On Universal Approximation and Error Bounds
for Fourier Neural Operators

Nikola Kovachki   NKOVACHKI@CALTECH.EDU
Computing and Mathematical Sciences
California Institute of Technology
Pasadena, California 91125, USA

Samuel Lanthaler   SAMUEL.LANTHALER@MATH.ETHZ.CH
Seminar for Applied Mathematics
Eidgenössische Technische Hochschule Zürich (ETH Zürich)
Rämistrasse 101, 8092 Zürich, Switzerland

Siddhartha Mishra   SIDDHARTHA.MISHRA@MATH.ETHZ.CH
Seminar for Applied Mathematics
Eidgenössische Technische Hochschule Zürich (ETH Zürich)
Rämistrasse 101, 8092 Zürich, Switzerland

Abstract

Fourier neural operators (FNOs) have recently been proposed as an effective framework for learning operators that map between infinite-dimensional spaces. We prove that FNOs are universal, in the sense that they can approximate any continuous operator to desired accuracy. Moreover, we suggest a mechanism by which FNOs can approximate operators associated with PDEs efficiently. Explicit error bounds are derived to show that the size of the FNO, approximating operators associated with a Darcy type elliptic PDE and with the incompressible Navier-Stokes equations of fluid dynamics, only increases sub (log)-linearly in terms of the reciprocal of the error. Thus, FNOs are shown to efficiently approximate operators arising in a large class of PDEs.

Keywords: operator learning, universal approximation, error bounds, complexity bounds, Darcy flow, incompressible Navier-Stokes

1. Introduction

Deep neural networks have been extremely successful in diverse fields of science and engineering including image classification, speech recognition, natural language understanding, autonomous systems, game intelligence and protein folding, see LeCun et al. (2015) and references therein. Moreover, deep neural networks are being increasingly used successfully in scientific computing, particular in simulating physical and engineering systems modeled by partial differential equations (PDEs). Examples include the use of physics informed neural networks for solving forward and inverse problems for PDEs (Raissi and Karniadakis, 2018; Raissi et al., 2019; Mishra and Molinaro, 2020a,b) and supervised learning algorithms for high-dimensional parabolic PDEs (E et al., 2017) and parametric elliptic (Kutyniok et al., 2021; Schwab and Zech, 2019) and hyperbolic (Lye et al., 2020, 2021) PDEs, among others.

The success of deep neural networks at a wide variety of learning tasks can be attributed to a confluence of several factors such as the availability of massive labeled data sets, the design of novel architectures and training algorithms as well as the abundance of high-end computing platforms such as GPUs (Goodfellow et al., 2016). Still, it is fair to surmise that this edifice of success partly rests on the foundation of *universal approximation* (Barron, 1993; Cybenko, 1989; Hornik et al., 1989), i.e., the ability of neural networks to approximate any continuous (even measurable) function, mapping a finite-dimensional input space into another finite-dimensional output space, to arbitrary accuracy.

However, many interesting learning tasks entail learning *operators*, i.e., mappings between an infinite-dimensional input Banach space and (possibly) an infinite-dimensional output space. A prototypical example in scientific computing is provided by nonlinear operators that map the initial datum into the (time series of) solution of a nonlinear time-dependent PDE such as the Navier-Stokes equations of fluid dynamics. A priori, it is unclear if neural networks can be successfully employed for learning such operators from data, given that their universality only pertains to finite-dimensional functions.

The first successful use of neural networks in the context of such *operator learning* was provided in Chen and Chen (1995), where the authors proposed a novel neural network based learning architecture, which they termed as *operator networks* and proved that these operator networks possess a surprising universal approximation property for infinite-dimensional nonlinear operators. Operator networks are based on two different neural networks, a *branch net* and a *trunk net*, which are trained concurrently to learn from data. More recently, Lu et al. (2021) have proposed using deep, instead of shallow, neural networks in both the trunk and branch net and have christened the resulting architecture as a *DeepOnet*. In a recent article (Lanthaler et al., 2021), the universal approximation property of DeepOnets was extended, making it completely analogous to universal approximation results for finite-dimensional functions by neural networks. Lanthaler et al. (2021) were also able to show that DeepOnets can break the curse of dimensionality for a large variety of PDE learning tasks. Hence, in spite of the underlying infinite-dimensional setting, DeepOnets are capable of approximating a large variety of nonlinear operators efficiently. This is further validated by the success of DeepOnets in many interesting examples in scientific computing, see Mao et al. (2021); Cai et al. (2021); Lin et al. (2021) and references therein.

An alternative operator learning framework is provided by the concept of *neural operators*, first proposed in Li et al. (2020a). Just as canonical artificial neural networks are a concatenated composition of multiple hidden layers, with each hidden layer composing an affine function with a scalar nonlinear activation function, neural operators also compose multiple hidden layers, with each hidden layer composing an affine *operator* with a local, scalar nonlinear activation operator. The infinite-dimensional setup is reflected in the fact that the affine operator can be significantly more general than in the finite-dimensional case, where it is represented by a weight matrix and bias vector. On the other hand, for neural operators, one can even use *non-local* linear operators, such as those defined in terms of an integral kernel. The evaluation of such integral kernels can be performed either with graph kernel networks (Li et al., 2020a) or with multipole expansions (Li et al., 2020b).

More recently, Li et al. (2021) have proposed using convolution-based integral kernels within neural operators. Such kernels can be efficiently evaluated in the Fourier space, leading to the resulting neural operators being termed as *Fourier Neural Operators* (FNOs).

Li et al. (2021) discuss the advantages, in terms of computational efficiency, of FNOs over the other neural operators mentioned above. Moreover, they present several convincing numerical experiments to demonstrate that FNOs can very efficiently approximate a variety of operators that arise in simulating PDEs.

However, the theoretical basis for neural operators has not yet been properly investigated. In particular, it is unclear if neural operators such as FNOs are *universal*, i.e., if they can approximate a large class of nonlinear infinite-dimensional operators. Moreover in this infinite-dimensional setting, universality does not suffice to indicate computational viability or efficiency as the size of the underlying neural networks might grow exponentially with respect to increasing accuracy, see discussion in Lanthaler et al. (2021) on this issue. Hence in addition to universality, it is natural to ask if neural operators can *efficiently* approximate a large class of operators, such as those arising in the simulation of parametric PDEs.

The investigation of these questions is the main rationale for the current paper. We focus our attention here on FNOs as they appear to be the most promising of the neural operator based operator learning frameworks. Our main result in this paper is to show that FNOs are *universal* in possessing the ability to approximate a very large class of continuous nonlinear operators. This result highlights the potential of FNOs for operator learning.

As argued before, a universality result is only a first step and by itself, does not constitute evidence for efficient approximation by FNOs. In fact, we show that in the worst case, the network size might grow exponentially with respect to accuracy, when approximating general operators. Hence, there is a need to derive explicit bounds on the network size in terms of the desired error tolerance. In this context, we consider a concrete computational realization of FNOs, that we term as *pseudospectral FNO* or Ψ-FNO (for short). In addition to proving universality for Ψ-FNOs, we will suggest a mechanism through which Ψ-FNOs can approximate operators arising from PDEs, efficiently. We also derive explicit error bounds for this architecture in approximating PDEs, for two widely used prototypical examples of PDEs i.e, a Darcy type elliptic equation and the incompressible Navier-Stokes equations of fluid dynamics. In particular, we prove that the size of Ψ-FNOs in approximating the underlying operators for both these PDEs, under suitable hypotheses, only scales polynomially (log-linearly) in the error. Thus, FNOs can approximate these operators efficiently and these results validate some of the computational findings of Li et al. (2021). Together, these results constitute the first theoretical justification for the use of FNOs.

The rest of the paper is organized as follows: in Section 2, we introduce FNOs and state the universality result. We also introduce Ψ-FNOs in this section. In Section 3, we show that Ψ-FNOs can efficiently approximate operators, stemming from the Darcy-type elliptic equation as well as the incompressible Navier-Stokes equations. In Section 4, we compare FNOs with DeepOnets and the results of the article are discussed in Section 5. The mathematical notation, used in this paper, is summarized in Appendix A and we present all the technical details and proofs in other appendices.

2. Approximation by Fourier Neural Operators

In this section, we present Fourier Neural Operators (FNOs) and discuss their approximation of a class of nonlinear operators specified below:

2.1 Setting for Operator Learning

Setting 1 *We fix a spatial dimension $d \in \mathbb{N}$, and denote by $D \subset \mathbb{R}^d$ a domain in $\mathbb{R}^d$. We consider the approximation of operators $\mathcal{G} : \mathcal{A}(D; \mathbb{R}^{d_a}) \to \mathcal{U}(D; \mathbb{R}^{d_u})$, $a \mapsto u := \mathcal{G}(a)$, where the input $a \in \mathcal{A}(D; \mathbb{R}^{d_a})$, $d_a \in \mathbb{N}$, is a function $a : D \to \mathbb{R}^{d_a}$ with d_a components, and the output $u \in \mathcal{U}(D; \mathbb{R}^{d_u})$, $d_u \in \mathbb{N}$, is a function $u : D \to \mathbb{R}^{d_u}$ with d_u components. Here $\mathcal{A}(D; \mathbb{R}^{d_a})$ and $\mathcal{U}(D; \mathbb{R}^{d_u})$ are Banach spaces (or suitable subsets of Banach spaces). Typical examples of $\mathcal{A}$ and $\mathcal{U}$ include the space of continuous functions $C(D; \mathbb{R}^{d_u})$, or Sobolev spaces $H^s(D; \mathbb{R}^{d_u})$ of order $s \geq 0$ (see Appendix B for definitions.).*

Concrete examples for operators $\mathcal{G}$, involving solution operators of PDEs, are given in Section 3.

2.2 Neural Operators

With the above Setting 1 and as defined in Li et al. (2020a), a neural operator $\mathcal{N} : \mathcal{A}(D; \mathbb{R}^{d_a}) \to \mathcal{U}(D; \mathbb{R}^{d_u})$, $a \mapsto \mathcal{N}(a)$ is a mapping of the form

$$\mathcal{N}(a) = \mathcal{Q} \circ \mathcal{L}_L \circ \mathcal{L}_{L-1} \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}(a),$$

for a given depth $L \in \mathbb{N}$, where $\mathcal{R} : \mathcal{A}(D; \mathbb{R}^{d_a}) \to \mathcal{U}(D; \mathbb{R}^{d_v})$, $d_v \geq d_u$, is a *lifting* operator (acting locally), of the form

$$\mathcal{R}(a)(x) = Ra(x), \quad R \in \mathbb{R}^{d_v \times d_a}, \tag{1}$$

and $\mathcal{Q} : \mathcal{U}(D; \mathbb{R}^{d_v}) \to \mathcal{U}(D; \mathbb{R}^{d_u})$ is a local *projection* operator, of the form

$$\mathcal{Q}(v)(x) = Qv(x), \quad Q \in \mathbb{R}^{d_u \times d_v}. \tag{2}$$

Remark 2 *In practice, it has been found that improved results can be obtained if the simple linear lifting and projection operators $\mathcal{R}$ (1) and $\mathcal{Q}$ (2) are replaced instead by non-linear mappings of the form*

$$\widetilde{\mathcal{R}}(a)(x) = \widetilde{R}(a(x), x), \quad \widetilde{\mathcal{Q}}(v)(x) = \widetilde{Q}(v(x), x),$$

where $\widetilde{R} : \mathbb{R}^{d_a} \times D \to \mathbb{R}^{d_v}$ and $\widetilde{Q} : \mathbb{R}^{d_v} \times D \to \mathbb{R}^{d_u}$ are neural networks with activation function σ. Our error estimates will rely on the (more restrictive) linear choice of lifting and projection operators, given by (1), (2). The linear choice has the theoretical benefit of ensuring compositionality, *i.e. that a composition of neural operators can again be represented by a neural operator (cf. Lemma 46). Despite this technical distinction, we emphasize that all of our error and complexity estimates continue to hold also for neural operators with non-linear lifting and projections, since linear operators can always be approximated by non-linear ones (cp. Lemma 42). In fact, in the non-linear case, our results imply that $\widetilde{Q}$, $\widetilde{R}$ can be chosen to be* shallow *networks.*

In analogy with canonical finite-dimensional neural networks, Li et al. (2020a) define the layers $\mathcal{L}_1, \ldots, \mathcal{L}_L$ to be non-linear operator layers, $\mathcal{L}_\ell : \mathcal{U}(D; \mathbb{R}^{d_v}) \to \mathcal{U}(D; \mathbb{R}^{d_v})$, $v \mapsto \mathcal{L}_\ell(v)$, of the form

$$\mathcal{L}_\ell(v)(x) = \sigma\left(W_\ell v(x) + b_\ell(x) + \big(\mathcal{K}(a; \theta_\ell)v\big)(x) \right), \quad \forall \, x \in D.$$

4

Here, the weight matrix $W_\ell \in \mathbb{R}^{d_v \times d_v}$ and bias $b_\ell(x) \in \mathcal{U}(D; \mathbb{R}^{d_v})$ define an affine pointwise mapping $W_\ell v(x) + b_\ell(x)$. The richness of linear operators in the infinite-dimensional setting can partly be realized by defining the following *non-local* linear operator,

$$\mathcal{K} : \mathcal{A} \times \Theta \to L\left(\mathcal{U}(D; \mathbb{R}^{d_v}), \mathcal{U}(D; \mathbb{R}^{d_v})\right),$$

that maps the input field a and a parameter $\theta \in \Theta$ in the parameter-set Θ to a bounded linear operator $\mathcal{K}(a, \theta) : \mathcal{U}(D; \mathbb{R}^{d_v}) \to \mathcal{U}(D; \mathbb{R}^{d_v})$, and the non-linear activation function $\sigma : \mathbb{R} \to \mathbb{R}$ is applied component-wise. As proposed in Li et al. (2020a), the linear operators $\mathcal{K}(a, \theta)$ are integral operators of the form

$$\left(\mathcal{K}(a; \theta)v\right)(x) = \int_D \kappa_\theta(x, y; a(x), a(y))v(y)\, dy, \quad \forall x \in D. \tag{3}$$

Here, the integral kernel $\kappa_\theta : \mathbb{R}^{2(d+d_a)} \to \mathbb{R}^{d_v \times d_v}$ is a neural network parametrized by $\theta \in \Theta$. Specific examples of the integral kernel (3) include those evaluated with a graph kernel network as in Li et al. (2020a) or with a multipole expansion in Li et al. (2020b). The explicit dependence on the input a of each kernel turns this construction into a RNN-like architecture where the single input is turned into a sequence simply by repetition across the layers. While we do not theoretically study the effect of this architectural choice, it was empirically shown in Li et al. (2020a) that it leads to more efficient approximation on certain problems.

2.3 Fourier Neural Operators

As defined in (Li et al., 2021), Fourier Neural operators (FNOs) are a special case of general neural operators (3), in which the kernel $\kappa_\theta(x, y; a(x), a(y))$ is of the form $\kappa_\theta = \kappa_\theta(x - y)$. In this case, (3) can be written as a convolution

$$\left(\mathcal{K}(\theta)v\right)(x) = \int_D \kappa_\theta(x - y)v(y)\, dy, \quad \forall x \in D. \tag{4}$$

For concreteness, we consider the periodic domain $D = \mathbb{T}^d$ (which we identify with the standard torus $\mathbb{T}^d = [0, 2\pi]^d$), although non-periodic, rectangular domains D can also be handled in a straightforward manner.

Given this periodic framework, the convolution operator in (4) can be computed using the Fourier transform $\mathcal{F}$ and the inverse Fourier transform $\mathcal{F}^{-1}$ (see Appendix B, equations 40 and 41 for notation and definitions), resulting in the following equivalent representation of the kernel (3),

$$(\mathcal{K}(\theta)v)(x) = \mathcal{F}^{-1}\left(P_\theta(k) \cdot \mathcal{F}(v)(k)\right)(x), \quad \forall x \in \mathbb{T}^d. \tag{5}$$

Here, $P_\theta(k) \in \mathbb{C}^{d_v \times d_v}$ is a full matrix indexed by $k \in \mathbb{Z}^d$, and is related to the integral kernel $\kappa_\theta(x)$ in (4) via the Fourier transform, $P_\theta(k) = \mathcal{F}(\kappa_\theta)(k)$. Note that we must impose that $P_\theta(-k) = P_\theta(k)^\dagger$ coincides with the Hermitian transpose for all $k \in \mathbb{Z}^d$, to ensure that the image function $(\mathcal{K}(\theta)v)(x)$ is a real-valued function for real-valued $v(x)$. Consequently, the

form of Fourier neural operators (FNOs) for the periodic domain $\mathbb{T}^d$ is that of a mapping $\mathcal{N} : \mathcal{A}(D; \mathbb{R}^{d_a}) \to \mathcal{U}(D; \mathbb{R}^{d_u})$, of the form

$$\mathcal{N}(a) := \mathcal{Q} \circ \mathcal{L}_L \circ \mathcal{L}_{L-1} \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}(a), \tag{6}$$

where the lifting and projection operators $\mathcal{R}$ and $\mathcal{Q}$ are given by (1) and (2), respectively, and where the non-linear layers $\mathcal{L}_\ell$ are of the form

$$\mathcal{L}_\ell(v)(x) = \sigma\left(W_\ell v(x) + b_\ell(x) + \mathcal{F}^{-1}\Big(P_\ell(k) \cdot \mathcal{F}(v)(k)\Big)(x)\right). \tag{7}$$

Here, $W_\ell \in \mathbb{R}^{d_v \times d_v}$ and $b_\ell(x)$ define a pointwise affine mapping (corresponding to weights and biases), and $P_\ell : \mathbb{Z}^d \to \mathbb{C}^{d_v \times d_v}$ defines the coefficients of a non-local, linear mapping via the Fourier transform.

Remark 3 *The simplest example for a FNO, as defined by (6),(4) is as follows; let $\Phi : \mathbb{R}^{d_a} \to \mathbb{R}^{d_u}$ be a conventional finite-dimensional neural network with activation function σ. We can associate to Φ the mapping $\mathcal{N} : L^2(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L^2(\mathbb{T}^d; \mathbb{R}^{d_u})$, given by $a(x) \mapsto \Phi(a(x))$. We easily observe that $\mathcal{N}$ is a FNO as we can write it in the form*

$$\mathcal{N}(a)(x) = \Phi(a(x)) = \mathcal{Q} \circ \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}(a(x)),$$

where $\mathcal{R}(y) = Ry$ with $R \in \mathbb{R}^{d_v \times d_a}$, and each layer $\mathcal{L}_\ell$ is of the form $\mathcal{L}_\ell(y) = \sigma(W_\ell y + b_\ell)$ for some $W_\ell \in \mathbb{R}^{d_v \times d_v}$, $b_\ell \in \mathbb{R}^{d_v}$, with $\mathcal{Q}$ being an affine output layer of the form $\mathcal{Q}(y) = Qy + q$ with $Q \in \mathbb{R}^{d_u \times d_v}$, $q \in \mathbb{R}^{d_u}$. Replacing the input y by a function $v(x)$, these layers clearly are a special case of the FNO lifting layer (1), the non-linear layers (7) (with $P_\ell \equiv 0$ and constant bias $b_\ell(x) \equiv b_\ell$), and the projection layer (2). Thus, any finite-dimensional neural network can be identified with a FNO as defined above.

For the remainder of this work, we make the following assumption,

Assumption 4 (Activation function) *Unless explicitly stated otherwise, the activation function $\sigma : \mathbb{R} \to \mathbb{R}$ in (7) is assumed to be non-polynomial, (globally) Lipschitz continuous and $\sigma \in C^\infty$.*

2.4 Universal Approximation by FNOs

Next, we will show that FNOs (6) are *universal*, i.e., given a large class of operators, as defined in Setting 1, one can find an FNO that approximates it to desired accuracy. To be more precise, we have the following theorem.

Theorem 5 (Universal approximation) *Let $s, s' \geq 0$. Let $\mathcal{G} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$ be a continuous operator. Let $K \subset H^s(\mathbb{T}^d; \mathbb{R}^{d_a})$ be a compact subset. Then for any $\epsilon > 0$, there exists a FNO $\mathcal{N} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$, of the form (6), continuous as an operator $H^s \to H^{s'}$, such that*

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}(a)\|_{H^{s'}} \leq \epsilon.$$

Proof [Sketch of proof] The detailed proof of this universal approximation theorem is provided in Appendix D.4 and we outline it here. For notational simplicity, we set $d_a = d_u = 1$, and first observe the following lemma, proved in Appendix D.1:

Lemma 6 *Assume that the universal approximation Theorem 5 holds for $s' = 0$. Then it holds for arbitrary $s' \geq 0$.*

The main objective is thus to prove Theorem 5 for the special case $s' = 0$; i.e. given a continuous operator $\mathcal{G} : H^s(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$, $K \subset H^s(\mathbb{T}^d)$ compact, and $\epsilon > 0$, we wish to construct a FNO $\mathcal{N} : H^s(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$, such that $\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}(a)\|_{L^2} \leq \epsilon$.

To this end, we start by defining the following operator,

$$\mathcal{G}_N : H^s(\mathbb{T}^d) \to L^2(\mathbb{T}^d), \quad \mathcal{G}_N(a) := P_N \mathcal{G}(P_N a), \tag{8}$$

with P_N being the orthogonal Fourier projection operator, defined in Appendix B (45). Thus, $\mathcal{G}_N$ can be thought of loosely as the *Fourier projection* of the continuous operator $\mathcal{G}$.

Next, we can show that for any given $\epsilon > 0$, there exists $N \in \mathbb{N}$, such that

$$\|\mathcal{G}(a) - \mathcal{G}_N(a)\|_{L^2} \leq \epsilon, \quad \forall a \in K. \tag{9}$$

Thus, the proof boils down to finding a FNO (6) that can approximate the operator $\mathcal{G}_N$ to any desired accuracy.

To this end, we introduce a set of Fourier wavenumbers $k \in \mathcal{K}_N$, by

$$\mathcal{K}_N := \left\{ k \in \mathbb{Z}^d \,\middle|\, |k|_\infty \leq N \right\}, \tag{10}$$

and define a *Fourier conjugate* or *Fourier dual* operator of the form $\widehat{\mathcal{G}}_N : \mathbb{C}^{\mathcal{K}_N} \to \mathbb{C}^{\mathcal{K}_N}$,

$$\widehat{\mathcal{G}}_N(\widehat{a}_k) := \mathcal{F}_N \left(\mathcal{G}_N \left(\mathrm{Re} \left(\mathcal{F}_N^{-1}(\widehat{a}_k) \right) \right) \right), \tag{11}$$

such that the identity

$$\mathcal{G}_N(a) = \mathcal{F}_N^{-1} \circ \widehat{\mathcal{G}}_N \circ \mathcal{F}_N(P_N a), \tag{12}$$

holds for all *real-valued* $a \in L^2(\mathbb{T}^d)$. Here, $\mathcal{F}_N$ is the discrete Fourier transform and $\mathcal{F}_N^{-1}$ is the discrete inverse Fourier transform, with both being defined in Appendix B (52) and (53), respectively.

The next steps in the proof are to leverage the natural decomposition of the projection $\mathcal{G}_N$ in (12) in terms of the discrete Fourier transform $\mathcal{F}_N \circ P_N$, the discrete inverse Fourier transform $\mathcal{F}_N^{-1}$ and the Fourier conjugate operator $\widehat{\mathcal{G}}_N$ and approximate each of these operators by Fourier neural operators.

We start by denoting,

$$\mathbb{R}^{2\mathcal{K}_N} = \left(\mathbb{R}^2 \right)^{\mathcal{K}_N} (\simeq \mathbb{C}^{\mathcal{K}_N}), \tag{13}$$

as the set consisting of coefficients $\{(v_{1,k}, v_{2,k})\}_{k \in \mathcal{K}_N}$, where $v_{1,k}, v_{2,k} \in \mathbb{R}$ are indexed by the wavenumber $k \in \mathcal{K}_N$, i.e. such that $|k|_\infty \leq N$, and we interpret the operator $\mathcal{F}_N \circ P_N$ as a mapping $\mathcal{F}_N \circ P_N : a \mapsto \{(\mathrm{Re}(\widehat{a}_k), \mathrm{Im}(\widehat{a}_k))\}_{|k| \leq N}$, with input $a \in L^2(\mathbb{T}^d)$ and the output $\{\mathrm{Re}(\widehat{a}_k), \mathrm{Im}(\widehat{a}_k)\}_{|k| \leq N} \in \mathbb{R}^{2\mathcal{K}_N}$ is viewed as a *constant* function in $L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$. The approximation of this operator is a straightforward consequence of the following Lemma, proved in Appendix D.2,

Lemma 7 *Let $B > 0$ and $N \in \mathbb{N}$ be given. For all $\epsilon > 0$, there exists a FNO $\mathcal{N} : L^2(\mathbb{T}^d) \to L^2(T^d; \mathbb{R}^{2\mathcal{K}_N})$, $v \mapsto \{\mathcal{N}(v)_{\ell,k}\}$, with* constant *output functions (constant as a function of $x \in \mathbb{T}^d$), and such that*

$$
\left.\begin{array}{l}
\|\mathrm{Re}(\widehat{v}_k) - \mathcal{N}(v)_{1,k}\|_{L^\infty} \leq \epsilon \\
\|\mathrm{Im}(\widehat{v}_k) - \mathcal{N}(v)_{2,k}\|_{L^\infty} \leq \epsilon
\end{array}\right\} \quad \forall\, k \in \mathbb{Z}^d,\ |k|_\infty \leq N,
$$

for all $\|v\|_{L^2} \leq B$, and where $\widehat{v}_k \in \mathbb{C}$ denotes the k-th Fourier coefficient of v.

In the next step, we approximate the (discrete) inverse Fourier transform $\mathcal{F}_N^{-1}$ by an FNO. We recall that FNOs act on functions rather than on constants. Therefore, to connect $\mathcal{F}_N^{-1}$ and FNOs, we are going to interpret the mapping

$$
\mathcal{F}_N^{-1} : [-R, R]^{2\mathcal{K}_N} \subset \mathbb{R}^{2\mathcal{K}_N} \to L^2(\mathbb{T}^d),
$$

as a mapping

$$
\mathcal{F}_N^{-1} : \begin{cases} L^2(\mathbb{T}^d; [-R,R]^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d), \\ \{\mathrm{Re}(\widehat{v}_k), \mathrm{Im}(\widehat{v}_k)\}_{|k| \leq N} \mapsto v(x), \end{cases}
$$

where the input $\{\mathrm{Re}(\widehat{v}_k), \mathrm{Im}(\widehat{v}_k)\}_{|k| \leq N} \in [-R, R]^{2\mathcal{K}_N}$ is identified with a *constant* function in $L^2(\mathbb{T}^d; [-R,R]^{2\mathcal{K}_N})$. The existence of a FNO of the form (6) that can approximate (73) to desired accuracy is a consequence of the following lemma, proved in Appendix D.3,

Lemma 8 *Let $B > 0$ and $N \in \mathbb{N}$ be given. For all $\epsilon > 0$, there exists a FNO $\mathcal{N} : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d)$, such that for any $v \in L_N^2(\mathbb{T}^d)$ with $\|v\|_{L^2} \leq B$, we have*

$$
\|v - \mathcal{N}(w)\|_{L^2} \leq \epsilon,
$$

where $w(x) := \{(\mathrm{Re}(\widehat{v}_k), \mathrm{Im}(\widehat{v}_k))\}_{k \in \mathcal{K}_N}$, i.e. $w \in L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$ is a constant function collecting the real and imaginary parts of the Fourier coefficients $\widehat{v}_k$ of v.

Finally, by setting $\widehat{K} := \mathcal{F}_N(P_N K) \subset \mathbb{C}^{\mathcal{K}_N}$ as the (compact) image of K under the continuous mapping $\mathcal{F}_N \circ P_N : L^2(\mathbb{T}^d) \to \mathbb{C}^{\mathcal{K}_N}$ and identifying $\mathbb{C}^{\mathcal{K}_N} \simeq \mathbb{R}^{2\mathcal{K}_N}$, where $\widehat{v}_{1,k} := \mathrm{Re}(\widehat{v}_k)$ and $\widehat{v}_{2,k} := \mathrm{Im}(\widehat{v}_k)$ for $k \in \mathcal{K}_N$, we can view $\widehat{\mathcal{G}}_N$ as a continuous mapping

$$
\widehat{\mathcal{G}}_N : \widehat{K} \subset \mathbb{R}^{2\mathcal{K}_N} \to \mathbb{R}^{2\mathcal{K}_N},
$$

on a compact subset. Hence, by the universal approximation theorem for finite-dimensional neural networks (Barron, 1993; Hornik et al., 1989), one can readily show that there exists an FNO, with only *local* weights (see Remark 3), which will approximate this continuous mapping $\widehat{G}_N$ on compact subsets to desired accuracy.

Hence, each of the component operators of the decomposition (12) can be approximated to desired accuracy by FNOs and the universal approximation theorem follows by composing these FNOs and estimating the resulting error, with details provided in Appendix D.4. ∎

In the following theorem, we will show that the universal approximation theorem 5 can be extended to include operators mapping function spaces, which are defined with respect

to arbitrary Lipschitz regular domains $D \subset \mathbb{R}^d$. In fact, even the Lipschitz condition on the domain can be relaxed further to include all locally uniform domains using ideas from Rogers (2006); we will, however, not pursue this for simplicity of the exposition. We show that one can construct a periodic extension of the input function and a FNO so that the restriction of the FNO's periodic output to the domain of interest provides a suitable approximation to any continuous operator.

Theorem 9 *Let $s, s' \geq 0$ and let $\Omega \subset \mathbb{R}^d$ be a domain with Lipschitz boundary, and such that the closure $\overline{\Omega} \subset (0, 2\pi)^d$. Let $\mathcal{G} : H^s(\Omega; \mathbb{R}^{d_a}) \to H^{s'}(\Omega; \mathbb{R}^{d_u})$ be a continuous operator. Let $K \subset H^s(\Omega; \mathbb{R}^{d_a})$ be a compact subset. Then there exists a continuous, linear operator $\mathcal{E} : H^s(\Omega; \mathbb{R}^{d_a}) \to H^s(\mathbb{T}^d; \mathbb{R}^{d_a})$ such that $\mathcal{E}(a)|_\Omega = a$ for all $a \in H^s(\Omega; \mathbb{R}^{d_a})$. Furthermore, for any $\epsilon > 0$, there exists a FNO $\mathcal{N} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$ of the form (6), such that*

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N} \circ \mathcal{E}(a)|_\Omega\|_{H^{s'}} \leq \epsilon.$$

Proof Since, by assumption, the closure $\overline{\Omega}$ is contained in the open set $(0, 2\pi)^d$, we have that $\mathrm{dist}(\Omega, \partial[0, 2\pi]^d) > 0$. Hence the conclusion of Lemma 41 in Appendix B follows with the hypercube $B = [0, 2\pi]^d$; in particular, there exists a continuous, linear operator $\mathcal{E} : H^s(\Omega; \mathbb{R}^{d_a}) \to H^s([0, 2\pi]^d; \mathbb{R}^{d_a})$ such that $\mathcal{E}(a)|_\Omega = a$ and $\mathcal{E}(a)$ is periodic on $[0, 2\pi]^d$ for all $a \in H^s(\Omega; \mathbb{R}^{d_a})$. Therefore $\mathcal{E} : H^s(\Omega; \mathbb{R}^{d_a}) \to H^s(\mathbb{T}^d; \mathbb{R}^{d_a})$. Similarly, we can construct an extension operator $\mathcal{E}' : H^{s'}(\Omega; \mathbb{R}^{d_u}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$ on $H^{s'}$.

We can then associate to $\mathcal{G} : H^s(\Omega; \mathbb{R}^{d_a}) \to H^{s'}(\Omega; \mathbb{R}^{d_u})$ another continuous operator $\overline{\mathcal{G}} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$, by defining $\overline{\mathcal{G}}(a) := \mathcal{E}' \circ \mathcal{G} \circ \mathcal{R}(a)$. Here $\mathcal{R}(a) := a|_\Omega$ denotes the restriction to Ω which is clearly linear and continuous. By the continuity of $\mathcal{E}$, we have that $K' := \mathcal{E}(K)$ is compact in $H^s(\mathbb{T}^d; \mathbb{R}^{d_a})$. By the universal approximation theorem 5, for any $\epsilon > 0$, there exists a FNO $\mathcal{N} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$, such that

$$\sup_{a' \in K'} \|\overline{\mathcal{G}}(a') - \mathcal{N}(a')\|_{H^{s'}} \leq \epsilon.$$

But then, using the fact that $\mathcal{R} \circ \mathcal{E} = \mathrm{Id}$, $\mathcal{R} \circ \mathcal{E}' = \mathrm{Id}$, the mapping $\mathcal{R} \circ \mathcal{N} \circ \mathcal{E} : H^s(\Omega; \mathbb{R}^{d_a}) \to H^{s'}(\Omega; \mathbb{R}^{d_u})$, given by $a \mapsto \mathcal{N} \circ \mathcal{E}(a)|_\Omega$, satisfies

$$
\begin{aligned}
\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N} \circ \mathcal{E}(a)|_\Omega\|_{H^{s'}} &= \sup_{a \in K} \|\mathcal{R} \circ \mathcal{E}' \circ \mathcal{G} \circ \mathcal{R} \circ \mathcal{E}(a) - \mathcal{R} \circ \mathcal{N} \circ \mathcal{E}(a)\|_{H^{s'}} \\
&= \sup_{a \in K} \|\mathcal{R} \circ \overline{\mathcal{G}} \circ \mathcal{E}(a) - \mathcal{R} \circ \mathcal{N} \circ \mathcal{E}(a)\|_{H^{s'}} \\
&\leq \sup_{a \in K} \|\overline{\mathcal{G}} \circ \mathcal{E}(a) - \mathcal{N} \circ \mathcal{E}(a)\|_{H^{s'}} \\
&= \sup_{a' \in K'} \|\overline{\mathcal{G}}(a') - \mathcal{N}(a')\|_{H^{s'}} \\
&\leq \epsilon.
\end{aligned}
$$

$\blacksquare$

Remark 10 *Similar ideas as in Theorem 9 have been pursued in the design of numerical algorithms for solving PDEs and usually go by the name of Fourier continuations (Bruno and Lyon, 2010; Lyon and Bruno, 2010). A major challenge for these methods is designing a suitable periodic function whose restriction gives the solution of interest. Theorem 9 shows that FNOs are expressive enough to automatically learn such a representation. It should be noted that Theorem 9 does not suggest a way of computing the extension operator $\mathcal{E}$ and designing suitable methods remains an interesting direction for computational research.*

2.5 Ψ-Fourier Neural Operators

In practice, one needs to compute the FNO, of form (6), both during training as well as for the evaluation of the neural operator. Thus, given any input function a, one should be able to readily calculate the FNO $\mathcal{N}(a)$, requiring the efficient computation of the Fourier transform $\mathcal{F}$ (40) and the inverse Fourier transform $\mathcal{F}^{-1}$ (41). In general, this is not possible as evaluating the Fourier transform (40) entails computing an integral exactly. Therefore, approximations are necessary to realize the action of FNOs on functions. Following Li et al. (2021), one can efficiently approximate the Fourier transform and its inverse by the discrete Fourier transform (52) and the discrete inverse Fourier transform (53), respectively. This amounts to performing a pseudo(Ψ)-spectral Fourier projection between successive layers of the FNO and leading to the following precise definition,

Definition 11 (Ψ-FNO) *A $\boldsymbol{\Psi}$-**FNO** (or Ψ-spectral FNO) is a mapping*

$$\mathcal{N}^* : \mathcal{A}(\mathbb{T}^d; \mathbb{R}^{d_a}) \to \mathcal{U}(\mathbb{T}^d; \mathbb{R}^{d_u}), \quad a \mapsto \mathcal{N}^*(a),$$

of the form

$$\mathcal{N}^*(a) = \mathcal{Q} \circ \mathcal{I}_N \circ \mathcal{L}_L \circ \mathcal{I}_N \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{I}_N \circ \mathcal{R}(a), \tag{14}$$

where $\mathcal{I}_N$ denotes the pseudo-spectral Fourier projection onto trigonometric polynomials of degree $N \in \mathbb{N}$ (49), the lifting operator $\mathcal{R} : \mathcal{A}(\mathbb{T}^d; \mathbb{R}^{d_a}) \to \mathcal{U}(\mathbb{T}^d; \mathbb{R}^{d_v})$, the projection $\mathcal{Q} : \mathcal{U}(\mathbb{T}^d; \mathbb{R}^{d_v}) \to \mathcal{U}(\mathbb{T}^d; \mathbb{R}^{d_u})$ are defined as in (1), (2), and the non-linear layers $\mathcal{L}_\ell$, for $\ell = 1, \ldots, N$, are of the form

$$\mathcal{L}_\ell(v)(x) = \sigma\left(W_\ell v(x) + b_\ell(x) + \mathcal{F}^{-1}\Big(P_\ell(k) \cdot \mathcal{F}(v)(k) \Big)(x) \right).$$

Here, $W_\ell \in \mathbb{R}^{d_v \times d_v}$ and $b_\ell(x) \in \mathcal{U}(\mathbb{T}^d; \mathbb{R}^{d_v})$ define a pointwise affine mapping $v \mapsto W_\ell v(x) + b_\ell(x)$, and the coefficients $P_\ell(k) \in \mathbb{R}^{d_v \times d_v}$ ($k \in \mathcal{K}_N$) define a (non-local) convolution operator via the Fourier transform.

