KANO: KOLMOGOROV–ARNOLD NEURAL OPERATOR

Anonymous authors
Paper under double-blind review

ABSTRACT

We introduce *Kolmogorov–Arnold Neural Operator* (KANO), a dual-domain neural operator jointly parameterized by both spectral and spatial bases with intrinsic symbolic interpretability. We theoretically demonstrate that KANO overcomes the pure-spectral bottleneck of Fourier Neural Operator (FNO): KANO remains expressive over a generic position-dependent dynamics for any physical input, whereas FNO stays practical only to spectrally sparse operators and strictly imposes fast-decaying input Fourier tail. We verify our claims empirically on position-dependent differential operators, for which KANO robustly generalizes but FNO fails to. In the quantum Hamiltonian learning benchmark, KANO reconstructs ground-truth Hamiltonians in closed-form symbolic representations accurate to the fourth decimal place in coefficients and attains $\approx 6 \times 10^{-6}$ state infidelity from projective measurement data, substantially outperforming that of the FNO trained with ideal full wave function data, $\approx 1.5 \times 10^{-2}$, by orders of magnitude.

1 INTRODUCTION

Contemporary science and engineering increasingly operate in regimes where the effective dimensionality and complexity of phenomena and data overwhelm human-designed calibrations and approximations. This motivates data-centric modeling of governing dynamics from observations (Karniadakis et al., 2021; Wang et al., 2023; Carleo et al., 2019). For a learned model to be constituted as a scientific law, it should first *generalize* universally over a well-defined domain, and also should be *interpretable* so that the learned representations can be extracted and reused for verification, testing, and downstream simulation. Mathematically, physical dynamics are generalized as operators as they are often formalized as PDEs (Courant et al., 1963; Evans, 2022), which are predominantly *position-dependent*: many canonical PDEs involve coefficients varying with location. Representative position-dependent PDEs include the heat and mass transport equation (Incropera et al., 1990), the advection-diffusion-reaction equation (Kundu et al., 2024), Fokker–Planck equation (Pathria, 2017), Maxwell and Helmholtz equation (Jackson & Fox, 1999), and Schrödinger and Gross–Pitaevskii equation (Sakurai & Napolitano, 2020). Scientific AI such as operator networks (Kovachki et al., 2023; Lu et al., 2021) should therefore efficiently approximate a generic position-dependent operator with tractable interpretability, which we recognize the absence and aim to fill the gap in this work.

An operator network approximates an arbitrary mapping between infinite-dimensional function spaces by first encoding functions into finite latent vectors and then learning the latent-to-latent map that represents the target operator (Lanthaler et al., 2022). DeepONet of Lu et al. (2021; 2019) implements the most general dense operator network where two neural networks learn both encoding and latent mapping directly from data, based on the theoretical foundation laid by Chen & Chen (1995). Fourier Neural Operator (FNO) of Li et al. (2020), on the other hand, hard-codes the encoding as pseudo-spectral projection with its spectrally diagonal kernels. FNO is provably and empirically superior when its hard-coded sparsity is optimal (Li et al., 2020; Kovachki et al., 2021), but this spectral sparsity becomes maladaptive for position-dependent or otherwise spectrally dense dynamics (Koshizuka et al., 2024; Qin et al., 2024). In such cases, the model size required for a target accuracy can grow super-exponentially (Kovachki et al., 2021), hence realistic FNO may only converge on an in-sample mapping that fails outside the training distribution. Numerous variants of FNO attempted to break this spectral bottleneck. Some have broadened spectral coverage by exploiting factorized (Tran et al., 2021) or multi-scale (You et al., 2024) spectral kernels, and others have injected local spatial kernels alongside the original spectral ones (Wen et al., 2022; Liu-Schiaffini et al., 2024; Liu et al., 2025). Yet all prior works still privilege the spectral basis and cannot achieve optimal sparsity in the spatial basis.

