DNN with a d-dimensional input and N_L dimensional output is a map $\Phi : \mathbb{R}^d \to \mathbb{R}^{N_L}$ such that

$$\Phi(x) = W_L(\sigma(W_{L-1}(\sigma(\dots \sigma(W_1(x)) \dots)))),$$

where W_l is a linear map from $\mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}$ such that $W_l(x) = A_l x + b_l$ where $A_l \in \mathbb{R}^{N_l \times N_{l-1}}$ and $b_l \in \mathbb{R}^{N_l}$. Here, $x \in \mathbb{R}^d$ with $N_0 = d$ being the input dimension and N_L being the output dimension. We shall say that the DNN has L layers and $L - 1$ hidden layers. Let M be the total nonzero entries of the matrices A_l and vectors b_l, M is the connectivity of the network. We have that $M \le LW(W + 1)$

where $W = \max_l N_l$ denotes the width of the network. We denote the class of these DNNs as $\mathcal{NN}_{L,M,d,N_L}$.

To develop a DNN approximation for the solution u^ε of the two-scale, homogenization problem (2.2), we will use the following results.

The first is a (by now widely used) result on the approximate realization of the product of two real numbers by a suitable ReLU NN, due to [30]

Lemma 3.1. *There is a constant c such that for $m > 0$ and $\delta > 0$, there is a NN $\widetilde{\times}(a,b) \in \mathcal{NN}_{L,M,2,1}$ with $L \le c\log(m^2/\delta)$, and $M \le c\log(m^2/\delta)$ and width $W \le 12$ such that*

$$\|\widetilde{\times}(a,b) - ab\|_{L^\infty((-m,m)^2)} \le \delta.$$

The next result, which is a key technical step in the DNN emulation of the fine scale corrector in elliptic homogenization, addresses the ReLU NN expression of oscillatory functions, and is, in the form used here, proved in [10, Thm. IV.1].

Lemma 3.2. *There is a positive constant c such that, for every $a, m > 0$, and for every $\delta \in (0,1]$ there exist a ReLU NN $\widetilde{\cos} \in \mathcal{NN}_{L,M,1,1}$, a ReLU NN $\widetilde{\sin} \in \mathcal{NN}_{L,M,1,1}$ with depth $L \le c((\log\delta)^2 + \log(am))$ and of size $M \le c((\log\delta)^2 + \log(am))$ with fixed width $W \le 16$ such that there holds,*

$$\|\cos(at) - \widetilde{\cos}(at)\|_{L^\infty(-m,m)} \le \delta,$$

and

$$\|\sin(at) - \widetilde{\sin}(at)\|_{L^\infty(-m,m)} \le \delta.$$

4. DNN expression rates for two-scale elliptic equations

We recapitulate results from the expression of first order, Lagrangean Finite Element spaces on regular, simplicial partitions of convex polyhedra $\mathrm{D} \subset \mathbb{R}^d$ with Lipschitz boundary $\partial\mathrm{D}$ consisting of a finite number of $d-1$ planes for $d \ge 2$. We consider in particular a nested family of shape regular partitions $\mathcal{T}^l$, $l = 0, 1, 2, \ldots$ of D into simplices $T \in \mathcal{T}^l$ of meshsize $h_l := \max\{\mathrm{diam}(T) : T \in \mathcal{T}^l\} = O(2^{-l})$. The simplices in $\mathcal{T}^l$ are obtained by dividing each simplex $T \in \mathcal{T}^{l-1}$ into 2^d congruent subsimplices (so-called uniform, or also called "red", mesh refinement).

Let $\mathcal{N}^l$ denote the set of all nodes of the triangulation $\mathcal{T}^l$. For each node $x_i \in \mathcal{N}^l$, we denote by $G(i)$ the union of all simplices that contain x_i. We assume that the partition is locally convex for all l, i.e. $G(i)$ is a convex set for all $x_i \in \mathcal{N}^l$. Let V^l be the space of all continuous functions which are linear in each simplex in $\mathcal{T}^l$. We then have the following result, from [11].

Lemma 4.1. *For the locally convex shape regular triangulation family $\mathcal{T}^l$ of $\mathrm{D} \subset \mathbb{R}^d$, each function in V^l can be represented exactly by a ReLU DNN in $\mathcal{NN}_{L,M,d,1}$ where $L = O(d)$ and $M = O(2^{dL})$.*

For approximating functions with Sobolev regularity, we use the following result (Ern and Guermond [8, Theorem 6.4 and Lemma 6.3]).

Lemma 4.2. *There is a linear map $\mathcal{J}_l : L^1(\mathrm{D}) \to V^l$ such that*

$$\|\mathcal{J}_l\phi\|_{L^\infty(\mathrm{D})} \le c\|\phi\|_{L^1(\mathrm{D})} \quad \forall\phi \in L^1(\mathrm{D})$$

and

$$\|\phi - \mathcal{J}_l\phi\|_{L^2(\mathrm{D})} \le c \inf_{\phi_l \in V_l} \|\phi - \phi_l\|_{L^2(\mathrm{D})} \quad \forall\phi \in L^2(\mathrm{D})$$

where the contant c is independent of l and ϕ.

From this, we will construct a ReLU DNN for approximating the solution of the two scale elliptic problem (2.2). We have the following result.

Theorem 4.3. *Assume that* $u_0 \in H^2(\mathrm{D})$ *and* $u_1 \in H^1(\mathrm{D}, H^p_\#(\mathrm{Y}))$ *with* $p \geq 2+d/2$. *Then, there exists a constant* $c > 0$ *such that, for every* $0 < \delta \leq 1/2$ *and for every* $0 < \epsilon \leq 1/2$ *there is a ReLU DNN* $\widetilde{Du_1^\varepsilon} \in \mathcal{NN}_{L,M,d,d}$ *with depth and width bounded by*

$$L \leq c(d + (\log\delta)^2 + |\log\varepsilon|)\ , \quad M \leq c(\delta^{-d} + \delta^{-2/(2p-2-d)}(d + |\log\varepsilon|)),$$

respectively, and a ReLU DNN $\widetilde{u}_0 \in \mathcal{NN}_{L,M,d,1}$ *with* $L \leq cd$ *and* $M \leq c\delta^{-d}$ *such that*

$$\left\|\nabla u^\varepsilon - \nabla\widetilde{u}_0 - \widetilde{Du_1^\varepsilon}\right\|_{L^2(\mathrm{D})^d} \leq c(\delta + \varepsilon^{1/2}).$$

Proof We consider the Fourier expansion of $u_1(x,y)$

$$u_1(x,y) = \sum_{k\in\mathbb{Z}^d} [a_k(x)\cos(2\pi k\cdot y) + b_k(x)\sin(2\pi k\cdot y)]. \tag{4.1}$$