Note that a Ψ-FNO $\mathcal{N}^*$ is uniquely defined, as an operator, by its restriction to the finite-dimensional subspace $L^2_N(\mathbb{T}^d; \mathbb{R}^{d_a}) \subset \mathcal{A}(\mathbb{T}^d; \mathbb{R}^{d_a})$ (see Appendix B for the definition of L^2_N). Furthermore, we have that the image $\mathrm{Im}(\mathcal{N}) \subset L^2_N(\mathbb{T}^d; \mathbb{R}^{d_u})$. To indicate that a Ψ-FNO is of the form (14), for some $N \in \mathbb{N}$, we shall thus more simply say that "$\mathcal{N}^* : L^2_N(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L^2_N(\mathbb{T}^d; \mathbb{R}^{d_u})$ is a Ψ-FNO".

At the level of numerical implementation, a Ψ-FNO can be naturally identified with a *finite-dimensional* mapping

$$\widetilde{\mathcal{N}}^* : \mathbb{R}^{d_a \times \mathcal{J}_N} \to \mathbb{R}^{d_u \times \mathcal{J}_N}, \quad \boldsymbol{a} \mapsto \widetilde{\mathcal{N}}^*(\boldsymbol{a}),$$

with input $\boldsymbol{a} = \{a_j\}_{j \in \mathcal{J}_N} \in \mathbb{R}^{d_a \times \mathcal{J}_N}$ corresponding to the point-values $a_j = a(x_j)$ on the grid $\{x_j\}_{j \in \mathcal{J}_N}$, and $\mathcal{J}_N := \{0, \ldots, 2N\}^d$. Here, $\widetilde{\mathcal{N}}^*$ is of the form

$$\widetilde{\mathcal{N}}^*(\boldsymbol{a}) = \widetilde{\mathcal{Q}} \circ \widetilde{\mathcal{L}}_L \circ \widetilde{\mathcal{L}}_{L-1} \circ \cdots \circ \widetilde{\mathcal{L}}_1 \circ \widetilde{\mathcal{R}}(\boldsymbol{a}), \tag{15}$$

where the lifting operator $\widetilde{\mathcal{R}} : \mathbb{R}^{d_a \times \mathcal{J}_N} \to \mathbb{R}^{d_v \times \mathcal{J}_N}$, $\boldsymbol{a} \mapsto \widetilde{\mathcal{R}}(\boldsymbol{a})$, the projection $\widetilde{\mathcal{Q}} : \mathbb{R}^{d_v \times \mathcal{J}_N} \to \mathbb{R}^{d_u \times \mathcal{J}_N}$, $\boldsymbol{v} \mapsto \widetilde{\mathcal{Q}}(\boldsymbol{v})$, are given by

$$\widetilde{\mathcal{R}}(\boldsymbol{a}) = \{Ra_j\}_{j \in \mathcal{J}_N}, \quad (R \in \mathbb{R}^{d_v \times d_a}),$$
$$\widetilde{\mathcal{Q}}(\boldsymbol{v}) = \{Qv_j\}_{j \in \mathcal{J}_N}, \quad (Q \in \mathbb{R}^{d_u \times d_v}),$$

and the non-linear layers $\widetilde{\mathcal{L}}_\ell$, for $\ell = 1, \ldots, N$, are of the form

$$\widetilde{\mathcal{L}}_\ell(\boldsymbol{v})_j = \sigma \left(W_\ell v_j + b_{\ell,j} + \mathcal{F}_N^{-1} \Big(P_\ell(k) \cdot \mathcal{F}_N(\boldsymbol{v})(k) \Big)_j \right) \tag{16}$$

for $j \in \mathcal{J}_N$. Here, $W_\ell \in \mathbb{R}^{d_v \times d_v}$, $b_{\ell,j} = b_\ell(x_j) \in \mathbb{R}^{d_v \times \mathcal{J}_N}$ defines a pointwise affine mapping $W_\ell v_j + b_{\ell,j}$, the coefficients $P_\ell(k) \in \mathbb{C}^{d_v \times d_v}$ $(k \in \mathcal{K}_N)$ satisfy the Hermitian conjugacy condition $P_\ell(-k) = P_\ell(k)^\dagger$ and define a (non-local) convolution operator via the discrete Fourier transform, and the non-linear activation function $\sigma : \mathbb{R} \to \mathbb{R}$ is extended componentwise to a function $\mathbb{R}^{d_v \times \mathcal{J}_N} \to \mathbb{R}^{d_v \times \mathcal{J}_N}$. Comparing $\mathcal{N}^*$ with the corresponding discretization $\widetilde{\mathcal{N}}^*$, it is easy to see that, for $a \in \mathcal{A}(\mathbb{T}^d; \mathbb{R}^{d_a})$ and denoting $\boldsymbol{a} := (a(x_j))_{j \in \mathcal{J}_N} \in \mathbb{R}^{d_a \times \mathcal{J}_N}$, we have

$$\widetilde{\mathcal{N}}^*(\boldsymbol{a})_j = \mathcal{N}^*(a)(x_j), \quad \forall j \in \mathcal{J}_N.$$

In particular, this implies that $\mathcal{N}^*(a)(x)$ can in practice be computed for any $x \in \mathbb{T}^d$ via the Fourier interpolation of the grid values $\widetilde{\mathcal{N}}^*(\boldsymbol{a})$. In contrast to general FNOs, Ψ-FNOs therefore allow for efficient numerical implementation. Furthermore, the discrete (inverse) Fourier transforms in each hidden layer in (16) can be very efficiently computed using the fast Fourier transform (FFT).

The above discussion also leads to a very natural definition of the size of a Ψ-FNO below:

Definition 12 (Depth, lift and size) *The **depth** of a Ψ-FNO $\mathcal{N}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_u})$ (cp. Definition 11), is defined by* $\mathrm{depth}(\mathcal{N}^*) := L$. *We refer to the dimension d_v, as the **lifting dimension** of $\mathcal{N}^*$, denoted* $\mathrm{lift}(\mathcal{N}^*) := d_v$. *We define the **size** of a Ψ-FNO $\mathcal{N}^*$ as the total number of degrees of freedom in its numerical implementation $\widetilde{\mathcal{N}}^*$ (cp. equation 15). A simple calculation shows that*

$$\mathrm{size}(\mathcal{N}^*) = \underbrace{d_u d_v}_{\mathrm{size}(\mathcal{Q})} + L \underbrace{\left(d_v^2 + d_v |\mathcal{J}_N| + d_v^2 |\mathcal{J}_N| \right)}_{\mathrm{size}(\mathcal{L}_\ell)} + \underbrace{d_a d_v}_{\mathrm{size}(\mathcal{R})} .$$

The precise size of a Ψ-FNO will not be of any particular relevance for our asymptotic complexity estimates. Instead, we will usually content ourselves with the simple estimate

$$\mathrm{size}(\mathcal{N}^*) \lesssim L d_v^2 N^d,$$

where we assume that $\max(d_a, d_u) \leq d_v$.

11

Given our discussion, it is natural to ask whether any FNO $\mathcal{N} = \mathcal{Q} \circ \mathcal{L}_L \circ \mathcal{L}_{L-1} \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}$ can be approximated to arbitrary accuracy by an associated Ψ-FNO $\mathcal{N}^* : L_N^2 \to L_N^2$,

$$\mathcal{N}^* = \mathcal{Q} \circ \mathcal{I}_N \circ \mathcal{L}_L \circ \mathcal{I}_N \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{I}_N \circ \mathcal{R},$$

for sufficiently large $N \in \mathbb{N}$? An affirmative answer can be given for a natural class of FNOs of *finite width*, defined as follows.

Definition 13 *A FNO $\mathcal{N} : \mathcal{A}(\mathbb{T}^d; \mathbb{R}^{d_a}) \to \mathcal{U}(\mathbb{T}^d; \mathbb{R}^{d_u})$ is said to be **of finite width**, if $\mathcal{N}$ is a composition $\mathcal{N} = \mathcal{Q} \circ \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}$, with layers $\mathcal{L}_\ell$ of the form (7), and for which there exists a "width" $w \in \mathbb{N}$, such that the Fourier multiplier $P_\ell(k) \equiv 0$, for $|k|_\infty > w$.*

We can now state the following theorem, which shows that Ψ-FNOs $\mathcal{N}^*$ provide an arbitrarily close approximation of a given FNO $\mathcal{N}$:

Theorem 14 *Assume that the activation function $\sigma \in C^m$ is globally Lipschitz continuous. Let $\mathcal{N} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L^2(\mathbb{T}^d; \mathbb{R}^{d_u})$ be a FNO of finite width, with $s > d/2$. and assume that $m > s$. Then for any $\epsilon, B > 0$, there exists $N \in \mathbb{N}$ and a Ψ-FNO $\mathcal{N}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_u})$, such that*

$$\sup_{\|a\|_{H^s} \leq B} \|\mathcal{N}(a) - \mathcal{N}^*(a)\|_{L^2} \leq \epsilon.$$

For the proof, we refer to Appendix D.5. In particular, the last theorem implies an extension of the universal approximation theorem 5 to Ψ-*FNOs*, provided that the input functions have sufficient regularity for the pseudo-spectral projection (49) to be well-defined:

Theorem 15 (Universal approximation for Ψ-FNOs) *Let $s > d/2$, and let $s' \geq 0$. Let $\mathcal{G} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$ be a continuous operator. And let $K \subset H^s(\mathbb{T}^d; \mathbb{R}^{d_a})$ be a compact subset. Then for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ and a Ψ-FNO $\mathcal{N}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_u})$, such that*

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}^*(a)\|_{H^{s'}} \leq \epsilon.$$

Proof Similar to the proof of the universal approximation theorem for FNOs, we again note that the general case $s' \geq 0$ can be deduced from the statement of Theorem 15 for the special case $s' = 0$. This is the content of the following lemma, whose proof is provided in Appendix D.6:

Lemma 16 *Assume that Theorem 15 holds for $s' = 0$. Then it holds for arbitrary $s' \geq 0$.*

The special case $s' = 0$ follows immediately from Theorem 14 and the observation that the FNO approximation constructed in the proof of the universal approximation theorem for FNOs, Theorem 5, has finite width. ∎

Remark 17 *The form of the universal approximation theorem 5 stated above shows that any continuous operator $\mathcal{G} : H^s \to H^{s'}$, $s \geq d/2$, can be approximated to arbitrary accuracy by a Ψ-FNO, on a given compact subset $K \subset H^s$. In practical applications, the parameters*

of neural networks are usually trained from data, i.e. based on the minimization of an empirical loss of the form

$$\frac{1}{M} \sum_{m=1}^{M} \|\mathcal{N}^*(a_m) - \mathcal{G}(a_m)\|_{H^{s'}}^2, \tag{17}$$

where $a_1, \ldots, a_M \sim \mu$ are (iid) samples from a probability measure $\mu \in \mathcal{P}(H^s)$ on the input function space. The empirical loss is a discretization of the following continuous quantity:

$$\int_{H^s} \|\mathcal{N}^*(a) - \mathcal{G}(a)\|_{H^{s'}}^2 \, d\mu(a). \tag{18}$$

The form of the continuous loss (18) might suggest to measure the distance between $\mathcal{N}^*$ and $\mathcal{G}$ in the $L^2_\mu(H^s; H^{s'})$-metric (the square root of equation 18), rather than the sup-norm. This L^2_μ-approach has been followed in recent work on DeepONets by Lanthaler et al. (2021), to explain some observations of the numerical experiments of Lu et al. (2021). In the present work, we focus instead on the sup-norm over compact subsets $K \subset H^s$:

$$\sup_{a \in K} \|\mathcal{N}^*(a) - \mathcal{G}(a)\|_{H^{s'}}. \tag{19}$$

Clearly, if the support of μ is a compact subset, with $\operatorname{supp}(\mu) = K$, then our sup-estimates immediately imply corresponding bounds on the L^2_μ-distance, since the sup-norm is a considerably stronger distance-measure than the L^2_μ-norm. On the other hand, the restriction to compact subsets $K \subset H^s$ does not allow for popular choices of the probability measure μ, such as Gaussian measures. Furthermore, the sup-estimates do not allow for discontinuous operators $\mathcal{G}$.

Expanding on the last remark, we would like to point out that the universal approximation theorem, Theorem 15, combined with a technical cut-off argument, can be used to prove a corresponding universal approximation theorem in the L^2_μ-norm, following an idea in Lanthaler et al. (2021). We state this extension without proof:

Theorem 18 *Assume that the activation function $\sigma \in C^\infty(\mathbb{R})$ is strictly monotonically increasing and satisfies $\lim_{x \to -\infty} \sigma(x) > -\infty$. Let $\mu \in \mathcal{P}(H^s(\mathbb{T}^d; \mathbb{R}^{d_a}))$ be a probability measure. Let $\mathcal{G} : H^s(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$ be a μ-measurable operator, for $s \geq d/2$, $s' \geq 0$, and assume that $\int_{H^s} \|\mathcal{G}(a)\|_{H^{s'}}^2 \, d\mu(a) < \infty$. Then, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$, and a Ψ-FNO $\mathcal{N}^* : L^2_N(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L^2_N(\mathbb{T}^d; \mathbb{R}^{d_u})$, such that*

$$\left(\int_{H^s} \|\mathcal{G}(a) - \mathcal{N}^*(a)\|_{H^{s'}}^2 \, d\mu(a) \right)^{1/2} \leq \epsilon.$$

Theorem 18 removes the compactness assumption from the support of μ, and allows for *discontinuous* operators $\mathcal{G}$—albeit at the expense of measuring the distance in the less informative L^2_μ-norm.

Before moving on to a detailed discussion of the structure of Ψ-FNOs in the next section, which will form the basis of our quantitative estimates for two concrete model problems in the remainder of this work, we would like to provide a final remark on the differences between the sup-estimates which form the focus of the present work and similar L^2_μ-estimates derived in Lanthaler et al. (2021):

Remark 19 *A benefit of the sup-estimates in the present work is that they are independent of the precise properties of the input measure μ, depending only on its topological support. This independence of the details of the input measure may make the estimates more readily applicable to a wide range of possible training distributions μ (with compact support) than similar results in Lanthaler et al. (2021), which depend on the finer properties of μ, such as the decay rates of the eigenvalues of the covariance operator of μ. On the other hand, taking into account the fine properties of μ in Lanthaler et al. (2021) has allowed to explain certain empirical observations by Lu et al. (2021), such as the exponential decay of the encoding error as a function of the number of sensor points used for the DeepONet encoding of the input function. As shown in Lanthaler et al. (2021), the explanation of this empirical finding depends crucially on the fine properties of the input measure μ, and in particular the decay of the eigenvalues, which cannot generally be derived by considering only the operator $\mathcal{G}$ together with the topological support, $\mathrm{supp}(\mu)$, of μ.*

In the remainder of this work, our goal will be to derive quantitative error and complexity estimates for two concrete non-linear operators, arising from the stationary Darcy equations and Navier-Stokes equations, respectively. To avoid additional technical difficulties, these results will be derived for input functions $a \in K$, taken from a suitable compact subset $K \subset H^s$. Given the discussion of Remarks 17, 19, we will thus focus on deriving estimates in the sup-norm over K.

2.5.1 Structure and Properties of Ψ-FNOs

We conclude this section by pointing out some aspects of the structure of Ψ-FNOs (14) that will be relevant in the following. To start with, we can simplify Ψ-FNOs by viewing them in terms of two types of layers. which we will refer to as σ- and $\mathcal{F}$-layers, respectively. A **σ-layer** $\mathcal{L}^*$ of a Ψ-FNO is a **local, non-linear** layer of the form $\mathcal{L}^*(v)(x) = \mathcal{I}_N \sigma \left(A \mathcal{I}_N v(x) + b \right)$, or, in the numerical implementation (cp. equation 16)

$$\widetilde{\mathcal{L}}^*(v)_j = \sigma \left(A v_j + b_j \right), \quad \forall j \in \mathcal{J}_N,$$

with $A \in \mathbb{R}^{d_v \times d_v}$, and $b_j \in \mathbb{R}^{\mathcal{J}_N \times d_v}$ defining an affine mapping. A **$\mathcal{F}$-layer** $\mathcal{L}^*$ of a Ψ-FNO is a **non-local, linear** layer of the form $\mathcal{L}^*(v)(x) = \mathcal{F}^{-1}(P(k) \cdot \mathcal{F}(\mathcal{I}_N v)(k))(x)$, which in a practical implementation corresponds to

$$\widetilde{\mathcal{L}}^*(v)_j = \mathcal{F}_N^{-1} \Big(P(k) \cdot \mathcal{F}_N(v)(k) \Big)_j, \quad \forall j \in \mathcal{J}_N,$$

where $P : \mathcal{K}_N \to \mathbb{C}^{d_v \times d_v}$ is a collection of complex weights, with $P(-k) = P(k)^\dagger$ the Hermitian transpose of $P(k)$, and $\mathcal{F}_N$ ($\mathcal{F}_N^{-1}$) denotes the discrete (inverse) Fourier transform.

The main point of these definitions is that each Ψ-FNO can be decomposed into a finite number of σ-layers and $\mathcal{F}$-layers, and that the converse is also true; i.e. any composition of σ-layers and $\mathcal{F}$-layers can be represented by a Ψ-FNO. These statements are made precise in a series of technical Lemmas, which are stated and proved in Appendix D.7.

3. Approximation of PDEs by Ψ-FNOs

We have shown in the previous section that FNOs (6) as well as their computational realizations (Ψ-FNOs, see equation 14) are universal, i.e., they approximate any continuous

operator, defined in the Setting 1, to desired accuracy. However, as repeatedly discussed in the introduction, universality alone does not suffice to claim that FNOs can approximate operators efficiently. In particular, it could happen that the size of the FNO is unfeasibly large to ensure a given accuracy of the approximation. In fact, the following result implies that (Ψ-)FNOs provides inefficient approximations in certain cases, where the underlying output functions are very *rough*:

Theorem 20 *Let* $\mathcal{G} : \mathcal{A}(\mathbb{T}^d; \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d; \mathbb{R}^{d_u})$ *be an operator defined on an arbitrary function space* $\mathcal{A}$*, and let* $K \subset \mathcal{A}(\mathbb{T}^d; \mathbb{R}^{d_a})$ *be a compact set. Then* any Ψ-FNO approximation $\mathcal{N}^* : L_N^2 \to L_N^2$ *satisfies the following lower bound*

$$\sup_{a \in K} \|\mathcal{N}^*(a) - \mathcal{G}(a)\|_{H^{s'}} \geq \sup_{a \in K} \|(1 - P_N)\mathcal{G}(a)\|_{H^{s'}}.$$

Proof The proof is almost trivial: Since $\mathcal{N}^*(a) \in L_N^2$, then for any $a \in \mathcal{A}$, we have

$$\|\mathcal{N}^*(a) - \mathcal{G}(a)\|_{H^{s'}} \geq \inf_{u \in L_N^2} \|u - \mathcal{G}(a)\|_{H^{s'}} = \|P_N \mathcal{G}(a) - \mathcal{G}(a)\|_{H^{s'}}.$$

$\blacksquare$

Remark 21 *In particular, the last Theorem 20 implies that if the image* $\mathcal{G}(K)$ *contains functions whose Fourier series converges only at a logarithmic rate* $\log(N)^{-1}$*, then the size of any* Ψ-FNO *achieving an approximation accuracy* ϵ *must scale exponentially,* $\text{size}(\mathcal{N}^*) \gtrsim \exp(N)$*. A concrete example is provided by the* Ψ-FNO *approximation of the identity operator* $\mathcal{G} = \text{Id} : H^s(\mathbb{T}) \to H^s(\mathbb{T})$ *(*$s \geq 0$*), on*

$$K := \left\{ a \in H^s(\mathbb{T}) \,\middle|\, \text{with Fourier coeff.:} \ |\widehat{a}_k| \leq \frac{1}{|k|^{s+1/2} \log(1+k)}, \ k \in \mathbb{Z} \right\}.$$

Thus, one reason for the potentially large sizes and consequently the inefficient approximation due to FNOs lies in the roughness of output functions. In the following remark, we point out an independent reason why the operators constructed in our proof of the universal approximation theorem may not yield an efficient approximation:

Remark 22 *We observe from the proof of Theorem 5 that the desired FNO, approximating the operator* $\mathcal{G}$*, is constructed as* $\mathcal{N}_{\text{IFT}} \circ \widehat{\mathcal{N}} \circ \mathcal{N}_{\text{FT}}$*, with* $\mathcal{N}_{\text{FT}}, \mathcal{N}_{\text{IFT}}$ *approximating the Fourier and Inverse Fourier transforms, respectively, whereas* $\widehat{\mathcal{N}} : \mathbb{R}^{2\mathcal{K}_N} \to \mathbb{R}^{2\mathcal{K}_N}$ *is a canonical finite-dimensional neural network approximation of the "Fourier conjugate operator" (11):* $\widehat{\mathcal{G}}_N : \mathbb{R}^{2\mathcal{K}_N} \to \mathbb{R}^{2\mathcal{K}_N}$*. We note that* N *herein has to be chosen sufficiently large in order to yield the desired error tolerance of* ϵ*. By Theorems 40, 20, this depends on the smoothness of the input and output spaces, i.e., if the input (output)* $a \in K \subset H^s$ *(*$\mathcal{G}(a) \in H^{s'}$*), for some* $s, s' > 0$*, then we need to choose* N *such that* $N^{-s}, N^{-s'} \sim \epsilon$*. For simplicity, we assume* $s = s'$*. Further assuming that the mapping* $\mathcal{G}$ *is Lipschitz continuous, implies that the Fourier conjugate operator* $\widehat{\mathcal{G}}$ *is also Lipschitz continuous as a mapping from* $\mathbb{R}^{2\mathcal{K}_N}$ *to* $\mathbb{R}^{2\mathcal{K}_N}$*. Hence, neural network approximation results, such as those of Yarotsky (2017) for ReLU activations or De Ryck et al. (2021) for tanh activations, yield that the size of*

the approximating neural network $\widehat{\mathcal{N}}$ scales as $\mathrm{size}(\widehat{\mathcal{N}}) \gtrsim \epsilon^{-D}$, *where D is the dimension of the domain of $\widehat{\mathcal{G}}_N$. The crucial observation is that, in the present case, the dimension $D = |\mathcal{K}_N| \sim N^d \sim \epsilon^{-d/s}$, is* not fixed. *Therefore, this yields an estimate*

$$\mathrm{size}(\widehat{\mathcal{N}}) \gtrsim \epsilon^{-\epsilon^{-d/s}}. \tag{20}$$

This scaling represents a super-exponential *growth in the size of the FNO $\mathcal{N}$, with respect to the error ϵ, incurred in approximating the underlying operator $\mathcal{G}$.*

Given the above remarks, we infer that in the worst case, a FNO approximating a generic Lipschitz continuous operator $\mathcal{G}$ might require extremely large sizes to achieve the desired accuracy, making it unfeasible in practice. The same holds for Ψ-FNOs of the form (14). This super-exponential growth appears as a form of *curse of dimensionality*, i.e., exponential growth of complexity (measured here in the size of the FNO), with respect to the error.

Hence, it is reasonable to ask how these extremely pessimistic complexity bounds on FNOs (Ψ-FNOs), can be reconciled to their robust numerical performance for approximating PDEs, as reported in Li et al. (2021). The rest of the section investigates this fundamental question.

The starting point of our explanation for the robustness of FNOs in approximating PDEs is the observation that operators which arise in the context of PDEs have a special structure and are not merely generic continuous operators mapping one infinite-dimensional function space to another. To see this, we point out that many time-dependent PDEs arising in physics can be written in the general abstract form,

$$\partial_t u + \nabla \cdot F(u, \nabla u) = 0, \tag{21}$$

where for any $(t, x) \in [0, T] \times D \subset \mathbb{R}^d$, $u(t, x) \in \mathbb{R}^{d_u}$ is a vector of physical quantities, describing e.g. density, velocity or temperature of a fluid or other material at a given point $x \in D$ in the domain D and at time $t \in [0, T]$. Equation (21) describes the general form of a conservation law for the physical quantities u with a flux function $F(u, \nabla u)$, which is typically *non-linear*, and can e.g. represent advection or diffusion terms. The flux function $F(u, \nabla u)$ may also depend on u in a *non-local* manner. For example, for the incompressible Navier-Stokes equations in $\mathbb{R}^d$,

$$u(x, t) = (u_1(x, t), \ldots, u_d(x, t)) \in \mathbb{R}^d,$$

represent the fluid velocity at (x, t), and the flux is defined by

$$F(u, \nabla u) = -u \otimes u - p + \nu \nabla u,$$

where $p = p(u)$ depends on u in a non-local manner:

$$p = \mathcal{R} : (u \otimes u), \quad \mathcal{R} := (-\Delta)^{-1}(\nabla \otimes \nabla),$$

where $\mathcal{R}$ is a (non-local) operator closely related to the Riesz transform.

A popular numerical method for time-dependent PDEs, of the form (21), particularly on periodic domains $D = \mathbb{T}^d$, is the *pseudo-spectral method* (Canuto et al., 2007), wherein (21) is discretized as,

$$\partial_t u_N + \nabla \cdot \mathcal{I}_N F(u_N, \nabla u_N) = 0, \tag{22}$$

where $u_N \in L_N^2$ is a trigonometric polynomial of degree $\leq N$.

The resulting system of ODEs (22) can be further discretized in time using a time-marching scheme. For simplicity, the forward Euler discretization with time step τ leads to,

$$u_N^{n+1} = u_N^n - \tau \nabla \cdot \mathcal{I}_N F(u_N^n, \nabla u_N^n). \tag{23}$$

One might prove that the system (23) provides a convergent approximation for the underlying time-dependent PDE (21) for many different choices of the flux F. In order to connect the approximation (23) with FNOs, we decompose the right hand side of (23) as,

$$
u_N^n \xmapsto{\mathcal{R}} \begin{Bmatrix} u_N^n \\ u_N^n \end{Bmatrix} \xmapsto{\mathcal{F}} \begin{Bmatrix} u_N^n \\ \nabla u_N^n \end{Bmatrix}
$$
$$
\xmapsto{\sigma} \ldots \xmapsto{\sigma} \begin{Bmatrix} u_N^n \\ F_N^\sigma(u_N^n, \nabla u_N^n) \end{Bmatrix}
$$
$$
\xmapsto{\mathcal{F}} \begin{Bmatrix} u_N^n \\ \nabla \cdot F_N^\sigma(u_N^n, \nabla u_N^n) \end{Bmatrix} \longmapsto u_N^n - \tau \nabla \cdot F_N^\sigma(u_N^n, \nabla u_N^n).
$$

Here, $\mathcal{R}$ is the lifting operator and σ, $\mathcal{F}$ are the σ- and $\mathcal{F}$-layers, respectively, of a Ψ-FNO, that are defined in Section 2.5.1. The above representation suggests that the Fourier $\mathcal{F}$-layers of a Ψ-FNO allow us to take *exact* derivatives, and a composition of σ-layers of a Ψ-FNO allows us to approximate continuous functions to any desired accuracy (via the standard universal approximation theorem for finite-dimensional neural networks); in particular, a composition of σ-layers can provide an approximation

$$(u, \nabla u) \mapsto F_N^\sigma(u, \nabla u) \approx \mathcal{I}_N F(u, \nabla u).$$

Thus, by a suitable composition of σ- and $\mathcal{F}$-layers, Ψ-FNOs can *emulate* pseudo-spectral methods, providing a mechanism by which such neural operators can approximate solution operators for a large class of PDEs, efficiently.

We will make this intuition precise in the following. However, instead of considering a generic abstract form of PDEs, we focus on two PDEs, often encountered in physics, which serve as prototypes for a wide variety of PDEs. We note that pseudo-spectral methods have a long and successful history in the numerical approximation of PDEs (Gottlieb and Orszag, 1977). Unfortunately, for the particular applications of the current work, the required detailed convergence estimates are not available in the literature, and therefore we include an extended discussion of suitable pseudo-spectral methods for the considered PDEs, as well as a detailed analysis of their convergence in the appendices. We start with an elliptic PDE below.

17

3.1 Stationary Darcy Flow

We consider the elliptic PDE of the form,

$$-\nabla \cdot (a\nabla u) = f. \tag{24}$$

Here, $u \in H^1(\mathbb{T}^d)$, can correspond to the steady-state pressure for a fluid, flowing according to the Darcy's law, in a porous medium with the positive coefficient $a \in L^\infty(\mathbb{T}^d)$, denoting the rock permeability. Another model for (24) is that of a diffusion equation, with u modeling the temperature and a the thermal conductivity of the medium.

For simplicity, we assume periodic boundary conditions on $\mathbb{T}^d$, and we impose that $\int_{\mathbb{T}^d} f\,dx = \int_{\mathbb{T}^d} u\,dx = 0$. Employing a suitable rescaling, we will furthermore assume that a can be written in the form:

$$a = 1 + \widetilde{a}, \quad \widetilde{a} \in H^s(\mathbb{T}^d), \ s > d/2.$$

We note that the assumption $s > d/2$ ensures that $\|\widetilde{a}\|_{L^\infty} < \infty$, via the Sobolev embedding $H^s(\mathbb{T}^d) \hookrightarrow L^\infty(\mathbb{T}^d)$ (cp. Theorem 39). To ensure that (24) is well-posed, we assume the following coercivity condition: There exists $\lambda > 0$, such that $1 + \widetilde{a} \geq \lambda$. In fact, we shall assume the slightly stronger condition that

$$\|\widetilde{a}\|_{L^\infty} \leq C\|\widetilde{a}\|_{H^s} \leq 1 - \lambda, \tag{25}$$

where C is the norm of the embedding $H^s(\mathbb{T}^d) \hookrightarrow L^\infty(\mathbb{T}^d)$. The condition (25) clearly implies the λ-coercivity of (24).

The underlying operator $\mathcal{G} : L^\infty(\mathbb{T}^d) \to H^1(\mathbb{T}^d)$, maps a coefficient $a \in L^\infty(\mathbb{T}^d)$ to the solution $u \in H^1(\mathbb{T}^d)$ of (24). Our aim is to *learn* this operator efficiently using a Ψ-FNO. To this end, we will follow the program discussed above and first present a pseudo-spectral method that approximates the Darcy flow PDE (24) accurately. Then, this pseudo-spectral method will be emulated by a suitable Ψ-FNO.

3.1.1 A FOURIER-GALERKIN APPROXIMATION OF (24)

We fix $N \in \mathbb{N}$ and assume that the coefficient field and right-hand side $a = 1 + \widetilde{a}, f \in H^s(\mathbb{T}^d)$, for some $s > d/2$, such that the pseudo-spectral projections $\mathcal{I}_N \widetilde{a}$ and $\mathcal{I}_N f$ are well-defined. We can now define a Fourier-Galerkin approximation of (24) as the (unique) solution $u_N \in L^2_N(\mathbb{T}^d)$, with $\int u_N\,dx = 0$, and such that

$$-\dot{P}_N \nabla \cdot (a_N \nabla u_N) = f_N, \tag{26}$$

where we set $a_N := 1 + \widetilde{a}_N$ and

$$\begin{cases} \widetilde{a}_N := \dot{P}_N \mathcal{I}_{2N} \widetilde{a}, \\ f_N := \dot{P}_N \mathcal{I}_{2N} f. \end{cases} \tag{27}$$

See notation for $P_N, \dot{P}_N$ and $\mathcal{I}_N$ in Appendix B. We observe from (27) that $\widetilde{a}_N$ and f_N are obtained by first carrying out a pseudo-spectral projection of $\widetilde{a}$, f on the regular grid $\{x_j\}_{j \in \mathcal{J}_{2N}}$ (with $2N$ grid points in each direction), yielding a representation of the form

$$\mathcal{I}_{2N}\widetilde{a} = \sum_{|k| \leq 2N} \widehat{a}_k e^{i\langle k, x\rangle}, \quad \mathcal{I}_{2N} f = \sum_{|k| \leq 2N} \widehat{f}_k e^{i\langle k, x\rangle},$$

18

for coefficients $\widehat{a}_k, \widehat{f}_k \in \mathbb{C}$, and then projecting these expressions onto $\dot{L}_N^2(\mathbb{T}^d)$:

$$\dot{P}_N \mathcal{I}_{2N} \widetilde{a} = \sum_{0 < |k| \le N} \widehat{a}_k e^{i\langle k, x \rangle}, \quad \dot{P}_N \mathcal{I}_{2N} f = \sum_{0 < |k| \le N} \widehat{f}_k e^{i\langle k, x \rangle}.$$

The reason for this particular choice of the projection (27) lies in the fact that the mapping

$$L_N^2(\mathbb{T}^d) \times L_N^2(\mathbb{T}^d) \to L_N^2(\mathbb{T}^d), \quad (a_N, u_N) \mapsto -P_N \nabla \cdot (a_N \nabla u_N),$$

can be exactly represented by a pseudo-spectral method on $\{x_j\}_{j \in \mathcal{J}_{2N}}$, i.e., on a denser grid of $2N$ grid points in each direction. Indeed, it is easy to see that if $a_N, u_N \in L_N^2$, then $\nabla u_N \in L_N^2$, and the values $\nabla u_N(x_j)$ can be computed exactly via the discrete Fourier transform from the values $u_N(x_j)$, $j \in \mathcal{J}_{2N}$. Since the product $a_N \nabla u_N \in L_{2N}^2$ is a trigonometric polynomial of degree $\le 2N$, the Fourier coefficients of $a_N \nabla u_N$ can also be recovered from knowledge of the point-values $a_N(x_j) \nabla u_N(x_j)$ at the grid points $x_j \in \mathcal{J}_{2N}$. Finally, since the divergence and projection P_N can be evaluated exactly via discrete Fourier transforms, we conclude that the mapping $(a_N, u_N) \mapsto -P_N \nabla \cdot (a_N \nabla u_N)$ can be computed based on knowledge of the grid values $a_N(x_j)$, and $u_N(x_j)$. The above procedure of computing a product of two trigonometric polynomials of degree $\le N$ *exactly*, based on the point-values on a finer grid of size $2N + 1$ in each direction is well-known in the context of pseudo-spectral methods, and is usually referred in the numerical analysis literature as *de-aliasing* (cf. section 3.4.2 of the textbook by Canuto et al. (2007)).