In parallel, interpretability has recently pivoted around Kolmogorov–Arnold Network (KAN) (Liu et al., 2024b;a), whose edges are trainable univariate functions and thus amenable to human inspection. Several works demonstrate data-driven scientific modeling with KAN: Chiparova & Popov (2025); Gashi et al. (2025) use KAN for system identification, and Koenig et al. (2024) replace the MLPs in Neural ODEs (Chen et al., 2018) with KANs, each reporting symbolic recovery of benchmark equations and parameters. KANs have also been explored within operator networks: Abueidda et al. (2025) employed KANs instead of MLPs in DeepONet and Xiao et al. (2024); Wang et al. (2025) augmented FNO with KANs. Despite performance gains however, prior KAN-based operator networks have not reported symbolic recovery of the learned operator, leaving the avenue of an interpretable operator network largely unexplored.

To address these research gaps, we introduce the **Kolmogorov–Arnold Neural Operator (KANO)**, an interpretable operator network jointly parameterized in both spatial and spectral bases with KAN sub-networks embedded in a pseudo-differential operator framework (Hörmander, 2007; Kohn & Nirenberg, 1965). The key insight is to represent each component of the operator in the basis where it is sparse: differential terms spectrally, localized terms spatially, to achieve the most compact and tractable representation. Our work offers three main contributions to the scientific AI community.

- First, we demonstrate the pure-spectral bottleneck of FNO with an illustrative example and theoretically analyze why FNO cannot converge closely as desired to a generic position-dependent dynamics with a practical model size.
- Second, we propose a novel framework of KANO that is expressive over a generic position-dependent dynamics with intrinsic symbolic interpretability. We provide theoretical analysis on KANO's dual-domain (spatial and spectral) expressivity along with the empirical evidences of KANO robustly generalizing on unseen input subspace when FNO fails to.
- Finally we validate the performance of KANO on some synthetic operators and a quantum simulation benchmark. KANO successfully recovered the closed-form formula accurately to the fourth decimal place in coefficients. Compared to the FNO baseline, KANO used only 0.03% of the model parameters, but achieved an order lower relative loss ℓ_2 in our synthetic operator benchmarks, and a four-order lower state infidelity in the quantum Hamiltonian learning benchmark.

To the best of our knowledge, our work is the first to demonstrate and quantify the symbolic recovery via KAN in operator learning. We shift the paradigm from mere *universal approximation* in operator learning toward the *universal generalization* of an operator network. Different from DeepOKAN (Abueidda et al., 2025) which replaces MLPs with KANs in DeepONet, our work achieves generalization over disjoint out-of-distribution subspace via a novel architecture design.

2 Background

2.1 Operator Learning and Fourier Neural Operator

Operator learning approximates mapping between infinite-dimensional function spaces, $\boldsymbol{\mathcal{G}} : \mathcal{A} \to \mathcal{U}$,[1] from function pairs $\{(\mathbf{a}_i \in \mathcal{A}, \mathbf{u}_i = \boldsymbol{\mathcal{G}}(\mathbf{a}_i) \in \mathcal{U})\}_{i=1}^{N}$[2](Kovachki et al., 2023; 2024b). An operator network $\boldsymbol{\mathcal{G}}_\theta$ first encodes input $\mathbf{a}_i$ via *encoder* $\boldsymbol{\mathcal{E}}_m : \mathcal{A} \to \mathbb{C}^m$ into a latent vector, then learns the *latent map* $\mathbf{T}_\theta : \mathbb{C}^m \to \mathbb{C}^{m'}$ which the output is reconstructed to approximate the label $\mathbf{u}_i$ via *reconstructor* $\boldsymbol{\mathcal{R}}_{m'} : \mathbb{C}^{m'} \to \mathcal{U}$: i.e. $\boldsymbol{\mathcal{G}}_\theta = \boldsymbol{\mathcal{R}}_{m'} \circ \mathbf{T}_\theta \circ \boldsymbol{\mathcal{E}}_m$ (Lanthaler et al., 2022). For fixed $(\boldsymbol{\mathcal{E}}_m, \boldsymbol{\mathcal{R}}_{m'})$, we can define the projection $\boldsymbol{\Pi}$ of an operator $\boldsymbol{\mathcal{G}}$ as

$$\boldsymbol{\Pi}(\boldsymbol{\mathcal{G}}) = \boldsymbol{\mathcal{R}}_{m'} \circ \widehat{\mathbf{T}} \circ \boldsymbol{\mathcal{E}}_m \quad \text{where} \quad \widehat{\mathbf{T}} \in \underset{\mathbf{T}:\, \mathbb{C}^m \to \mathbb{C}^{m'}}{\arg\min} \left\| \boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{R}}_{m'} \circ \mathbf{T} \circ \boldsymbol{\mathcal{E}}_m \right\|. \tag{1}$$