Without loss of generality, we assume that $\int_\mathrm{Y} u_1(x,y)dy = 0$ for all $x \in \mathrm{D}$. As $u_1 \in H^1(\mathrm{D}, H^p_\#(\mathrm{Y}))$,

$$\sum_{k\in\mathbb{Z}^d} |k|^{2p}(\|a_k\|^2_{H^1(\mathrm{D})} + \|b_k\|^2_{H^1(\mathrm{D})}) < \infty\ .$$

We consider a finitely truncated approximation of u_1, and define for $K \in \mathbb{N}$

$$u_1^K := \sum_{k\in\mathbb{Z}^d, |k|\leq K} [a_k(x)\cos(2\pi k\cdot y) + b_k(x)\sin(2\pi k\cdot y)].$$

Then, there exists a constant $c > 0$ such that for every $K \in \mathbb{N}$ holds

$$\begin{aligned}\|\nabla_y u_1(\cdot,\frac{\cdot}{\varepsilon}) - \nabla_y u_1^K(\cdot,\frac{\cdot}{\varepsilon})\|_{L^2(\mathrm{D})} &\leq c \sum_{k\in\mathbb{Z}^d,|k|>K} |k|(\|a_k\|_{L^2(\mathrm{D})} + \|b_k\|_{L^2(\mathrm{D})})\\ &\leq c \sum_{k\in\mathbb{Z}^d,|k|>K} |k|^{1-p}(|k|^p\|a_k\|_{L^2(\mathrm{D})} + |k|^p\|b_k\|_{L^2(\mathrm{D})})\end{aligned}$$

$$\leq c\left(\sum_{k\in\mathbb{Z}^d,|k|>K} |k|^{2(1-p)}\right)^{1/2} \left(\sum_{k\in\mathbb{Z}^d,|k|>K} |k|^{2p}\|a_k\|^2_{L^2(\mathrm{D})} + |k|^{2p}\|b_k\|^2_{L^2(\mathrm{D})}\right)^{1/2} \leq cK^{1+d/2-p}.$$

We choose K so that $K^{1+d/2-p} \approx \delta$, i.e., $K = K_\delta := \lceil \delta^{-2/(2p-2-d)} \rceil$.

To approximate a_k and b_k, we consider the quasiinterpolant $\mathcal{J}_l$ introduced in Lemma 4.2. For $k \in \mathbb{Z}^d$, we consider a resolution level $l_k \in \mathbb{N}$ which will be specified exactly below. We then have

$$\|a_k - \mathcal{J}_{l_k} a_k\|_{L^2(\mathrm{D})} \leq c2^{-l_k}\|a_k\|_{H^1(\mathrm{D})},$$

and

$$\|b_k - \mathcal{J}_{l_k} b_k\|_{L^2(\mathrm{D})} \leq 2^{-l_k}\|b_k\|_{H^1(\mathrm{D})}.$$

By Lemma 4.1, $\mathcal{J}_{l_k}a_k$ and $\mathcal{J}_{l_k}b_k$ can be exactly represented by DNNs $\widetilde{a}_k \in \mathcal{NN}_{L,M,1,1}$ and $\widetilde{b}_k \in \mathcal{NN}_{L,M,1,1}$ where $L = O(d)$ and $M = O(2^{dl_k})$. We then have

$$\begin{aligned}
&\| \sum_{k\in\mathbb{Z}^d,|k|\le K} a_k k\cos(2\pi k\cdot\frac{x}{\varepsilon}) - \sum_{k\in\mathbb{Z}^d,|k|\le K}\widetilde{a}_k k\cos(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^2(\mathrm{D})}\\
&\le \sum_{k\in\mathbb{Z}^d,|k|\le K}|k|\|a_k-\widetilde{a}_k\|_{L^2(\mathrm{D})} \le c\sum_{k\in\mathbb{Z}^d,|k|\le K}|k|2^{-l_k}\|a_k\|_{H^1(\mathrm{D})}\\
&\le c\sum_{k\in\mathbb{Z}^d,|k|\le K}2^{-l_k}|k|^{1-p}|k|^p\|a_k\|_{H^1(\mathrm{D})}\\
&\le c\left(\sum_{k\in\mathbb{Z},|k|\le K}2^{-2l_k}|k|^{2(1-p)}\right)^{1/2}\left(\sum_{|k|\le\mathbb{Z}^d,|k|\le K}|k|^{2p}\|a_k\|^2_{H^1(\mathrm{D})}\right)^{1/2}\\
&\le c\left(\sum_{k\in\mathbb{Z}^d,|k|\le K}2^{-2l_k}|k|^{2(1-p)}\right)^{1/2}.
\end{aligned}$$

We choose l_k so that $2^{-l_k}|k|^{1-p} = \delta|k|^{-s}$ where $s > d/2$, i.e. $l_k := \lceil -\log_2(\delta|k|^{p-1-s})\rceil$. We then have

$$\| \sum_{k\in\mathbb{Z}^d,|k|\le K} a_k k\cos(2\pi k\cdot\frac{x}{\varepsilon}) - \sum_{k\in\mathbb{Z}^d,|k|\le K}\widetilde{a}_k k\cos(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^2(\mathrm{D})} \le c\delta.$$

Similarly,

$$\| \sum_{k\in\mathbb{Z}^d,|k|\le K} b_k k\sin(2\pi k\cdot\frac{x}{\varepsilon}) - \sum_{k\in\mathbb{Z}^d,|k|\le K}\widetilde{b}_k k\sin(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^2(\mathrm{D})} \le c\delta.$$

Let $\widetilde{\cos}$ be a ReLU NN such that

$$\|\cos(2\pi k\cdot\frac{x}{\varepsilon}) - \widetilde{\cos}(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^\infty(\mathrm{D})} \le \delta_k \tag{4.2}$$

where δ_k is to be chosen later. We have

$$\begin{aligned}
&\| \sum_{k\in\mathbb{Z}^d,|k|\le K}\widetilde{a}_k k\cos(2\pi k\cdot\frac{x}{\varepsilon}) - \widetilde{a}_k k\widetilde{\cos}(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^2(\mathrm{D})}\\
&\le \sum_{k\in\mathbb{Z}^d,|k|\le K}\|\widetilde{a}_k\|_{L^2(\mathrm{D})}|k|\delta_k = \sum_{k\in\mathbb{Z}^d,|k|\le K}\|\widetilde{a}_k\|_{L^2(\mathrm{D})}|k|^p\delta_k|k|^{1-p}\\
&\le \left(\sum_{k\in\mathbb{Z}^d,|k|\le K}\|\widetilde{a}_k\|^2_{L^2(\mathrm{D})}|k|^{2p}\right)^{1/2}\left(\sum_{k\in\mathbb{Z}^d,|k|\le K}\delta_k^2|k|^{-2(p-1)}\right)^{1/2}.
\end{aligned}$$