In order to computationally realize the Fourier-Galerkin method (26), we are going to recast it as a fixed point problem $u_N = F_N(u_N; a_N, f_N)$, where $F_N(\,\cdot\,; a_N, f_N) : L_N^2(\mathbb{T}^d) \to L_N^2(\mathbb{T}^d)$ is defined by

$$F_N(u_N; a_N, f_N) = \dot{P}_N(-\Delta)^{-1} \nabla \cdot (\widetilde{a}_N \nabla u_N) + (-\Delta)^{-1} f_N. \tag{28}$$

In Lemma 48 Appendix E, we show that the map (28) is a *contraction* and can be efficiently realized by a Picard type fixed-point iteration scheme. This leads to the following algorithm for realizing the Fourier-Galerkin method (26) computationally,

Algorithm 23 (Approximate solution of Darcy equation)
Input: $N \in \mathbb{N}$, $a \in H^s(\mathbb{T}^d)$, $f \in \dot{H}^{k-1}(\mathbb{T}^d)$, with $s > d/2 + k > d$, $k \in \mathbb{N}$, $k \ge 1$.
Output: $u_N \in H_N^1(\mathbb{T}^d)$, *such that* $u_N \approx u$, *where* $u \in H^1(\mathbb{T}^d)$ *solves* (24) *with coefficient field* a *and right-hand side* f.

1. *Compute pseudo-spectral projections* $\widetilde{a}_N, f_N \in L_N^2(\mathbb{T}^d)$, *defined via the values on "de-aliased" grid* $\{x_j\}_{j \in \mathcal{J}_{2N}}$ *(cp. equation 27):*

$$\left.\begin{cases} \widetilde{a}_N := \dot{P}_N \mathcal{I}_{2N}(a - 1), \\ f_N := \dot{P}_N \mathcal{I}_{2N} f \end{cases}\right\} \in L_N^2(\mathbb{T}^d).$$

2. *Set*

$$K := \left\lceil \frac{\log\left(\lambda^{-1} N^{-k}\right)}{\log\left(1 - \lambda/2\right)} \right\rceil.$$

19

3. *Set $u_N^0 := 0 \in L_N^2(\mathbb{T}^d)$.*

4. *For $k = 1, 2, \dots, K$: compute*

$$u_N^k := \dot{P}_N (-\Delta)^{-1} \nabla \cdot \left(\widetilde{a}_N \nabla u_N^{k-1} \right) + (-\Delta)^{-1} f_N.$$

5. *Set $u_N := u_N^K$.*

We have the following theorem on the convergence of the Algorithm 23,

Theorem 24 *Let u be the unique solution of (24). Let $k \in \mathbb{N}$ be given, with $k > d/2+1$, and assume that the coefficient $a \in H^s(\mathbb{T}^d)$ for $s > d/2+k$ satisfies the coercivity condition (25). Assume furthermore that $f \in H^{k-1}(\mathbb{T}^d)$. Then there exists $N_0 = N_0(s, d, \|a\|_{H^s}, \lambda) \in \mathbb{N}$, such that for any integer $N \geq N_0$, there exists a unique solution of the discretized elliptic equation (26), and there exists a constant $C = C(s, d, \lambda, \|a\|_{H^s}, \|f\|_{H^{k-1}}) > 0$, such that*

$$\|u - u_N\|_{H^1} \leq C N^{-k},$$

where u_N is the output of Algorithm 23.

The proof of Theorem 24 relies only on standard techniques of numerical analysis and is presented in Appendix E.1.

3.1.2 Ψ-FNO APPROXIMATION OF THE DARCY EQUATIONS

We will emulate the pseudo-spectral fixed point Algorithm 23 by a Ψ-FNO, allowing us to derive approximation bounds. We consider the following setting,

Setting 25 *Let $s \geq d/2 + k + \delta$ for some $k \in \mathbb{N}$, $\delta \in (0, 1)$ be a given Sobolev regularity parameter, and let $\lambda \in (0, 1)$ be a (coercivity) constant. Denote by $\mathcal{A}_\lambda^s(\mathbb{T}^d) \subset H^s(\mathbb{T}^d)$ the set of λ-coercive coefficients $a \in H^s(\mathbb{T}^d)$ of the form $a = 1 + \widetilde{a}$, and such that*

$$\|a\|_{H^s} \leq \lambda^{-1}, \quad \|\widetilde{a}\|_{L^\infty} \leq C \|\widetilde{a}\|_{H^{d/2+\delta}} \leq 1 - \lambda.$$

Here $C > 0$ denotes the norm of the embedding $H^{d/2+\delta}(\mathbb{T}^d) \hookrightarrow L^\infty(\mathbb{T}^d)$. We consider the operator $\mathcal{G} : \mathcal{A}_\lambda^s(\mathbb{T}^d) \to \dot{H}^1(\mathbb{T}^d)$, $a \mapsto u$, where u solves the Darcy equation

$$-\nabla \cdot (a \nabla u) = f, \quad \fint_{\mathbb{T}^d} u(x)\, dx = 0,$$

on the periodic torus $\mathbb{T}^d$, with right-hand side $f \in \dot{H}^{k-1}$.

With this setting in place, we can now state our main FNO approximation theorem for the solution operator $\mathcal{G} : \mathcal{A}_\lambda^s(\mathbb{T}^d) \to H^1(\mathbb{T}^d)$ of the Darcy problem:

Theorem 26 *Assume Setting 25, and assume that the activation function σ is non-linear and $\sigma \in C^3(\mathbb{R})$. Then there exists a constant $C = C(s, \lambda, d) > 0$, such that for any $N \in \mathbb{N}$, there exists a Ψ-FNO $\mathcal{N} : \mathcal{A}_\lambda^s(\mathbb{T}^d) \to H^1(\mathbb{T}^d)$, such that*

$$\sup_{a \in \mathcal{A}_\lambda^s} \|\mathcal{G}(a) - \mathcal{N}(a)\|_{H^1(\mathbb{T}^d)} \leq C N^{-k}$$

and

$$\mathrm{depth}(\mathcal{N}) \leq C\log(N), \quad \mathrm{lift}(\mathcal{N}) = d_v \leq C.$$

In particular, we have

$$\mathrm{size}(\mathcal{N}) \lesssim N^d \log(N).$$

The proof of this theorem, presented in Appendix E.3, relies crucially on the efficient approximation of quadratic nonlinearities by shallow neural networks with smooth activation functions, see Lemma 50 in Appendix E for the precise statement and proof.

Remark 27 *To achieve a FNO approximation error of order ϵ for the Darcy problem, Theorem 26 shows that a Ψ-FNO $\mathcal{N}$ with*

$$\mathrm{size}(\mathcal{N}) \lesssim \left(\frac{1}{\epsilon}\right)^{d/k} \log(\epsilon^{-1}), \quad \mathrm{depth}(\mathcal{N}) \lesssim \log(\epsilon^{-1}), \tag{29}$$

is sufficient. Furthermore, the lifting dimension d_v can be kept uniformly bounded, independently of ϵ. In particular, for $k > d$, the required total size of the Ψ-FNO scales sub log-linearly in the approximation accuracy ϵ, indicating that Ψ-FNOs may provide a very efficient approximation to the solution operator of the Darcy problem, in this case.

As a concrete example for illustrating the approximation of Darcy equations by FNOs, we consider the following example.

Example 1 *A possible model for coefficients a with a typical length scale $\ell > 0$, is to assume an expansion of $a = a(x; Y)$ in terms of random variables $Y = (Y_1, Y_2, \dots) \in [-1,1]^{\mathbb{N}}$ (not necessarily i.i.d.), of the following form, similar to the ansatz in (Schwab and Zech, 2019):*

$$a(x;Y) = 1 + \sum_{k \in \mathbb{Z}^d \setminus \{0\}} b_k Y_k e^{i\langle k,x\rangle},$$

where we assume that the coefficients b_k satisfy a decay condition of the form $|b_k| \leq C_b \exp(-\ell|k|)$ for a constant $C_b > 0$. We also assume that $a(x; Y)$ satisfy the coercivity condition

$$\|a(\,\cdot\,;Y) - 1\|_{L^\infty} \leq C\|a(\,\cdot\,;Y) - 1\|_{H^{d/2+\delta}(\mathbb{T}^d)} \leq 1 - \lambda,$$

for some $\delta, \lambda \in (0,1)$, and that the source term $f \in C^\infty(\mathbb{T}^d)$. Under these assumptions, we have $a \in H^s$ for any $s > 0$, and hence it follows from Theorem 26 that for any $r \in \mathbb{N}$, there exists a constant $C = C(r, \ell, \|f\|_{C^k}, \delta)$, with the following property (cp. also Remark 27): For any $\epsilon > 0$, there exists a Ψ-FNO $\mathcal{N}$, such that

$$\sup_{Y \in [-1,1]^{\mathbb{N}}} \|\mathcal{G}(a(\,\cdot\,,Y)) - \mathcal{N}(a(\,\cdot\,,Y))\|_{H^1} < \epsilon,$$

and

$$\mathrm{size}(\mathcal{N}) \leq C\epsilon^{-1/r}, \quad \mathrm{depth}(\mathcal{N}) \leq C\log(\epsilon^{-1}). \tag{30}$$

The complexity bounds (29) and (30) suggest that the size of a Ψ-FNO approximating the operator $\mathcal{G}$ stemming from the Darcy equations, scales only sub-linearly (or even better) in the desired accuracy ϵ. This should be contrasted with the fact that the size of a Ψ-FNO for a generic Lipschitz operator grows super-exponentially in the desired error (20). Thus, we are able to show that a Ψ-FNO can approximate this PDE solution operator much more efficiently than it might a generic infinite-dimensional operator.

3.2 Incompressible Navier-Stokes Equations

The motion of a viscous, incompressible Newtonian fluid is modeled by the incompressible Navier-Stokes equations,

$$\begin{cases} \partial_t u + u \cdot \nabla u + \nabla p = \nu \Delta u, \\ \operatorname{div}(u) = 0, \ u(t = 0) = u_0, \end{cases} \tag{31}$$

Here, $u \in \mathbb{R}^d$ is the fluid velocity and $p \in \mathbb{R}$ is the pressure of the fluid, acting as a Lagrange multiplier to enforce the divergence-free constraint $\operatorname{div}(u) = 0$. The initial fluid velocity is denoted by u_0. For simplicity, we assume periodic boundary conditions in the domain $\mathbb{T}^d$. The viscosity is denoted by $\nu \geq 0$ and we would like to state that the subsequent analysis also applies for $\nu = 0$, where (31) reduces to the incompressible Euler equations modeling an ideal fluid. For definiteness, we recall the following well-known theorem for the well-posedness of the Navier-Stokes equations (31),

Theorem 28 (see e.g. Thm 3.4 of Majda and Bertozzi (2001)) *Let $r > d/2 + 2$. For any $u_0 \in H^r$, there exists $T > 0$ and a unique classical solution of the Navier-Stokes equations (31), such that $u \in C([0, T]; H^r) \cap C^1([0, T]; H^{r-2})$ with $u(t = 0) = u_0$.*

It is well known that in two space dimensions $d = 2$, the time interval for existence of solutions $[0, T]$ can be extended to any finite T as long as $u_0 \in H^r$, whereas the corresponding finite-time well-posedness result for three space-dimensions is an outstanding open problem (Majda and Bertozzi, 2001).

We recall that if the initial data u_0 of (31) belongs to $\dot{L}^2(\mathbb{T}^d; \mathbb{R}^d)$, i.e. if

$$\int_{\mathbb{T}^d} u_0(x) \, dx = 0,$$

then we also have that the corresponding solution $u(x, t) \in \dot{L}^2(\mathbb{T}^d; \mathbb{R}^d)$ (reflecting momentum conservation). Next, we introduce the **Leray-projection operator** $\mathbb{P} : L^2(\mathbb{T}^d; \mathbb{R}^d) \to \dot{L}^2(\mathbb{T}^d; \operatorname{div})$, as the L^2-orthogonal projection onto the subspace $\dot{L}^2(\mathbb{T}^d; \operatorname{div}) \subset \dot{L}^2(\mathbb{T}^d; \mathbb{R}^d)$, consisting of divergence-free vector fields; i.e. we have $u \in \dot{L}^2(\mathbb{T}^d; \operatorname{div})$ if, and only if, $u \in \dot{L}^2(\mathbb{T}^d; \mathbb{R}^d)$ and

$$\int_{\mathbb{T}^d} u(x) \cdot \nabla \phi(x) \, dx = 0, \quad \forall \phi \in C^\infty(\mathbb{T}^d).$$

In terms of Fourier series, the Leray projection $\mathbb{P} : \dot{L}^2(\mathbb{T}^d; \mathbb{R}^d) \to \dot{L}^2(\mathbb{T}^d; \operatorname{div})$ is explicitly given by

$$\mathbb{P}\left(\sum_{k \in \mathbb{Z}^d} \widehat{u}_k e^{i\langle k, x \rangle} \right) = \sum_{k \in \mathbb{Z}^d \setminus \{0\}} \left(1 - \frac{k \otimes k}{|k|^2} \right) \widehat{u}_k e^{i\langle k, x \rangle}. \tag{32}$$

In terms of the Leray projection $\mathbb{P}$, we can now equivalently write the incompressible Navier-Stokes equations (31) as the following equation on the Hilbert space $\dot{L}^2(\mathbb{T}^d; \operatorname{div})$ as,

$$\begin{cases} \partial_t u = -\mathbb{P}\left(u \cdot \nabla u \right) + \nu \Delta u, \\ \qquad\qquad u(t = 0) = u_0. \end{cases} \tag{33}$$

Given this background, our main objective in this section is to construct a Ψ-FNO that will approximate the operator $\mathcal{G}$ which maps the initial data u_0 to the solution $u(\cdot, T)$ (at the final time T) of the incompressible Navier-Stokes equations (31), (33). To this end, we will follow the general program outlined at the beginning of this section and introduce a suitable pseudo-spectral method for approximating the Navier-Stokes equations. Then, we construct a Ψ-FNO that can efficiently emulate this pseudo-spectral method.

3.2.1 A FULLY-DISCRETE Ψ-SPECTRAL APPROXIMATION OF THE NAVIER-STOKES EQUATIONS (31)

The form of the Leray-projected Navier-Stokes equations (33) naturally suggests the following fully-discrete approximation of (31):

$$\begin{cases} \dfrac{u_N^{n+1} - u_N^n}{\tau} + \mathbb{P}_N\left(u_N^n \cdot \nabla u_N^{n+1}\right) = \nu \Delta u_N^{n+1}, \\ \qquad\qquad u_N^0 = \mathcal{I}_N u(t=0). \end{cases} \tag{34}$$

Here, we fix $N \in \mathbb{N}$ and introduce the space, $\dot{L}_N^2(\mathbb{T}^d; \mathrm{div}) := \dot{L}^2(\mathbb{T}^d; \mathrm{div}) \cap \dot{L}_N^2(\mathbb{T}^d; \mathbb{R}^d)$. We fix a time-step $\tau > 0$ and let $u_N^n \in \dot{L}_N^2(\mathbb{T}^d; \mathrm{div})$, for all $n = 0, \ldots, n_T$, with n_T such that $\tau n_T = T$. Moreover, we use the following *finite-dimensional* Leray-Fourier projection operator $\mathbb{P}_N : \dot{L}^2(\mathbb{T}^d; \mathbb{R}^d) \to \dot{L}_N^2(\mathbb{T}^d; \mathrm{div})$ in analogy with (32):

$$\mathbb{P}_N\left(\sum_{k \in \mathbb{Z}^d} \widehat{u}_k e^{i\langle k, x\rangle}\right) := \sum_{0 < |k|_\infty \leq N} \left(1 - \frac{k \otimes k}{|k|^2}\right) \widehat{u}_k e^{i\langle k, x\rangle}, \tag{35}$$

to complete the description of the scheme (34).

We observe that the scheme (34) is *implicit* i.e., at each time step n, one has to solve an operator equation to compute the velocity field u_N^{n+1} at the next time step. Thus, one needs to show the solvability of this operator equation in order to ensure that the scheme (34) is *well-defined*. Under the following CFL condition for choosing a small enough time step τ,

$$\tau \|u_N^n\|_{L^\infty} N \leq \frac{1}{2}, \tag{36}$$

we prove in Appendix F.1 that the scheme (34) is well-defined.

Next, in practice, one has to numerically approximate the solutions of the implicit equation (34) for evaluating the velocity field u_N^{n+1}, at the next time-step. We choose to do so by recasting the solution of the implicit equation (34) to finding a *fixed point* for the mapping,

$$w_N \mapsto F(w_N) := (1 - \nu\tau\Delta)^{-1} u_N^n - \tau(1 - \nu\tau\Delta)^{-1}\mathbb{P}_N(u_N^n \cdot \nabla w_N). \tag{37}$$

In Appendix F.1 Lemma 54, we show that a standard Picard-type iteration converges to a fixed point for the map (37). This suggests the following numerical algorithm for approximating strong solutions of the incompressible Navier-Stokes equations (31),

Algorithm 29 (Pseudo-spectral approximation of the Navier-Stokes equations (31))

Input: $U > 0$, $N \in \mathbb{N}$, $T > 0$, a time-step $\tau > 0$, such that $n_T = T/\tau \in \mathbb{N}$, and $\tau U N^{d/2+1} \le \frac{1}{2e}$, initial data $u_N^0 \in L_N^2(\mathbb{T}^d; \mathrm{div})$, such that $\|u_N^0\|_{L^2} \le U$.

Output: $u_N^{n_T} \in L_N^2(\mathbb{T}^d; \mathrm{div})$ an approximation of the solution $u_N^{n_T} \approx u(t = T)$ of (31) at time $t = T$.

1. Set

$$\kappa_0 := \left\lceil \frac{\log\left(T^2/\tau^2\right)}{\log(2)} \right\rceil \in \mathbb{N}.$$

2. For $n = 0, \dots, n_T - 1$:

   (a) Set $w_N^{n,0} := 0$,

   (b) For $k = 1, \dots, \kappa_0$: Compute

   $$w_N^{n,k} := (1 - \nu\tau\Delta)^{-1} u_N^n - \tau(1 - \nu\tau\Delta)^{-1} \mathbb{P}_N \left(u_N^n \cdot \nabla w_N^{n,k-1} \right),$$

   (c) Set $u_N^{n+1} := w_N^{n,\kappa_0}$,

The convergence of the Algorithm 29, together with a convergence rate, to the strong solution of the Navier-Stokes equations is summarized in the following theorem,

Theorem 30 Let $U, T > 0$. Consider the Navier-Stokes equations on $\mathbb{T}^d$, for $d \ge 2$. Assume that $r \ge d/2 + 2$, and let $u \in C([0,T]; H^r) \cap C^1([0,T]; H^{r-2})$ be a solution of the Navier-Stokes equations (31), such that $\|u\|_{L^2} \le U$. Choose a time-step τ, such that $\tau U N^{d/2+1} \le (2e)^{-1}$. There exists a constant

$$C = C(T, d, r, \|u\|_{C_t(H_x^r)}, \|u\|_{C_t^1(H_x^{r-2})}) > 0,$$

such that with $u_N^0 := \mathcal{I}_N u(0)$, and for the sequence $u_N^1, \dots, u_N^{n_T} \in L_N^2(\mathbb{T}^d; \mathrm{div})$ generated by Algorithm 29, we have

$$\max_{n=0,\dots,n_T} \|u_N^n - u(t^n)\|_{L^2} \le C \left(\tau + N^{-r} \right),$$

where $n_T \tau = T$. In particular, choosing $\tau \sim N^{-r}$, we have

$$\max_{n=0,\dots,n_T} \|u_N^n - u(t^n)\|_{L^2} \le C N^{-r},$$

with $n_T \sim N^r$ (and enlarging the constant $C > 0$ by a constant factor).

The proof of this theorem relies on several techniques from numerical analysis and is presented in detail in Appendix F.2.

3.2.2 APPROXIMATION OF ALGORITHM 29 BY Ψ-FNOS

Next, we are going to construct a Ψ-FNO of the form (14), which can efficiently emulate the pseudo-spectral Algorithm 29. To this end, we have the following result (proved in Appendix F.3) on the efficient approximation of the non-linear term in the Navier-Stokes equations by FNOs,

Lemma 31 *Assume that the activation function $\sigma \in C^3$ is three times continuously differentiable and non-linear. There exists a constant $C > 0$, such that for any $N \in \mathbb{N}$, and for any $\epsilon, B > 0$, there exists a Ψ-FNO $\mathcal{N} : L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d) \times L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d) \to L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d)$, with*

$$\mathrm{depth}(\mathcal{N}), \ \mathrm{lift}(\mathcal{N}) = d_v \leq C,$$

such that we have

$$\|\mathbb{P}_N (u_N \cdot \nabla w_N) - \mathcal{N}(u_N, w_N)\|_{L^2_N} \leq \epsilon,$$

for all trigonometric polynomials $u_N, w_N \in L^2_N(\mathbb{T}^d; \mathbb{R}^d) \subset L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d)$ of degree $|k|_\infty \leq N$, satisfying the bound $\|u_N\|_{L^2}, \|w_N\|_{L^2} \leq B$.

Thus, from the preceding Lemma, we have that the nonlinearities in Algorithm 29 can be efficiently approximated by Ψ-FNOs. This paves the way for the following theorem on the emulation of the pseudo-spectral Algorithm 29 by Ψ-FNOs,

Theorem 32 *Let $U, T > 0$ and viscosity $\nu \geq 0$. Consider the Navier-Stokes equations on $\mathbb{T}^d$, for $d \geq 2$. Assume that $r \geq d/2 + 2$, and let $\mathcal{V} \subset C([0, T]; H^r) \cap C^1([0, T]; H^{r-2})$ be a set of solutions of the Navier-Stokes equations (31), such that $\sup_{u \in \mathcal{V}} \|u\|_{L^2} \leq U$, and*

$$\overline{U} := \sup_{u \in \mathcal{V}} \left\{ \|u\|_{C_t(H^r_x)} + \|u\|_{C^1_t(H^{r-2}_x)} \right\} < \infty.$$

For $t \in [0, T]$, denote $\mathcal{V}_t := \{u(t) \, | \, u \in \mathcal{V}\}$. Let $\mathcal{G} : \mathcal{V}_0 \to \mathcal{V}_T$ denote the solution operator of (31), mapping initial data $u_0 = u(t = 0)$, to the solution $u(T)$ at $t = T$ of the incompressible Navier-Stokes equations. There exists a constant

$$C = C(d, r, U, \overline{U}, T) > 0,$$

such that for $N \in \mathbb{N}$ there exists a Ψ-FNO $\mathcal{N} : L^2_N(\mathbb{T}^d; \mathbb{R}^d) \to L^2_N(\mathbb{T}^d; \mathbb{R}^d)$, such that

$$\sup_{u \in \mathcal{V}_0} \|\mathcal{G}(u) - \mathcal{N}(u)\|_{L^2} \leq C N^{-r},$$

and such that

$$\mathrm{depth}(\mathcal{N}) \leq C N^r \log(N), \quad \mathrm{lift}(\mathcal{N}) = d_v \leq C.$$

The proof of this theorem is provided in Appendix F.3.

Remark 33 *It is straightforward to observe from Theorem 32 that the size of a Ψ-FNO to achieve a desired error tolerance of $\epsilon > 0$, scales (neglecting log-terms) as*

$$\mathrm{size}(\mathcal{N}) \leq C \epsilon^{-\left(1 + \frac{d}{r}\right)}, \tag{38}$$

Given that we need $r \geq d/2 + 2$, we observe from (38) that the size of the Ψ-FNO approximating the initial data to solution operator $\mathcal{G}$ for the Navier-Stokes equations (31) scales at most sub-quadratically with respect to the error tolerance ϵ for the physically relevant values $d = 2, 3$. This polynomial scaling should be compared with the super-exponential growth (see Remark 22) of the size of FNOs in approximating a generic Lipschitz-continuous operator. Thus, we are able to demonstrate that Ψ-FNOs can approximate the solutions of Navier-Stokes equations far more efficiently than what the universal approximation theorem 15 suggests.

Remark 34 *From the convergence theorem 30, we observe that the underlying scheme (34) is* first-order *in time. This low accuracy of the scheme necessitates a large number of time steps and affects the overall complexity. We describe a second-order accurate time discretized version of the pseudo-spectral method for approximating the Navier-Stokes equations (31) in Appendix F.4 and in complete analogy with Theorem 32 (see Theorem 58 in Appendix F.4), we can construct a Ψ-FNO to emulate this second-order in time pseudo-spectral scheme, resulting in a Ψ-FNO of*

$$\text{size}(\mathcal{N}) \leq C\epsilon^{-\left(\frac{1}{2} + \frac{d}{r}\right)}, \tag{39}$$

to obtain a desired accuracy of ϵ. Thus, we can obtain a more efficient approximation of the underlying operator than Ψ-FNO emulating the first-order time scheme (34). In particular for $r \geq 2d$, we obtain that the size of a Ψ-FNO only grows sub-linearly in terms of the desired accuracy, making this FNO approximation comparable in complexity to the FNO approximation of the Darcy equation (see equation 29).

4. Comparison of FNOs with DeepOnets

In this section, we will compare Ψ-FNOs with another operator learning framework, namely DeepOnets of Chen and Chen (1995); Lu et al. (2021), defined as follows:

Definition 35 *Fix $m, p \in \mathbb{N}$. A **DeepOnet** $\mathcal{D}$ with output dimension p and **sensor points** $x_1, \ldots, x_m \in \mathbb{T}^d$ is a mapping $\mathcal{D} : C(\mathbb{T}^d; \mathbb{R}^{d_a}) \to C(\mathbb{T}^d; \mathbb{R}^{d_v})$ of the form*

$$\mathcal{D}(a)(x) = \sum_{k=1}^{p} \beta_k(a(x_1), \ldots, a(x_p))\tau_k(x),$$

*where $\beta : \mathbb{R}^{m \times d_a} \to \mathbb{R}^{p \times d_u}$, $\alpha \mapsto \beta(\alpha) = (\beta_1(\alpha), \ldots, \beta_p(\alpha))$, and $\tau : \mathbb{R}^d \mapsto \mathbb{R}^p$, $x \mapsto \tau(x) = (\tau_1(x), \ldots, \tau_p(x))$ are (ordinary) neural networks. We refer to β and τ as the **branch** and **trunk** nets, respectively.*

In the following theorem (proved in Appendix G), we show that a Ψ-FNO can naturally be viewed as a DeepOnet with a specific choice of the branch and trunk-nets; where the trunk net is fixed to represent a trigonometric basis, and the branch net is constrained by a specific choice of hidden layer architecture, which provides a more parsimonious representation compared to a canonical DeepOnet implementation based on dense layers.

Theorem 36 (DeepOnet approximation of Ψ-FNOs) *Let* $\mathcal{N} : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_u})$ *be a Ψ-FNO, and fix* $B > 0$*. For any* $\epsilon > 0$*, there exists* $p \in \mathbb{N}$ *and a DeepOnet* (β, τ)*, with branch net* β*, trunk net* τ*, and sensor points* $\{x_j\}_{j \in \mathcal{J}_N}$*, such that*

$$\sup_{\|a\|_{L^\infty} \leq B} \sup_{y \in \mathbb{T}^d} \left| \mathcal{N}(a)(y) - \sum_{k=1}^p \beta_k(a) \tau_k(y) \right| \leq \epsilon,$$

where the first supremum is taken over all $a \in C(\mathbb{T}^d)$*, such that* $\|a\|_{L^\infty} \leq B$*. Furthermore, we have*

$$\text{width}(\beta) \sim \text{lift}(\mathcal{N}) N^d = d_v N^d, \quad \text{depth}(\beta) = \text{depth}(\mathcal{N}),$$

and the trunk net τ *defines a mapping* $\tau : \mathbb{R}^d \to \mathbb{R}^{\mathcal{K}_N}$*, which approximates an (arbitrary) orthonormal trigonometric basis* $\{\mathbf{e}_k\}_{k \in \mathcal{K}_N}$ *with* $\text{span}\{\mathbf{e}_k\}_{k \in \mathcal{K}_N} = \text{span}\{e^{i\langle k, x\rangle}\}_{k \in \mathcal{K}_N}$*, such that*

$$\max_{k \in \mathcal{K}_N} \|\mathbf{e}_k - \tau_k\|_{L^\infty} \leq \epsilon / \overline{B},$$

where

$$\overline{B} := (2N+1)^d \left(\sup_{\|a\|_{L^\infty} \leq B} \|\mathcal{N}(a)\|_{L^2} \right).$$

Remark 37 *We note that due to the particular architecture of* $\Psi-$*Fourier neural operators, the total size of a Ψ-FNO* $\mathcal{N}$ *is upper bounded by*

$$\text{size}(\mathcal{N}) \lesssim \text{depth}(\mathcal{N}) d_v^2 N^d.$$

In contrast, since β *and* τ *in the DeepOnet approximation of Theorem 36 are conventional (dense) deep neural networks, the total number of degrees of freedom (i.e. the total number of weights and biases) is given, in general, by*

$$\text{size}(\beta) \sim \text{depth}(\beta)\text{width}(\beta)^2 \sim \text{depth}(\mathcal{N}) d_v^2 N^{2d},$$

and $\text{size}(\tau) \sim \text{depth}(\tau)\text{width}(\tau)^2$*. So that in particular, comparing the sizes of the Ψ-FNO and the corresponding (fully connected) DeepOnet approximation, we have*

$$\frac{\text{size}(\mathcal{N})}{\text{size}(\beta, \tau)} \lesssim N^{-d}.$$

In particular, this indicates that the Ψ-FNO provides a more parsimonious approximation of the underlying operator than the DeepOnet "emulator" constructed in Theorem 36, with $\text{size}(\mathcal{N}) \ll \text{size}(\beta, \tau)$*.*

Before summarizing and discussing the main results of this work in the next section, we would like to make a final remark on the different definition of "size" employed in the present work, compared to related work such as Lanthaler et al. (2021):

Remark 38 *In the approximation theory of neural networks, especially in the context of fully connected neural networks, it is customary to define the "size" of a neural network as the number of* non-zero *weights needed to approximate a given function (or operator) to*

accuracy ϵ. As a consequence, approximation estimates often depend crucially on considerable additional sparsity, *which may be present in a particular construction in the proof of such results. We would like to emphasize that due to the special (constrained) architecture of Ψ-FNOs considered in the present work, our results do not depend on such additional sparsity properties. Instead, the size of our neural networks is measured in terms of the* total number *of weights and biases (be they zero or not). This is another indication of the potential parsimony of FNOs over other architectures, such as DeepONets.*

5. Summary and Discussion

Many learning tasks, particularly, but not exclusively, in scientific computing, are naturally formulated as learning operators mapping one infinite-dimensional space to another. Neural operators have recently been proposed as a framework for operator learning. A particular form, the so-called Fourier Neural Operators (FNOs) (6), have been shown to be efficient in approximating a wide variety of operators that arise in PDEs (Li et al., 2021). Our main aim in this paper was to analyze FNOs and Ψ-FNOs (14), which is a concrete computational realization of FNOs. To this end, we have presented the following results,

- We showed in Theorem 5 and Theorem 15 that FNOs (resp. Ψ-FNOs) are *universal*, i.e., they can approximate any continuous operator to desired accuracy. Our proof relies heavily on the ability of FNOs to approximate the Fourier transform and its inverse, together with the neural network approximation of the finite-dimensional Fourier conjugate operator (11). Thus, FNOs have the same universal approximation property as canonical neural networks for finite-dimensional functions and DeepOnets for operators (Lanthaler et al., 2021). This universality result paves the way for the widespread use of FNOs in the context of operator learning.

- However as stated in Remark 22, in the worst case, the size of a FNO can grow *super-exponentially* in terms of the desired error for approximating a general Lipschitz continuous operator. This might inhibit the use of FNOs. On the other hand, we argue in the beginning of Section 3 that Ψ-FNOs, which are a concrete computational realization of FNOs, can approximate the nonlinearities and differential operators that define PDEs, very efficiently. Hence, one can think of Ψ-FNOs as a new form of pseudo-spectral methods for PDEs, which in practice are adapted to, and optimized based on the given training data. Thus, one can expect that Ψ-FNOs can approximate PDEs efficiently.

- We consider two widely used prototypical PDEs, namely the elliptic PDE (24) that arises in a stationary Darcy flow and the well-known incompressible Navier-Stokes equations for fluid dynamics. For both these PDEs, we prove rigorously that there exists a Ψ-FNO which can approximate the underlying nonlinear operators efficiently, as we can show that the size of the Ψ-FNO only needs to grow polynomially in terms of the error. In fact, we show that the size grows sub-linearly in terms of the error. Thus, FNOs can approximate these widely used PDEs efficiently, corroborating the empirical results presented in Li et al. (2021).

Hence, our analysis provides very strong theoretical evidence that FNOs are an effective framework for operator learning. Moreover, we also compare FNOs to an alternative operator learning framework, that of DeepOnets (Lu et al., 2021) and show that Ψ-FNOs can be thought of a special case of DeepOnets with a trunk net approximating trigonometric functions and sensor points being equi-distributed Cartesian grid points. Given its special architecture, we argue that a Ψ-FNO can allow for a more parsimonious representation of operators than a DeepOnet, enabling a cheaper approximation of certain operators.

The comparison with DeepOnets also brings out some obvious limitations of FNOs. In particular, FNOs are efficient on rectangular domains as the Ψ-FNO can be evaluated efficiently with FFT. Although one can use FNOs for operators defined on arbitrary domains using suitable extension and restriction operators (see Theorem 9), it is unclear if these operators can be realized computationally in an efficient manner. Moreover, FNOs fix the trigonometric basis as the trunk net in the underlying DeepOnet (see Theorem 36). On the other hand, a general DeepOnet can learn trunk nets from the data during training, allowing the possibility of learning a more suitable representation from data. These considerations call for a more thorough computational comparison between DeepOnets and FNOs, and possibly other operator learning frameworks such as the one from Bhattacharya et al. (2021).

Similarly, extending the analysis of this article to other neural operators can be readily envisaged. The use of FNOs for more general operators, particularly those arising in non-scientific computing settings, such as images, text and speech, also needs to be investigated.