DeepONet (Lu et al., 2021; 2019) learns $\boldsymbol{\mathcal{E}}_m$, $\boldsymbol{\mathcal{R}}_{m'}$, and $\mathbf{T}_\theta$ all with two sub-networks. FNO (Li et al., 2020), on the other hand, *hard-codes* $\boldsymbol{\mathcal{E}}_m$ to be the truncated Fourier transform and $\boldsymbol{\mathcal{R}}_m$ to be its band-limited inverse.

Fourier Neural Operator (FNO). Let the domain $D \subset \mathbb{R}^d$ be periodic and write the Fourier transform $\boldsymbol{\mathcal{F}}$ of function $\mathbf{a}(\mathbf{x})$ as $\hat{\mathbf{a}}(\boldsymbol{\xi})$:

$$[\boldsymbol{\mathcal{F}}\mathbf{a}](\boldsymbol{\xi}) \;=\; \hat{\mathbf{a}}(\boldsymbol{\xi}) \;=\; \int_D \mathbf{a}(\mathbf{x})\, e^{-i\,\boldsymbol{\omega}\mathbf{x}}\, d\mathbf{x}, \qquad \boldsymbol{\omega} = 2\pi\boldsymbol{\xi} \in \mathbb{Z}^d. \tag{2}$$

[1] $\mathcal{A}$ and $\mathcal{U}$ are Banach function spaces (e.g., Sobolev spaces) defined on a bounded domain $D \subset \mathbb{R}^d$.

[2] In practice, each function is sampled on a discretized grid in D and stored as a vector.

For fixed set of retained modes $\boldsymbol{\xi}_i \in \Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m\} \subset \mathbb{Z}^d$, truncated Fourier transform $\boldsymbol{\mathcal{F}}_m : \mathcal{A} \to \mathbb{C}^m$ and its band-limited inverse $\boldsymbol{\mathcal{F}}_m^{-1} : \mathbb{C}^m \to \mathcal{U}$ can be defined as:

$$\boldsymbol{\mathcal{F}}_m(\mathbf{a}) = [\widehat{\mathbf{a}}(\boldsymbol{\xi}_1), \ldots, \widehat{\mathbf{a}}(\boldsymbol{\xi}_m)], \qquad \boldsymbol{\mathcal{F}}_m^{-1}(\boldsymbol{\mathcal{F}}_m \mathbf{a})(\mathbf{x}) \;=\; \sum_{j=1}^{m} \widehat{\mathbf{a}}(\boldsymbol{\xi}_j)\, e^{2\pi i\, \boldsymbol{\xi}_j \cdot \mathbf{x}}, \tag{3}$$

with a slight abuse of notation. A single Fourier layer $\boldsymbol{\mathcal{L}}_{\text{FNO}}$ of FNO is written as:

$$\boldsymbol{\mathcal{L}}_{\text{FNO}}(\mathbf{a})(\mathbf{x}) \;=\; \sigma\Big(\boldsymbol{\mathcal{F}}_m^{-1}\Big(\mathbf{R}_{\theta_1}(\boldsymbol{\xi}) \cdot\; \boldsymbol{\mathcal{F}}_m(\mathbf{a})(\boldsymbol{\xi})\Big)(\mathbf{x}) \;+\; \mathbf{W}_{\theta_2} \cdot \mathbf{a}(\mathbf{x})\Big) \tag{4}$$