We choose $\delta_k = \delta|k|^{p-1-s}$ where $s > d/2$. Then

$$\| \sum_{k\in\mathbb{Z}^d,|k|\le K}\widetilde{a}_k k\cos(2\pi k\cdot\frac{x}{\varepsilon}) - \widetilde{a}_k k\widetilde{\cos}(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^2(\mathrm{D})} \le c\delta.$$

Similarly, let $\widetilde{\sin}$ be a DNN such that

$$\|\sin(2\pi k\cdot\frac{x}{\varepsilon}) - \widetilde{\sin}(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^\infty(\mathrm{D})} \le \delta_k. \tag{4.3}$$

Then

$$\| \sum_{k\in\mathbb{Z}^d,|k|\le K}\widetilde{b}_k k\sin(2\pi k\cdot\frac{x}{\varepsilon}) - \widetilde{b}_k k\widetilde{\sin}(2\pi k\cdot\frac{x}{\varepsilon})\|_{L^2(\mathrm{D})} \le c\delta.$$

We note that (4.2) and (4.3) are achieved if the NNs $\widetilde{\cos}$ and $\widetilde{\sin}$ belong to $\mathcal{NN}_{L,M,d,1}$ and if there exists a constant $c>0$ such that for all $k\in\mathbb{N}$, $\delta,\varepsilon>0$ it holds

$$L\le c((\log\delta_k)^2+\log(c|k|\frac{1}{\varepsilon}))=c(|\log\delta|^2+\log|k|+|\log\varepsilon|), \tag{4.4}$$

and

$$M\le c((\log\delta_k)^2+\log(c|k|\frac{1}{\varepsilon})+d)=c(|\log\delta|^2+\log|k|+|\log\varepsilon|+d). \tag{4.5}$$

With these choices of l_k and δ_k, we deduce that there exists a constant $c>0$ such that for $0<\delta\le 1/2$ and for $K\in\mathbb{N}$

$$\|\sum_{k\in\mathbb{Z}^d,|k|\le K}ka_k\cos(2\pi k\cdot\frac{\cdot}{\varepsilon})-k\widetilde{a}_k\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon})\|_{L^2(\mathrm{D})}\le c\delta.$$

Now we consider the multiplication $\widetilde{a}_k(x)\widetilde{\cos}(2\pi k\cdot\frac{x}{\varepsilon})$ and $\widetilde{b}_k(x)\widetilde{\sin}(2\pi k\cdot\frac{x}{\varepsilon})$. To realize it via DNNs, for each k, we consider the approximate ReLU multiplication network $\widetilde{\times}_k$ from [30] such that

$$\|\widetilde{a}_k(\cdot)\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon})-\widetilde{\times}_k(\widetilde{a}_k(\cdot),\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon}))\|_{L^\infty(\mathrm{D})}\le\eta_k,$$

and

$$\|\widetilde{b}_k(\cdot)\widetilde{\sin}(2\pi k\cdot\frac{\cdot}{\varepsilon})-\widetilde{\times}_k(\widetilde{b}_k(\cdot),\widetilde{\sin}(2\pi k\cdot\frac{\cdot}{\varepsilon}))\|_{L^\infty(\mathrm{D})}\le\eta_k, \tag{4.6}$$

where we choose $\eta_k=\delta|k|^{-r}$ for $r>d+1$. With this choice of η_k, we have that

$$\|\sum_{k\in\mathbb{Z}^d,|k|\le K}k\widetilde{a}_k(\cdot)\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon})-k\widetilde{\times}_k(\widetilde{a}_k(\cdot),\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon}))\|_{L^\infty(\mathrm{D})}\le\delta.$$

As a_k is uniformly bounded in $L^1(\mathrm{D})$ so $\widetilde{a}_k$ is uniformly bounded in $C(\bar{\mathrm{D}})$ with respect to k. Further, from (4.2), we find that $\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon})$ is uniformly bounded in $L^\infty(\mathrm{D})$ with respect to k. Therefore, (4.6) holds if $\widetilde{\times}_k\in\mathcal{NN}_{L,M,2,1}$ where $L\le c(1+|\log\delta|+\log|k|)$ and $M\le c(1+|\log\delta|+\log|k|)$. Let $d\widetilde{u}^\varepsilon$ be the DNN that performs the summation $\sum_{k\in\mathbb{Z}^d,|k|\le K}k\widetilde{\times}_k(\widetilde{a}_k(\cdot),\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon}))+k\widetilde{\times}_k(\widetilde{b}_k(\cdot),\widetilde{\sin}(2\pi k\cdot\frac{\cdot}{\varepsilon}))$. This network can be represented by

$$\widetilde{\sum}_{k\in\mathbb{Z}^d,|k|\le K}\left(\mathrm{Id}_{\mathbb{R}}\widetilde{\times}_k\bullet\left(\mathrm{Id}_{\mathbb{R}}\widetilde{a}_k(\cdot),\mathrm{Id}_{\mathbb{R}}\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon})\right),\mathrm{Id}_{\mathbb{R}}\widetilde{\times}_k\bullet\left(\mathrm{Id}_{\mathbb{R}}\widetilde{b}_k(\cdot),\mathrm{Id}_{\mathbb{R}}\widetilde{\sin}(2\pi k\cdot\frac{\cdot}{\varepsilon})\right)\right)$$

where $\widetilde{\sum}$ denotes the DNN for summation, and the round parenthesis denotes parellelization. The number of layers in the identity networks $\mathrm{Id}_{\mathbb{R}}$ is chosen so that the networks are of the same depths.