Appendix A. Glossary of Mathematical Notation

Symbol	Description	Page
σ	activation function	
$\mathbb{T}^d$	periodic torus, identified with $[0, 2\pi]^d$	
d	spatial dimension of domain	p. 4
d_a, d_u, d_v	number of components of input, output and lifting	p. 4
$\mathcal{A}(D; \mathbb{R}^{d_a})$	input function space	p. 4
$\mathcal{U}(D; \mathbb{R}^{d_u})$	output function space	p. 4
$\mathcal{F}, \mathcal{F}^{-1}$	Fourier transform and inverse Fourier transform	p. 31
$\mathcal{F}_N, \mathcal{F}_N^{-1}$	discrete Fourier transform and inverse	p. 33
$\{x_j\}_{j \in \mathcal{J}_N}$	regular periodic grid, $x_j = 2\pi j/(2N+1)$	p. 33
$\mathcal{J}_N$	grid point indices, $\mathcal{J}_N = \{0, \ldots, 2N\}^d$	p. 33
$\mathcal{K}_N$	Fourier wavenumbers $\mathcal{K}_N = \{ k \in \mathbb{Z}^d \,\big|\, |k|_\infty \le N \}$	
$\mathcal{R}$	lifting operator	p. 4
$\mathcal{L}_\ell$	neural operator layer	p. 6
$\mathcal{Q}$	projection operator	p. 4
$\mathcal{F}$-layer	linear, non-local layer; $v(x) \mapsto \mathcal{F}^{-1}(P \mathcal{F} v)(x)$	p. 14
σ-layer	non-linear, local layer; $v(x) \mapsto \sigma(W v(x) + b(x))$	p. 14

Banach spaces

L^2	Space of square-integrable functions	
$\dot{L}^2$	$\dot{L}^2 \subset L^2$ square-integrable functions with zero mean	p. 31
L_N^2	$L_N^2 \subset L^2$ trigonometric polynomials of degree $\le N$	p. 31
$\dot{L}_N^2$	$\dot{L}_N^2 = \dot{L}^2 \cap L_N^2$ trigonometric polynomials with zero mean	p. 31
H^s	Sobolev space of smoothness s, with norm $\|\cdot\|_{H^s}$	p. 31
$\dot{H}^s$	Sobolev space with zero mean, with norm $\|\cdot\|_{\dot{H}^s}$	p. 31

Projection operators

P_N	L^2-orthogonal Fourier projection $P_N : L^2 \to L_N^2$	p. 31
$\dot{P}_N$	Fourier projection $\dot{P}_N : L^2 \to \dot{L}_N^2$ with zero mean	p. 32
$\mathcal{I}_N$	Pseudo-spectral Fourier projection, i.e. trigonometric interpolation on regular grid $\{x_j\}_{j \in \mathcal{J}_N}$	p. 32
$\mathbb{P}$	Leray projection onto divergence-free vector fields	p. 22
$\mathbb{P}_N$	Leray projection followed by projection P_N; $\mathbb{P}_N = P_N \circ \mathbb{P}$	p. 23

Appendix B. Notation and Technical Preliminaries

In this section, we introduce frequently used notation in the article main text and recall some essential facts about Fourier analysis.

In the main text, our focus is functions defined on the periodic torus $\mathbb{T}^d$, identified as $\mathbb{T}^d = [0, 2\pi]^d$. Following standard practice, we denote by $L^2(\mathbb{T}^d)$ the space of square-integrable functions. For any such function $v \in L^2(\mathbb{T}^d)$, we can define the *Fourier Transform*

as,

$$\mathcal{F}(v)(k) := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} v(x) e^{-i\langle k, x\rangle} \, dx, \quad \forall \, k \in \mathbb{Z}^d. \tag{40}$$

For any $k \in \mathbb{Z}^d$, the k-th Fourier coefficient of v is denoted by $\widehat{v}_k = \mathcal{F}(v)(k)$.

Given a set of Fourier coefficients $\{\widehat{v}_k\}_{k \in \mathbb{Z}^d}$, the *inverse Fourier Transform* is defined as,

$$\mathcal{F}^{-1}(\widehat{v})(x) := \sum_{k \in \mathbb{Z}^d} \widehat{v}_k e^{i\langle k, x\rangle}, \quad \forall \, x \in \mathbb{T}^d. \tag{41}$$

Using the Fourier transform (40) and for $s \geq 0$, one can denote by $H^s(\mathbb{T}^d)$ the **Sobolev space** of functions $v \in L^2(\mathbb{T}^d)$, with Fourier coefficients $\{\widehat{v}_k\}_{k \in \mathbb{Z}^d}$, having a finite H^s-norm:

$$\|v\|_{H^s}^2 := \frac{(2\pi)^d}{2} \sum_{k \in \mathbb{Z}^d} (1 + |k|^{2s}) |\widehat{v}_k|^2 < \infty. \tag{42}$$

Note that with this definition, we have from Parseval's identity, that $\|v\|_{H^0} = \|v\|_{L^2}$, so that $H^0(\mathbb{T}^d) = L^2(\mathbb{T}^d)$.

We also introduce the corresponding **homogeneous Sobolev spaces** $\dot{H}^s(\mathbb{T}^d)$ (and $\dot{L}^2(\mathbb{T}^d) := \dot{H}^0(\mathbb{T}^d)$), consisting of functions $v(x) \in H^s(\mathbb{T}^d)$ and with zero mean $\fint_{\mathbb{T}^d} v(x) \, dx = \widehat{v}_0 = 0$, and with norm

$$\|v\|_{\dot{H}^s} := \left((2\pi)^d \sum_{k \in \mathbb{Z}^d \setminus \{0\}} |k|^{2s} |\widehat{v}_k|^2 \right)^{1/2}. \tag{43}$$

Given $N \in \mathbb{N}$, throughout this work, we will denote by $L_N^2(\mathbb{T}^d)$, the space of trigonometric polynomials $v_N : \mathbb{T}^d \to \mathbb{R}$, of the form

$$v_N(x) = \sum_{|k|_\infty \leq N} c_k e^{i\langle x, k\rangle}, \tag{44}$$

where the summation is over all $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$ such that

$$|k|_\infty := \max_{i=1,\ldots,d} |k_i| \leq N.$$

The space $L_N^2(\mathbb{T}^d)$ is viewed as a normed vector space with norm $\|\cdot\|_{L^2}$. Similarly, for $s \geq 0$, we denote by $H_N^s(\mathbb{T}^d)$ the normed vector space of trigonometric polynomials v_N of degree $\leq N$, with norm $\|\cdot\|_{H^s}$.

We note that in order to ensure that $v_N(x) \in \mathbb{R}$ is real-valued for all $x \in \mathbb{T}^d$, the coefficients $c_k \in \mathbb{C}$ must satisfy the relations $c_{-k} = \overline{c_k}$ for all $|k|_\infty \leq N$, and where $\overline{c_k}$ denotes the complex conjugate of c_k.

We denote by

$$P_N : L^2(\mathbb{T}^d) \to L_N^2(\mathbb{T}^d), \quad v \mapsto P_N v, \tag{45}$$

the **L^2-orthogonal projection onto $L_N^2(\mathbb{T}^d)$**; or more explicitly,

$$P_N \left(\sum_{k \in \mathbb{Z}^d} c_k e^{i\langle k, x \rangle} \right) = \sum_{|k|_\infty \leq N} c_k e^{i\langle k, x \rangle}, \quad \forall (c_k)_{k \in \mathbb{Z}^d} \in \ell^2(\mathbb{Z}^d).$$

In fact, the mapping P_N defines a projection $H^s(\mathbb{T}^d) \to H_N^s(\mathbb{T}^d)$ for any $s \geq 0$. We have the following spectral approximation estimate: Let $s > 0$ be given. There exists a constant $C = C(s, d) > 0$, such that for any $v \in H^s(\mathbb{T}^d)$, we have

$$\|v - P_N v\|_{H^\varsigma} \leq C N^{-(s-\varsigma)} \|v\|_{H^s}, \quad \text{for any } \varsigma \in [0, s]. \tag{46}$$

We also define a natural projection

$$\dot{P}_N : L^2(\mathbb{T}^d) \to \dot{L}_N^2(\mathbb{T}^d), \tag{47}$$

by removing the mean, i.e. $\dot{P}_N v = P_N v - \fint_{\mathbb{T}^d} v(x)\,dx$, or equivalently:

$$\dot{P}_N \left(\sum_{k \in \mathbb{Z}^d} c_k e^{i\langle k, x \rangle} \right) = \sum_{0 < |k|_\infty \leq N} c_k e^{i\langle k, x \rangle}, \quad \forall (c_k)_{k \in \mathbb{Z}^d} \in \ell^2(\mathbb{Z}^d).$$

Furthermore, we denote by by

$$\mathcal{I}_N : C(\mathbb{T}^d) \mapsto L_N^2(\mathbb{T}^d), \quad u \mapsto \mathcal{I}_N u, \tag{48}$$

the **pseudo-spectral projection** onto $L_N^2(\mathbb{T}^d)$; we recall that the pseudo-spectral projection $\mathcal{I}_N v$ of a continuous function v is defined as the unique trigonometric polynomial $\mathcal{I}_N v \in L_N^2(\mathbb{T}^d)$, such that

$$\mathcal{I}_N v(x_j) = v(x_j), \quad \forall j \in \mathcal{J}_N, \tag{49}$$

where $\{x_j\}_{j \in \mathcal{J}_N}$ denotes the set of all regular grid points $x_j \in \mathbb{Z}^d$ of the form $x_j = 2\pi j/(2N+1) \in \mathbb{T}^d$, $j \in \mathbb{Z}^d$ (cp. equation 50).

We also recall the following embedding theorem for the Sobolev spaces $H^s(\mathbb{T}^d)$:

Theorem 39 (Sobolev embedding) *Let $d \in \mathbb{N}$. For any $s > d/2$, we have a compact embedding $H^s(\mathbb{T}^d) \hookrightarrow C(\mathbb{T}^d)$ into the space of continuous functions. In particular, there exists a constant $C = C(s, d) > 0$, such that*

$$\|v\|_{L^\infty} \leq C\|v\|_{H^s}, \quad \forall v \in H^s(\mathbb{T}^d).$$

The Sobolev embedding theorem implies in particular that the pseudo-spectral projection $\mathcal{I}_N$ is well-defined as an operator $\mathcal{I}_N : H^s(\mathbb{T}^d) \to L_N^2(\mathbb{T}^d)$ for $s > d/2$. In the following theorem, we recall a well-known approximation error estimate for the pseudo-spectral projection $\mathcal{I}_N$:

Theorem 40 (Pseudo-spectral approximation estimate) *Let $d \in \mathbb{N}$. For any $s > d/2$ and $N \in \mathbb{N}$, the spectral interpolation operator $\mathcal{I}_N : H^s(\mathbb{T}^d) \to L^2_N(\mathbb{T}^d)$ is well-defined. Furthermore, there exists a constant $C = C(s, d) > 0$, such that the following approximation error estimate holds*

$$\|(1 - \mathcal{I}_N)v\|_{H^\varsigma} \leq C N^{-(s-\varsigma)} \|v\|_{H^s}, \quad \forall v \in H^s(\mathbb{T}^d),$$

for any $\varsigma \in [0, s]$.

Finally, for $N \in \mathbb{N}$, we fix a regular grid $\{x_j\}_{j \in \mathcal{J}_N}$ of values

$$x_j = \frac{2\pi j}{2N + 1}, \tag{50}$$

where the index $j \in \mathcal{J}_N$ belong to the index set

$$\mathcal{J}_N := \{0, \dots, 2N\}^d. \tag{51}$$

Recall the set of Fourier wave numbers (10) and we define the **discrete Fourier transform** $\mathcal{F}_N : \mathbb{R}^{\mathcal{J}_N} \to \mathbb{C}^{\mathcal{K}_N}$ by

$$\mathcal{F}_N(v)(k) := \frac{1}{(2N + 1)^d} \sum_{j \in \mathcal{J}_N} v_j e^{-2\pi i \langle j, k \rangle / N}, \tag{52}$$

with inverse $\mathcal{F}_N^{-1} : \mathbb{C}^{\mathcal{K}_N} \to \mathbb{R}^{\mathcal{J}_N}$,

$$\mathcal{F}_N^{-1}(\widehat{v})(j) := \sum_{k \in \mathcal{K}_N} \widehat{v}_k e^{2\pi i \langle j, k \rangle / N}. \tag{53}$$

Lemma 41 (Periodic extension operator) *Let $\Omega \subset \mathbb{R}^d$ be a bounded, Lipschitz domain. There exists a continuous, linear operator $\mathcal{E} : W^{m,p}(\Omega) \to W^{m,p}(B)$ for any $m \geq 0$ and $1 \leq p \leq \infty$ where $B \Subset \mathbb{R}^d$ is a hypercube with $\Omega \subset B$ such that, for any $u \in W^{m,p}(\Omega)$,*

1. *$\mathcal{E}(u)|_\Omega = u$,*

2. *$\mathcal{E}(u)$ is periodic on B (including its derivatives).*

Proof By (Stein, 1970, Chapter 6, Theorem 5), there exists a continuous, linear operator $\widetilde{\mathcal{E}} : W^{m,p}(\Omega) \to W^{m,p}(\mathbb{R}^d)$ such that $\widetilde{\mathcal{E}}(u)|_\Omega = u$ for any $u \in W^{m,p}(\Omega)$. Let $\Phi \in C^\infty(\mathbb{R}^d)$ be a mapping whose zeroth level set defines a curve $\partial\Omega'$ such that $\Omega \subset \Omega'$ and $\mathrm{dist}(\partial\Omega', \partial\Omega) > 0$ where Ω' is a bounded domain with boundary $\partial\Omega'$. Furthermore, suppose that the first level set of Φ defines a curve $\partial\Lambda$ such that $\Omega' \subset \Lambda$ and $\mathrm{dist}(\partial\Lambda, \partial\Omega') > 0$ where, again, Λ is a bounded domain with boundary $\partial\Lambda$. For example, we may take $\Phi(x) = x_1^2 + \cdots + x_d^2 - r$ for some $r > 0$ large enough then Ω' is d-ball of radius $\sqrt{r}$ enclosing Ω, and Λ is a d-ball of radius $\sqrt{1+r}$ clearly enclosing Ω'. We will now follow the approach of (Boyd, 2005) and construct a windowing function $\rho \in C_c^\infty(\mathbb{R}^d)$. In particular, we define

$$\rho(x) = H\big(1 - 2\Phi(x)\big), \qquad \forall x \in \mathbb{R}^d$$

where

$$H(t) = \frac{1}{2}\big(1 + S(t)\big), \qquad \forall t \in \mathbb{R}$$

and

$$S(t) = \begin{cases} -1, & t < -1 \\ \mathrm{erf}\left(\frac{t}{\sqrt{1-t^2}}\right), & -1 \le t \le 1 \\ 1, & t > 1 \end{cases}, \qquad \forall t \in \mathbb{R}$$

with erf, the Gauss error function,

$$\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-z^2} \, \mathrm{d}z, \qquad \forall t \in \mathbb{R}.$$

Define $B \Subset \mathbb{R}^d$ to be a large enough d-cube such that $\Lambda \subset B$ and $\mathrm{dist}(\partial\Lambda, \partial B) > 0$. It is easy to check that $\rho \in C_c^\infty(\mathbb{R}^d)$ and we have that $\rho|_\Omega = \rho|_{\Omega'} = 1$ and $\rho|_{B\setminus\overline{\Lambda}} = 0$ with ρ and all of its derivatives vanishing on $\partial\Lambda$. Furthermore,

$$\|\rho\|_{C^m(B)} < \infty$$

for any $m \ge 0$. We can thus define $\mathcal{E} : W^{m,p}(\Omega) \to W^{m,p}(B)$ by

$$\mathcal{E}(u) = \rho\widetilde{\mathcal{E}}(u), \qquad \forall u \in W^{m,p}(\Omega).$$

By Leibniz's rule and the generalized triangle inequality, there exists a constant $C_1 = C_1(d, m, p) > 0$ (independent of p in the case $p = \infty$) such that

$$\|\mathcal{E}(u)\|_{W^{m,p}(B)} \le C_1 \|\rho\|_{C^m(B)} \|\widetilde{\mathcal{E}}(u)\|_{W^{m,p}(B)}.$$

Since $\widetilde{\mathcal{E}}$ is bounded, there is a constant $C_2 > 0$ such that

$$\begin{aligned} \|\mathcal{E}(u)\|_{W^{m,p}(B)} &\le C_1 \|\rho\|_{C^m(B)} \|\widetilde{\mathcal{E}}(u)\|_{W^{m,p}(B)} \\ &\le C_1 \|\rho\|_{C^m(B)} \|\widetilde{\mathcal{E}}(u)\|_{W^{m,p}(\mathbb{R}^d)} \\ &\le C_1 C_2 \|\rho\|_{C^m(B)} \|u\|_{W^{m,p}(\Omega)} \end{aligned}$$

hence $\mathcal{E}$ is a continuous, linear operator. Since $\rho|_\Omega = 1$, we immediately have $\mathcal{E}(u)|_\Omega = \widetilde{\mathcal{E}}(u)|_\Omega = u$. Since $\mathrm{supp}(\partial^\alpha \rho) \subseteq \overline{\Lambda}$ for any multi-index $0 \le |\alpha| \le m$, we conclude that $\partial^\alpha \mathcal{E}(u)|_{B\setminus\overline{\Lambda}} = 0$ hence $\mathcal{E}(u)$ is periodic on B as desired.

∎

Appendix C. Non-Linear Lifting and Projection Operators

Lemma 42 *Assume that the activation function $\sigma \in C^2$ is non-linear and that $\|\sigma\|_{C^2(\mathbb{R})} < \infty$. Let $\mathcal{N} : L^2(D; \mathbb{R}^{d_a}) \to L^2(D; \mathbb{R}^{d_u})$ be a neural operator of the form $\mathcal{N} = \mathcal{Q} \circ \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}$, with Lipschitz continuous layers $\mathcal{L}_\ell : L^2(D; \mathbb{R}^{d_v}) \to L^2(D; \mathbb{R}^{d_v})$, and with linear lifting/projection operators $\mathcal{R}, \mathcal{Q}$ of the form (1) and (2), respectively. For any $K \subset L^2(D; \mathbb{R}^{d_v})$ compact and $\epsilon > 0$, there exist neural networks $\widehat{R} : \mathbb{R}^{d_a} \times D \to \mathbb{R}^{d_v}$,*

$\widehat{Q} : \mathbb{R}^{d_v} \times D \to \mathbb{R}^{d_u}$ *with a single hidden layer and* $\mathrm{width}(\widehat{R}), \mathrm{width}(\widehat{Q}) \le 2d_v$, *such that the mapping* $\widehat{\mathcal{N}} : \mathcal{A}(D; \mathbb{R}^{d_a}) \to \mathcal{U}(D; \mathbb{R}^{d_u})$ *defined by*

$$\widehat{\mathcal{N}}(a) := \widehat{\mathcal{Q}} \circ \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1 \circ \widehat{\mathcal{R}}(a),$$

where $\mathcal{R}(a)(x) := \widehat{R}(a(x), x)$, $\mathcal{Q}(v) := \widehat{Q}(v(x), x)$, *approximates* $\mathcal{N}$ *to order* ϵ:

$$\sup_{a \in K} \|\mathcal{N}(a) - \widehat{\mathcal{N}}(a)\|_{L^2} \le \epsilon.$$

Proof We recall that, by our definition of a neural operator, we have $\mathcal{R}(a)(x) = Ra(x)$, $\mathcal{Q}(v)(x) = Qv(x)$, where $R \in \mathbb{R}^{d_v \times d_a}$, $Q \in \mathbb{R}^{d_u \times d_v}$. By assumption on the non-linearity of σ, there exists $z_0 \in \mathbb{R}$, such that $\sigma'(z_0) \ne 0$. For any $h > 0$, we now define

$$\begin{cases} \widehat{Q}_h(v, x) := \dfrac{\sigma(z_0 + hQv) - \sigma(z_0 - hQv)}{2h\sigma'(z_0)}, \\ \widehat{R}_h(a, x) := \dfrac{\sigma(z_0 + hRa) - \sigma(z_0 - hRa)}{2h\sigma'(z_0)}. \end{cases}$$

We remark that in the above expressions, the addition of $z_0 \in \mathbb{R}$ and $hQv \in \mathbb{R}^{d_u}$, $hRa \in \mathbb{R}^{d_v}$ is carried out componentwise, as is the evaluation of σ.

Since the interior layers $\mathcal{L}_1, \ldots, \mathcal{L}_L$ are fixed, we introduce the short-hand notation $\mathcal{L} := \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1$. By assumption on the activation function and the layers, we have that $\mathcal{L} : L^2(D; \mathbb{R}^{d_v}) \to L^2(D; \mathbb{R}^{d_v})$ is a Lipschitz continuous mapping. Our goal is to show that for any $K \subset L^2$ compact and $\epsilon > 0$, there exists a sufficiently small $h > 0$, such that

$$\sup_{a \in K} \|\mathcal{N}(a) - \widehat{\mathcal{N}}_h(a)\|_{L^2} \le \epsilon,$$

where $\widehat{\mathcal{N}}_h(a) := \widehat{\mathcal{Q}}_h \circ \mathcal{L} \circ \widehat{\mathcal{R}}_h(a)$. To this end, we note that for any $a \in K$, we have

$$\|\mathcal{N}(a) - \widehat{\mathcal{N}}_h(a)\|_{L^2} = \|\mathcal{Q} \circ \mathcal{L} \circ \mathcal{R}(a) - \widehat{\mathcal{Q}}_h \circ \mathcal{L} \circ \widehat{\mathcal{R}}_h(a)\|_{L^2}$$
$$\le \|\mathcal{Q} \circ \mathcal{L} \circ \mathcal{R}(a) - \widehat{\mathcal{Q}}_h \circ \mathcal{L} \circ \mathcal{R}(a)\|_{L^2}$$
$$+ \|\widehat{\mathcal{Q}}_h \circ \mathcal{L} \circ \mathcal{R}(a) - \widehat{\mathcal{Q}}_h \circ \mathcal{L} \circ \widehat{\mathcal{R}}_h(a)\|_{L^2}.$$

Introducing $\widetilde{K} := \mathcal{L} \circ \mathcal{R}(K)$, we note that $\widetilde{K} \subset L^2$ is compact, and

$$\sup_{a \in K} \|\mathcal{N}(a) - \widehat{\mathcal{N}}_h(a)\|_{L^2} \le \sup_{v \in \widetilde{K}} \|\mathcal{Q}(v) - \widehat{\mathcal{Q}}_h(v)\|_{L^2}$$
$$+ \mathrm{Lip}(\widehat{\mathcal{Q}}_h)\mathrm{Lip}(\mathcal{L}) \sup_{a \in K} \|\mathcal{R}(a) - \widehat{\mathcal{R}}_h(a)\|_{L^2}.$$

The proof of this lemma thus follows form the following three claims:

Claim 1: There exists $C > 0$, independent of h, such that $\mathrm{Lip}(\widehat{\mathcal{Q}}_h) \le C$ for all $h > 0$.

Claim 2: For any $\epsilon > 0$ and $\widetilde{K} \subset L^2$ compact, there exists $h > 0$, such that

$$\sup_{v \in \widetilde{K}} \|\widehat{\mathcal{Q}}_h(v) - \mathcal{Q}(v)\|_{L^2} \le \epsilon.$$

Claim 3: For any $\epsilon > 0$ and $K \subset L^2$ compact, there exists $h > 0$, such that

$$\sup_{a \in K} \|\widehat{\mathcal{R}}_h(a) - \mathcal{R}(a)\|_{L^2} \leq \epsilon.$$

Clearly, the only difference between Claims 2 and 3 is notational, hence it suffices show Claims 1 and 2 to conclude the proof of the present lemma.

Proof of Claim 1: We note that for any $v(x), v'(x) \in \mathbb{R}^{d_v}$, we have

$$|\widehat{Q}_h(v(x), x) - \widehat{Q}_h(v'(x), x)| \leq \frac{1}{2h|\sigma'(z_0)|} 2\text{Lip}(\sigma)|hQ(v(x) - v'(x))| \tag{54}$$

$$\leq \frac{\text{Lip}(\sigma)\|Q\|}{|\sigma'(z_0)|}|v(x) - v'(x)|, \tag{55}$$

where $|\cdot|$ denotes the Euclidean norm and $\|Q\|$ denotes the operator norm of Q. Hence, it follows that for any $v, v' \in L^2(D; \mathbb{R}^{d_v})$, we have

$$\|\widehat{\mathcal{Q}}_h(v) - \widehat{\mathcal{Q}}_h(v')\|_{L^2} \leq \frac{\text{Lip}(\sigma)\|Q\|}{|\sigma'(z_0)|}\|v - v'\|_{L^2},$$

and thus,

$$\text{Lip}(\widehat{\mathcal{Q}}_h) \leq \frac{\text{Lip}(\sigma)\|Q\|}{|\sigma'(z_0)|},$$

is bounded independently of $h > 0$.

Proof of Claim 2: It follows form Taylor expansion that $\widehat{Q}_h(v(x), x) = Qv(x) + hR(h; Qv(x))$, where the remainder R satisfies a uniform bound

$$|R(h; Qv(x))| \leq \frac{C\|\sigma\|_{C^2}}{|\sigma'(z_0)|}|Qv(x)|^2, \quad \forall h \in (0, 1], \ v(x) \in \mathbb{R}^{d_v}.$$

In particular, considering the L^2-norm of the left-hand side and assuming $x \mapsto v(x)$ to be *bounded*, we conclude that there exists a constant $C = C(\sigma, \|Q\|) > 0$, depending only on the activation function σ and the (Euclidean) operator norm $\|Q\|$ of $Q : (\mathbb{R}^{d_v}, |\cdot|) \to (\mathbb{R}^{d_u}, |\cdot|)$, such that

$$\|\widehat{\mathcal{Q}}_h(v) - \mathcal{Q}(v)\|_{L^2} = \|R(h; Qv)\|_{L^2} \leq C\|v\|_{L^\infty}\|v\|_{L^2}, \quad \forall v \in L^2 \cap L^\infty. \tag{56}$$

Next, for $M > 0$ we introduce a cut-off operator $p_M : L^2 \to L^2 \cap L^\infty$, by

$$p_M(v(x)) := \begin{cases} v(x), & (|v(x)| \leq M), \\ 0, & (|v(x)| > M). \end{cases}$$

Since $\widetilde{K} \subset L^2$ is compact, it follows that

$$\lim_{M \to \infty} \sup_{v \in \widetilde{K}} \|v - p_M(v)\|_{L^2} = 0.$$

We can thus choose $M > 0$ sufficiently large, such that

$$\sup_{v \in \widetilde{K}} \|v - p_M(v)\|_{L^2} \leq \frac{\epsilon}{2\left(\sup_{h \in (0,1]} \text{Lip}(\widehat{\mathcal{Q}}_h) + \text{Lip}(\mathcal{Q})\right)},$$

where we note that $\sup_{h \in (0,1]} \mathrm{Lip}(\widehat{\mathcal{Q}}_h) < \infty$, by Claim 1. Furthermore, and again by compactness of $\widetilde{K} \subset L^2$, there exists a constant $\widetilde{M} > 0$, such that

$$\sup_{v \in \widetilde{K}} \|v\|_{L^2} \le \widetilde{M}.$$

With this $\widetilde{M}$, our choice of M and the estimate (56) on the remainder R, it follows that for any $v \in \widetilde{K}$:

$$\begin{aligned}
\|\widehat{\mathcal{Q}}_h(v) - \mathcal{Q}(v)\|_{L^2} &\le \|\widehat{\mathcal{Q}}_h(v) - \widehat{\mathcal{Q}}_h(p_M(v))\|_{L^2} \\
&\quad + \|\widehat{\mathcal{Q}}_h(p_M(v)) - \mathcal{Q}(p_M(v))\|_{L^2} \\
&\quad + \|\mathcal{Q}(p_M(v)) - \mathcal{Q}(v)\|_{L^2} \\
&\le \left(\mathrm{Lip}(\widehat{\mathcal{Q}}_h) + \mathrm{Lip}(\mathcal{Q}) \right) \|v - p_M(v)\|_{L^2} \\
&\quad + Ch\|p_M(v)\|_{L^\infty}\|p_M(v)\|_{L^2} \\
&\le \epsilon/2 + CM\widetilde{M}h.
\end{aligned}$$

Finally, choosing $h = \epsilon/(2CM\widetilde{M})$ implies that

$$\sup_{v \in \widetilde{K}} \|\widehat{\mathcal{Q}}_h(v) - \mathcal{Q}(v)\|_{L^2} \le \epsilon.$$

This concludes the proof.

$\blacksquare$

Appendix D. Proofs and Technical Details for Section 2

D.1   Proof of Lemma 6

The proof of Lemma 6 will rely on the following technical lemma:

Lemma 43 *Let $s' \ge 0$, $N \in \mathbb{N}$ be given. Let $K \subset H^{s'}$ be compact, and assume that $\sigma \in C^m$, where $m > s'$ is integer. Then for any $\epsilon > 0$, there exists a single-layer FNO $\mathcal{L} : H^{s'} \to H^{s'}$, such that*

$$\sup_{v \in K} \|P_N v - \mathcal{L}(v)\|_{H^{s'}} \le \epsilon.$$

Proof [Proof of Lemma 43] First, we note that the Fourier projection $P_N : H^{s'} \to H^{s'}$ is a continuous operator, and hence the image $P_N K \subset H^{s'}$ is compact. Furthermore, P_N maps into a finite-dimensional subspace of $H^{s'}$, spanned by smooth (trigonometric) functions. Due to the norm-equivalence on finite-dimensional spaces, there thus exists $C_0 = C_0(N, K) > 0$, such that

$$\sup_{v \in K} \|P_N v\|_{L^\infty} \le C_0, \quad \sup_{v \in K} \|P_N v\|_{H^m} \le C_0. \tag{57}$$

Let $x_0 \in \mathbb{R}$ be such that $\sigma'(x_0) \neq 0$. We define, for $h > 0$,

$$\psi_h(x) := \frac{\sigma(x_0 + hx) - \sigma(x_0 - hx)}{2h\sigma'(x_0)}. \tag{58}$$

One readily shows that $\psi_h \in C^m$, and that there exists a constant $C_1 = C_1(\sigma, C_0) > 0$, such that

$$\|\psi_h\|_{C^m([-C_0, C_0])} \leq C_1, \quad \forall\, h \in (0, 1]. \tag{59}$$

Furthermore, by Taylor expansion, we have

$$|\psi_h(x) - x| \leq Ch, \quad \forall\, x \in [-C_0, C_0], \ \forall\, h \in (0, 1]. \tag{60}$$

By the composition rule for Sobolev functions, we have that $\psi_h \circ P_N v \in H^m$, for $P_N v \in H^m$, and there exists a constant $C_2 = C_2(C_1, C_0) > 0$, such that

$$\|\psi_h(P_N v)\|_{H^m} \leq C_2, \quad \forall\, v \in K. \tag{61}$$

We finally observe that the mapping $v \mapsto \mathcal{L}_h(v) := \psi_h(P_N v)$ can be represented by a single-layer FNO, and by (61), we have

$$\|\mathcal{L}_h(v)\|_{H^m} \leq C_2, \quad \forall\, v \in K. \tag{62}$$

From the interpolation inequality between Sobolev spaces, it follows that

$$\|\mathcal{L}_h(v) - P_N v\|_{H^{s'}} \leq \|\mathcal{L}_h(v) - P_N v\|_{L^2}^{\theta} \|\mathcal{L}_h(v) - P_N v\|_{H^m}^{1-\theta},$$

where $\theta = 1 - s'/m > 0$. By (62) and (57), the second factor can be bounded independently of $h > 0$. By (60) and (57), we have

$$\|\mathcal{L}_h(v) - P_N v\|_{L^2} = \|\psi_h(P_N v) - P_N v\|_{L^2} \leq Ch,$$

for a constant $C > 0$, independent of h and $v \in K$. We conclude that

$$\|\mathcal{L}_h(v) - P_N v\|_{H^{s'}} \leq Ch^{\theta} \to 0,$$

as $h \to 0$, for some constant $C > 0$, independent of h. This proves the claim. ∎

We can now prove:

Proof [Proof of Lemma 6] Let $\mathcal{G} : H^s \to H^{s'}$ be a continuous operator, and let $K \subset H^s$ be compact. We assume that FNOs are universal approximators of operators $H^s \to L^2$, and we wish to show that for any $\epsilon > 0$, there exists a FNO approximation of $\mathcal{G} : H^s \to H^{s'}$ to accuracy ϵ. We first note that by the compactness of $\mathcal{G}(K) \subset H^{s'}$, there exists $N \in \mathbb{N}$, such that

$$\sup_{a \in K} \|\mathcal{G}(a) - P_N \mathcal{G}(a)\|_{H^{s'}} \leq \epsilon/2. \tag{63}$$

Fix $\delta > 0$ for the moment. A suitable choice of δ will be specified at the end of this proof. By assumption on the universal approximation of operators $H^s \to L^2$, there exists a FNO $\widetilde{\mathcal{N}} : H^s \to L^2$, continuous as an operator $H^s \to L^2$, such that

$$\sup_{a \in K} \|P_N \mathcal{G}(a) - \widetilde{\mathcal{N}}(a)\|_{L^2} \leq \delta. \tag{64}$$

One difficulty in the present construction is that there is no guarantee that $\widetilde{\mathcal{N}}$ defines a mapping $H^s \to H^{s'}$, and indeed for $s' > s$ this is not generally the case. We circumvent this issue by composing with an additional FNO layer $\widetilde{\mathcal{L}} : L^2 \to H^{s'}$. By Lemma 43, there exists a single-layer FNO $v \mapsto \widetilde{\mathcal{L}}(v)$, satisfying the identity

$$\widetilde{\mathcal{L}}(v) = \widetilde{\mathcal{L}}(P_N v), \tag{65}$$

for all v, and defining a continuous operator $H^{s'} \to H^{s'}$, such that

$$\sup_{v \in K'} \|P_N v - \widetilde{\mathcal{L}}(v)\|_{H^{s'}} \leq \delta, \tag{66}$$

where $K' := P_N \widetilde{\mathcal{N}}(K) \subset H^{s'}$ is a compact subset of $H^{s'}$.[1] Next, we define a new FNO by the composition $\mathcal{N} := \widetilde{\mathcal{L}} \circ \widetilde{\mathcal{N}} : H^s \to H^{s'}$. $\mathcal{N}$ is a continuous operator $H^s \to H^{s'}$, since it can be written as the composition

$$H^s \xrightarrow{\widetilde{\mathcal{N}}} L^2 \xrightarrow{P_N} H^{s'} \xrightarrow{\widetilde{\mathcal{L}}} H^{s'},$$

of continuous operators. We observe that for any $a \in K$, we have the following bound:

$$\|P_N \mathcal{G}(a) - \mathcal{N}(a)\|_{H^{s'}} \leq \|P_N \mathcal{G}(a) - P_N \widetilde{\mathcal{N}}(a)\|_{H^{s'}} + \|P_N \widetilde{\mathcal{N}}(a) - \mathcal{N}(a)\|_{H^{s'}}$$
$$\leq C N^{s'} \|P_N \mathcal{G}(a) - \widetilde{\mathcal{N}}(a)\|_{L^2} + \|P_N \widetilde{\mathcal{N}}(a) - \widetilde{\mathcal{L}}(P_N \widetilde{\mathcal{N}}(a))\|_{H^{s'}},$$

having made use of the inequality $\|P_N v\|_{H^{s'}} \leq C N^{s'} \|P_N v\|_{L^2}$ for a constant $C = C(\mathbb{T}^d, s') > 0$ independent of N, and the fact that $\mathcal{N}(a) = \widetilde{\mathcal{L}}(\widetilde{\mathcal{N}}(a)) = \widetilde{\mathcal{L}}(P_N \widetilde{\mathcal{N}}(a))$ (cp. equation 65). Using (64), we can estimate

$$C N^{s'} \|P_N \mathcal{G}(a) - \widetilde{\mathcal{N}}(a)\|_{L^2} \leq C N^{s'} \delta.$$

The bound (66) implies that $\|P_N \widetilde{\mathcal{N}}(a) - \widetilde{\mathcal{L}}(P_N \widetilde{\mathcal{N}}(a))\|_{H^{s'}} = \|P_N v - \widetilde{\mathcal{L}}(v)\|_{H^{s'}} \overset{\overset{(66)}{\downarrow}}{\leq} \delta$, with $v := P_N \mathcal{N}(a) \in K'$. We thus obtain

$$\|P_N \mathcal{G}(a) - \mathcal{N}(a)\|_{H^{s'}} \leq (C N^{s'} + 1)\delta, \tag{67}$$

where $C = C(\mathbb{T}^d, s') > 0$ is independent of δ. Since $\delta > 0$ was arbitrary, we can ensure that $(C N^{s'} + 1)\delta \leq \epsilon/2$. From this estimate, and the bound (63), we conclude that there exists