with learnable spectral block-diagonal multiplier $\mathbf{R}(\boldsymbol{\xi})$, parametrized linear transformation $\mathbf{W}$, and point-wise nonlinear activation σ. FNO is comprised of iterative $\boldsymbol{\mathcal{L}}_{\text{FNO}}$ between lift-up ($\boldsymbol{\mathcal{P}}$) and projection ($\boldsymbol{\mathcal{Q}}$) networks:

$$\boldsymbol{\mathcal{G}}_\theta^{\text{FNO}}(\mathbf{a}) \;=\; \boldsymbol{\mathcal{Q}} \circ \boldsymbol{\mathcal{L}}_{\text{FNO}}^{(\ell)} \circ \cdots \circ \boldsymbol{\mathcal{L}}_{\text{FNO}}^{(1)} \circ \boldsymbol{\mathcal{P}}(\mathbf{a})\,. \tag{5}$$

In the perspective of the operator network formulation (1) (Lanthaler et al., 2022), FNO hard-codes its encoder $\boldsymbol{\mathcal{E}}_m$ as $\boldsymbol{\mathcal{F}}_m$ and reconstructor $\boldsymbol{\mathcal{R}}_m$ as $\boldsymbol{\mathcal{F}}_m^{-1}$, then learns the latent map $\mathbf{T}$ by its iterative layers of parametrized linear kernels interleaved by non-linear activations (Kovachki et al., 2021).

2.2 Kolmogorov–Arnold Network

KAN (Liu et al., 2024b;a) replaces fixed node activations of traditional MLP with simple sum operations and train the learnable univariate 1D functions ϕ on edges. With layer width $n_l \to n_{l+1}$ and input field $\mathbf{x}^{(l)} \to \mathbf{x}^{(l+1)}$, a KAN layer yields a function matrix $\mathbf{\Phi}^{(l)}$ at l^{th} layer as

$$\mathbf{x}^{(l+1)} = \mathbf{\Phi}^{(l)} \mathbf{x}^{(l)}{}^{3}, \qquad \mathbf{\Phi}^{(l)} = \Big[\, \phi_{q,p}^{(l)}(\,\cdot\,) \,\Big]_{q=1,\ldots,n_{l+1}}^{p=1,\ldots,n_l}, \qquad x_q^{(l+1)} = \sum_{i=1}^{n_l} \phi_{q,p}^{(l)}(x_p^{(l)}), \tag{6}$$

so each output channel is a sum of edgewise transforms of the inputs (Liu et al., 2024b;a). In the original KAN each edge function is a spline expansion

$$\phi_{q,p}^{(\ell)}(t) \;=\; c_{q,p,0}^{(\ell)}\, b(t) \;+\; \sum_{i=1}^{g} c_{q,p,i}^{(\ell)}\, B_i(t), \tag{7}$$

with learnable coefficients for a fixed base 1D function $b(\cdot)$ and 1D B–spline basis $\{B_i\}$. Because every $\phi_{q,p}$ is a 1D curve, KANs are directly inspectable and amenable to visualization followed by symbolic regression. On expressivity, Wang et al. (2024) theoretically prove that KANs match MLPs up to constant depth and width factors; empirically, with appropriate optimization recipes, KANs and MLPs exhibit comparable scaling on PDE and operator benchmarks (Shukla et al., 2024). Hence, swapping a latent MLP for a KAN preserves expressivity while enabling symbolic readout.

3 Theoretical Analysis on FNO's Pure-Spectral Bottleneck

This section first illustrates the pure-spectral bottleneck of FNO. Then we provide a theoretical analysis and prove that FNO suffers from the curse of dimensionality for position-dependent dynamics.