The depth of this network is

$$\begin{aligned}O(\max_k\{&\mathrm{depth}(\widetilde{\times}_k),\mathrm{depth}(\widetilde{a}_k),\mathrm{depth}(\widetilde{b}_k),\mathrm{depth}(\widetilde{\cos}(2\pi k\cdot\frac{\cdot}{\varepsilon}),\mathrm{depth}(\widetilde{\sin}(2\pi k\cdot\frac{\cdot}{\varepsilon})\})\\&=O(d+(\log\delta)^2+|\log\varepsilon|+\log K)=O(d+(\log\delta)^2+|\log\varepsilon|).\end{aligned}$$

The number of nonzero weights of this nework is

$$O\Big(\sum_{|k|\le K}\delta^{-d}|k|^{d(1+s-p)}+(\log\delta)^2+\log|k|+|\log\varepsilon|+d\Big).$$

When p satisfies $2+s-p<0$, with $K=O(\delta^{-2/(2p-2-d)})$, the size of this network is

$$\begin{aligned} &O(\delta^{-d}+\delta^{-2/(2p-2-d)}((\log\delta)^2+|\log\varepsilon|+d)+\delta^{-2d/(2p-2-d)}|\log\delta|) \\ &\quad = O(\delta^{-d}+\delta^{-2d/(2p-2-d)}|\log\delta|+\delta^{-2/(2p-2-d)}(|\log\varepsilon|+d)). \end{aligned}$$

When $p>2+d/2$, we can choose $s>d/2$ such that $2+s-p<0$. In this case, $2/(2p-2-d)<1$. Thus the size of the network is

$$O(\delta^{-d}+\delta^{-2/(2p-2-d)}(|\log\varepsilon|+d)).$$

□

Remark 4.4. *The regularity conditions required in Theorem 4.3 are satisfied when the coefficient a and the right hand side f in* (2.2) *are sufficiently smooth. In particular, if the two scale coefficient $a\in C^1(\bar{\mathrm{D}},C^0(\bar{\mathrm{Y}}))$, then $a^0\in C^1(\bar{\mathrm{D}})$. If furthermore* D *is a convex domain and $f\in L^2(\mathrm{D})$, then $u_0\in H^2(\mathrm{D})$ (*[9]*). Further, if $a\in C^1(\bar{\mathrm{D}},H^p_\#(\bar{\mathrm{Y}})\cap C^{p-1}_\#(\bar{\mathrm{Y}}))$, then the solution of the cell problem* (2.5) *$N^i\in C^1(\bar{\mathrm{D}},H^p(\bar{\mathrm{Y}}))$ (*[28]*, Theorem 20.1) which implies $u_1\in H^1(\mathrm{D},H^p_\#(\mathrm{Y}))$.*

5. Conclusion, Generalization and Further Results

We analyzed the DNN expression rate of the solution family $\{u_\varepsilon:0<\varepsilon\le 1/2\}\subset H^1_0(\mathrm{D})$ for the linear, elliptic 2-scale homogenization problem. Under the given (weak) regularity assumptions on the problem data (i.e. coefficient a , source term f) we showed that the parametric family of solutions admits corresponding DNN approximations with error bounded (in the $H^1(\mathrm{D})$-norm in the physical domain D) by $\delta>0$ with ReLU DNNs of depth $O(d+(\log\delta)^2+|\log\varepsilon|)$ and connectivity $O(\delta^{-d}+\delta^{-2/(2p-2-d)}(d+|\log\varepsilon|))$ under the regularity condition $u_0\in H^2(\mathrm{D})$ and $u_1\in H^1(\mathrm{D},H^p_\#(\mathrm{Y}))$.

Let us now indicate further results and generalizations of the presently obtained results.

5.1. Stronger Expression Rate Bounds.

In the present work, we obtained DNN expression rate bounds for linear, elliptic second order PDEs. We proved that ReLU DNNs are able to express solutions within a prescribed accuracy at complexity bounds which increase logarithmically with respect to the scale parameter ε. The mathematical approach consisted in reapproximation of the first order corrector term in the 2-scale asymptotic expansion of the solution u^ε. The expression rate bounds were obtained using finite Sobolev regularity of the 2-scale corrector function. Similar results could probably be obtained by expressing the 2-scale corrector resulting from the so-called 2-scale convergence approach, where approximation rates by sparse grid approximations have been considered in [14, 13]. Solutions to two scale linear elasticity problems can also be obtained in the same way [29].

5.2. Multiple Scales and Nonlinear Problems.

The results of the present paper were derived for *linear, elliptic divergence form PDEs with two separated scales.* The 2-scale convergence result and the asymptotic structure (2.4) of the solution with the cell-problems (2.5) are also available for problems with multiple, separated scales. We refer for examples to [14, 15], [13] , [29] and the references there. The analysis in these references differs from the presently proposed one in that here, the (Y-periodic) "scale interaction function" $u_1(x,y)$ in (2.6), (2.7) is expanded into Fourier series, whereas in [14, 15], [13] it is approximated by so-called sparse grid approximations. There is a corresponding DNN counterpart to such approximations, see e.g. [18]. However, the approximation results in [18] can not be directly used in the present context, as they assume "isotropic" sparse approximations and sparse grids.

5.3. **Nonperiodic Coefficients.** The present results and their generalizations and extensions in the mentioned references are valid subject to the assumptions of scale-separation and of periodic fine-scale dependence of the coefficient $a(x, y)$ in (2.2). The presently obtained DNN expression rate bounds were based on the asymptotic structure (2.4) of the parametric solution family $\{u^\varepsilon : 0 < \varepsilon \le 1/2\} \subset H^1_0(\mathrm{D})$ and on the regularity of the fine-scale function $u_1(x, y)$ in (2.4). Let us here indicate that corresponding results can also be expected in considerably more general settings.

To this end, consider $\mathrm{D} = (0, 1)$, $f \in L^2(\mathrm{D})$ and the *nonperiodic diffusion coefficient* $a(x)$ which satisfies

$$0 < a_- \le a \le a_+ < \infty \quad a.e.\ x \in \mathrm{D}\ . \tag{5.1}$$

We consider the two-point boundary value problem

$$-\,(a(x)u')' = f \quad \text{in} \quad \mathrm{D}\ , \qquad u(0) = 0,\ \ a(1)u'(1) = g. \tag{5.2}$$

Under these assumptions, (5.2) admits, for every $f \in L^2(\mathrm{D})$, a unique solution $u \in V := \{v \in H^1(\mathrm{D}) : v(0) = 0\}$. This solution $u \in V$ admits the representation

$$u'(x) = \frac{1}{a}(x)(F(x) + c)\ , \quad x \in \mathrm{D} \tag{5.3}$$

where $F \in H^1(\mathrm{D})$ denotes the antiderivative of $f \in L^2(\mathrm{D})$ and $c \in \mathbb{R}$ denotes a constant of integration which depends on the Neumann datum g in (5.2).