---

1. We note that $\mathcal{N}(K) \subset L^2$ is compact as the continuous image of K, and that $P_N : L^2 \to H^{s'}$ defines a continuous mapping for fixed $N \in \mathbb{N}$.

a FNO $\mathcal{N} : H^s \to H^{s'}$, such that

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}(a)\|_{H^{s'}} \leq \sup_{a \in K} \|\mathcal{G}(a) - P_N \mathcal{G}(a)\|_{H^{s'}}$$
$$+ \sup_{a \in K} \|P_N \mathcal{G}(a) - \mathcal{N}(a)\|_{H^{s'}}$$
$$\leq \epsilon,$$

This concludes our proof. ∎

D.2 Proof of Lemma 7

Proof Step 1: In this first step, for any $\epsilon > 0$, we will construct a FNO

$$\mathcal{N}_1 : L^2(\mathbb{T}^d; \mathbb{R}) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}),$$

such that

$$\begin{cases} \|\mathcal{N}_1(v)_{1,k} - P_N v(x) \cos(\langle k, x \rangle)\|_{L^\infty} < \epsilon, \\ \|\mathcal{N}_1(v)_{2,k} - P_N v(x) \sin(\langle k, x \rangle)\|_{L^\infty} < \epsilon, \end{cases} \quad \forall k \in \mathcal{K}_N. \tag{68}$$

To see how to construct such $\mathcal{N}_1$, we first define a lifting

$$\mathcal{R}_1 : L^2(\mathbb{T}^d; \mathbb{R}) \to L^2(\mathbb{T}^d; \mathbb{R}^{4\mathcal{K}_N}), \quad v(x) \mapsto \widehat{w}(x),$$

where $\widehat{w}(x) := \{(v(x), 0, v(x), 0)\}_{k \in \mathcal{K}_N} \in \mathbb{R}^{4\mathcal{K}_N}$ for any $x \in \mathbb{T}^d$. In the following, we will identify $\mathbb{R}^{4\mathcal{K}_N} \simeq \mathbb{R}^{d_v}$, with $d_v = 4|\mathcal{K}_N|$. Next, we define the inner part of the first FNO layer (i.e. the matrix W, multiplier P, bias $b(x)$ in (7)), such that $P(k) \equiv 1_{|k| \leq N} \mathbf{1}_{d_v \times d_v}$ either vanishes (for $|k| > N$) or is the unit matrix (for $|k| \leq N$), $W \equiv 0$ is zero and the bias function $b(x) := \{(0, b_{2,k}(x), 0, b_{4,k}(x))\}_{k \in \mathcal{K}_N}$, $b_{2,k}(x) = \cos(\langle k, x \rangle)$, $b_{4,k}(x) = \sin(\langle k, x \rangle)$, such that

$$\widehat{\mathcal{L}}_1(\widehat{v})(x) := W\widehat{w}(x) + b(x) + \mathcal{F}^{-1}(P\mathcal{F}\widehat{w})(x)$$
$$= \left\{ \left(P_N \widehat{w}_{1,k}(x), \cos(\langle k, x \rangle), P_N \widehat{w}_{3,k}(x), \sin(\langle k, x \rangle) \right) \right\}_{k \in \mathcal{K}_N}.$$

Next, we recall that by assumption, we have $\|v\|_{L^2} \leq B$, and by construction, we have $\widehat{w}_{1,k}(x) = \widehat{w}_{3,k}(x) = v(x)$, implying that

$$\|P_N \widehat{w}_{1,k}\|_{L^\infty} = \|P_N \widehat{w}_{3,k}\|_{L^\infty} = \|P_N v\|_{L^\infty} \leq C\|P_N v\|_{L^2} \leq CB,$$

where $C = C(N) \propto N^{d/2}$ is a constant depending on N. By the universal approximation theorem for (ordinary) neural networks, there exists a neural network

$$\widehat{\mathcal{N}} : [-CB, CB] \times [-1, 1] \times [-CB, CB] \times [-1, 1] \to \mathbb{R}^2$$

with activation function σ, such that $\max_{(a,b,c,d)} |\widehat{\mathcal{N}}(a, b, c, d) - (ab, cd)| < \epsilon$, and where the maximum is taken over all $a, c \in [-CB, CB]$, $b, d \in [-1, 1]$. But then, the point-wise mapping

$$\mathcal{N}_1 = \widehat{\mathcal{N}} \circ \widehat{\mathcal{L}} \circ \mathcal{R}_1 : \begin{cases} L^2(\mathbb{T}^d; \mathbb{R}) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}), \\ v(x) \mapsto \widetilde{w}(x) := \widehat{\mathcal{L}}_1(\mathcal{R}_1(v))(x) \mapsto \widehat{\mathcal{N}}(\widetilde{w}(x)), \end{cases}$$

satisfies (68), and $\mathcal{N}_1$ can be represented by a FNO (cp. Remark 3). **Step 2:** By the last step, we have $\|\mathcal{N}_1(v)_{1,k} - P_N v(x) \cos(\langle k, x\rangle)\|_{L^\infty} < \epsilon$ and $\|\mathcal{N}_1(v)_{2,k} - P_N v(x) \sin(\langle k, x\rangle)\|_{L^\infty} < \epsilon$ for all $v \in L^2(\mathbb{T}^d)$ with $\|v\|_{L^2} \le B$. We next note that, since

$$P_N v(x) = \sum_{|k|_\infty \le N} \widehat{v}_k e^{i\langle k, x\rangle},$$

and since $v(x)$ is real-valued, the Fourier coefficients $\widehat{v}_k$ satisfy $\widehat{v}_{-k} = \overline{\widehat{v}_k}$, where $\overline{\widehat{v}_k}$ denotes the complex conjugate of $\widehat{v}_k$, and

$$\fint_{\mathbb{T}^d} P_N v(x) \cos(\langle k, x\rangle)\, dx = \mathrm{Re}(\widehat{v}_k), \quad \fint_{\mathbb{T}^d} P_N v(x) \sin(\langle k, x\rangle)\, dx = -\mathrm{Im}(\widehat{v}_k),$$

In particular, this implies that the 0-th Fourier modes of $P_N v(x) \cos(\langle k, x\rangle)$, and $P_N v(x) \sin(\langle k, x\rangle)$, respectively, are given by

$$\mathcal{F}\Big[P_N v \cos(\langle k, \cdot\rangle)\Big](0) = \mathrm{Re}(\widehat{v}_k), \quad \mathcal{F}\Big[P_N v \sin(\langle k, \cdot\rangle)\Big](0) = -\mathrm{Im}(\widehat{v}_k),$$

and, as a consequence, we have with the Fourier multiplier

$$\delta_0(k') = \begin{cases} 1, & (k' = 0) \\ 0, & (k' \ne 0), \end{cases} \quad \forall\, k' \in \mathbb{Z}^d,$$

and with $w(x) := \mathcal{N}_1(v)(x) \in L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$, written as $w(x) = \{(w_{1,k}(x), w_{2,k}(x))\}_{k \in \mathcal{K}_N}$, that

$$\left\| \mathcal{F}^{-1}\Big(\delta_0(k')\mathcal{F}(w_{1,k})(k')\Big)(x) - \mathrm{Re}(\widehat{v}_k) \right\|_{L^\infty} < \epsilon,$$
$$\left\| \mathcal{F}^{-1}\Big(\delta_0(k')\mathcal{F}(w_{2,k})(k')\Big)(x) + \mathrm{Im}(\widehat{v}_k) \right\|_{L^\infty} < \epsilon,$$

for all $v \in L^2(\mathbb{T}^d)$ with $\|v\|_{L^2} \le B$. We use this observation to define a suitable FNO layer; with local matrix $W = 0$, bias $b(x) = 0$ and a Fourier multiplier matrix $P : \mathbb{Z}^d \to \mathbb{C}^{d_v \times d_v}$, $k' \mapsto P(k')$ (where $d_v = 2|\mathcal{K}_N|$, and $\mathbb{C}^{d_v} \simeq \mathbb{C}^{2\mathcal{K}_N}$), with entries

$$[P(k')]_{(\ell,k),(\widetilde{\ell},\widetilde{k})} := \delta_0(k')\left\{\delta_{\ell=1}(\ell) - \delta_{\ell=2}(\ell)\right\} \mathbf{1}_{d_v \times d_v}.$$

With this definition of the Fourier multiplier P, we define $\widehat{\mathcal{L}}_2 : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$, by

$$\widehat{\mathcal{L}}_2(w) := W w(x) + b(x) + \mathcal{F}^{-1}\Big(P\mathcal{F}w\Big)(x).$$

Then, by construction, we have for any $v \in L^2(\mathbb{T}^d)$ with $\|v\|_{L^2} \le B$ that the output $w := \widehat{\mathcal{L}}_2(\mathcal{N}_1(v))$, $w(x) = \{(w_{1,k}(x), w_{2,k}(x))\}_{k \in \mathcal{K}_N}$, satisfies

$$\begin{cases} \|w_{1,k}(x) - \mathrm{Re}(\widehat{v}_k)\|_{L^\infty} < \epsilon, \\ \|w_{2,k}(x) - \mathrm{Im}(\widehat{v}_k)\|_{L^\infty} < \epsilon. \end{cases}$$

We also note that $\widehat{\mathcal{L}}_2$ outputs only constant functions, by construction. This is almost the desired result, except for the fact that the composition $\widehat{\mathcal{L}}_2 \circ \mathcal{N}_1$ does not define a FNO, since

the linear layer $\widehat{\mathcal{L}}_2$ is missing the non-linearity σ. This can be rectified by composition with a suitable (ordinary) σ-neural network $\widetilde{\mathcal{N}}$, which approximates the identity: Indeed, the last estimate implies in particular that $|w_{\ell,k}| \leq |\widehat{v}_k| + \epsilon \leq B + \epsilon$ for all ℓ, k. By the (ordinary) universal approximation theorem, there exists a neural network $\widetilde{\mathcal{N}} : \mathbb{R}^{2\mathcal{K}_N} \to \mathbb{R}^{2\mathcal{K}_N}$, such that

$$\|\widetilde{\mathcal{N}}(w) - w\|_{\ell^\infty} < \epsilon, \quad \forall w \in \mathbb{R}^{2\mathcal{K}_N}, \; \|w\|_{\ell^\infty} \leq B + \epsilon.$$

Then the composition $\mathcal{N}_2 := \widetilde{\mathcal{N}} \circ \widehat{\mathcal{L}}_2 : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$, given by $w(x) \mapsto \widetilde{\mathcal{N}}\left(\widehat{\mathcal{L}}_2(w)(x)\right)$, *does* define a FNO.

Step 3: We finally observe that since both $\mathcal{N}_1$ and $\mathcal{N}_2$ can be represented by FNOs, then also their composition

$$\mathcal{N} := \mathcal{N}_2 \circ \mathcal{N}_1 : L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}),$$

can be represented by a FNO with $\mathrm{depth}(\mathcal{N}) = \mathrm{depth}(\mathcal{N}_1) + \mathrm{depth}(\mathcal{N}_2)$, $\mathrm{lift}(\mathcal{N}) = \max_{j=1,2} \mathrm{lift}(\mathcal{N}_j)$. Furthermore, the mapping $v \mapsto \mathcal{N}(v)$ maps to constant functions, and for any $v \in L^2(\mathbb{T}^d)$, $\|v\|_{L^2} \leq B$, we have

$$\begin{cases} \|\mathrm{Re}(\widehat{v}_k) - \mathcal{N}(v)_{1,k}\|_{L^\infty} < 2\epsilon, \\ \|\mathrm{Im}(\widehat{v}_k) - \mathcal{N}(v)_{2,k}\|_{L^\infty} < 2\epsilon. \end{cases}$$

Since $\epsilon > 0$ was arbitrary, the claim follows. ∎

D.3 Proof of Lemma 8

Proof Step 1: We first construct a FNO $\mathcal{N}_1 : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$, such that

$$\begin{cases} \left\|\mathcal{N}_1(w)_{1,k} - P_N w_{1,k}(x) \cos(\langle k, x \rangle)\right\|_{L^\infty} < \left(2|\mathcal{K}_N||\mathbb{T}^d|^{1/2}\right)^{-1} \epsilon, \\ \left\|\mathcal{N}_1(w)_{2,k} - P_N w_{2,k}(x) \sin(\langle k, x \rangle)\right\|_{L^\infty} < \left(2|\mathcal{K}_N||\mathbb{T}^d|^{1/2}\right)^{-1} \epsilon, \end{cases} \tag{69}$$

for all $w(x) = (w_{1,k}(x), w_{2,k}(x))$, such that $\|w_{1,k}\|_{L^\infty}, \|w_{2,k}\|_{L^\infty} \leq B$, for all $k \in \mathcal{K}_N$. Here $|\mathbb{T}^d| = (2\pi)^d$ denotes the Lebesgue measure of $\mathbb{T}^d$ and $|\mathcal{K}_N| = (2N + 1)^d$ the number of elements of $\mathcal{K}_N$.

To construct such $\mathcal{N}_1$, we first define a lifting $\mathcal{R} : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d; \mathbb{R}^{4\mathcal{K}_N})$, $\{(w_{1,k}(x), w_{2,k}(x))\}_{k \in \mathcal{K}_N} \mapsto \{(w_{1,k}(x), 0, w_{2,k}(x), 0)\}_{k \in \mathcal{K}_N}$, followed by a linear layer $\mathcal{L} : L^2(\mathbb{T}^d; \mathbb{R}^{4\mathcal{K}_N}) \to L^2(\mathbb{T}^d; \mathbb{R}^{4\mathcal{K}_N})$, which only introduces a bias (setting $W \equiv 0$, $b(x) = \{(0, \cos(\langle k, x \rangle), 0, \sin(\langle k, x \rangle))\}_{k \in \mathcal{K}_N}$, $P \equiv 0$), to yield

$$\mathcal{L} \circ \mathcal{R}(w) = \{(w_{1,k}(x), \cos(\langle k, x \rangle), w_{2,k}(x), \sin(\langle k, x \rangle))\}_{k \in \mathcal{K}_N},$$

for all $w \in L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$. There exists an ordinary neural network $\widehat{\mathcal{N}} : [-B, B] \times [-1, 1] \times [-B, B] \times [-1, 1] \to \mathbb{R}^2$, such that $\max_{a,b,c,d} |\widehat{\mathcal{N}}(a, b, c, d) - (ab, cd)| < (2|\mathcal{K}_N||\mathbb{T}^d|^{1/2})^{-1}\epsilon$, where the maximum is taken over $a, c \in [-B, B]$ and $b, d \in [-1, 1]$. Since, by assumption, we have $\|w_{1,k}\|_{L^\infty}, \|w_{2,k}\|_{L^\infty} \leq B$ for the inputs of interest, we conclude that $\mathcal{N}_1(w) := \widehat{\mathcal{N}} \circ \mathcal{L} \circ \mathcal{R}(w)$ satisfies (69). Furthermore, by Lemma 44 and Lemma 46, $\mathcal{N}_1$ is represented by a FNO.

Step 2: We define a projection $\mathcal{Q}$ (cp. the definition of neural operators, equation 2), mapping $w(x) \in L^2(\mathbb{T}^d); \mathbb{R}^{2\mathcal{K}_N})$ to a scalar-valued function, by

$$\mathcal{Q} : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d), \quad w(x) \mapsto \sum_{k \in \mathcal{K}_N} (w_{1,k}(x) - w_{2,k}(x)).$$

Then $\mathcal{N} := \mathcal{Q} \circ \mathcal{N}_1 : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d)$ is an FNO, and for any $v \in L^2_N(\mathbb{T}^d)$, and $w = \{(\text{Re}(\widehat{v}_k), \text{Im}(\widehat{v}_k))\}_{k \in \mathcal{K}_N}$ defined as in the statement of this lemma, we have

$$
\begin{aligned}
\|v - \mathcal{N}(w)\|_{L^2} &= \left\| v - \sum_{k \in \mathcal{K}_N} \left(\mathcal{N}_1(w)_{1,k} - \mathcal{N}_1(w)_{2,k} \right) \right\|_{L^2} \\
&\leq \sum_{k \in \mathcal{K}_N} \|\mathcal{N}_1(w)_{1,k} - \text{Re}(\widehat{v}_k) \cos(\langle k, x \rangle)\|_{L^2} \\
&\quad + \sum_{k \in \mathcal{K}_N} \|\mathcal{N}_1(w)_{2,k} - \text{Im}(\widehat{v}_k) \sin(\langle k, x \rangle)\|_{L^2} \\
&\leq \sum_{k \in \mathcal{K}_N} \|\mathcal{N}_1(w)_{1,k} - \text{Re}(\widehat{v}_k) \cos(\langle k, x \rangle)\|_{L^2} \\
&\quad + \sum_{k \in \mathcal{K}_N} \|\mathcal{N}_1(w)_{2,k} - \text{Im}(\widehat{v}_k) \sin(\langle k, x \rangle)\|_{L^2} \\
&\leq |\mathcal{K}_N||\mathbb{T}^d|^{1/2} \max_{k \in \mathcal{K}_N} \|\mathcal{N}_1(w)_{1,k} - w_{1,k} \cos(\langle k, x \rangle)\|_{L^\infty} \\
&\quad\quad |\mathcal{K}_N||\mathbb{T}^d|^{1/2} \max_{k \in \mathcal{K}_N} \|\mathcal{N}_1(w)_{2,k} - w_{2,k} \sin(\langle k, x \rangle)\|_{L^\infty} \\
&< \epsilon.
\end{aligned}
$$

∎

D.4 Proof of the Universal Approximation Theorem 5

As mentioned in the sketch of the proof of Theorem 5, we only need to consider the special case $s' = 0$, due to Lemma 6; i.e. given a continuous operator $\mathcal{G} : H^s(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$, $K \subset H^s(\mathbb{T}^d)$ compact, and $\epsilon > 0$, we wish to construct a FNO $\mathcal{N} : H^s(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$, such that $\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}(a)\|_{L^2} \leq \epsilon$.

Throughout this proof, we set $d_a = d_u = 1$ for notational convenience. The general case with $d_a, d_u > 1$ follows analogously. For $N \in \mathbb{N}$, let P_N denote the orthogonal Fourier projection (45). First, we note that since $K \subset H^s(\mathbb{T}^d)$ is compact, one can prove by elementary arguments that the set $\widetilde{K}$ given by

$$\widetilde{K} := K \cup \bigcup_{N \in \mathbb{N}} P_N K,$$

is compact. Since $\mathcal{G}$ is continuous, its restriction to $\widetilde{K}$ is uniformly continuous, i.e. there exists a modulus of continuity $\omega : [0, \infty) \to [0, \infty)$, such that

$$\|\mathcal{G}(a) - \mathcal{G}(a')\|_{L^2} \leq \omega \left(\|a - a'\|_{H^s} \right), \quad \forall\, a, a' \in \widetilde{K}.$$

43

From the definition of the projection $\mathcal{G}_N$ (8), we have,

$$
\begin{aligned}
\|\mathcal{G}(a) - \mathcal{G}_N(a)\|_{L^2} &\leq \|\mathcal{G}(a) - P_N\mathcal{G}(a)\|_{L^2} + \|P_N\mathcal{G}(a) - P_N\mathcal{G}(P_Na)\|_{L^2} \\
&\leq \|\mathcal{G}(a) - P_N\mathcal{G}(a)\|_{L^2} + \|\mathcal{G}(a) - \mathcal{G}(P_Na)\|_{L^2} \\
&\leq \sup_{v \in \mathcal{G}(\widetilde{K})} \|(1 - P_N)v\|_{L^2} + \omega\left(\sup_{a \in \widetilde{K}} \|(1 - P_N)a\|_{H^\sigma}\right).
\end{aligned}
\tag{70}
$$

Since $\widetilde{K}$ is compact, also the image $\mathcal{G}(\widetilde{K})$ is compact leading to,

$$
\limsup_{N \to \infty} \sup_{u \in \mathcal{G}(\widetilde{K})} \|(1 - P_N)v\|_{L^2} = 0 = \limsup_{N \to \infty} \sup_{a \in \widetilde{K}} \|(1 - P_N)a\|_{H^s}.
$$

In particular, there exists $N \in \mathbb{N}$, such that

$$
\|\mathcal{G}(a) - \mathcal{G}_N(a)\|_{L^2} \leq \epsilon, \quad \forall a \in K \subset \widetilde{K}.
\tag{71}
$$

In the remainder of this proof, we will construct an FNO approximation of $\mathcal{G}_N$. In fact, we note that $\mathcal{G}_N$ defines a continuous operator $\mathcal{G}_N : L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$, via $a \mapsto P_N\mathcal{G}(P_Na)$, and the compact set K remains compact also as a subset of $L^2(\mathbb{T}^d)$. We will show that there exists a FNO $\mathcal{N} : L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$, such that

$$
\sup_{a \in K} \|\mathcal{G}_N(a) - \mathcal{N}(a)\|_{L^2} < \epsilon.
$$

Then the restriction of $\mathcal{N}$ to $H^s(\mathbb{T}^d) \subset L^2(\mathbb{T}^d)$ provides an approximation of $\mathcal{G}$, such that

$$
\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}(a)\|_{L^2} < 2\epsilon.
$$

Since $\epsilon > 0$ was arbitrary, the claim follows from this.

As outlined in the sketch, our proof for the existence of a FNO approximating $\mathcal{G}_N$ relies on the decomposition (12), which in turn is defined via the Fourier conjugate operator $\widehat{G}_N$ (11). Our aim is to show that each operator in the decomposition (12) can be approximated by FNOs to desired accuracy. To this end, let $\epsilon > 0$ be given, and choose $R_K, R_{\widehat{K}}, R_{\widehat{\mathcal{G}}} > 0$, such that

$$
\begin{cases}
K \subset B_{R_K}(0) := \{\|u\|_{L^2} \leq R_K\} \subset L^2(\mathbb{T}^d), \\[2mm]
\mathcal{F}_N \circ P_N\left(B_{R_K}(0)\right) \subset \left[-\dfrac{R_{\widehat{K}}}{2}, \dfrac{R_{\widehat{K}}}{2}\right]^{2\mathcal{K}_N}, \\[3mm]
\widehat{\mathcal{G}}_N\left([-R_{\widehat{K}}, R_{\widehat{K}}]^{2\mathcal{K}_N}\right) \subset \left[-\dfrac{R_{\widehat{\mathcal{G}}}}{2}, \dfrac{R_{\widehat{\mathcal{G}}}}{2}\right]^{2\mathcal{K}_N}.
\end{cases}
\tag{72}
$$

The reason for introducing $R_K, R_{\widehat{K}}, R_{\widehat{\mathcal{G}}}$ lies in the fact that in order to approximate $\mathcal{G}_N$ by a composition of FNOs, we need to ensure that each FNO in this composition maps its own domain into the domain of the next FNO. The FNO approximations of the individual steps in the composition $\mathcal{G}_N = \mathcal{F}_N^{-1} \circ \widehat{\mathcal{G}}_N \circ (\mathcal{F}_N \circ P_N)$ are constructed below:

44

FNO approximation of $\mathcal{F}_N^{-1}$: We start with our construction of a FNO approximation of the last step in the composition. To this end, we are going to interpret the mapping

$$\mathcal{F}_N^{-1} : [-R, R]^{2\mathcal{K}_N} \subset \mathbb{R}^{2\mathcal{K}_N} \to L^2(\mathbb{T}^d),$$

as a mapping

$$\mathcal{F}_N^{-1} : \begin{cases} L^2(\mathbb{T}^d; [-R, R]^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d), \\ \{\mathrm{Re}(\widehat{v}_k), \mathrm{Im}(\widehat{v}_k)\}_{|k| \leq N} \mapsto v(x), \end{cases} \tag{73}$$

where the input $\{\mathrm{Re}(\widehat{v}_k), \mathrm{Im}(\widehat{v}_k)\}_{|k| \leq N} \in [-R, R]^{2\mathcal{K}_N}$ is identified with a *constant* function in $L^2(\mathbb{T}^d; [-R, R]^{2\mathcal{K}_N})$ (for non-constant input function $v(x)$, we apply $\mathcal{F}_N^{-1}$ to the constant function $x \mapsto \fint_{\mathbb{T}^d} v(\xi) \, d\xi$ to define the mapping in equation 73 for general inputs). By Lemma 8, the mapping (73) can be approximated to any desired accuracy by an FNO $\mathcal{N}_{\mathrm{IFT}} : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d)$, such that

$$\|\mathcal{N}_{\mathrm{IFT}}(\widehat{v}) - \mathcal{F}_N^{-1}(\widehat{v})\|_{L^2} \leq \epsilon/3, \tag{74}$$

for all *constant* input functions $\widehat{v} \in L^2(\mathbb{T}^d; [-R, R]^{2\mathcal{K}_N})$.

FNO approximation of $\widehat{\mathcal{G}}_N$: We view the Fourier conjugate operator $\widehat{\mathcal{G}}_N$ (11) as a continuous mapping

$$\widehat{\mathcal{G}}_N : [-R_{\widehat{K}}, R_{\widehat{K}}]^{2\mathcal{K}_N} \subset \mathbb{R}^{2\mathcal{K}_N} \to \mathbb{R}^{2\mathcal{K}_N}.$$

Since $[-R_{\widehat{K}}, R_{\widehat{K}}]^{2\mathcal{K}_N}$ is compact, there exists a finite-dimensional canonical neural network $\widehat{\mathcal{N}} : \mathbb{R}^{2\mathcal{K}_N} \to \mathbb{R}^{2\mathcal{K}_N}$, such that

$$\sup_{\widehat{v} \in [-R_{\widehat{K}}, R_{\widehat{K}}]^{2\mathcal{K}_N}} \|\widehat{\mathcal{G}}_N(\widehat{v}) - \widehat{\mathcal{N}}(\widehat{v})\|_{\ell^2} \leq \epsilon/3. \tag{75}$$

Furthermore, by (72), we have

$$\widehat{G}_N\left([-R_{\widehat{K}}, R_{\widehat{K}}]^{2\mathcal{K}_N}\right) \subset \left[-\frac{R_{\widehat{\mathcal{G}}}}{2}, \frac{R_{\widehat{\mathcal{G}}}}{2}\right]^{2\mathcal{K}_N}.$$

Therefore, by choosing a neural network approximation $\widehat{\mathcal{N}}$ with sufficiently high accuracy, we can also ensure that

$$\widehat{\mathcal{N}}\left([-R_{\widehat{K}}, R_{\widehat{K}}]^{2\mathcal{K}_N}\right) \subset \left[-R_{\widehat{\mathcal{G}}}, R_{\widehat{\mathcal{G}}}\right]^{2\mathcal{K}_N},$$

in addition to (75). Finally, we note that for $v \in L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$, the corresponding mapping

$$\widehat{\mathcal{N}} : L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}), \quad v(x) \mapsto \widehat{\mathcal{N}}(v(x)),$$

is in fact an FNO, with only *local* layers of the form

$$v_\ell(x) \mapsto \sigma\left(A_\ell v_\ell(x) + b_\ell\right), \quad (A_\ell \in \mathbb{R}^{d_v \times d_v}, \; b_\ell \in \mathbb{R}^{d_v}),$$

and where $d_v := 2|\mathcal{K}_N|$, i.e. a FNO for which all $P_\ell \equiv 0$ (cp. Remark 3). We shall thus identify $\widehat{\mathcal{N}}$ with this particular FNO, in the following.

FNO approximation of $\mathcal{F}_N \circ P_N$: Finally, we can similarly interpret

$$\mathcal{F}_N \circ P_N : B_{R_K}(0) \subset L^2(\mathbb{T}^d) \to \mathbb{R}^{2\mathcal{K}_N},$$

as a mapping

$$\mathcal{F}_N \circ P_N : \begin{cases} B_{R_K}(0) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N}), \\ v \mapsto \{\mathrm{Re}(\widehat{v}_k), \mathrm{Im}(\widehat{v}_k)\}_{|k| \leq N}, \end{cases} \tag{76}$$

where the output $\{\mathrm{Re}(\widehat{v}_k), \mathrm{Im}(\widehat{v}_k)\}_{|k| \leq N} \in \mathbb{R}^{2\mathcal{K}_N}$ is a *constant* function in $L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$. By Lemma 7, the mapping (76) can be approximated to any desired accuracy by an FNO $\mathcal{N}_{\mathrm{FT}} : B_{R_K}(0) \to L^2(\mathbb{T}^d; \mathbb{R}^{2\mathcal{K}_N})$ (with constant output functions). In particular, denoting $\mathrm{Lip}(\widehat{\mathcal{N}})$ the Lipschitz constant of the FNO constructed in the previous step, we can ensure that

$$\mathrm{Lip}(\widehat{\mathcal{N}}) \, \|\mathcal{F}_N P_N v - \mathcal{N}_{\mathrm{FT}}(v)\|_{\ell^2} \leq \epsilon/3, \quad \forall\, v \in B_{R_K}(0), \tag{77}$$

and furthermore, since by (72), we have

$$\mathcal{F}_N \circ P_N(B_{R_K}(0)) \subset \left[-\frac{R_{\widehat{K}}}{2}, \frac{R_{\widehat{K}}}{2} \right]^{2\mathcal{K}_N},$$

we can in addition ensure that

$$\mathcal{N}_{\mathrm{FT}}(B_{R_K}(0)) \subset \left[-R_{\widehat{K}}, R_{\widehat{K}} \right]^{2\mathcal{K}_N}.$$

Error estimate for resulting FNO approximation: We now define a FNO $\mathcal{N}(a) := \mathcal{N}_{\mathrm{IFT}} \circ \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}}(a)$, where the right-hand side terms have been constructed above. We note that

$$\sup_K \|\mathcal{G}_N - \mathcal{N}\|_{L^2}$$

$$\leq \sup_{B_{R_K}(0)} \left\| \mathcal{F}_N^{-1} \circ \widehat{\mathcal{G}}_N \circ \mathcal{F}_N \circ P_N - \mathcal{N}_{\mathrm{IFT}} \circ \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}} \right\|_{L^2}$$

$$\leq \sup_{B_{R_K}(0)} \left\| \mathcal{F}_N^{-1} \circ \widehat{\mathcal{G}}_N \circ \mathcal{F}_N \circ P_N - \mathcal{F}_N^{-1} \circ \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}} \right\|_{L^2}$$

$$+ \sup_{B_{R_K}(0)} \left\| \mathcal{F}_N^{-1} \circ \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}} - \mathcal{N}_{\mathrm{IFT}} \circ \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}} \right\|_{L^2}$$

$$\leq \sup_{B_{R_K}(0)} \left\| \widehat{\mathcal{G}}_N \circ \mathcal{F}_N \circ P_N - \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}} \right\|_{L^2}$$

$$+ \sup_{\widehat{\mathcal{N}}(\mathcal{N}_{\mathrm{FT}}(B_{R_K}(0)))} \left\| \mathcal{F}_N^{-1} - \mathcal{N}_{\mathrm{IFT}} \right\|_{L^2}$$

$$=: (I) + (II).$$

For the second term (II), we note that

$$\widehat{\mathcal{N}}\left(\mathcal{N}_{\mathrm{FT}}\left(B_{R_K}(0)\right)\right) \subset \widehat{\mathcal{N}}\left([-R_{\widehat{K}}, R_{\widehat{K}}]^{2\mathcal{K}_N}\right) \subset \left[-R_{\widehat{\mathcal{G}}}, R_{\widehat{\mathcal{G}}}\right]^{2\mathcal{K}_N},$$

and hence, by (74), we can bound

$$(II) \le \sup_{\left[-R_{\widehat{\mathcal{G}}}, R_{\widehat{\mathcal{G}}}\right]^{2\mathcal{K}_N}} \left\|\mathcal{F}_N^{-1} - \mathcal{N}_{\mathrm{IFT}}\right\|_{L^2} \le \epsilon/3.$$

To estimate the first term (I), we note that

$$\begin{aligned}
(I) &= \sup_{B_{R_K}(0)} \left\|\widehat{\mathcal{G}}_N \circ \mathcal{F}_N \circ P_N - \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}}\right\|_{L^2} \\
&\le \sup_{B_{R_K}(0)} \left\|\widehat{\mathcal{G}}_N \circ \mathcal{F}_N \circ P_N - \widehat{\mathcal{N}} \circ \mathcal{F}_N \circ P_N\right\|_{L^2} \\
&\quad + \sup_{B_{R_K}(0)} \left\|\widehat{\mathcal{N}} \circ \mathcal{F}_N \circ P_N - \widehat{\mathcal{N}} \circ \mathcal{N}_{\mathrm{FT}}\right\|_{L^2} \\
&=: (Ia) + (Ib).
\end{aligned}$$

To estimate (Ia), we note that

$$\mathcal{F}_N\left(P_N\left(B_{R_K}(0)\right)\right) \subset \left[-R_{\widehat{K}}, R_{\widehat{K}}\right]^{2\mathcal{K}_N},$$

and hence

$$(Ia) \le \sup_{\left[-R_{\widehat{K}}, R_{\widehat{K}}\right]^{2\mathcal{K}_N}} \left\|\widehat{\mathcal{G}}_N - \widehat{\mathcal{N}}\right\|_{L^2} \le \epsilon/3,$$

by (75). Finally, to estimate (Ib), we note that

$$(Ib) \le \mathrm{Lip}(\widehat{\mathcal{N}}) \sup_{B_{R_K}(0)} \left\|\mathcal{F}_N \circ P_N - \mathcal{N}_{\mathrm{FT}}\right\|_{L^2} \le \epsilon/3,$$

by (77). Combining these estimates, we conclude that

$$\sup_{a \in K} \left\|\mathcal{G}_N(a) - \mathcal{N}(a)\right\|_{L^2} \le \epsilon,$$

This shows that the continuous operator $\mathcal{G}_N$ can be approximated by a FNO $\mathcal{N}$ to any desired accuracy $\epsilon > 0$, and together with (9) concludes our proof of the universal approximation theorem 5 for the special case $s' = 0$. The general case with $s' \ge 0$ now follows from Lemma 6.

D.5 Proof of Theorem 14

Proof The proof involves 4 steps.

Step 1: We may wlog assume that all biases $b_\ell(x) \in C^\infty$.