3.1 The Pure-Spectral Bottleneck of FNO

We consider the 1D quantum harmonic oscillator Hamiltonian as an example:

$$\boldsymbol{\mathcal{H}} a(x) \;=\; -\partial_{xx} a(x) \;+\; x^2 \cdot a(x). \tag{8}$$

We define the class of operators that involve spatial multiplication, such as $\boldsymbol{\mathcal{H}}$, as *position-dependent dynamics*. Multiplication and differentiation have a dual relationship under the Fourier transform:

$$\boldsymbol{\mathcal{F}}[(-\partial_{xx} a)](\xi) \;=\; \xi^2 \cdot \hat{a}(\xi), \qquad \boldsymbol{\mathcal{F}}[(x^2 \cdot a)](\xi) \;=\; -\,\partial_{\xi\xi} \hat{a}(\xi). \tag{9}$$

In spectral domain, the spatial differential ∂_{xx} is a spectral multiplier ξ^2, whereas the spatial multiplier x^2 becomes a spectral differential $\partial_{\xi\xi}$. Consider a truncated polynomial basis $\{1, x, x^2, \ldots, x^{n-1}\}$

[3] Akin to matrix-vector multiplication but follows the third equation Eq. 6 instead of row-vector inner product.

and a truncated Fourier basis $e_k(\theta) = e^{ik\theta}$, $k = 0, \dots, n-1$, on a periodic domain. In the spatial (polynomial) basis, the map $a(x) \mapsto x^2 \cdot a(x)$ acts as a two-step up-shift sparse matrix

$$\mathbf{S}_n^{(2)} := \begin{bmatrix} 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{10}$$

while in the spectral (Fourier) basis it is a dense Toeplitz matrix (Morrison, 1995)

$$\mathbf{T}_n[x^2] := \begin{bmatrix} c_0 & c_{-1} & c_{-2} & \cdots & c_{-n+1} \\ c_1 & c_0 & c_{-1} & \cdots & c_{-n+2} \\ c_2 & c_1 & c_0 & \cdots & c_{-n+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \cdots & c_0 \end{bmatrix}, \qquad c_m = \frac{1}{2\pi} \int_0^{2\pi} \theta^2 e^{-im\theta}\, d\theta. \tag{11}$$

Thus each term in $\boldsymbol{\mathcal{H}}$ is sparse in one basis and dense in the other (Morrison, 1995).

An FNO layer $\boldsymbol{\mathcal{L}}_{\text{FNO}}$ (4) can easily parametrize $-\partial_{xx}$ by taking $\mathbf{R}(\xi) \approx \xi^2$. However, approximating the dense *off*-diagonals in $\mathbf{T}_n[x^2]$ to parametrize x^2 must rely on the non-linear activation $\sigma(\cdot)$ since $\mathbf{R}(\xi)$ and $\mathbf{W}$ are spectrally diagonal and hence incapable of mixing modes. Let $\mathbf{z}(\mathbf{u})$ denote the pre-activation for input $\mathbf{u}$, then the Jacobian of $\boldsymbol{\mathcal{L}}_{\text{FNO}}$ at $\mathbf{u}$ gives the first-order approximation of the learned map and its Fourier transform reveals the spectral off-diagonals of itself as

$$\boldsymbol{\mathcal{F}}[\mathbf{J}(\mathbf{u})](\boldsymbol{\xi}, \boldsymbol{\xi}') \;=\; \Big(\boldsymbol{\mathcal{F}}[\sigma'\big(\mathbf{z}(\mathbf{u}, \cdot)\big)]\,[\,\boldsymbol{\xi} - \boldsymbol{\xi}'\,]\Big) \cdot \Big(\mathbf{W} \;+\; \mathbf{R}_\theta(\boldsymbol{\xi}')\Big). \tag{12}$$

Therefore, *all* off-diagonals arise from the spectrum of the $\mathbf{u}$-dependent gate $\sigma'(\mathbf{z}(\mathbf{u}, \cdot))$: FNO's nonlinearity *can* create off-diagonals, but they are tied to the input distribution of $\mathbf{u}$. This is the pure-spectral bottleneck of FNO: *spectral off-diagonals of a learned FNO are tied to the train subspace, hence FNO can converge only on the in-sample mapping that fails outside the train distribution*[4].