DNNs $\tilde{u}$ which approximate the exact solution u of (5.2) can be based on the formula (5.3). To this end, for any $0 < \delta \le 1/2$ we assume at hand DNNs $\tilde{F}$ and $\tilde{a}$ such that

$$\|a - \tilde{a}\|_{L^\infty(\mathrm{D})} \le \delta\ , \qquad \|F - \tilde{F}\|_{L^2(\mathrm{D})} \le \delta\ , \tag{5.4}$$

(concrete choices will be discussed below) and which are such that there exist constants $0 < \tilde{a}_- \le \tilde{a}_+ < \infty$ with

$$\tilde{a}_- \le \tilde{a}(x) \le \tilde{a}_+ \quad \text{for}\ \ x \in \overline{\mathrm{D}}\ . \tag{5.5}$$

Define a ReQU DNN via the ReLU approximation

$$\widetilde{u'} := \tilde{\times}(\mathrm{inv} \circ \tilde{a}, \tilde{F} + c) \tag{5.6}$$

of u' and subsequent integration: for $x \in \mathrm{D}$ and with $u(0) = 0$ in (5.2) we obtain

$$\tilde{u}(x) := \int_0^x \widetilde{u'}(\xi) d\xi\ , \quad x \in \mathrm{D}\ . \tag{5.7}$$

Evidently, $\tilde{u}$ is a piecewise quadratic spline in $C^1(\overline{\mathrm{D}})$ with $\tilde{u}(0) = 0$.

In (5.6), the operation $\tilde{\times}$ denotes the approximate ReLU multiplication network of [30] and the operation inv denotes the ReLU NN approximation of the inversion map $\mathrm{inv} : x \mapsto 1/x$. Concrete ReLU DNN constructions $\{\tilde{f}_n\}_{n\ge1}$ which approximate this reciprocal map inv were given in [12, Appendix C]. There holds

Lemma 5.1. [12, Lemma C.1] *Let $0 < a < b < \infty$. There exists $\kappa > 0$ and constants C_0, $C_1 > 0$ that are independent of a, b such that for every $n \in \mathbb{N}$, there exists a ReLU NN $\tilde{f}_n$ such that*

$$\sup_{x\in[a,b]} \left|\frac{1}{x} - \tilde{f}_n(x)\right| \le C_0 \frac{\lceil \log(b/a)\rceil}{a}\left(1+\frac{1}{b-a}\right)\exp\left(-\frac{\kappa}{\sqrt{\lceil \log(b/a)\rceil}}n\right).$$

Furthermore, the size and depth of the ReLU DNN $\tilde{f}_n$ is bounded as

$$\begin{aligned} &\mathrm{depth}(\tilde{f}_n) \le C_1(1+n\log(n)+\log^3(n)) \\ \text{and} \quad & \\ &\mathrm{size}(\tilde{f}_n) \le C_1[1+n^2(\log(n)+\log(\sqrt{\lceil\log(b/a)\rceil}))] \,. \end{aligned} \tag{5.8}$$

Based on this result, which we use to emulate the map $a(x) \mapsto 1/a(x)$ with $a := \tilde{a}_-$ and $b := \tilde{a}_+$, we are in position to bound the DNN expression error $e := u - \tilde{u}$ in $H^1(\mathrm{D})$. Due to $e(0) = 0$, it suffices to bound the seminorm. To this end, we write with the ReLU NN $\tilde{f}_n$ from Lemma 5.1, for $n \in \mathbb{N}$ to be selected

$$e' = \frac{F-\tilde{F}}{a} + \left(\frac{1}{a}-\frac{1}{\tilde{a}}\right)\tilde{F} + \left(\frac{1}{\tilde{a}} - \tilde{f}_n\circ\tilde{a}\right)\tilde{F} + \left(\times(\tilde{f}_n\circ\tilde{a},\tilde{F}) - \tilde{\times}(\tilde{f}_n\circ\tilde{a},\tilde{F})\right) \,. \tag{5.9}$$

Here, the binary operation $\times(\cdot,\cdot)$ denotes the exact multiplication of (ReLU) NNs (which is not a ReLU NN), wherefore we approximate $\times$ by the approximate (still binary) operation $\tilde{\times}$ from [24, Prop. 3.1]. We estimate

$$\begin{aligned} \|e'\|_{L^2(\mathrm{D})} \quad &\le \quad \frac{1}{a_-}\|F-\tilde{F}\|_{L^2(\mathrm{D})} + \left\|\frac{1}{a}-\frac{1}{\tilde{a}}\right\|_{L^\infty(\mathrm{D})}\|\tilde{F}\|_{L^2(\mathrm{D})} \\ &\quad +\|(\xi^{-1}-\tilde{f}_n)\circ\tilde{a}\|_{L^\infty(\mathrm{D})} + \|(\tilde{f}_n\circ\tilde{a})\tilde{F} - \tilde{\times}(\tilde{f}_n\circ\tilde{a},\tilde{F})\|_{L^2(\mathrm{D})} \\ &=: \quad I + II + III + IV \,. \end{aligned} \tag{5.10}$$

To prove DNN expression rate bounds, we estimate terms I through IV. From (5.4) and (5.1), we have $I \le \delta/a_-$. To bound II, we write $1/a - 1/\tilde{a} = (a-\tilde{a})/a\tilde{a}$. Therefore, with (5.4),

$$\left\|\frac{1}{a}-\frac{1}{\tilde{a}}\right\|_{L^\infty(\mathrm{D})} \le \frac{1}{a_-\tilde{a}_-}\|a-\tilde{a}\|_{L^\infty(\mathrm{D})} \le \frac{1}{a_-\tilde{a}_-}\delta \,.$$

To bound III, we use Lemma 5.1 with $a := \tilde{a}_-$ and $b := \tilde{a}_+$. We obtain that there exist constants $C(\tilde{a}_-,\tilde{a}_+) > 0$ and $b(\tilde{a}_-,\tilde{a}_+) > 0$ such that for every $n \in \mathbb{N}$ exists a ReLU DNN $\tilde{f}_n$ with

$$III \le C(\tilde{a}_-,\tilde{a}_+)\exp(-b(\tilde{a}_-,\tilde{a}_+)n) \,. \tag{5.11}$$

Thus, $III \le \delta \le 1/2$ is ensured by choosing $n(\delta) \ge c|\log\delta|$ for some $c > 0$ sufficiently large (but independent of δ). The size and the depth of the DNNs $\tilde{f}_n$ are bounded as in (5.8), with $n(\delta)$ in place of n.