Indeed, it is easy to see that under the finite width assumption and for Lipschitz continuous σ, each layer $\mathcal{L}_\ell$ defines a Lipschitz continuous mapping $L^2 \to L^2$. Replacing $b_\ell(x)$ by it's δ-mollification $b_\ell^\delta(x)$ we obtain a new layer $\mathcal{L}_\ell^\delta$, i.e.

$$\mathcal{L}_\ell^\delta(v) = \sigma\left(W_\ell v(x) + b_\ell^\delta(x) + \mathcal{F}^{-1}\left(P_\ell \mathcal{F} v\right)(x)\right).$$

Then we have

$$\|\mathcal{L}_\ell(v) - \mathcal{L}_\ell^\delta(v)\|_{L^2} \lesssim_{\mathrm{Lip}(\sigma),\ell} \|b_\ell - b_\ell^\delta\|_{L^2} \to 0, \tag{78}$$

and uniformly in $\delta > 0$, we have

$$\|\mathcal{L}_\ell^\delta(v) - \mathcal{L}_\ell^\delta(v')\|_{L^2} \lesssim_{\mathrm{Lip}(\sigma),\ell} \|v - v'\|_{L^2}. \tag{79}$$

In particular, this implies that $\mathrm{Lip}(\mathcal{L}_\ell^\delta) \le C(\mathrm{Lip}(\sigma), \ell)$ is uniformly bounded in $\delta \ge 0$. Thus, there exists $\overline{M} \ge 1$, such that $\mathrm{Lip}(\mathcal{Q}) \le \overline{M}$ and $\mathrm{Lip}(\mathcal{L}_\ell^\delta) \le \overline{M}$, for all $\ell = 1, \dots, L$ and $\delta > 0$. Properties (78) and (79) are sufficient to show that with a sufficiently small choice of $\delta > 0$, we have

$$\sup_{\|a\|_{H^s} \le B} \|\mathcal{N}(a) - \mathcal{N}^\delta(a)\|_{L^2} \le \epsilon, \tag{80}$$

where $\mathcal{N}^\delta := \mathcal{Q} \circ \mathcal{L}_L^\delta \circ \cdots \circ \mathcal{L}_1^\delta \circ \mathcal{R}$: To see this, we introduce $\mathcal{N}_\ell^\delta := \mathcal{Q} \circ \mathcal{L}_L^\delta \circ \dots \mathcal{L}_{\ell+1}^\delta \circ \mathcal{L}_\ell \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}$, and we note that for any $a \in L^2$, we have

$$\|\mathcal{N}_\ell^\delta(a) - \mathcal{N}_{\ell-1}^\delta(a)\|_{L^2} \le \mathrm{Lip}(\mathcal{Q} \circ \mathcal{L}_L^\delta \circ \cdots \circ \mathcal{L}_{\ell+1}^\delta)\|\mathcal{L}_\ell(a') - \mathcal{L}_\ell^\delta(a')\|_{L^2},$$

where $a' := \mathcal{L}_{\ell-1} \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}(a)$. Using the uniform Lipschitz bound $\overline{M}$ and the estimate (78), we obtain

$$\|\mathcal{N}_\ell^\delta(a) - \mathcal{N}_{\ell-1}^\delta(a)\|_{L^2} \le \overline{M}^L \|b_\ell - b_\ell^\delta\|_{L^2}.$$

It now follows from the telescoping sum $\mathcal{N}(a) - \mathcal{N}^\delta(a) = \mathcal{N}_L^\delta(a) - \mathcal{N}_0^\delta(a) = \sum_{\ell=1}^L \left(\mathcal{N}_\ell^\delta(a) - \mathcal{N}_{\ell-1}^\delta(a)\right)$ that

$$\|\mathcal{N}(a) - \mathcal{N}^\delta(a)\|_{L^2} \le \sum_{\ell=1}^L \|\mathcal{N}_\ell^\delta(a) - \mathcal{N}_{\ell-1}^\delta(a)\|_{L^2}$$
$$\le \overline{M}^L \sum_{\ell=1}^L \|b_\ell - b_\ell^\delta\|_{L^2}.$$

Since the bound on the right-hand side is independent of a, and since $\lim_{\delta \to 0} \|b_\ell - b_\ell^\delta\|_{L^2} = 0$, choosing $\delta > 0$ sufficiently small, we obtain (80).

Step 2: Using Step 1, we will assume wlog that all biases are smooth, in the following. Given $a \in H^s$, let $a_\ell := \mathcal{L}_\ell \circ \dots \mathcal{L}_1 \circ \mathcal{R}(a)$. We claim that there exists a constant $B' > 0$, depending only on B, the activation function σ and the biases $b_\ell(x)$, such that $\|a_\ell\|_{H^s} \le B'$, provided that $\|a\|_{H^s} \le B$ and $\ell = 1, \dots, L$.

Indeed, it is a well-known fact that if $v \in H^s$, $s > d/2$, and if $\sigma \in C^{\lfloor s \rfloor + 1}$ (satisfied by assumption), then $\sigma \circ v \in H^s$.[2] In fact, there exists a monotonically increasing function C_σ, such that $\|\sigma \circ v\|_{H^s} \leq C_\sigma(\|v\|_{H^s})$. The claimed existence of B' now follows immediately from this result on compositions with σ, and the observation that the linear part of each layer define a continuous (bounded, affine) mapping $H^s \to H^s$, due to the finite width assumption, and the assumption that $b_\ell \in C^\infty \subset H^s$.

Step 3: Let ς_1, ς_2 be given such that $s \geq \varsigma_1 > \varsigma_2 > d/2$. Each layer defines a Hölder continuous mapping $\mathcal{L}_\ell : D_\ell \subset H^{\varsigma_1} \to H^{\varsigma_2}$, with Hölder exponent $\alpha = 1 - \varsigma_2/s$, and where D_ℓ denotes the image of the set $\{\|a\|_{H^s} \leq B\}$ under the previous $\ell - 1$ layers.

By Step 2, we have that $\|\mathcal{L}_\ell(v)\|_{H^s} \leq B'$ for all $v \in D_\ell$. Furthermore, as observed in Step 1, (79), $\mathcal{L}_\ell : L^2 \to L^2$ is Lipschitz continuous. It follows that for any $v \in D_\ell$, we have, by the interpolation inequality, that

$$
\begin{aligned}
\|\mathcal{L}_\ell(v) - \mathcal{L}_\ell(v')\|_{H^{\varsigma_2}} &\lesssim \|\mathcal{L}_\ell(v) - \mathcal{L}_\ell(v')\|_{L^2}^{1-\varsigma_2/s} \|\mathcal{L}_\ell(v) - \mathcal{L}_\ell(v')\|_{H^s}^{\varsigma_2/s} \\
&\lesssim_{\varsigma_2, B'} \|\mathcal{L}_\ell(v) - \mathcal{L}_\ell(v')\|_{L^2}^{\alpha} \\
&\lesssim_{\varsigma_2, B', \sigma, \ell} \|v - v'\|_{L^2}^{\alpha} \\
&\lesssim \|v - v'\|_{H^{\varsigma_1}}^{\alpha}.
\end{aligned}
$$

Step 4: We use Steps 1-3 to conclude that for a sufficiently large choice of $N \in \mathbb{N}$, the Ψ-FNO $\mathcal{N}^* : L_N^2 \to L_N^2$, given by

$$
\mathcal{N}^*(a) := \mathcal{Q} \circ \mathcal{I}_N \circ \mathcal{L}_L \circ \mathcal{I}_N \circ \mathcal{L}_{L-1} \circ \mathcal{I}_N \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{I}_N \circ \mathcal{R},
$$

satisfies

$$
\sup_{\|a\|_{H^s} \leq B} \|\mathcal{N}(a) - \mathcal{N}^*(a)\|_{L^2} \leq \epsilon.
$$

In the following, we will denote by $\mathcal{L}_\ell^* := \mathcal{I}_N \circ \mathcal{L}_\ell \circ \mathcal{I}_N$ the "pseudo-spectral projection" of the layer $\mathcal{L}_\ell$, and we observe that

$$
\mathcal{N}^* = \mathcal{Q} \circ \mathcal{L}_L^* \circ \cdots \circ \mathcal{L}_1^* \circ \mathcal{R}.
$$

Fix a sequence $s = \varsigma_0 > \varsigma_1 > \varsigma_2 > \cdots > \varsigma_L > d/2$. By Step 3, we may view the ℓ-th layer $\mathcal{L}_\ell$ as a Hölder continuous mapping $\mathcal{L}_\ell : H^{\varsigma_{\ell-1}} \to H^{\varsigma_\ell}$. Furthermore, since $\varsigma_\ell > d/2$ for all ℓ, we also have that $\mathcal{I}_N : H^{\varsigma_\ell} \to H^{\varsigma_\ell}$ is a bounded linear operator, with operator norm that is uniformly bounded in N. In particular, this implies that there exists a constant $\overline{C} \geq 1$, independent of N, such that, for all $a, a' \in D_\ell$:

$$
\begin{aligned}
\|\mathcal{L}_\ell^*(a) - \mathcal{L}_\ell^*(a')\|_{H^{\varsigma_\ell}} &= \|\mathcal{I}_N \mathcal{L}_\ell(\mathcal{I}_N a) - \mathcal{I}_N \mathcal{L}_\ell(\mathcal{I}_N a')\|_{H^{\varsigma_\ell}} \\
&\leq \|\mathcal{I}_N\| \|\mathcal{L}_\ell\|_{C^{1-\varsigma_\ell/s}} \|\mathcal{I}_N\|^{1-\varsigma_\ell/s} \|a - a'\|_{H^{\varsigma_{\ell-1}}}^{1-\varsigma_\ell/s} \\
&\leq \overline{C} \|a - a'\|_{H^{\varsigma_{\ell-1}}}^{1-\varsigma_\ell/s}.
\end{aligned}
$$

---

2. This is well-known in the integer case $s \in \mathbb{N}$, but remains true also for fractional Sobolev spaces (Brezis and Mironescu, 2001).

We now introduce $\mathcal{N}_\ell^* := \mathcal{Q} \circ \mathcal{L}_L^* \circ \cdots \circ \mathcal{L}_{\ell+1}^* \circ \mathcal{L}_\ell \circ \ldots \mathcal{L}_1 \circ \mathcal{R}$, for any $\ell = 0, \ldots, L$. Then, $\mathcal{N}(a) - \mathcal{N}^*(a) = \sum_{\ell=1}^L \left(\mathcal{N}_\ell^*(a) - \mathcal{N}_{\ell-1}^*(a)\right)$, and using the Hölder regularity of the $\mathcal{L}_\ell^*$, we obtain

$$\|\mathcal{N}_\ell^*(a) - \mathcal{N}_{\ell-1}^*(a)\|_{L^2} \le \|\mathcal{N}_\ell^*(a) - \mathcal{N}_{\ell-1}^*(a)\|_{H^{\varsigma_L}}$$
$$\le \text{Lip}(\mathcal{Q})\overline{C}^L \|\mathcal{L}_\ell^*(a') - \mathcal{L}_\ell^*(a')\|_{H^{\varsigma_\ell}}^{\beta_\ell},$$

where $a' := \mathcal{L}_{\ell-1} \circ \ldots \mathcal{L}_1 \circ \mathcal{R}(a)$ belongs to $D_\ell \subset \{a' \in H^s \,|\, \|a'\|_{H^s} \le B'\}$ (cp. Step 3), and $\beta_\ell := \prod_{k=\ell+1}^L \left(1 - \frac{\varsigma_k}{s}\right)$ is the Hölder exponent of the composition of the layers $\mathcal{L}_L^* \circ \cdots \circ \mathcal{L}_{\ell+1}^* : H^{\varsigma_{\ell+1}} \to H^{\varsigma_L}$. To estimate the last difference, we recall that if $\|a\|_{H^s} \le B$, then $\|a'\|_{H^s} \le B'$, by Step 2. In particular, it follows from the uniform Hölder continuity of the layers $\mathcal{L}_\ell$, established in Step 3 and the pseudo-spectral approximation estimate, that

$$\|\mathcal{L}_\ell(a') - \mathcal{L}_\ell^*(a')\|_{H^{\varsigma_\ell}} = \|\mathcal{L}_\ell(a') - \mathcal{I}_N \mathcal{L}_\ell(\mathcal{I}_N a')\|_{H^{\varsigma_\ell}}$$
$$\le \|\mathcal{L}_\ell(a') - \mathcal{I}_N \mathcal{L}_\ell(a')\|_{H^{\varsigma_\ell}} + \|\mathcal{I}_N\| \|\mathcal{L}_\ell(a') - \mathcal{L}_\ell(\mathcal{I}_N a')\|_{H^{\varsigma_\ell}}$$
$$\lesssim \frac{B'}{N^{-(s-\varsigma_\ell)}} + \|a' - \mathcal{I}_N a'\|_{H^{\varsigma_\ell-1}}^{1-\varsigma_\ell/s}$$
$$\lesssim \frac{B'}{N^{-(s-\varsigma_\ell)}} + \left(\frac{B'}{N^{-(s-\varsigma_{\ell-1})}}\right)^{(1-\varsigma_\ell/s)},$$

where the implied constant is independent of N. In particular, we have

$$\lim_{N\to\infty} \sup_{a'\in D_\ell} \|\mathcal{L}_\ell(a') - \mathcal{L}_\ell^*(a')\|_{H^{\varsigma_\ell}} = \lim_{N\to\infty} \sup_{a'\in D_\ell} \|\mathcal{L}_\ell(a') - \mathcal{L}_\ell(\mathcal{I}_N a')\|_{H^{\varsigma_\ell}} = 0,$$

for all $\ell = 1, \ldots, L$. Choosing N sufficiently large, we can thus ensure that

$$\sup_{\|a\|_{H^s}\le B} \|\mathcal{N}(a) - \mathcal{N}^*(a)\|_{L^2} \le \sum_{\ell=1}^L \sup_{\|a\|_{H^s}\le B} \|\mathcal{N}_\ell^*(a) - \mathcal{N}_{\ell-1}^*(a)\|_{L^2}$$
$$\le \text{Lip}(\mathcal{Q})\overline{C}^L \sum_{\ell=1}^L \sup_{a'\in D_\ell} \|\mathcal{L}_\ell(a') - \mathcal{L}_\ell^*(a')\|_{H^{\varsigma_\ell}}^{\beta_\ell}$$
$$\le \epsilon.$$

This concludes the proof. ∎

D.6 Proof of Lemma 16

Proof Let $\epsilon > 0$ be given. Let $\mathcal{G} : H^s \to H^{s'}$ be continuous operator, and $K \subset H^s$ a compact subset. We wish to show that there exists $N_0 \in \mathbb{N}$, such that for any $N \ge N_0$, there exists a Ψ-FNO $\mathcal{N}^* : L_N^2 \to L_N^2$ such that

$$\sup_{a\in K} \|\mathcal{G}(a) - \mathcal{N}^*(a)\|_{H^{s'}} \le \epsilon. \tag{81}$$

Choose $M \in \mathbb{N}$, such that

$$\sup_{a \in K} \|P_M \mathcal{G}(a) - \mathcal{G}(a)\|_{H^{s'}} \leq \epsilon/2. \tag{82}$$

Fix $\delta > 0$ for the moment. We will specify a suitable choice of $\delta = \delta(M, s', \mathbb{T}^d, \epsilon) > 0$ at the end of this proof. We emphasize that δ will depend only on parameters already introduced at this point of the proof. By assumption of the validity of the universal approximation theorem for Ψ-FNOs for $s' = 0$, we can choose $\widetilde{N}_0 = \widetilde{N}_0(\mathcal{G}, K, \delta) \in \mathbb{N}$, depending on $\mathcal{G}$, K and δ, such that for any $N \geq \widetilde{N}_0$, there exists a Ψ-FNO $\widetilde{\mathcal{N}}^* : L_N^2 \to L_N^2$, such that

$$\sup_{a \in K} \|\mathcal{G}(a) - \widetilde{\mathcal{N}}^*(a)\|_{L^2} \leq \delta.$$

We define $N_0 := \max(\widetilde{N}_0, M)$. By our choice of $\widetilde{N}_0$ and M, the constant N_0 depends only on the underlying operator $\mathcal{G}$, the compact set K and the parameter $\delta > 0$. Let $N \geq N_0$. We claim that if $\delta = \delta(M, s', \mathbb{T}^d, \epsilon) > 0$ is been chosen sufficiently small, then for any $N \geq N_0$, there exists a Ψ-FNO $\mathcal{N}^* : L_N^2 \to L_N^2$ satisfying (81).

To see this, let $N \geq N_0$, and let $\widetilde{\mathcal{N}}^* : L_N^2 \to L_N^2$ be a Ψ-FNO such that $\sup_{a \in K} \|\mathcal{G}(a) - \widetilde{\mathcal{N}}^*(a)\|_{L^2} \leq \delta$. Then we have

$$\sup_{a \in K} \|P_M \mathcal{G}(a) - P_M \widetilde{\mathcal{N}}^*(a)\|_{H^{s'}} \leq CM^{s'}\delta, \tag{83}$$

where $C = C(\mathbb{T}^d, s') > 0$ is independent of M, N and δ. Let $m \in \mathbb{N}$ be the smallest natural number strictly bigger than s' and $d/2$, i.e. $m > \max(s', d/2)$. Let

$$K' := \widetilde{\mathcal{N}}^*(K) \subset L_N^2, \tag{84}$$

be the compact image of $K \subset H^s$ under the (continuous) mapping $\widetilde{\mathcal{N}}^* : H^s \to L_N^2$. We note that since L_N^2 is a finite-dimensional space, the L^2-norm is equivalent to the H^m-norm on L_N^2. This implies that $K' \subset H^m$ is also compact when considered as a subset of H^m, with respect to norm $\|\cdot\|_{H^m}$. By Lemma 43, there exists a single-layer FNO $\mathcal{L} : H^m \to H^m$, such that

$$\sup_{v \in K'} \|P_M v - \mathcal{L}(v)\|_{H^m} \leq \delta. \tag{85}$$

We define a Ψ-FNO $\mathcal{N}^* : L_N^2 \to L_N^2$ by the following composition

$$\mathcal{N}^* := \mathcal{I}_N \circ \mathcal{L} \circ \mathcal{I}_N \circ \widetilde{\mathcal{N}}^* = \mathcal{I}_N \circ \mathcal{L} \circ \widetilde{\mathcal{N}}^*,$$

where $\mathcal{I}_N$ is the pseudo-spectral projection operator. We note that, since $N \geq N_0 \geq M$, we have $\mathcal{I}_N P_M = P_M$, and recall that for $m > d/2$, the mapping $\mathcal{I}_N : H^m \to H^m$ is continuous, with an operator norm that can be bounded independently of N. Hence, we can estimate, for any $a \in K$,

$$\|P_M \widetilde{\mathcal{N}}^*(a) - \mathcal{N}^*(a)\|_{H^{s'}} = \|\mathcal{I}_N \circ P_M \widetilde{\mathcal{N}}^*(a) - \mathcal{I}_N \circ \mathcal{L} \circ \widetilde{\mathcal{N}}^*(a)\|_{H^{s'}}$$
$$\overset{\underset{(m>s')}{\downarrow}}{\leq} \|\mathcal{I}_N \circ P_M \widetilde{\mathcal{N}}^*(a) - \mathcal{I}_N \circ \mathcal{L} \circ \widetilde{\mathcal{N}}^*(a)\|_{H^m}$$
$$\leq C\|P_M v - \mathcal{L}(v)\|_{H^m},$$

where $C = C(\mathbb{T}^d, m) > 0$ is independent of M and N, and where $v := \widetilde{\mathcal{N}}^*(a) \in K'$, with $K' = \widetilde{\mathcal{N}}^*(K)$ defined in (84). From (85), we conclude that

$$\sup_{a \in K} \|P_M \widetilde{\mathcal{N}}^*(a) - \mathcal{N}^*(a)\|_{H^{s'}} \leq C\delta, \tag{86}$$

where $C = C(\mathbb{T}^d, m) > 0$ is independent of M, N and δ. In fact, since m is defined as the smallest integer $> \max(s', d/2)$, the above constant only depends on s' and the dimension d of the domain, i.e. we have $C = C(\mathbb{T}^d, s', d)$. Combining (82), (83) and (86), we find that for $N \geq N_0$ and for any $\delta > 0$, there exists a Ψ-FNO $\mathcal{N}^* : L_N^2 \to L_N^2$, such that

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}^*(a)\|_{H^{s'}} \leq \epsilon/2 + C(1 + M^{s'})\delta,$$

where $C = C(\mathbb{T}^d, s', d) > 0$ is a constant independent of δ. Thus, if we choose $\delta = \delta(M, s', d, \mathbb{T}^d, \epsilon) > 0$ at the beginning of this proof sufficiently small to ensure that

$$C(1 + M^{s'})\delta \leq \epsilon/2,$$

then we conclude that for any $N \geq N_0$, there exists a Ψ-FNO $\mathcal{N}^* : L_N^2 \to L_N^2$, such that

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}^*(a)\|_{H^{s'}} \leq \epsilon.$$

This concludes the proof. ■

D.7 Technical Results on the Structure of Ψ-FNOs

Recall Section 2.5.1 where we introduced σ layers and $\mathcal{F}$-layers and claimed that Ψ-FNOs can be decomposed in terms of these layers. We have the following series of lemmas, which make this observation precise.

Lemma 44 *Let $\Phi : \mathbb{R}^{d_a} \to \mathbb{R}^{d_u}$ be a (ordinary) neural network with activation function σ. For any $N \in \mathbb{N}$, the mapping*

$$\mathcal{N}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_a}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_u}), \quad a(x) \mapsto \mathcal{N}^*(a)(x) := \mathcal{I}_N \Phi(a(x)),$$

can be represented by a Ψ-FNO, with

$$\mathrm{depth}(\mathcal{N}^*) = \mathrm{depth}(\Phi), \quad \mathrm{lift}(\mathcal{N}^*) = \mathrm{width}(\Phi).$$

The proof is straight-forward and completely analogous to the same statement for FNOs in Remark 3. We also note the following lemma, which allows us in practice to replace $\mathcal{F}$-layers by proper FNO layers:

Lemma 45 (Linear approximation lemma) *Assume that the activation function $\sigma \in C^2$ is twice continuously differentiable and non-constant. Let $\mathcal{L}^*$ be a linear Ψ-FNO layer of the form*

$$\mathcal{L}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_v}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_v}), \ \mathcal{L}^*(v_N) := W v_N(x_j) + b_j + \mathcal{F}_N^{-1}(P\mathcal{F}_N v_N)_j.$$

There exists a constant $C > 0$, independent of N, such that for any ϵ, $B > 0$, there exists a Ψ-FNO $\mathcal{N}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_v}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_v})$, such that

$$\sup_{\|v_N\|_{L^2} \leq B} \|\mathcal{N}^*(v_N) - \mathcal{L}^*(v_N)\|_{L^2} \leq \epsilon,$$

and

$$\mathrm{depth}(\mathcal{N}^*) \leq C, \quad \mathrm{lift}(\mathcal{N}^*) = d_v \leq C.$$

Proof This follows from the observation that, by assumption, there exists $x_0 \in \mathbb{R}$, such that $\sigma'(x_0) \neq 0$. By Taylor expansion, we have

$$\frac{\sigma(x_0 + hy) - \sigma(x_0 - hy)}{h\sigma'(x_0)} = y + R(h, y),$$

where $|R(h, y)| \leq C\overline{B}^2 h$ for any $y \in [-\overline{B}, \overline{B}]$ and $0 < h \leq 1$. Replacing y by the output of the linear layer $\widehat{\mathcal{L}}$, and choosing $\overline{B}$ sufficiently large, we obtain with

$$\mathcal{N}_h^*(v_n) := \frac{\sigma(x_0 + h\mathcal{L}^*(v_N)) - \sigma(x_0 - h\mathcal{L}^*(v_N))}{h\sigma'(x_0)},$$

that

$$\|\mathcal{L}^*(v_N) - \mathcal{N}_h^*(v_N)\|_{L^2} = O(h),$$

uniformly for all $v_N \in L_N^2(\mathbb{T}^d; \mathbb{R}^{d_v})$, such that $\|v_N\|_{L^2} \leq B$, and for $0 < h \leq 1$. In particular, choosing $h = h(\epsilon) > 0$ sufficiently small, we can ensure that

$$\|\mathcal{L}^*(v_N) - \mathcal{N}_h^*(v_N)\|_{L^2} \leq \epsilon,$$

for all $v_n \in L_N^2(\mathbb{T}^d; \mathbb{R}^{d_v})$, such that $\|v_N\|_{L^2} \leq B$. The proof is concluded by observing that the mapping $v_N \mapsto \mathcal{N}_h^*(v_N)$ defines a Ψ-FNO, and that the size (depth, lift) is uniformly bounded in h. ∎

According to Lemma 44, any composition of σ-layers can be identified with a Ψ-FNO. Lemma 45 shows that we can approximate $\mathcal{F}$-layers to arbitrary accuracy with a Ψ-FNO of uniformly bounded size. In the following lemma, we record the simple fact that a composition of Ψ-FNOs is again representable by a Ψ-FNO:

Lemma 46 (Composition Lemma) *For $p \in \mathbb{N}$, let $\mathcal{N}_1^*, \ldots, \mathcal{N}_p^*$ be Ψ-FNOs, defined with respect to the same grid $\{x_j\}_{j \in \mathcal{J}_N}$, and defining operators $\mathcal{N}_k^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_{k-1}}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_k})$ for $d_0, \ldots, d_p \in \mathbb{N}$, such that the composition $\mathcal{N}_p^* \circ \ldots \mathcal{N}_1^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_0}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_p})$ is well-defined. Then there exists a Ψ-FNO $\mathcal{N}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_0}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_p})$, such that $\mathcal{N}^* = \mathcal{N}_p^* \circ \cdots \circ \mathcal{N}_1^*$, and such that $\mathrm{depth}(\mathcal{N}^*) \leq \sum_{k=1}^p \mathrm{depth}(\mathcal{N}_k^*)$, and $\mathrm{lift}(\mathcal{N}^*) \leq \max_{k=1, \ldots, p} \mathrm{lift}(\mathcal{N}_k^*)$.*

The proof of the previous lemma is straight-forward, and only requires a padding of each layer with zeros to achieve a uniform lifting dimension d_v across all layers. The following lemma will be central for our approximation and complexity estimates:

Lemma 47 (Replacement Lemma) *Assume that σ is Lipschitz continuous. For $p \in \mathbb{N}$, let $\widetilde{\mathcal{N}}_1, \ldots, \widetilde{\mathcal{N}}_p$ be continuous operators $\widetilde{\mathcal{N}}_k : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_{k-1}}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_k})$ for $d_0, \ldots, d_p \in \mathbb{N}$, such that the composition $\widetilde{\mathcal{N}} := \widetilde{\mathcal{N}}_p \circ \cdots \circ \widetilde{\mathcal{N}}_1 : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_0}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_p})$ is well-defined. Assume that there exist constants $D_k, L_k > 0$, $k = 1, \ldots, p$, such that for any $\epsilon, M > 0$, there exists a Ψ-FNO $\mathcal{N}_k^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_{k-1}}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_k})$, such that*

$$\sup_{\|u\|_{L^2} \leq M} \|\widetilde{\mathcal{N}}_k(u) - \mathcal{N}_k^*(u)\|_{L_N^2} \leq \epsilon,$$

and $\operatorname{depth}(\mathcal{N}_k^*) \leq D_k$, $\operatorname{lift}(\mathcal{N}_k^*) \leq L_k$. *Then for any $\epsilon > 0$ and $M > 0$, there exists a Ψ-FNO $\mathcal{N}^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_0}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_p})$, such that*

$$\sup_{\|u\|_{L^2} \leq M} \|\widetilde{\mathcal{N}}(u) - \mathcal{N}^*(u)\|_{L_N^2} \leq \epsilon,$$

and $\operatorname{depth}(\mathcal{N}^*) \leq \sum_{k=1}^p D_k$, *and* $\operatorname{lift}(\mathcal{N}^*) \leq \max_{k=1,\ldots,p} L_k$.

Proof We prove the statement by induction on $p \in \mathbb{N}$. We start the induction at $p = 2$: Let $\epsilon > 0$ and $M > 0$ be given, and let $\widetilde{\mathcal{N}}_1, \widetilde{\mathcal{N}}_2$ satisfy the hypotheses of this lemma. Since $B_M(0) \subset L_N^2(\mathbb{T}^d; \mathbb{R}^{d_0})$, the closed ball of radius M, is compact in the finite-dimensional space $L_N^2(\mathbb{T}^d; \mathbb{R}^{d_0})$, and since $\widetilde{\mathcal{N}}_1$ is continuous, it follows that the image $\widetilde{\mathcal{N}}_1(B_M(0))$ is also compact; in particular, there exists $M_1 > 0$, such that $\widetilde{\mathcal{N}}_1(B_M(0)) \subset B_{M_1}(0)$. By assumption, there exists a Ψ-FNO $\mathcal{N}_2^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_1}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_2})$, such that

$$\sup_{\|u\|_{L^2} \leq 2M_1} \|\widetilde{\mathcal{N}}_2(u) - \mathcal{N}_2^*(u)\|_{L_N^2} \leq \epsilon/2.$$

We note that (for Lipschitz continuous σ) $\widetilde{\mathcal{N}}_2$ is Lipschitz continuous on $B_{2M_1}(0) \subset L_N^2$. Let $\operatorname{Lip}(\widetilde{\mathcal{N}}_2)$ denote the corresponding Lipschitz constant. By assumption on $\widetilde{\mathcal{N}}_1$, there exists a Ψ-FNO $\mathcal{N}_1^* : L_N^2(\mathbb{T}^d; \mathbb{R}^{d_0}) \to L_N^2(\mathbb{T}^d; \mathbb{R}^{d_1})$, such that

$$\sup_{\|u\|_{L^2} \leq M} \|\widetilde{\mathcal{N}}_1(u) - \mathcal{N}_1^*(u)\|_{L_N^2} \leq \min\left(\frac{\epsilon}{2\operatorname{Lip}(\widetilde{\mathcal{N}}_2)}, M_1\right).$$

Note that this estimate implies in particular that

$$\|\mathcal{N}_1^*(u)\|_{L^2} \leq \|\widetilde{\mathcal{N}}_1(u) - \mathcal{N}_1^*(u)\|_{L^2} + \|\widetilde{\mathcal{N}}_1(u)\|_{L^2} \leq 2M_1.$$

Thus, we can estimate

$$\sup_{\|u\|_{L^2} \leq M} \|\widetilde{\mathcal{N}}_2 \circ \widetilde{\mathcal{N}}_1(u) - \mathcal{N}_2^* \circ \mathcal{N}_1^*(u)\|_{L^2}$$

$$\leq \sup_{\|u\|_{L^2} \leq M} \|\widetilde{\mathcal{N}}_2 \circ \widetilde{\mathcal{N}}_1(u) - \widetilde{\mathcal{N}}_2 \circ \mathcal{N}_1^*(u)\|_{L^2}$$

$$+ \sup_{\|u\|_{L^2} \leq M} \|\widetilde{\mathcal{N}}_2 \circ \mathcal{N}_1^*(u) - \mathcal{N}_2^* \circ \mathcal{N}_1^*(u)\|_{L^2}$$

$$\leq \operatorname{Lip}(\widetilde{\mathcal{N}}_2) \sup_{\|u\|_{L^2} \leq M} \|\widetilde{\mathcal{N}}_1(u) - \mathcal{N}_1^*(u)\|_{L^2}$$

$$+ \sup_{\|v\| \leq 2M_1} \|\widetilde{\mathcal{N}}_2(v) - \mathcal{N}_2^*(v)\|_{L^2}$$

$$< \epsilon.$$

The estimate on the depth and lift is immediate (cp. Lemma 46). This proves the base case $p = 2$ of the induction.

Let now $p > 2$, and assume that the claim holds for the composition of $p - 1$ operators. Given $\widetilde{\mathcal{N}}_1, \ldots, \widetilde{\mathcal{N}}_p$ as in the statement of the lemma, we can apply the induction hypothesis for the composition of $p - 1$ operators, to see that the two continuous operators $\widehat{\mathcal{N}}_1 := \widetilde{\mathcal{N}}_1$ and $\widehat{\mathcal{N}}_2 := (\widetilde{\mathcal{N}}_p \circ \cdots \circ \widetilde{\mathcal{N}}_2)$ fulfill all assumptions of the lemma (with $p = 2$), and with depths and lifts of the approximating FNOs given by $\widehat{D}_1 = D_1$, $\widehat{L}_1 = L_1$, and $\widehat{D}_2 = \sum_{k=2}^{p} D_k$, $\widehat{L}_2 = \max_{k=2,\ldots,p} L_k$. The proof of the induction step now follows from the base case (with $p = 2$) already considered above. ∎

Appendix E. Technical Results and Proofs for Section 3.1

In this section, we collect some technical results and proofs for the material in Section 3.1 of the main text. We start with the following Lemma on the contraction property of the map (28).

Lemma 48 *Let $k \in \mathbb{N}$, $k > d/2 + 1$, and let $s > d/2 + k$ be given. Assume that the coefficient $a \in H^s(\mathbb{T}^d)$ satisfies the λ-coercivity condition (25) for some $\lambda \in (0,1)$, and that the right-hand side of the stationary Darcy equation (24) belongs to the Sobolev space $f \in H^{k-1}$. Then there exists $N_0 = N_0(s, d, \|a\|_{H^s}, \lambda) \in \mathbb{N}$, such that for any $N \geq N_0$, we have $a_N \geq \lambda/2$ and the mapping F_N defined by (28) is a contraction, with*

$$\mathrm{Lip}\left(F_N : \dot{H}^1 \to \dot{H}^1\right) \leq 1 - \frac{\lambda}{2}.$$

Proof As $u_N \mapsto F_N(u_N)$ is an affine mapping, we can express the Lipschitz constant in terms of the following supremum

$$\mathrm{Lip}\left(F_N : \dot{H}^1 \to \dot{H}^1\right) = \sup_{u_N \in \dot{H}^1 \backslash \{0\}} \frac{\left\|\dot{P}_N(-\Delta)^{-1} \nabla \cdot (\widetilde{a}_N \nabla u_N)\right\|_{\dot{H}^1}}{\|u_N\|_{\dot{H}^1}}.$$

We now observe that

$$\left\|\dot{P}_N(-\Delta)^{-1} \nabla \cdot (\widetilde{a}_N \nabla u_N)\right\|_{\dot{H}^1} \leq \left\|(-\Delta)^{-1} \nabla \cdot (\widetilde{a}_N \nabla u_N)\right\|_{\dot{H}^1}$$
$$= \|\widetilde{a}_N \nabla u_N\|_{L^2}$$
$$\leq \|\widetilde{a}_N\|_{L^\infty} \|u_N\|_{\dot{H}^1},$$

for all $u_N \in \dot{H}^1_N(\mathbb{T}^d)$. To finish the proof, we note that

$$\|\widetilde{a}_N\|_{L^\infty} \leq \|\widetilde{a}\|_{L^\infty} + \|\widetilde{a} - \widetilde{a}_N\|_{L^\infty},$$

where the first term is bounded by $1 - \lambda$ by the coercivity assumption. By our definition of $\widetilde{a}_N$ (cp. equation 27), we have

$$\|\widetilde{a} - \widetilde{a}_N\|_{L^\infty} \leq \left\|\widetilde{a} - \dot{P}_N \widetilde{a}\right\|_{L^\infty} + \left\|\dot{P}_N \widetilde{a} - \dot{P}_N \mathcal{I}_{2N} \widetilde{a}\right\|_{L^\infty}.$$

But by the spectral/pseudo-spectral approximation estimates, we now have

$$\left\| \widetilde{a} - \dot{P}_N \widetilde{a} \right\|_{L^\infty} \lesssim_{d,\delta} \left\| \widetilde{a} - \dot{P}_N \widetilde{a} \right\|_{H^{d/2+\delta}}$$
$$\lesssim_{s,d,\delta} N^{-(s-d/2-\delta)} \|a\|_{H^s},$$

and

$$\left\| \dot{P}_N \widetilde{a} - \dot{P}_N \mathcal{I}_{2N} \widetilde{a} \right\|_{L^\infty} \lesssim_{d,\delta} \left\| \dot{P}_N \widetilde{a} - \dot{P}_N \mathcal{I}_{2N} \widetilde{a} \right\|_{H^{d/2+\delta}}$$
$$\leq \| \widetilde{a} - \mathcal{I}_{2N} \widetilde{a} \|_{H^{d/2+\delta}}$$
$$\lesssim_{s,d,\delta} N^{-(s-d/2-\delta)} \|a\|_{H^s}.$$

To be definite, let us choose $\delta := (s - d/2)/2$. In particular, it follows that there exists a constant $C = C(s, d)$, and $N_0 = N_0(s, d, \|a\|_{H^s}, \lambda) \in \mathbb{N}$, such that

$$\|\widetilde{a} - \widetilde{a}_N\|_{L^\infty} \leq \frac{\lambda}{2},$$

for $N \geq N_0$, and hence $a_N = 1 + \widetilde{a}_N \geq 1 + \widetilde{a} - \lambda/2 \geq \lambda/2$, and

$$\mathrm{Lip}(F_N) \leq \|\widetilde{a}_N\|_{L^\infty} \leq \|\widetilde{a}\|_{L^\infty} + \|\widetilde{a} - \widetilde{a}_N\|_{L^\infty} \leq 1 - \frac{\lambda}{2},$$

for all $N \geq N_0$. The claim follows. ∎

Next, we provide the detailed proof of Theorem 24 which guarantees convergence of the Algorithm 23 to the solutions of the Darcy equation (24).

E.1 Proof of Theorem 24

Proof Choose $N_0 \in \mathbb{N}$ as in Lemma 48.

Step 1: Well-posedness and error estimate of the FG approximation (26).

We consider the bilinear form $B : \dot{H}_N^1 \times \dot{H}_N^1 \to \mathbb{R}$, given by

$$B(u_N, w_N) := \int_{\mathbb{T}^d} a_N(x) \nabla u_N(x) \cdot \nabla w_N(x) \, dx.$$

We note that u_N solves (26), if and only if,

$$B(u_N, w_N) = \langle f_N, w_N \rangle_{L^2}, \tag{87}$$

for all $w_N \in \dot{H}_N^1$. By Lemma 48, we have $a_N \geq \lambda/2$ for all $N \geq N_0$, and hence

$$B(w_N, w_N) \geq \lambda/2 \|w_N\|_{\dot{H}^1}^2.$$

Thus, B is a coercive bilinear form. The existence and uniqueness of a solution $u_N^* \in \dot{H}_N^1$ of (87) follows from this for any right-hand side f_N. Furthermore, we have the estimate

$$\|u_N^*\|_{H^1}^2 \leq 2\|u_N^*\|_{\dot{H}^1}^2 \leq \frac{4}{\lambda} B(u_N^*, u_N^*) = \frac{4}{\lambda} \langle u_N^*, f_N \rangle_{L^2} \leq \frac{4}{\lambda} \|u_N^*\|_{H^1} \|f_N\|_{H^{-1}},$$

and thus, $\|u_N^*\|_{H^1} \leq 4\lambda^{-1}\|f_N\|_{H^{-1}}$.

It is also straightforward to show that by inductive differentiation of (24), one can also obtain higher-order estimates for $k \in \mathbb{N}$, $k \geq 1$ for the (non-discretized) elliptic equation (24), of the form:

$$\|u\|_{H^{k+1}} \leq C\|f\|_{H^{k-1}}, \tag{88}$$

where $C = C(k, d, \|a\|_{C^k}) > 0$.

To prove the claimed error estimate, we note that if u solves (24), then $P_N u \in \dot{H}_N^1$ solves

$$-P_N \nabla \cdot (a_N \nabla P_N u) = P_N f + P_N \nabla \cdot ((a - a_N)\nabla P_N u) \\ + P_N \nabla \cdot (a\nabla(1 - P_N)u).$$

It thus follows that $w_N = P_N u - u_N^*$ solves

$$-P_N \nabla \cdot (a_N \nabla w_N) = (P_N f - f_N) + P_N \nabla \cdot ((a - a_N)\nabla P_N u) \\ + P_N \nabla \cdot (a\nabla(1 - P_N)u) \\ =: (I) + (II) + (III).$$

The stability estimate then implies that $\|w_N\|_{H^1} \leq 4\lambda^{-1}\|(I) + (II) + (III)\|_{H^{-1}}$, can be bounded in terms of the H^{-1}-norm of the right-hand side. We can now estimate

$$\|(I)\|_{H^{-1}} = \|(P_N f - f_N)\|_{H^{-1}} \\ \leq \|(1 - P_N)f\|_{H^{-1}} + \|f - f_N\|_{H^{-1}} \lesssim_{d,s} N^{-k}\|f\|_{H^{k-1}}.$$

The last inequality follows from the fact that $f \in H^{k-1}(\mathbb{T}^d)$, for $k - 1 > d/2$, and the (pseudo-)spectral approximation estimate. For the second term, we fix a small $\delta > 0$, and obtain

$$\|(II)\|_{H^{-1}} \leq \|(a - a_N)\nabla P_N u\|_{L^2} \\ \lesssim_d \|a - a_N\|_{L^\infty}\|u\|_{H^1} \\ \lesssim_{d,\delta,\lambda} \|a - a_N\|_{H^{d/2+\delta}}\|f\|_{H^{-1}},$$

where we have used the embedding $H^{d/2+\delta} \hookrightarrow L^\infty$ in the last step. Assuming that $\delta > 0$ is chosen sufficiently small, so that $s \geq d/2 + \delta$, we can then estimate

$$\|(II)\|_{H^{-1}} \lesssim_{s,d} \lambda^{-1} N^{-(s-d/2-\delta)}\|a\|_{H^s}\|f\|_{H^{-1}}.$$

Finally, the third term can be bounded as follows:

$$\|(III)\|_{H^{-1}} \lesssim_d \|a\nabla(1 - P_N)u\|_{L^2} \\ \lesssim_d \|a\|_{L^\infty}\|(1 - P_N)u\|_{H^1} \\ \lesssim_{s,d} N^{-k}\|a\|_{H^s}\|u\|_{H^{k+1}} \\ \lesssim_{s,d,\lambda,k,\|a\|_{C^k}} N^{-k}\|a\|_{H^s}\|f\|_{H^{k-1}}.$$

In the last step, we used the higher regularity estimate (88). We note that for this step we assume that $a \in H^s$, $s > d/2 + k$, so that we have an embedding $H^s \hookrightarrow C^k$. Under this condition, we can further estimate (II) by

$$\|(II)\|_{H^{-1}} \lesssim_{s,d} N^{-(s-d/2-\delta)} \lambda^{-1} \|a\|_{H^s} \|f\|_{H^{-1}}$$
$$\lesssim_{s,d,\lambda,\|a\|_{H^s},\|f\|_{H^{-1}}} N^{-k}.$$

Combining the above estimates for (I), (II) and (III), we conclude that for any $s \in \mathbb{R}$ and $k \in \mathbb{N}$ satisfying $s > d/2 + k$, we have

$$\|w_N\|_{H^1} = \|u_N^* - P_N u\|_{H^1} \leq C N^{-k},$$

provided that $a \in H^s$, $f \in H^{k-1}$, where $C = C\left(s, d, \lambda, k, \|a\|_{H^s}, \|f\|_{H^{k-1}}\right) > 0$ is independent of N.

The proof is now finished by observing that

$$\|u - u_N^*\|_{H^1}^2 = \|(1 - P_N)u\|_{H^1}^2 + \|P_N u - u_N^*\|_{H^1}^2,$$

and that, by the higher-regularity estimate (88), we have

$$\|(1 - P_N)u\|_{H^1} \lesssim_{k,d} N^{-k} \|u\|_{H^{k+1}} \lesssim_{s,k,d,\lambda,\|a\|_{H^s},\|f\|_{H^{k-1}}} N^{-k}.$$

Thus, it follows that there exists a constant $C = C(s, k, d, \lambda, \|a\|_{H^s}, \|f\|_{H^{k-1}}) > 0$, such that

$$\|u - u_N^*\|_{H^1} \leq C N^{-k}, \tag{89}$$

where u_N^* is the solution of the Fourier-Galerkin discretization (26).

Step 2: Picard iteration estimate.

The output of Algorithm 23 is obtained by Picard iteration, i.e. by K-fold application of the mapping $F_N : \dot{H}_N^1 \to \dot{H}_N^1$, yielding a recursively defined sequence $u_N^0 := 0$, and $u_N^k = F_N(u_N^{k-1})$, for $k = 1, \ldots, K$. Since F_N is a contraction with $\text{Lip}(F_N) \leq 1 - \lambda/2$ for $N \geq N_0$, by Lemma 48, and since u_N^* is the unique fixed point of F_N, this implies that

$$\|u_N^K - u_N^*\|_{\dot{H}^1} \leq (1 - \lambda/2)^K \|u_N^*\|_{\dot{H}^1}.$$

Using the definition of H^1, $\dot{H}^1$, and the regularity estimate (88), we can further estimate

$$\|u_N^K - u_N^*\|_{H^1} \leq 2\|u_N^K - u_N^*\|_{\dot{H}^1} \leq 8\lambda^{-1}(1 - \lambda/2)^K \|f_N\|_{H^{-1}}.$$

By definition of K, we have

$$K \geq \frac{\log(\lambda N^{-k})}{\log(1 - \lambda/2)},$$

and hence $(1 - \lambda/2)^K \leq \lambda N^{-k}$. This yields

$$\|u_N^K - u_N^*\|_{H^1} \leq 8 N^{-k} \|f_N\|_{H^{-1}}.$$

Combining Steps 1 and 2, implies that there exists a constant $C > 0$, depending on s, k, d, λ, $\|a\|_{H^s}$ and $\|f\|_{H^{k-1}}) > 0$, such that

$$\|u - u_N\|_{H^1} \leq CN^{-k},$$

for all $N \geq N_0$, where $u_N := u_N^K$ is the output of Algorithm 23. This is the claimed estimate. ∎

E.2 Neural Network Approximation of Quadratic Non-Linearities

Our next aim is to prove Theorem 26. As stated in the main text, the proof relies crucially on the following Lemmas, which show that neural networks can efficiently approximate certain quadratic non-linearities. We start with the following result,

Lemma 49 *Let $\sigma \in C^3$ be a activation function. Let $d \in \mathbb{N}$. There exists a constant $C = C(d) > 0$, such that for any $\epsilon > 0$ and $B > 0$, there exists*

1. *a neural network $\Phi_1 : \mathbb{R}^2 \to \mathbb{R}$, such that*

$$\sup_{|a|,|b| \leq B} |\Phi_1(a,b) - ab| \leq \epsilon,$$

   and $\mathrm{width}(\Phi_1) \leq C$, $\mathrm{depth}(\Phi_1) \leq C$,

2. *a neural network $\Phi_2 : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$, such that*

$$\sup_{|a|,\|v\|_{\ell^2} \leq B} |\Phi_2(a,v) - av| \leq \epsilon,$$

   and $\mathrm{width}(\Phi_2) \leq C$, $\mathrm{depth}(\Phi_2) \leq C$,

3. *a neural network $\Phi_3 : \mathbb{R}^d \times \mathbb{R}^{d \times d} \to \mathbb{R}^d$, such that*

$$\sup_{\|v\|_{\ell^2},\|U\|_{\ell^2 \to \ell^2} \leq B} |\Phi_3(v,U) - v \cdot U| \leq \epsilon,$$

   and $\mathrm{width}(\Phi_3) \leq C$, $\mathrm{depth}(\Phi_3) \leq C$.

Proof Points (2) and (3) easily follow from (1), by parallelizing multiple networks Φ_1. To see the first claim (1), we note that the quadratic function $y \mapsto y^2$ can be approximated for $y \in [-B, B]$, $B > 0$, to arbitrary precision by finite differences

$$y^2 = \underbrace{\frac{\sigma(x+hy) - 2\sigma(x) + \sigma(x-hy)}{h^2 \sigma^{(2)}(x)}}_{\mathrm{sq}_h(y)} + R(h;y),$$

where we assume that x is chosen so that the second derivative $\sigma^{(2)}(x) \neq 0$, and $|R(h,y)| \leq Ch$ for all $y \in [-B, B]$, and $C = C(B)$. Finally, following (Yarotsky, 2017), we observe that the product ab of two numbers a, b can be expressed in the form

$$ab = \frac{1}{2}\left((a+b)^2 - a^2 - b^2\right) = \frac{1}{2}\left(\mathrm{sq}_h(a+b) - \mathrm{sq}_h(a) - \mathrm{sq}_h(b)\right) + \widetilde{R}(h;a,b),$$

where $\widetilde{R}$ is related to R, and there exists a constant $C = C(B)$, such that $|\widetilde{R}(h; a, b)| \leq Ch$, for all $a, b \in [-B, B]$. Since sq_h is a neural network of finite width and depth (independent of h), and since the last expression

$$\Phi_h(a, b) := \frac{1}{2} \left(\mathrm{sq}_h(a + b) - \mathrm{sq}_h(a) - \mathrm{sq}_h(b) \right),$$

is simply a linear combination of sq_h, we conclude that $\widehat{\mathcal{N}}_h$ is a neural network of uniformly bounded width and depth (uniform in h), and for sufficiently small $h > 0$, we have

$$\sup_{a,b \in [-B,B]} |\Phi_h(a, b) - ab| \leq \epsilon.$$

This concludes the proof. ∎

Using the above lemma, one can prove the following result,

Lemma 50 *Assume that the activation function $\sigma \in C^3$ is three times continuously differentiable and non-linear. There exists a constant $C > 0$, such that for any $N \in \mathbb{N}$, and for any $\epsilon, B > 0$, there exists a Ψ-FNO $\mathcal{N}^* : L^2_{2N}(\mathbb{T}^d; \mathbb{R}) \times L^2_{2N}(\mathbb{T}^d; \mathbb{R}) \to L^2_{2N}(\mathbb{T}^d; \mathbb{R})$, with*

$$\mathrm{depth}(\mathcal{N}^*), \ \mathrm{lift}(\mathcal{N}^*) = d_v \leq C,$$

such that we have

$$\|P_N (a_N \nabla u_N) - \mathcal{N}^*(a_N, u_N)\|_{L^2_N} \leq \epsilon,$$

for all trigonometric polynomials $a_N, u_N \in L^2_N(\mathbb{T}^d; \mathbb{R}) \subset L^2_{2N}(\mathbb{T}^d; \mathbb{R})$ of degree $|k|_\infty \leq N$, satisfying the bound $\|a_N\|_{L^2}, \|u_N\|_{L^2} \leq B$.

Proof First, we observe that there exists a linear FNO layer $\mathcal{L}^* : L^2_{2N} \to L^2_{2N}$, with a suitable choice of the Fourier multiplier matrix P, such that

$$\mathcal{L}^*(u_N) = \mathcal{F}_{2N}^{-1}(P\mathcal{F}(u_N)) = \mathcal{F}_{2N}^{-1} \left(\sum_{|k|_\infty \leq 2N} ik \widehat{u}_k e^{i\langle k, x \rangle} \right) = \nabla u_N,$$

is satisfied *exactly* for all $u_N \in L^2_{2N}(\mathbb{T}^d; \mathbb{R})$. We also note that if $\|u_N\|_{L^2} \leq B$, then $\|\nabla u_N\|_{L^2} \leq NB$. By the fact that all norms are equivalent on the finite-dimensional space $L^2_{2N}(\mathbb{T}^d; \mathbb{R}^{d \times d})$, there exists a constant $C' > 0$ (depending on N), such that

$$\sup_{x \in \mathbb{T}^d} \|\nabla u_N(x)\|_{\ell^2} \leq C' \|\nabla u_N\|_{L^2} \leq C'NB,$$

whenever $\|u_N\|_{L^2} \leq B$. We similarly see that by norm equivalence on $L^2_{2N}(\mathbb{T}^d; \mathbb{R})$, there exists a constant $C'' > 0$ (depending on N), such that we also have

$$\sup_{x \in \mathbb{T}^d} |a_N(x)| \leq C'' \|a_N\|_{L^2} \leq C''B,$$

for any $\|a_N\|_{L^2} \leq B$. Let $\overline{B} := \max(C'NB, C''B)$. By Lemma 49, there exists an ordinary neural network $\Phi : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ with $\mathrm{width}(\Phi), \mathrm{depth}(\Phi) \leq C(d)$ (with $C = C(d)$ *independent* of N), such that

$$\sup_{|a|, \|v\|_{\ell^2} \leq \overline{B}} \|\Phi(a, v) - av\|_{\ell^2} \leq \epsilon.$$

By Lemmas 44, 45 and the composition lemma 46, the composition

$$(a_N, u_N) \mapsto (a_N, \mathcal{L}^*(u_N)) = (a_N, \nabla u_N) \mapsto \Phi(a_N, \nabla u_N),$$

can be represented by a Ψ-FNO $\widetilde{\mathcal{N}}^* : L_{2N}^2(\mathbb{T}^d; \mathbb{R}^2) \to L_{2N}^2(\mathbb{T}^d; \mathbb{R}^d)$, and by construction, we have

$$\sup_{x \in \mathbb{T}^d} \|a_N(x)\nabla u_N(x) - \widetilde{\mathcal{N}}^*(a_N, u_N)(x)\|_{\ell^2} \leq \epsilon, \tag{90}$$

for all a_N, $u_N \in L_{2N}^2$, with $\|a_N\|_{L^2}, \|u_N\|_{L^2} \leq B$. Furthermore, since $\mathcal{L}^*$ is a linear layer, and Φ is an ordinary neural network with width(Φ), depth$(\Phi) \leq C = C(d)$, we in fact conclude that for some new constant $C = C(d) > 0$, we have

$$\text{depth}(\widetilde{\mathcal{N}}^*) \leq C, \quad \text{lift}(\widetilde{\mathcal{N}}^*) = d_v \leq C.$$

Finally, we note that the projection $P_N : L_{2N}^2 \to L_{2N}^2$ onto Fourier modes with wavenumbers $|k|_\infty \leq N$ can again be represented exactly by a linear Ψ-FNO layer $\widehat{\mathcal{L}}^*$, and by Lemmas 45, 46, there exists a Ψ-FNO $\mathcal{N}^* : L_{2N}^2(\mathbb{T}^d; \mathbb{R}^2) \to L_{2N}^2(\mathbb{T}^d; \mathbb{R}^d)$, such that $\|\mathcal{N}^*(a_N, u_N) - \widehat{\mathcal{L}}^* \circ \widetilde{\mathcal{N}}^*(a_N, u_N)\|_{L^2} \leq \epsilon$, for all $\|a_N\|_{L^2}, \|u_N\|_{L^2} \leq B$, and such that

$$\text{depth}(\mathcal{N}^*) \leq C, \quad \text{lift}(\mathcal{N}^*) = d_v \leq C,$$

where $C = C(d) > 0$ depends only on d. Combining this with (90), we conclude that $\mathcal{N}^*$ satisfies

$$
\begin{aligned}
\|P_N(a_N \nabla u_N) - \mathcal{N}^*(a_N, u_N)\|_{L^2} &\leq \|\widehat{\mathcal{L}}^* \circ \widetilde{\mathcal{N}}^*(a_N, u_N) - \mathcal{N}^*(a_N, u_N)\|_{L^2} \\
&\quad + \|P_N(a_N \nabla u_N) - \widehat{\mathcal{L}}^* \circ \widetilde{\mathcal{N}}^*(a_N, u_N)\|_{L^2} \\
&= \|\widehat{\mathcal{L}}^* \circ \widetilde{\mathcal{N}}^*(a_N, u_N) - \mathcal{N}^*(a_N, u_N)\|_{L^2} \\
&\quad + \|P_N(a_N \nabla u_N) - P_N \widetilde{\mathcal{N}}^*(a_N, u_N)\|_{L^2} \\
&\leq \epsilon + \|a_N \nabla u_N - \widetilde{\mathcal{N}}^*(a_N, u_N)\|_{L^2} \\
&\leq \epsilon + |\mathbb{T}^d|^{1/2} \sup_{x \in \mathbb{T}^d} \|a_N(x)\nabla u_N(x) - \widetilde{\mathcal{N}}^*(a_N, u_N)(x)\|_{\ell^2} \\
&\leq \left(1 + (2\pi)^{d/2}\right)\epsilon,
\end{aligned}
$$

for all $\|a_N\|_{L^2}, \|u_N\|_{L^2} \leq B$. Since $\epsilon > 0$ was arbitrary, the claim follows. ∎

E.3 Proof of Theorem 26

The stage is now set for the proof of Theorem 26 in the following,

Proof [Proof of Theorem 26] Since the claim is an asymptotic statement, it suffices to consider $N \geq N_0$, where N_0 is the constant of Lemma 48. Indeed, the exceptional cases $N < N_0$ can be handled by suitably enlarging the constant C. We will thus assume that $N \geq N_0$, and $f \in \dot{H}^{k-1}$ are given. We fix $F_N := (-\Delta)f_N \in L_N^2(\mathbb{T}^d)$ for the rest of this proof, where $f_N := \dot{P}_N \mathcal{I}_{2N} f$ is defined as in Algorithm 23.

We define two operators $\widehat{\mathcal{N}}_1$, $\widehat{\mathcal{N}}_2$ as follows: We let

$$\widehat{\mathcal{N}}_1 : L^2_{2N}(\mathbb{T}^d; \mathbb{R}^2) \to L^2_{2N}(\mathbb{T}^d; \mathbb{R} \times \mathbb{R}^d), \quad \widehat{\mathcal{N}}_1(a, u) := (a, P_N(a\nabla u)),$$

and define $\widehat{\mathcal{N}}_2 : L^2_{2N}(\mathbb{T}^d; \mathbb{R} \times \mathbb{R}^d) \to L^2_{2N}(\mathbb{T}^d; \mathbb{R}^2)$ by

$$\widehat{\mathcal{N}}_2(a, U) := \left(a, \dot{P}_N(-\Delta)^{-1}\nabla \cdot U + F_N \right).$$

In terms of $\widehat{\mathcal{N}}_1$, $\widehat{\mathcal{N}}_2$, Algorithm 23, which defines a mapping $a \mapsto u_N = \widehat{\mathcal{N}}(a)$, can be written in the form

$$\widehat{\mathcal{N}}(a) = \widehat{\mathcal{Q}} \circ \underbrace{\widehat{\mathcal{N}}_2 \circ \widehat{\mathcal{N}}_1 \circ \cdots \circ \widehat{\mathcal{N}}_2 \circ \widehat{\mathcal{N}}_1}_{K\text{-fold composition}} \circ \widehat{\mathcal{R}}(a),$$

where $\widehat{\mathcal{R}}(a) := (a, 0)$, $\widehat{\mathcal{Q}}(a, u) := u$ and where $K \lesssim \log(N)$. By the composition lemma 46, to prove the claim of this theorem, it therefore suffices to show the following

Claim: For any $B > 0$, there exists $C > 0$, such that for any $\epsilon > 0$, there exist Ψ-FNOs $\mathcal{N}_1^*$ and $\mathcal{N}_2^*$, with $\text{depth}(\mathcal{N}_1^*)$, $\text{depth}(\mathcal{N}_2^*)$, $\text{lift}(\mathcal{N}_1^*)$, $\text{lift}(\mathcal{N}_2^*) \le C$, and such that

$$\|\widehat{\mathcal{N}}_1(u) - \mathcal{N}_1^*(u)\|_{L^2}, \ \|\widehat{\mathcal{N}}_2(u) - \mathcal{N}_2^*(u)\|_{L^2} \le \epsilon,$$

for all $u \in L^2_{2N}(\mathbb{T}^d)$ with bounded norm $\|u\|_{L^2} \le B$.

For $\widehat{\mathcal{N}}_1$, the claim follows from Lemma 50. For $\widehat{\mathcal{N}}_2$, we note that $\widehat{\mathcal{N}}_2$ can be represented exactly by a linear FNO layer with $W = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, bias $b_j = (0, F_N(x_j))$, and with Fourier multiplier matrix $P(k) = \begin{pmatrix} 0 & 0 \\ 0 & \widetilde{P}(k) \end{pmatrix}$, where $\widetilde{P}(k) := 1_{[|k|_\infty \le N]} \frac{ik^T}{|k|^2}$, so that

$$\widehat{\mathcal{N}}_2(v)_j = \widehat{\mathcal{N}}_2(v)(x_j) = Wv_j + b_j + \mathcal{F}_N^{-1}(P\mathcal{F}_N v)_j,$$

for $v_j = (a(x_j), U(x_j))$. The claim for $\widehat{\mathcal{N}}_2$ thus follows from the linear approximation lemma 45. ∎

Appendix F. Technical Results and Proofs from Section 3.2

F.1 Properties of the Pseudo-Spectral Scheme (34)

Our first aim is to show that the implicit operator equation that defines the scheme (34). To this end, we have following lemmas,

Lemma 51 *If $v \in L^2_N(\mathbb{T}^d; \mathbb{R}^d)$ and $\tau \|v\|_{L^\infty} N \le \frac{1}{2}$, then we have*

$$\|\tau \mathbb{P}_N(v \cdot \nabla w)\|_{L^2} \le \frac{1}{2} \|w\|_{L^2},$$

for all $w \in L^2_N(\mathbb{T}^d; \mathbb{R}^d)$. In particular, this estimate holds provided that

$$\tau \|v\|_{L^2} N^{d/2+1} \le \frac{1}{2}.$$

Proof We first note that for any $v \in L^2_N$, we have

$$\tau \|v\|_{L^\infty} N \leq \tau \|v\|_{L^2} N^{d/2+1} \leq \frac{1}{2}.$$

The claim is now an immediate consequence of the fact that

$$
\begin{aligned}
\|\tau \mathbb{P}_N(v \cdot \nabla w)\|_{L^2} &\leq \|\tau v \cdot \nabla w\|_{L^2} \\
&\leq \tau \|v\|_{L^\infty} \|\nabla w\|_{L^2} \\
&\leq \tau \|v\|_{L^\infty} N \|w\|_{L^2},
\end{aligned}
$$

for any $w \in L^2_N(\mathbb{T}^d; \mathbb{R}^d)$. ∎

We can now state the following lemma on the well-posedness:

Lemma 52 *Let $U > 0$. If $\|u^n_N\|_{L^2} \leq U$ for $n = 0, \dots, n_T$, and if the CFL condition*

$$\tau U N^{d/2+1} \leq \frac{1}{2}, \tag{91}$$

is satisfied, then the recursion (34) is well-defined.

Proof The recursion (34) can be written in the form

$$\mathcal{T}_n u^{n+1}_N = u^n, \tag{92}$$

where the operator $\mathcal{T}_n : \dot{L}^2_N(\mathbb{T}^d; \mathrm{div}) \to \dot{L}^2_N(\mathbb{T}^d; \mathrm{div})$ is given by

$$\mathcal{T}_n w_N := w_N + \tau \mathbb{P}_N(u^n_N \cdot \nabla w_N) - \nu \tau \Delta w_N.$$

The claimed well-posedness of the recursion follows from the fact that, under the CFL assumption (91), the operator $\mathcal{T}_n$ is invertible: Indeed, by Lemma 51, this implies that for any $w_N \in L^2_N(\mathbb{T}^d; \mathrm{div})$, we have

$$\|\tau \mathbb{P}_N(u^n_N \cdot \nabla w_N)\|_{L^2} \leq \frac{1}{2} \|w_N\|_{L^2}.$$

But then, we have for any $w_N \in L^2_N(\mathbb{T}^d; \mathrm{div})$, that

$$
\begin{aligned}
\|\mathcal{T}_n w_N\|_{L^2} &\geq \|(1 - \tau \nu \Delta) w_N\|_{L^2} - \|\tau \mathbb{P}_N(u^n_N \cdot \nabla w_N)\|_{L^2_N} \\
&\geq \|w_N\|_{L^2} - \frac{1}{2} \|w_N\|_{L^2} \\
&= \frac{1}{2} \|w_N\|_{L^2},
\end{aligned}
$$

where, in the first step, we have used the fact that $-\tau \nu \Delta$ is a non-negative operator. In particular, the estimate $\|\mathcal{T}_n w_N\|_{L^2} \geq \frac{1}{2} \|w_N\|_{L^2}$ implies that $\mathcal{T}_n : L^2_N(\mathbb{T}^d; \mathrm{div}) \to L^2_N(\mathbb{T}^d; \mathrm{div})$ is injective. Since $L^2_N(\mathbb{T}^d; \mathrm{div})$ is finite-dimensional, we conclude that $\mathcal{T}_n$ is actually invertible, and hence the recursion (34) is well-defined, i.e. it can be solved for u^{n+1}_N, given u^n_N. ∎

By Lemma 52, to prove the well-posedness of the scheme (34), it remains to be shown that with a suitable choice of the time-step τ, we have a uniform L^2-energy bound of the form $\|u^n_N\|_{L^2} \leq U$, for some $U > 0$. This is the content of the following lemma:

Lemma 53 *Assume that $\|u_N^0\|_{L^2} \leq U$, then the first-order scheme (34) is well-defined for any time-step satisfying the CFL condition $\tau U N^{d/2+1} \leq \frac{1}{2}$, and we have $\|u_N^n\|_{L^2} \leq \|u_N^0\|_{L^2} \leq U$ for all $n = 0, \ldots, n_T$.*

Proof We show inductively that if $\|u_N^n\|_{L^2} \leq U$, then $\|u_N^{n+1}\|_{L^2} \leq \|u_N^n\|_{L^2} \leq U$. The well-posedness then follows from Lemma 52. To see that $\|u_N^{n+1}\|_{L^2} \leq \|u_N^n\|_{L^2}$, we integrate (34) against u_N^{n+1}, to find

$$
\begin{aligned}
\|u_N^{n+1}\|_{L^2}^2 &= \langle u_N^{n+1}, u_N^n \rangle - \tau \left\langle u_N^{n+1}, \mathbb{P}_N \left(u_N^n \cdot \nabla u_N^{n+1} \right) \right\rangle - \nu\tau \|\nabla u_N^{n+1}\|_{L^2}^2 \\
&= \langle u_N^{n+1}, u_N^n \rangle - \tau \underbrace{\left\langle u_N^{n+1}, u_N^n \cdot \nabla u_N^{n+1} \right\rangle}_{=0} - \nu\tau \|\nabla u_N^{n+1}\|_{L^2}^2 \\
&\leq \left| \langle u_N^{n+1}, u_N^n \rangle \right| \leq \|u_N^{n+1}\|_{L^2} \|u_N^n\|_{L^2}.
\end{aligned}
$$

This proves the claim. ■

Next, we have the following Lemma on the convergence of the iterations in Algorithm 29,

Lemma 54 *Given $\|u_N^n\|_{L^2} \leq U$, assume that the CFL condition $\tau U N^{d/2+1} \leq \frac{1}{2}$ is satisfied. Define a recursive sequence $w_N^{n,k} \in \dot{L}_N^2(\mathbb{T}^d; \mathrm{div})$, $k \in \mathbb{N}$, by $w_N^{n,0} := 0$, and $w_N^{n,k+1} := F(w_N^{n,k})$, where F is defined by (37). Then, we have $\|u_N^{n+1} - w_N^{n,k}\|_{L^2} \leq 2^{-k} \|u_N^n\|_{L^2}$, for $k \in \mathbb{N}$, and $\|w_N^{n,k}\|_{L^2} \leq (1 + 2^{-k})\|u_N^n\|_{L^2}$.*

Proof By Lemma 51, and the fact that $(1 - \nu\tau\Delta)^{-1}$ is a contraction, it follows that $\mathrm{Lip}(F) \leq \frac{1}{2}$. By Picard iteration, and recalling that u_N^{n+1} is the unique fixed point of the recursion $w_N^{n,k}$ ($k \in \mathbb{N}$), it immediately follows that

$$
\|u_N^{n+1} - w_N^{n,k}\|_{L^2} \leq \frac{1}{2^k} \|u_N^{n+1} - w_N^{n,0}\|_{L^2} \leq \frac{\|u_N^n\|_{L^2}}{2^k}.
$$

The last step is a consequence of the a priori L^2-bound $\|u_N^{n+1}\|_{L^2} \leq \|u_N^n\|_{L^2}$ proven in Lemma 53. In particular, this estimate implies that

$$
\|w_N^{n,k}\|_{L^2} \leq \|u_N^{n+1}\|_{L^2} + \|w_N^{n,k} - u_N^{n+1}\|_{L^2} \leq (1 + 2^{-k})\|u_N^n\|_{L^2}.
$$

■

Remark 55 *Note that as a consequence of Lemma 54, we recursively find that*

$$
\|u_N^{n+1}\|_{L^2} = \|w_N^{n,\kappa_0}\|_{L^2} \leq (1 + 2^{-\kappa_0})\|u_N^n\|_{L^2} \leq \left(1 + \frac{\tau^2}{T^2} \right) \|u_N^n\|_{L^2}
$$

$$
\leq \cdots \leq \left(1 + \frac{\tau^2}{T^2} \right)^n \|u_N^0\|_{L^2} \leq \exp\left(\frac{\tau}{T} \right) U \leq eU,
$$

for $n = 0, \ldots, n_T$. In particular, this ensures that the CFL condition (91) is satisfied for all u_N^n, generated by Algorithm 29.

Finally, we provide the proof of the convergence Theorem 30.

F.2 Proof of Theorem 30

In this appendix, we provide a detailed proof of the convergence estimate of Theorem 30, for the first-order scheme defined by Algorithm 29.

To this end, let $u(t)$ be an exact solution of (31), satisfying the assumptions of Theorem 30, and let $u_N^0, \dots, u_N^{n_T}$ denote the sequence generated by Algorithm 29. For any $n = 0, \dots, n_T - 1$, we denote by $u_N^{n+1,*}$ the solution of a single time-step with the semi-implicit scheme (34), starting from u_N^n, i.e. satisfying

$$\frac{u_N^{n+1,*} - u_N^n}{\tau} + \mathbb{P}_N \left(u_N^n \cdot \nabla u_N^{n+1,*} \right) = \nu \Delta u_N^{n+1,*}. \tag{93}$$

We recall that by Lemma 54 and Remark 55, we have a uniform bound $\|u_N^n\|_{L^2} \le eU$. Since by assumption, the time-step $\tau eU N^{d/2+1} \le \frac{1}{2}$ satisfies the relevant CFL condition, it follows that a unique solution $u_N^{n+1,*}$ exists for all n. We also recall that by Lemma 54, and by our choice of the number of iteration steps κ in Algorithm 29, we have

$$\|u_N^{n+1,*} - u_N^{n+1}\|_{L^2} \le 2^{-k} eU \le C\tau^2,$$

where $C > 0$ depends only on U and the final time T.