For term IV, we use [24, Prop. 3.1] or Lemma 3.1. We obtain for $0 < \delta \le 1/2$

$$\|(\tilde{f}_n\circ\tilde{a})\tilde{F} - \tilde{\times}_\delta(\tilde{f}_n\circ\tilde{a},\tilde{F})\|_{L^2(\mathrm{D})} \le \delta$$

with $\mathrm{size}(\tilde{\times}_\delta) = O(\log(1/\delta))$ and $\mathrm{depth}(\tilde{\times}_\delta) = O(\log(1/\delta))$. This verifies that the ReLU DNN $\tilde{u'}$ defined in (5.6) satisfies the consistency

$$\|u' - \tilde{u'}\|_{L^2(\mathrm{D})} \le \delta \,. \tag{5.12}$$

Bounds on size and depth of the NN $\tilde{u}$ can be obtained as follows:

$$L(\tilde{u}) \leq L(\tilde{\times}_\delta) + L(\tilde{a}) + L(\tilde{f}_{n(\delta)}) + L(\tilde{F}), \tag{5.13}$$

and

$$M(\tilde{u}) \leq M(\tilde{\times}_\delta) + M(\tilde{a}) + M(\tilde{f}_{n(\delta)}) + M(\tilde{F}). \tag{5.14}$$

We next discuss several particular hypothesis-classes on a and on f in (5.2), (5.1) which allow obtaining quantitative DNN expression rate bounds for the corresponding solution sets.

1. Finite (Low) Sobolev Regularity. Under the sole assumption $f \in L^2(\mathrm{D})$ (as we assumed here), $F \in H^1(\mathrm{D})$. Denoting by $I_h F$ the continuous, piecewise polynomial nodal interpolation of degree 1 on a uniform partition of D of width $0 < h < 1$, we have

$$\|F - I_h F\|_{L^2(\mathrm{D})} \leq Ch\|F'\|_{L^2(\mathrm{D})} = Ch\|f\|_{L^2(\mathrm{D})} \ .$$

In the same way, for $a \in W^{1,\infty}(\mathrm{D})$,

$$\|a - I_h a\|_{L^\infty(\mathrm{D})} \leq Ch\|a'\|_{L^\infty(\mathrm{D})} \ .$$

On the other hand, $I_h F$ corresponds exactly to a (shallow) ReLU NN $\tilde{F}_h$ of depth 1 and width resp. size $M(\tilde{F}) = O(h^{-1})$. This verifies the second item in (5.4) with $\tilde{F} = \tilde{F}_h$, and ascertains accuracy $\delta > 0$ at fixed DNN depth 1 and at DNN size $M(\tilde{F}) = O(\delta^{-1})$. A generic Lipschitz coefficient a will likewise admit a shallow ReLU approximation $\tilde{a}_h$ with fixed DNN depth 1 and DNN size $M(\tilde{a}) = O(\delta^{-1})$.

2. Poly-log DNN expression of a and f.

It has been recently observed in [10] and [7] that ReLU DNNs allow approximating certain fractal functions $a(x)$ of low Besov regularity at exponential rates, in terms of NN size and depth; in particular, the so-called Weierstrass functions and the Takagi class of functions. Accuracy $\delta > 0$ is achieved in $L^\infty(\mathrm{D})$ with the NN size and depth being only polylogarithmic w.r. to δ. The self-similarity of such functions precludes scale-separation which we used in Theorem 4.3.

To study corresponding DNN expression rates for the solution u of (5.2)we assume here, generically, that a and F belong to the class $\mathrm{polylog}(k, X)$ defined for $k \in \mathbb{N}$ and a suitable Banach space X with norm $\| \circ \|_X$ as

$$\begin{aligned} \mathrm{polylog}(k, X) \quad := \quad & \big\{ v \in X : \forall 0 < \delta \leq 1 \ \exists \ \tilde{v}_\delta \in \mathcal{NN}^{ReLU}_{L,M,1,1} : \\ & \|v - \tilde{v}_\delta\|_X \leq \delta, \ L, M \leq c(1 + |\log \delta|^k) \big\} \ . \end{aligned} \tag{5.15}$$

For a we shall choose $X = L^\infty(\mathrm{D})$ and $X = L^2(\mathrm{D})$ for F.

Theorem 5.2. *Consider the model problem* (5.2)*with data*

$$a \in \mathrm{polylog}(k_a, L^\infty(\mathrm{D})), \ \ F \in \mathrm{polylog}(k_f, L^2(\mathrm{D})).$$

There exists $C > 0$ such that for all admissible a, F and for every prescribed accuracy $0 < \delta \leq 1/2$, there exists a ReLU DNN $\tilde{u}'_\delta \in \mathcal{NN}_{L,M,1,1}$ such that

$$\|u' - \tilde{u}'_\delta\|_{L^2(\mathrm{D})} \leq \delta;$$

and such that

$$L \leq C(|\log \delta|^{\max\{1,k_a,k_f\}} + |\log \delta| \log |\log \delta|),$$

$$M \leq C(|\log \delta|^{\max\{1,k_a,k_f\}} + (\log \delta)^2 \log |\log \delta|).$$

Examples:
(i) The class of Weierstrass functions

$$W_{p,a} = \sum_{k=0}^{\infty} p^k \cos(a^k \pi x)$$

belongs to polylog$(3, L^\infty(D))$ (see [10] page 52). We consider the positive coefficient of the form $A(x) + W_{p,a}(x)$ where ess-inf $A(x)$ is sufficiently large.
(ii) We consider the class of Tagaki functions

$$F = \sum_{k\geq 1} t^k g(\psi^{\circ k})$$

where $|t| < 1$; and $\psi^{\circ k}$ denotes the k-fold composition of the function ψ. Assuming that g and ψ can be represented exactly by ReLU DNN, then, the antiderivative $F \in$ polylog$(1, C([0, 1]))$ (see [7] page 28).

Further examples for fractal coefficients a with very low Sobolev regularity are developed in [10] and [7]. The preceding argument will provide DNN expression rate bounds for the corresponding solution families of the univariate boundary value problem (5.2).

3. Almost-periodic coefficient We now depart from the periodic setting considered above, and assume that the multiscale coefficient is almost periodic. In this case, a well-developed theory of homogenization is still available (see, e.g., [16, Chap. 7.4] and the references there). Assume given $\{a_n\}_{n\geq 0}, \{b_n\}_{n\geq 1} \subset W^{1,\infty}(\mathrm{D})$, $\{\xi_n\}_{n\geq 1}, \{\zeta_n\}_{n\geq 1} \subset \mathbb{R}$. We consider (5.2) with quasi-periodic coefficient of the form

$$a(x) = a_0(x) + \sum_{n=1}^{\infty} [a_n(x) \cos(\frac{x}{\varepsilon}\xi_n) + b_n(x) \sin(\frac{x}{\varepsilon}\zeta_n)].$$

To ensure convergence of the series in this representation, we assume the following bounds: there are real numbers $p > 1, q > 0, c_0 > 0$ such that for every $n \in \mathbb{N}$

$$\|a_n\|_{W^{1,\infty}(\mathrm{D})} \leq c_0/n^p, \;\; \|b_n\|_{W^{1,\infty}(\mathrm{D})} \leq c_0/n^p, \;\; |\xi_n| \leq c_0 n^q, \;\; |\zeta_n| \leq c_0 n^q. \tag{5.16}$$

We assume further that there is a positive constant a_- such that

$$\operatorname*{ess-inf}_{x\in \mathrm{D}} a_0 \geq \sum_{n=1}^{\infty} 2c_0 n^{-p} + a_-,$$

so that

$$\operatorname*{ess-inf}_{x\in \mathrm{D}} a(x) \geq a_-.$$

For $N \in \mathbb{N}$, we denote by

$$a_N(x) = a_0(x) + \sum_{n=1}^{N} [a_n(x) \cos(\frac{x}{\varepsilon}\xi_n) + b_n(x) \sin(\frac{x}{\varepsilon}\zeta_n)].$$

We then have

$$\|a - a_N\|_{L^\infty(\mathrm{D})} \leq 2 \sum_{n>N} c_0 n^{-p} \leq cN^{-p+1}.$$

Choosing $N = \lceil \delta^{-1/(p-1)} \rceil$, we then have $\|a - a_N\|_{L^\infty(\mathrm{D})} \leq c\delta$. For each n, we choose a resolution h_n for the piecewise linear nodal interpolation such that

$$\|a_n - I_{h_n} a_n\|_{L^\infty(\mathrm{D})} \leq ch_n \|a_n\|_{W^{1,\infty}(\mathrm{D})} \leq \delta n^{-s}, \;\; \|b_n - I_{h_n} b_n\|_{L^\infty(\mathrm{D})} \leq ch_n \|b_n\|_{W^{1,\infty}(\mathrm{D})} \leq \delta n^{-s} \tag{5.17}$$

for $s > 1$. Then the functions $I_{h_n} a_n$ and $I_{h_n} b_n$ can be represented exactly by the ReLU networks $\tilde a_n$ and $\tilde b_n$ with $O(1)$ depth and $O(\delta^{-1} n^{-p+s})$ connectivity. From Lemma 3.2, $\cos(\frac{\cdot}{\varepsilon}\xi_n)$ and $\sin(\frac{\cdot}{\varepsilon}\zeta_n)$ can each be approximated by ReLU NNs, which we denote by $\widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n)$ and $\widetilde{\sin}(\frac{\cdot}{\varepsilon}\zeta_n)$. These approximations have accuracy

$$\|\cos(\frac{\cdot}{\varepsilon}\xi_n) - \widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n)\|_{L^\infty(\mathrm{D})} \le \delta, \quad \|\sin(\frac{\cdot}{\varepsilon}\zeta_n) - \widetilde{\sin}(\frac{\cdot}{\varepsilon}\zeta_n)\|_{L^\infty(\mathrm{D})} \le \delta.$$

From Lemma 3.2, we have

$$L(\widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n)) \le c((\log\delta)^2 + \log n + |\log\varepsilon|), \quad M(\widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n)) \le c((\log\delta)^2 + \log n + |\log\varepsilon|),$$

and

$$L(\widetilde{\sin}(\frac{\cdot}{\varepsilon}\xi_n)) \le c((\log\delta)^2 + \log n + |\log\varepsilon|), \quad M(\widetilde{\sin}(\frac{\cdot}{\varepsilon}\xi_n)) \le c((\log\delta)^2 + \log n + |\log\varepsilon|).$$

Let

$$\bar a_N(x) = \sum_{n\le N} \tilde a_n(x)\widetilde{\cos}(\frac{x}{\varepsilon}\xi_n) + \tilde b_n(x)\widetilde{\sin}(\frac{x}{\varepsilon}\zeta_n).$$

We then have

$$\begin{aligned}\bar a_N(x) - a_N(x) = \sum_{n\le N} \Big[&(\tilde a_n(x) - a_n(x))\widetilde{\cos}(\frac{x}{\varepsilon}\xi_n) + (\tilde b_n(x) - b_n(x))\widetilde{\sin}(\frac{x}{\varepsilon}\zeta_n) + \\ &a_n(x)(\widetilde{\cos}(\frac{x}{\varepsilon}\xi_n) - \cos(\frac{x}{\varepsilon}\xi_n)) + b_n(x)(\widetilde{\sin}(\frac{x}{\varepsilon}\zeta_n) - \sin(\frac{x}{\varepsilon}\zeta_n))\Big].\end{aligned}$$

Then

$$\|\bar a_N - a_N\|_{L^\infty(\mathrm{D})} \le c\delta \sum_{n\le N}(n^{-s} + n^{-p}) \le c\delta.$$

Now we define the NN

$$\tilde a_N = \sum_{n\le N} \widetilde{\times}_1(\tilde a_n, \widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n)) + \widetilde{\times}_1(\tilde b_n, \widetilde{\sin}(\frac{\cdot}{\varepsilon}\zeta_n)),$$

where the multiplication ReLU NN $\widetilde{\times}_1$ achieves the accuracy $\delta^{-1-1/(p-1)}$ and has $O(|\log\delta|)$ layers and $O(|\log\delta|)$ connectivity. We then have $\|\tilde a_N - a\|_{L^\infty(\mathrm{D})} \le c\delta$ The networks $\tilde a_n, \tilde b_n, \widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n), \widetilde{\sin}(\frac{\cdot}{\varepsilon}\zeta_n)$ are performed in parallel. The depth of the network $\tilde u'$ is bounded by

$$\begin{aligned}&\max\{L(\tilde a_0), L(\tilde a_n), L(\tilde b_n), L(\widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n)), L(\widetilde{\sin}(\frac{\cdot}{\varepsilon}\zeta_n)\} + L(\widetilde{\times}_1) + L(\mathrm{inv}) + L(\widetilde{\times}) + L(\tilde F) \\ &\quad \le c(|\log(\delta)|^2 + \log N + |\log\varepsilon| + |\log\delta|^2\log|\log\delta|) \\ &\quad \le c(|\log\delta|^2\log|\log\delta| + |\log\varepsilon|).\end{aligned}$$

The connectivity of the network is bounded by

$$\begin{aligned}M(\tilde a_0) + \sum_{n=1}^N [M(\tilde a_n) + M(\tilde b_n) + M(\widetilde{\cos}(\frac{\cdot}{\varepsilon}\xi_n)) + M(\widetilde{\sin}(\frac{\cdot}{\varepsilon}\zeta_n))]& \\ + NM(\widetilde{\times}_1) + M(\widetilde{\times}) + M(\mathrm{inv}) + M(\tilde F)& \\ \le c(\delta^{-1} + \delta^{-1}\sum_{n=1}^N n^{-p+s} + N(\log\delta)^2 + N|\log\varepsilon|& \\ + \sum_{n=1}^N \log n + N|\log\delta|) + |\log\delta|^2\log|\log\delta|)&\ .\end{aligned}$$

Assume that $p > 2$, we can choose $s > 1$ in (5.17) such that $p - s - 1 > 0$.