Our first goal is to derive an estimate on the magnification of the approximation error due to a single timestep $u_N^n \mapsto u_N^{n+1}$. Let $u(t)$ be the exact solution of (31). Observing that

$$\|u_N^{n+1} - u(t^{n+1})\|_{L^2} \le \|u_N^{n+1} - u_N^{n+1,*}\|_{L^2} + \|u_N^{n+1,*} - u(t^{n+1})\|_{L^2}$$
$$\le C\tau^2 + \|u_N^{n+1,*} - u(t^{n+1})\|_{L^2}, \tag{94}$$

we only need to consider the error introduced by a single time-step $u_N^n \mapsto u_N^{n+1,*}$ of the semi-implicit scheme (34). To this end, we can write

$$\frac{u(t^{n+1}) - u(t^n)}{\tau} + \mathbb{P}_N \left(u(t^n) \cdot \nabla u(t^{n+1}) \right) = \nu \Delta u(t^{n+1}) + \mathcal{E}^n,$$

where $\mathcal{E}^n$ collects all error terms:

$$\mathcal{E}^n = \mathcal{E}_\tau^n + \mathcal{E}_{NL}^n + \mathcal{E}_P^n + \mathcal{E}_\nu^n,$$

where

$$\mathcal{E}_\tau^n = \frac{u(t^{n+1}) - u(t^n)}{\tau} - \partial_t u(t^n),$$
$$\mathcal{E}_{NL}^n = \mathbb{P} \left(u(t^n) \cdot \nabla \left(u(t^{n+1}) - u(t^n) \right) \right),$$
$$\mathcal{E}_P^n = (1 - \mathbb{P}_N) \left(u(t^n) \cdot \nabla u(t^n) \right),$$
$$\mathcal{E}_\nu^n = -\nu \Delta \left(u(t^{n+1}) - u(t^n) \right).$$

Subtracting (93), and introducing the short-hand notation $e^n := u(t^n) - u_N^n$, $e^{n+1} := u(t^{n+1}) - u_N^{n,*}$, we find

$$\frac{e^{n+1} - e^n}{\tau} = -\mathbb{P}_N \left(e^n \cdot \nabla u(t^{n+1}) \right) - \mathbb{P}_N \left(u_N^n \cdot \nabla e^{n+1} \right) + \nu \Delta e^{n+1} + \mathcal{E}^n.$$

Next, integrate against $e_N^{n+1} := P_N e^{n+1}$ to find

$$\frac{1}{2\tau} \left(\|e_N^{n+1}\|_{L^2}^2 + \|e_N^{n+1} - e_N^n\|_{L^2}^2 - \|e_N^n\|_{L^2}^2 \right)$$
$$\leq \|e^n\|_{L^2} \|\nabla u\|_{L_{t,x}^\infty} \|e_N^{n+1}\|_{L^2} - \langle \mathbb{P}_N(u_N^n \cdot \nabla e^{n+1}), e^{n+1} \rangle - \nu \|\nabla e^{n+1}\|_{L^2}^2 + \langle \mathcal{E}^n, e^{n+1} \rangle.$$

We note that

$$\begin{aligned}
\langle \mathbb{P}_N(u_N^n \cdot \nabla e^{n+1}), e^{n+1} \rangle &= \langle P_N(u_N^n \cdot \nabla e^{n+1}), e^{n+1} \rangle \\
&= \langle u_N^n \cdot \nabla e^{n+1}, e^{n+1} \rangle - \langle (1 - P_N) u_N^n \cdot \nabla e^{n+1}, e^{n+1} \rangle \\
&= 0 - \langle u_N^n \cdot \nabla e^{n+1}, (1 - P_N) e^{n+1} \rangle \\
&= -\langle u_N^n \cdot \nabla e^{n+1}, (1 - P_N) u(t^{n+1}) \rangle \qquad (95) \\
&= \langle e^{n+1}, u_N^n \cdot \nabla (1 - P_N) u(t^{n+1}) \rangle \\
&= -\langle e^{n+1}, e^n \cdot \nabla (1 - P_N) u(t^{n+1}) \rangle \\
&\quad + \langle e^{n+1}, u(t^n) \cdot \nabla (1 - P_N) u(t^{n+1}) \rangle
\end{aligned}$$

We now note that we can rewrite the last term as follows:

$$\langle e^{n+1}, u(t^n) \cdot \nabla (1 - P_N) u(t^{n+1}) \rangle = \left\langle (1 - P_N) \left(e^{n+1} \otimes u(t^n) \right), \nabla u(t^{n+1}) \right\rangle$$

Using the fact that

$$(1 - P_N) \left(e^{n+1} \otimes u(t^n) \right) = (1 - P_N) \left((1 - P_{N/2}) e^{n+1} \otimes (1 - P_{N/2}) u(t^n) \right),$$

it then follows that we have

$$\begin{aligned}
\langle P_N(u_N^n \cdot \nabla e^{n+1}), e^{n+1} \rangle &= -\langle e^{n+1}, e^n \cdot \nabla (1 - P_N) u(t^{n+1}) \rangle \\
&\qquad \left\langle (1 - P_N) \left(e^{n+1} \otimes u(t^n) \right), \nabla u(t^{n+1}) \right\rangle \\
&\leq \|e^{n+1}\|_{L^2} \|e^n\|_{L^2} \|(1 - P_N) \nabla u(t^{n+1})\|_{L^\infty} \\
&\qquad + \|e^{n+1}\|_{L^2} \|(1 - P_{N/2}) u(t^n)\|_{L^2} \|(1 - P_N) \nabla u(t^{n+1})\|_{L^\infty}.
\end{aligned}$$

For $r > d/2 + 1$, we have a continuous embedding $H^r \hookrightarrow W^{1,\infty}$, and an inequality of the form

$$\|(1 - P_{N/2}) u(t^n)\|_{L^2} \lesssim_{r,d} \frac{1}{N^r} \|u(t^n)\|_{H^r}.$$

Hence we can estimate

$$\begin{aligned}
\langle P_N(u_N^n \cdot \nabla e^{n+1}), e^{n+1} \rangle &\lesssim_{r,d} \|e^{n+1}\|_{L^2} \|e^n\|_{L^2} \|u(t^n)\|_{H^r} \\
&\qquad + \|e^{n+1}\|_{L^2} \frac{\|u(t^n)\|_{H^r}^2}{N^{2r}}.
\end{aligned}$$

Estimating the products on the right-hand side using the inequality $ab \leq \epsilon a^2 + \frac{1}{4\epsilon} b^2$ with suitable $\epsilon > 0$, it follows that there exists a constant $C > 0$ (independent of $\nu > 0$, N and n), such that

$$\begin{aligned}
\|e^{n+1}\|_{L^2}^2 &\leq \left(1 + C\tau \|u\|_{C_t(H_x^r)}^2 \right) \|e^n\|_{L^2}^2 - \nu \|\nabla e^{n+1}\|_{L^2}^2 \\
&\qquad + C\tau N^{-2r} \|u\|_{C_t(H_x^r)}^4 + C\tau |\langle \mathcal{E}^n, e^{n+1} \rangle|.
\end{aligned} \qquad (96)$$

F.2.1 TIME-DIFFERENCING ERROR

We note that

$$
\begin{aligned}
\langle \mathcal{E}_\tau^n, e^{n+1} \rangle &= \frac{1}{\tau} \int_{t^n}^{t^{n+1}} \int_{t^n}^{t} \langle \partial_t^2 u(s), e^{n+1} \rangle \, ds \, dt \\
&= \frac{1}{\tau} \int_{t^n}^{t^{n+1}} \int_{t^n}^{t} \langle -P \{ \partial_t u(s) \cdot \nabla u(s) + u(s) \cdot \nabla \partial_t u(s) \}, e^{n+1} \rangle \, ds \, dt \\
&\quad + \frac{1}{\tau} \int_{t^n}^{t^{n+1}} \int_{t^n}^{t} \langle \nu \Delta u(s), e^{n+1} \rangle \, ds \, dt \\
&=: (I)_\tau + (II)_\tau.
\end{aligned}
$$

The first term can be bounded from above by

$$
\begin{aligned}
(I)_\tau &\lesssim \tau \left(\|u\|_{C_t^1(L_x^2)} \|u\|_{C_t(W_x^{1,\infty})} + \|u\|_{C_t(L_x^\infty)} \|u\|_{C_t^1(H_x^1)} \right) \|e^{n+1}\|_{L^2} \\
&\lesssim_{r,d} \tau \|u\|_{C_t^1(H^{r-2})} \|u\|_{C_t(H^r)} \|e^{n+1}\|_{L^2},
\end{aligned}
$$

provided that $r > d/2 + 2 \geq 3$ (the last bound is automatic for $d \geq 2$). For the second term, we derive the bound

$$
(II)_\tau \lesssim \tau \nu \|u\|_{C_t(H_x^1)} \|\nabla e^{n+1}\|_{L^2}.
$$

Thus, for any $\epsilon > 0$ (to be specified later), we have for some constant $C = C(r,d) > 0$:

$$
\begin{aligned}
|\langle \mathcal{E}_\tau^n, e^{n+1} \rangle| &\leq \epsilon \left(\|e^{n+1}\|_{L^2} + \nu \|\nabla e^{n+1}\|_{L^2} \right) \\
&\quad + \frac{C\tau^2}{\epsilon} \left(\|u\|_{C_t^1(H^{r-2})}^2 \|u\|_{C_t(H^r)}^2 + \nu \|u\|_{C_t(H_x^1)}^2 \right).
\end{aligned}
\tag{97}
$$

F.2.2 NON-LINEAR TIME-DIFFERENCING ERROR

For the error associated with $\mathcal{E}_{NL}^n$, we simply estimate (for $\epsilon > 0$ to be determined later)

$$
|\langle \mathcal{E}_{NL}^n, e^{n+1} \rangle| \leq \epsilon \|e^{n+1}\|_{L^2}^2 + \epsilon^{-1} \|\mathcal{E}_{NL}^n\|_{L^2}^2,
$$

and we observe that

$$
\|\mathcal{E}_{NL}^n\|_{L^2} \leq \tau \|u\|_{C_t(L_x^\infty)} \|u\|_{C_t^1(H_x^1)} \leq \tau \|u\|_{C_t(H_x^r)} \|u\|_{C_t^1(H_x^{r-2})},
$$

assuming that $r > d/2 + 2 \geq 3$. This yields

$$
|\langle \mathcal{E}_{NL}^n, e^{n+1} \rangle| \leq \epsilon \|e^{n+1}\|_{L^2}^2 + \frac{\tau^2}{\epsilon} \|u\|_{C_t(H_x^r)}^2 \|u\|_{C_t^1(H_x^{r-2})}^2
\tag{98}
$$

F.2.3 NONLINEAR PROJECTION ERROR

Again, we estimate the error $\mathcal{E}_P^n$ using the simple estimate

$$
|\langle \mathcal{E}_P^n, e^{n+1} \rangle| \leq \epsilon \|e^{n+1}\|_{L^2}^2 + \epsilon^{-1} \|\mathcal{E}_P^n\|_{L^2}^2,
$$

with $\epsilon > 0$ to be specified later. We furthermore note that

$$(1 - P_N)\left\{u(t^n) \cdot \nabla u(t^{n+1})\right\} = (1 - P_N)\left\{(1 - P_{N/2})u(t^n) \cdot (1 - P_{N/2})\nabla u(t^{n+1})\right\},$$

which implies that

$$\|\mathcal{E}_P^n\|_{L^2} \le \|(1 - P_{N/2})u(t^n)\|_{L^2}\|(1 - P_{N/2})\nabla u(t^{n+1})\|_{L^\infty} \lesssim_{r,d} N^{-r}\|u\|_{C_t(H^r)}^2.$$

In the last step, we have used the fact that $\|(1 - P_{N/2})v\|_{L^2} \lesssim_{r,d} N^{-r}\|v\|_{H^r}$ for $r \ge 0$, and that by Sobolev embedding

$$\|(1 - P_{N/2})\nabla u(t^{n+1})\|_{L^\infty} \lesssim_d \|(1 - P_{N/2})u(t^{n+1})\|_{H^r} \le \|u\|_{C_t(H_x^r)},$$

for any $r > d/2 + 1$. Thus, there exists $C = C(d, r) > 0$, such that

$$|\langle \mathcal{E}_P^n, e^{n+1}\rangle| \le \epsilon\|e^{n+1}\|_{L^2}^2 + \frac{C}{\epsilon N^{2r}}\|u\|_{C_t(H_x^r)}^2. \tag{99}$$

F.2.4 VISCOSITY ERROR

Finally, we note that

$$
\begin{aligned}
|\langle \mathcal{E}_\nu^n, e^{n+1}\rangle| &\le \nu\|\nabla(u(t^{n+1}) - u(t^n))\|_{L^2}\|\nabla e^{n+1}\|_{L^2} \\
&\le \nu\epsilon\|\nabla e^{n+1}\|_{L^2}^2 + \frac{\nu\tau^2}{\epsilon}\|u\|_{C_t^1(H_x^1)}^2 \\
&\le \nu\epsilon\|\nabla e^{n+1}\|_{L^2}^2 + \frac{\nu\tau^2}{\epsilon}\|u\|_{C_t^1(H_x^{r-2})}^2,
\end{aligned}
\tag{100}
$$

for any $r > d/2 + 2 \ge 3$.

F.2.5 THE FINAL STABILITY ESTIMATE

Choosing $\epsilon = 1/4$, it follows from (97), (98), (99) and (100), that the total error term $\langle \mathcal{E}^n, e^{n+1}\rangle$ can be estimated by

$$|\langle \mathcal{E}^n, e^{n+1}\rangle| \le \|e^{n+1}\|_{L^2}^2 + \nu\|\nabla e^{n+1}\|_{L^2}^2 + C^*\left(\tau^2 + N^{-2r}\right). \tag{101}$$

for some constant $C^* > 0$, depending only on $r > d/2 + 2$, the spatial dimension d and the norms $\|u\|_{C_t(H^r)}$, $\|u\|_{C_t^1(H^r)}$ of the exact solution u.

Substitution of the error estimate (101) in (96) finally yields

$$(1 - \tau)\|e^{n+1}\|_{L^2}^2 \le (1 + C^*\tau)\|e^n\|_{L^2}^2 + \tau C^*\left(\tau^2 + N^{-2r}\right),$$

where the constant C^* depends only on $r \ge d/2 + 2$, the dimension d and the norms $\|u\|_{C_t(H_x^r)}$ and $\|u\|_{C_t^1(H_x^{r-2})}$. Assuming that $\tau \le 1/2$, dividing by $(1 - \tau)$, and noting that

$$\frac{1 + C^*\tau}{1 - \tau} = 1 + \tau\left(C^* + \frac{(1 + C^*\tau)}{1 - \tau}\right) \le 1 + \tau 2\left(C^* + 1\right),$$

and

$$\frac{C^*}{1 - \tau} \le 2C^*,$$

we can clearly absorb the additional factor of $(1-\tau)^{-1}$ by increasing the constant C^*, if necessary.

From this, we conclude that for a time-step $\tau \leq 1/2$ satisfying the CFL condition (91), there exists a constant $C^* = C^*(r, d, \|u\|_{C_t(H_x^r)}, \|u\|_{C_t^1(H_x^{r-2})}) > 0$, such that

$$\|u_N^{n+1,*} - u(t^{n+1})\|_{L^2}^2 \leq (1+C^*\tau)\|u_N^n - u(t^n)\|_{L^2}^2 + \tau C^* \left(\tau^2 + N^{-2r}\right). \tag{102}$$

In fact, recalling also that $\|u_N^{n+1} - u_N^{n+1,*}\|_{L^2} \leq C\tau^2$ by (94), we find that an inequality of the form (102) remains true with $u_N^{n+1,*}$ replaced by u_N^{n+1}. Indeed, we have

$$
\begin{aligned}
\|u_N^{n+1} - u(t^{n+1})\|_{L^2}^2 &\leq \|u_N^{n+1,*} - u(t^{n+1})\|_{L^2}^2 + 2C\tau^2\|u_N^{n+1,*} - u(t^{n+1})\|_{L^2} + C^2\tau^4 \\
&\leq (1+\epsilon)\|u_N^{n+1,*} - u(t^{n+1})\|_{L^2}^2 + C^2(1+4\epsilon^{-1})\tau^4 \\
&\overset{(\epsilon:=\tau)}{=} (1+\tau)\|u_N^{n+1,*} - u(t^{n+1})\|_{L^2}^2 + C^2(1+4\tau^{-1})\tau^4 \\
&\leq (1+C^*\tau)\|u_N^n - u(t^n)\|_{L^2}^2 + \tau C^* \left(\tau^2 + N^{-2r}\right),
\end{aligned}
\tag{103}
$$

where the last estimate follows from (102), and C^* has been suitably enlarged (but still only depends on $r, d, \|u\|_{C_t(H_x^r)}, \|u\|_{C_t^1(H_x^{r-2})}$). In particular, denoting $E^n := \|u_N^n - u(t^n)\|_{L^2}^2$, we have obtained

$$E^{n+1} \leq (1+C^*\tau)E^n + \tau C^* \left(\tau^2 + N^{-2r}\right),$$

and from Gronwall's inequality it now follows that

$$E^n \leq e^{C^*T}\left[E^0 + C^*T\left(\tau^2 + N^{-2r}\right)\right].$$

We note that $E^0 = \|(1-\dot{P}_N\mathcal{I}_{2N})u(t=0)\|_{L^2}^2 \leq N^{-2r}\|u\|_{H^r}^2$. And hence, we finally find, for $n = 0, \ldots, n_T$, that

$$\|u_N^n - u(t^n)\|_{L^2} = \sqrt{E^n} \leq C\left(\tau + N^{-r}\right), \tag{104}$$

where $C > 0$ depends only on T, r, d, $\|u\|_{C_t(H_x^r)}$ and $\|u\|_{C_t^1(H_x^{r-2})}$.

F.3 Proof of Theorem 32

In this appendix, we provide a proof for Theorem 32. At this proof relies on Lemma 31, we prove this lemma below.

Proof [Proof of Lemma 31] The proof of Lemma 31 is almost identical to the proof of Lemma 50; where in the present case, we replace $a_N \to u_N$ and refer to Lemma 49 point (3), rather than point (2). The only main difference being that in the last step of the proof, the projection P_N is now replaced by the Leray projection $\mathbb{P}_N$. However, also for $\mathbb{P}_N$, we observe that $\mathbb{P}_N : L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d) \to L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d)$ can again be represented *exactly* by a linear Ψ-FNO layer. Thus, replacing the linear Ψ-FNO layer $\widehat{\mathcal{L}}^*$ which represents P_N in the proof of Lemma 50 by a layer representing $\mathbb{P}_N$, an almost identical argument also applies in this case, and yields for any $\epsilon, B > 0$, a Ψ-FNO $\mathcal{N}^* : L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d) \to L^2_{2N}(\mathbb{T}^d; \mathbb{R}^d)$, such that

$$\|\mathbb{P}_N(u_N \cdot \nabla u_N) - \mathcal{N}^*(u_N)\|_{L^2} \leq \epsilon,$$

for all $\|u_N\|_{L^2} \le B$, with
$$\text{depth}(\mathcal{N}^*), \ \text{lift}(\mathcal{N}^*) = d_v \le C,$$
where $C = C(d)$. $\blacksquare$

Proof [Proof of Theorem 32] Given $N \in \mathbb{N}$, choose $\tau \sim N^{-r}$, such that the CFL condition $\tau N^{d/2+1} U \le \frac{1}{2}$ is satisfied. This is possible, since $r \ge d/2 + 2$, by assumption. It then follows that $n_T \sim N^r$, and we note that Algorithm 29 can be written as the composition of $O(n_T \log(n_T)) = O(N^r \log(N))$ mappings of the form

$$\widehat{\mathcal{N}}_1(u_N^n, w_N^{n,k}) := \begin{bmatrix} u_N^n \\ (1 - \nu\tau\Delta)^{-1} u_N^n - \tau(1 - \nu\tau\Delta)^{-1} \mathbb{P}_N \left(u_N^n \cdot \nabla w_N^{n,k} \right) \end{bmatrix},$$

where $\|u_N^n\|_{L^2}, \|w_N^{n,k}\|_{L^2} \le 2U$ for all k, n. Applying the replacement lemma, Lemma 47, the claim will thus follow if we can show that there exists a constant $C > 0$ independent of N, such that for any $\epsilon > 0$, there exists a Ψ-FNO $\mathcal{N}_1^* : L_{2N}^2(\mathbb{T}^d; \text{div}) \to L_{2N}^2(\mathbb{T}^d; \text{div})$, such that $\|\mathcal{N}_1^*(u_N, w_N) - \widehat{\mathcal{N}}_1(u_N, w_N)\|_{L^2} \le \epsilon$, and

$$N \sim C\epsilon^{-1/r}, \quad \text{depth}(\mathcal{N}_1^*) \le C, \quad \text{lift}(\mathcal{N}_1^*) \le C.$$

This is immediate for the approximation of the first component of $\widehat{\mathcal{N}}_1$. For the second component, we note that we can write it as a composition:

$$\begin{bmatrix} u_N^n \\ w_N^n \end{bmatrix} \mapsto \begin{bmatrix} u_N^n \\ \mathbb{P}_N \left(u_N^n \cdot \nabla w_N^{n,k} \right) \end{bmatrix} \mapsto (1 - \nu\tau\Delta)^{-1} \left\{ u_N^n - \tau \mathbb{P}_N \left(u_N^n \cdot \nabla w_N^{n,k} \right) \right\}.$$

To finish the proof, we note that the first mapping can be approximated to arbitrary accuracy $\epsilon > 0$ with a Ψ-FNO of uniformly bounded depth and lift, by Lemma 31. The second mapping can be represented *exactly* by a linear FNO layer. Thus, $\mathcal{N}_1^*$ can be obtained as the composition of a Ψ-FNO approximating the quadratic non-linearity, and a linear FNO layer, implying the claimed complexity estimate. $\blacksquare$

F.4 A Second-Order in Time Accurate Pseudo-Spectral Method for Approximating the Navier-Stokes Equation (31) and its Emulation by Ψ-FNOs

Our aim is to describe a second-order accurate (in time) version of the pseudo-spectral scheme (34). To this end, we propose the following scheme,

$$\frac{u_N^{n+1} - u_N^n}{\tau} + \mathbb{P}_N \left(\left[\frac{3}{2} u^n - \frac{1}{2} u^{n-1} \right] \cdot \nabla \frac{1}{2} \left[u^{n+1} + u^n \right] \right) = \nu\Delta \frac{1}{2} \left[u^{n+1} + u^n \right], \qquad (105)$$

In contrast to the first-order method (34) of the last section, to start the scheme (34), we now require two starting values $u_N^0 \approx u(t_0)$ and $u_N^1 \approx u(t_1)$. Given initial data $u_0 \in \dot{H}^r(\mathbb{T}^d; \text{div})$, $r \ge d/2$, we propose to define $u_N^0 := \mathcal{I}_N u_0$, where $\mathcal{I}_N$ denotes the pseudo-spectral projection,

and to generate u_N^1 by the first-order accurate Algorithm 29 applied on the time-interval $[0, \tau]$, with reduced time-step of size $\sim \tau^2$.

As in the case of (34), one needs to solve an implicit operator equation in the time update for (105). Analogously, we will use a fixed point iteration to approximate this implicit equation, resulting in the following algorithm,

Algorithm 56 (Second-order in time approximation of (31))
Input: $U > 0$, $N \in \mathbb{N}$, $T > 0$, $\nu \geq 0$, a time-step $\tau > 0$, such that $n_T = T/\tau \in \mathbb{N}$, and $\tau U N^{d/2+1} \leq \frac{1}{2e}$, initial data $u_N^0 \in L_N^2(\mathbb{T}^d; \mathrm{div})$, such that $\|u_N^0\|_{L^2} \leq U$.
Output: $u_N^{n_T} \in L_N^2(\mathbb{T}^d; \mathrm{div})$ an approximation of the solution $S_T(u_N^0)$ of (31) at time $t = T$.

1. Set
$$\kappa := \left\lceil \frac{\log\left(T^3/\tau^3\right)}{\log(2)} \right\rceil \in \mathbb{N}.$$

2. Compute $u_N^1 \approx u(t_1)$ by applying Algorithm 29 on the time-interval $[0, \tau]$, with n_T steps and with time-step $\tau' = \tau/n_T$.

3. For $n = 1, \ldots, n_T - 1$:

   (a) Set $w_N^{n,0} := 0$,

   (b) For $k = 1, \ldots, \kappa_0$: Given the values u_N^{n-1}, u_N^n from the previous steps, compute
   $$w_N^{n,k} := F_2^n(w_N^{n,k-1}),$$

   (c) Set $u_N^{n+1} := w_N^{n,\kappa_0}$,

Next, we have the following convergence theorem for the Algorithm 56,

Theorem 57 Let $U, T > 0$. Consider the Navier-Stokes equations on $\mathbb{T}^d$, for $d \geq 2$. Assume that $r \geq d + 2$, and let $u \in C([0,T]; H^r) \cap C^1([0,T]; H^{r-2}) \cap C^2([0,T]; H^1)$ be a solution of the Navier-Stokes equations (31), such that $\|u\|_{L^2} \leq U$. Choose a time-step τ, such that $\tau U N^{d/2+1} \leq (2e)^{-1}$. There exists a constant
$$C = C(T, d, r, \|u\|_{C_t(H_x^r)}, \|u\|_{C_t^1(H_x^{r-2})}, \|u\|_{C_t^2(H_x^1)}) > 0,$$

such that with $u_N^0 := \mathcal{I}_N u(0)$, if $\|u(t_1) - u_N^1\|_{L^2} \leq \delta$, and for the sequence $u_N^2, \ldots, u_N^{n_T} \in L_N^2(\mathbb{T}^d; \mathrm{div})$ generated by Algorithm 29, we have
$$\max_{n=0,\ldots,n_T} \|u_N^n - u(t^n)\|_{L^2} \leq C\left(\delta + \tau^2 + N^{-r}\right),$$

where $n_T \tau = T$. In particular, choosing a suitable time-step $\tau \sim N^{-r/2}$, and assuming that $\delta \leq C N^{-r}$, we have
$$\max_{n=0,\ldots,n_T} \|u_N^n - u(t^n)\|_{L^2} \leq 3C N^{-r},$$

with $n_T \sim N^{r/2}$.

71

The proof uses very similar techniques as the proof of Theorem 30 and we omit it here.

Finally and in complete analogy with the proof of Theorem 32, one can prove the following Theorem for the approximation of scheme (105) with a Ψ-FNO,

Theorem 58 *Let $U, T > 0$ and viscosity $\nu \geq 0$. Consider the Navier-Stokes equations on $\mathbb{T}^d$, for $d \geq 2$. Assume that $r \geq d + 2$, and let $\mathcal{V} \subset C([0,T]; H^r) \cap C^1([0,T]; H^{r-2}) \cap C^2([0,T]; H^1)$ be a set of solutions of the Navier-Stokes equations (31), such that $\sup_{u \in \mathcal{V}} \|u\|_{L^2} \leq U$, and*

$$\overline{U} := \sup_{u \in \mathcal{V}} \left\{ \|u\|_{C_t(H_x^r)} + \|u\|_{C_t^1(H_x^{r-2})} + \|u\|_{C_t^2(H_x^1)} \right\}.$$

For $t \in [0,T]$, denote $\mathcal{V}_t := \{u(t) \,|\, u \in \mathcal{V}\}$. Let $\mathcal{G} : \mathcal{V}_0 \to \mathcal{V}_T$ denote the solution operator of (31), mapping initial data $u_0 = u(t = 0)$, to the solution $u(T)$ at $t = T$ of the incompressible Navier-Stokes equations. Then there exists a constant

$$C = C(d, r, U, \overline{U}, T) > 0,$$

such that for any $N \in \mathbb{N}$ there exists a Ψ-FNO $\mathcal{N}^* : L_{2N}^2(\mathbb{T}^d; \mathbb{R}^d) \to L_{2N}^2(\mathbb{T}^d; \mathbb{R}^d)$, such that

$$\sup_{u \in \mathcal{V}_0} \|\mathcal{G}(u) - \mathcal{N}^*(u)\|_{L^2} \leq C N^{-r},$$

and such that

$$\mathrm{depth}(\mathcal{N}^*) \leq C N^{r/2} \log(N), \quad \mathrm{lift}(\mathcal{N}^*) = d_v \leq C.$$

Appendix G. Proof of Theorem 36

Proof First, we note that each layer

$$v(x_j) \mapsto \sigma \left(W v(x_j) + b_j + \mathcal{F}_N^{-1} \left(P \mathcal{F}_N v \right) \right),$$

is simply the composition of

- an affine mapping $\mathbb{R}^{d_v \times \mathcal{J}_N} \to \mathbb{R}^{d_v \times \mathcal{J}_N}$, and

- a componentwise application of the activation function σ.

In particular, the Ψ-FNO $\mathcal{N}^*$, interpreted as a mapping

$$\mathbb{R}^{\mathcal{J}_N \times d_a} \to \mathbb{R}^{\mathcal{J}_N \times d_v} \to \cdots \to \mathbb{R}^{\mathcal{J}_N \times d_v} \to \mathbb{R}^{\mathcal{J}_N \times d_u},$$

can be represented by an ordinary neural network $\widetilde{\beta} : \mathbb{R}^{\mathcal{J}_N \times d_a} \to \mathbb{R}^{\mathcal{J}_N \times d_u}$, with

$$\mathrm{width}(\widetilde{\beta}) = |\mathcal{J}_N| d_v \sim N^d d_v, \quad \mathrm{depth}(\widetilde{\beta}) = \mathrm{depth}(\mathcal{N}).$$

In fact, by suitably modifying the linear output layer of $\widetilde{\beta}$, we can map the grid values encoded in the output $\widetilde{\beta}_j(a) = \mathcal{N}(a)(x_j)$ to the corresponding coefficients in a (real) trigonometric basis $\{\mathbf{e}_k\}_{k \in \mathcal{K}_N}$ with $\mathrm{span}\{\mathbf{e}_k\}_{k \in \mathcal{K}_N} = \mathrm{span}\{e^{i\langle k, x \rangle}\}_{k \in \mathcal{K}_N}$; i.e. by modifying the linear output layer of $\widetilde{\beta}$ (and re-indexing the components of the output), we obtain another neural network $\beta : \mathbb{R}^{\mathcal{J}_N \times d_a} \to \mathbb{R}^{\mathcal{K}_N \times d_u}$, such that

$$\mathcal{N}(a)(x_j) = \sum_{k \in \mathcal{K}_N} \beta_k(a) \mathbf{e}_k(x_j), \quad \forall j \in \mathcal{J}_N. \tag{106}$$

Since $\widetilde{\beta}$ and β only differ in their output layers, we clearly have

$$\mathrm{width}(\beta) = \mathrm{width}(\widetilde{\beta}) \sim d_v N^d, \quad \mathrm{depth}(\beta) = \mathrm{depth}(\widetilde{\beta}) = \mathrm{depth}(\mathcal{N}).$$

Since $\{\mathbf{e}_k\}_{k \in \mathcal{K}_N}$ has the same span as $\{e^{i\langle k, x \rangle}\}_{k \in \mathcal{K}_N}$, it follows form (106), that

$$\mathcal{N}(a)(x) = \sum_{k \in \mathcal{K}_N} \beta_k(a) \mathbf{e}_k(x), \quad \forall \, x \in \mathbb{T}^d.$$

To prove the claim, it thus suffices to observe that there exists a neural network

$$\tau : \mathbb{R}^d \to \mathbb{R}^{\mathcal{K}_N}, \quad x \mapsto \{\tau_k(x)\}_{k \in \mathcal{K}_N},$$

such that the DeepOnet defined by (β, τ) satisfies

$$\sup_{\|a\|_{L^\infty} \leq B} \sup_{x \in \mathbb{T}^d} \left| \mathcal{N}(a)(x) - \sum_{k \in \mathcal{K}_N} \beta_k(a) \tau_k(x) \right|$$

$$= \sup_{\|a\|_{L^\infty} \leq B} \sup_{x \in \mathbb{T}^d} \left| \sum_{k \in \mathcal{K}_N} \beta_k(a) \left[\mathbf{e}_k(x) - \tau_k(x) \right] \right|$$

$$\leq (2N + 1)^d \left(\sup_{\|a\|_{L^\infty} \leq B} \max_{k \in \mathcal{K}_N} |\beta_k(a)| \right) \sup_{x \in \mathbb{T}^d} \max_{k \in \mathcal{K}_N} |\mathbf{e}_k(x) - \tau_k(x)|$$

$$\leq (2N + 1)^d \left(\sup_{\|a\|_{L^\infty} \leq B} \|\mathcal{N}(a)\|_{L^2} \right) \sup_{x \in \mathbb{T}^d} \max_{k \in \mathcal{K}_N} |\mathbf{e}_k(x) - \tau_k(x)|$$

$$= \overline{B} \sup_{x \in \mathbb{T}^d} \max_{k \in \mathcal{K}_N} |\mathbf{e}_k(x) - \tau_k(x)|.$$

∎

References

Andrew R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inform. Theory.*, 39(3):930–945, 1993.

Kaushik Bhattacharya, Bamdad Hosseini, Nikola B. Kovachki, and Andrew M. Stuart. Model reduction and neural networks for parametric PDEs. *The SMAI journal of computational mathematics*, 7:121–157, 2021. doi: 10.5802/smai-jcm.74. URL https://smai-jcm.centre-mersenne.org/articles/10.5802/smai-jcm.74/.

John P. Boyd. Fourier embedded domain methods: Extending a function defined on an irregular region to a rectangle so that the extension is spatially periodic and C infinity. *Appl. Math. Comput.*, 161(2):591–597, 2005.

Haim Brezis and Petru Mironescu. Composition in fractional Sobolev spaces. *Discrete & Continuous Dynamical Systems-A*, 7(2):241, 2001.

Oscar P. Bruno and Mark Lyon. High-order unconditionally stable FC-AD solvers for general smooth domains I. Basic elements. *Journal of Computational Physics*, 229(6): 2009–2033, 2010.

Shengze Cai, Zhicheng Wang, Lu Lu, Tamer A Zaki, and George Em Karniadakis. DeepM&Mnet: Inferring the electroconvection multiphysics fields based on operator approximation by neural networks. *Journal of Computational Physics*, 436:110296, 2021.

Claudio Canuto, M Yousuff Hussaini, Alfio Quarteroni, and Thomas A Zang. *Spectral methods: fundamentals in single domains*. Springer Science & Business Media, 2007.

Tianping Chen and Hong Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995.

George Cybenko. Approximations by superpositions of sigmoidal functions. *Approximation theory and its applications*, 9(3):17–28, 1989.

Tim De Ryck, Samuel Lanthaler, and Siddhartha Mishra. On the approximation of functions by tanh neural networks. *Neural Networks*, 143:732–750, 2021. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2021.08.015. URL https://www.sciencedirect.com/science/article/pii/S0893608021003208.

Weinan E, Jiequn Han, and Arnulf Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, 5(4):349–380, 2017.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

David Gottlieb and Steven A Orszag. *Numerical analysis of spectral methods: theory and applications*, volume 26. Siam, 1977.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Gitta Kutyniok, Philipp Petersen, Mones Raslan, and Reinhold Schneider. A theoretical analysis of deep neural networks and parametric PDEs. *Constructive Approximation*, pages 1–53, 2021.

Samuel Lanthaler, Siddhartha Mishra, and George Em Karniadakis. Error estimates for DeepOnets: A deep learning framework in infinite dimensions, 2021. URL https://arxiv.org/abs/2102.09618. Preprint, available from arXiv:2102.09618.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, 2015.

Zongyi Li, Nikola B Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *CoRR*, abs/2003.03485, 2020a.

Zongyi Li, Nikola B Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew M Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 6755–6766. Curran Associates, Inc., 2020b.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=c8P9NQVtmnO`.

Chensen Lin, Zhen Li, Lu Lu, Shengze Cai, Martin Maxey, and George Em Karniadakis. Operator learning for predicting multiscale bubble growth dynamics. *The Journal of Chemical Physics*, 154(10):104118, 2021.

Lu Lu, Pengzhan Jin, Guofei Pang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat Mach Intell*, 3:218–229, 2021.

Kjetil O Lye, Siddhartha Mishra, and Deep Ray. Deep learning observables in computational fluid dynamics. *Journal of Computational Physics*, page 109339, 2020.

Kjetil O Lye, Siddhartha Mishra, Deep Ray, and Praveen Chandrashekar. Iterative surrogate model optimization (ismo): An active learning algorithm for pde constrained optimization with deep neural networks. *Computer Methods in Applied Mechanics and Engineering*, 374:113575, 2021.

Mark Lyon and Oscar P. Bruno. High-order unconditionally stable FC-AD solvers for general smooth domains II. Elliptic, parabolic and hyperbolic PDEs; theoretical considerations. *Journal of Computational Physics*, 229(9):3358–3381, 2010.

Andrew J. Majda and Andrea L. Bertozzi. *Vorticity and Incompressible Flow*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2001. doi: 10.1017/CBO9780511613203.

Zhiping Mao, Lu Lu, Olaf Marxen, Tamer A Zaki, and George Em Karniadakis. Deepm&mnet for hypersonics: Predicting the coupled flow and finite-rate chemistry behind a normal shock using neural-network approximation of operators. *Journal of Computational Physics*, 447:110698, 2021.

S. Mishra and R. Molinaro. Estimates on the generalization error of physics informed neural networks (PINNs) for approximating PDEs. Preprint, available from arXiv:2006:16144v1, 2020a.

S. Mishra and R. Molinaro. Estimates on the generalization error of physics informed neural networks (PINNs) for approximating PDEs II: A class of inverse problems. Preprint, available from arXiv:2007:01138v1, 2020b.

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. *Journal of Computational Physics*, 378:686–707, 2019.

Maziar Raissi and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.

Luke G. Rogers. Degree-independent Sobolev extension on locally uniform domains. *Journal of Functional Analysis*, 235(2):619–665, 2006.

Christoph Schwab and Jakob Zech. Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ. *Analysis and Applications*, 17(01):19–55, 2019.

Elias M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, 1970.

Dmitry Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94:103–114, 2017.