We then have

$$M(\tilde{u}') \leq c\delta^{-1} + c\delta^{-1/(p-1)}|\log\varepsilon|.$$

We arrive at the following result.

Theorem 5.3. *Assume that functions a_n, b_n and the constants ξ_n and ζ_n satisfy conditions* (5.16). *Then for every $0 < \delta \leq 1/2$, there exists a NN $\tilde{u}$ with $O(|\log\delta|^2 \log|\log\delta| + |\log\varepsilon|)$ layers and $O(\delta^{-1} + \delta^{-1/(p-1)}|\log\varepsilon|)$ connectivity such that*

$$\|u' - \tilde{u}'\|_{L^2(\mathrm{D})} \leq c\delta.$$

In conclusion, for linear, second order elliptic boundary value problems in divergence form with multiscale coefficients, we analyzed the expressive power of ReLU neural networks for the corresponding solution families.

For the classical, periodic setting of homogenization, in $d \geq 1$ space dimension, we proved in Theorem 4.3 expressive power estimates for the 2-scale solutions which are explicit in the scale parameter ϵ and the expression accuracy $\delta > 0$. The main result is that the deep ReLU NNs achieve scale resolution with depth which increases logarithmic with respect to the scale parameter ϵ.

For several classes of nonperiodic coefficients of low Sobolev and Besov regularity, in particular for fractal coefficients from the Weierstrass and Takagi classes, we established likewise exponential expressivity.

References

[1] N. Bakhvalov and G. Panasenko. *Homogenisation: averaging processes in periodic media*, volume 36 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1989. Mathematical problems in the mechanics of composite materials, Translated from the Russian by D. Leĭtes.

[2] C. Beck, W. E, and A. Jentzen. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *J. Nonlinear Sci.*, 29(4):1563–1619, 2019.

[3] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-New York, 1978.

[4] J. Berg and K. Nyström. Data-driven discovery of PDEs in complex datasets. *ArXiv e-prints*, Aug. 2018.

[5] H. Bölcskei, P. Grohs, G. Kutyniok, and P. Petersen. Optimal approximation with sparsely connected deep neural networks. *SIAM Journal on Mathematics of Data Science*, 1(1):8–45, 2019. ArXiv:1705.01714.

[6] N. Cohen, O. Sharir, and A. Shashua. On the expressive power of deep learning: A tensor analysis. In *Proc. of 29th Ann. Conf. Learning Theory*, pages 698–728, 2016. ArXiv:1509.05009v3.

[7] I. Daubechies, R. A. DeVore, S. Foucart, B. Hanin, and G. Petrova. Nonlinear Approximation and (Deep) ReLU Networks. *CoRR*, abs/1905.02199, 2019.

[8] A. Ern and J.-L. Guermond. Finite element quasi-interpolation and best approximation. *arXiv e-prints*, page arXiv:1505.06931v4, May 2015.

[9] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[10] P. Grohs, D. Perekrestenko, D. Elbrächter, and H. Bölcskei. Deep neural network approximation theory. *CoRR*, abs/1901.02220, 2019.

[11] J. He, L. Li, J. Xu, and C. Zheng. ReLU Deep Neural Networks and Linear Finite Elements, Jul 2018.

[12] L. Herrmann, C. Schwab, and J. Zech. Deep ReLU Neural Network Expression Rates for Data-to-QoI Maps in Bayesian PDE Inversion. Technical Report 2020-02, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2020. (in review).

[13] V. H. Hoang. Sparse finite element method for periodic multiscale nonlinear monotone problems. *Multiscale Model. Simul.*, 7(3):1042–1072, 2008.

[14] V. H. Hoang and C. Schwab. High-dimensional finite elements for elliptic problems with multiple scales. *Multiscale Model. Simul.*, 3(1):168–194, 2004/05.

[15] V. H. Hoang and C. Schwab. Analytic regularity and polynomial approximation of stochastic, parametric elliptic multiscale PDEs. *Anal. Appl. (Singap.)*, 11(1):1350001, 50, 2013.

[16] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994. Translated from the Russian by G. A. Yosifian [G. A. Iosifyan].

[17] S. Liang and R. Srikant. Why deep neural networks for function approximation? In *Proc. of ICLR 2017*, pages 1 – 17, 2017. ArXiv:1610.04161.

[18] H. Montanelli and Q. Du. New error bounds for deep ReLU networks using sparse grids. *SIAM J. Math. Data Sci.*, 1(1):78–92, 2019.

[19] J. A. A. Opschoor, P. C. Petersen, and C. Schwab. Deep ReLU Networks and High-Order Finite Element Methods. Technical Report 2019-07, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2019. (to appear in Analysis and Applications 2020).

[20] G. Pang, L. Lu, and G. E. Karniadakis. fPINNs: fractional physics-informed neural networks. *SIAM J. Sci. Comput.*, 41(4):A2603–A2626, 2019.

[21] P. Petersen and F. Voigtlaender. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Networks*, 108:296–330, 2018.

[22] A. Pinkus. Approximation theory of the MLP model in neural networks. In *Acta numerica, 1999*, volume 8 of *Acta Numer.*, pages 143–195. Cambridge Univ. Press, Cambridge, 1999.

[23] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.

[24] C. Schwab and J. Zech. Deep learning in high dimension: neural network expression rates for generalized polynomial chaos expansions in UQ. *Anal. Appl. (Singap.)*, 17(1):19–55, 2019.

[25] J. Sirignano and K. Spiliopoulos. DGM: a deep learning algorithm for solving partial differential equations. *J. Comput. Phys.*, 375:1339–1364, 2018.

[26] R. K. Tripathy and I. Bilionis. Deep UQ: learning deep neural network surrogate models for high dimensional uncertainty quantification. *J. Comput. Phys.*, 375:565–588, 2018.

[27] N. Winovich, K. Ramani, and G. Lin. ConvPDE-UQ: convolutional neural networks with quantified uncertainty for heterogeneous elliptic partial differential equations on varied domains. *J. Comput. Phys.*, 394:263–279, 2019.

[28] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987. Translated from the German by C. B. Thomas and M. J. Thomas.

[29] B. Xia and V. H. Hoang. High-dimensional finite element method for multiscale linear elasticity. *IMA J. Numer. Anal.*, 35(3):1277–1314, 2015.

[30] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Netw.*, 94:103–114, 2017.

[1] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, 637371

[2] Seminar for Applied Mathematics, ETH, 8092 Zurich, Switzerland