3.2 FNO SUFFERS CURSE OF DIMENSIONALITY ON POSITION-DEPENDENT DYNAMICS

As explained previously, *position operator*, $a(x) \mapsto x \cdot a(x)$, is a highly dense Toeplitz map in the spectral basis (Morrison, 1995). Based on the Remark 21 & 22 of Kovachki et al. (2021), we prove that any position-dependent dynamics induces super-exponential scaling in FNO size by the desired error bound: *FNO cannot converge closely as desired on a generic position-dependent dynamics with practical model size, hence can only overfit on the in-sample mapping.* We provide Lemma 1, that a single position operator already spreads the input spectra too much for FNO to stay practical, and Theorem 1, expanding Lemma 1 to an arbitrary composition of position operators.

Following from the operator network formulation (1), the error estimate of an operator network $\boldsymbol{\mathcal{G}}_\theta$ approximating the ground-truth operator $\boldsymbol{\mathcal{G}}$ in an operator norm is bounded as

$$\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{G}}_\theta\| \;\le\; \underbrace{\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{R}}_{m'} \circ \widehat{\mathbf{T}} \circ \boldsymbol{\mathcal{E}}_m\|}_{\textbf{projection error}:\,\epsilon_{\text{proj}}} \;+\; \underbrace{\|\boldsymbol{\mathcal{R}}_{m'} \circ (\widehat{\mathbf{T}} - \mathbf{T}_\theta) \circ \boldsymbol{\mathcal{E}}_m\|}_{\textbf{latent network error}:\,\epsilon_{\text{net}}}, \tag{13}$$

by the triangle inequility. Latent network error ϵ_{net} follows the well-established scaling law of conventional neural networks (Hornik et al., 1989; Cybenko, 1989). Therefore, whether an operator network is efficient in model and sample size to achieve the desired accuracy hinges on the scalability of the projection error ϵ_{proj} (Lanthaler et al., 2022; Kovachki et al., 2021; 2024a).

Reviewing Remark 21 & 22 of Kovachki et al. (2021) ϵ_{proj} of FNO is governed by the Fourier tail, the sum of Fourier coefficients outside the retained spectrum Ξ truncated by width m (Gottlieb & Orszag, 1977; Trefethen, 2000): to achieve the desired ϵ_{proj} with practical m, both input and output Fourier tails must decay algebraically or faster. However, even in the optimal case of the band-limited input, if the ground-truth operator is spectrally dense to spread out the input spectra and induce algebraic or slower decay in output Fourier tail, m must scale at least polynomially to suppress ϵ_{proj}:

[4]This issue of out-of-distribution fragility from underspecification is well studied by D'Amour et al. (2022)

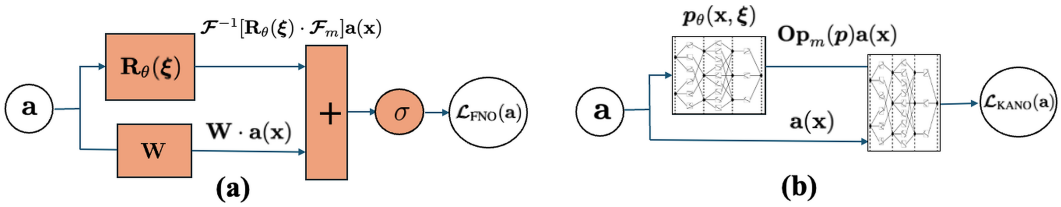

Figure 1: **(a)** $\mathcal{L}_{\text{FNO}}$ architecture. **(b)** $\mathcal{L}_{\text{KANO}}$ architecture.

$m \sim \mathcal{O}(\epsilon_{\text{proj}}^{-1/s})$ where s is a geometric constant. Meanwhile, as the latent mapping would be also dense, the size of the latent network, $\mathcal{N}_{\text{net}}$, follows the canonical polynomial neural scaling (Yarotsky, 2017; De Ryck et al., 2021) by the desired ϵ_{net} with its width m^d (d is the input domain dimension) as the exponent: $\mathcal{N}_{\text{net}} \sim \mathcal{O}(\epsilon_{\text{net}}^{-m^d})$. Consequently, this results as the super-exponential scaling in the latent network size $\mathcal{N}_{\text{net}} \sim \mathcal{O}(\epsilon_{\text{net}}^{-\epsilon_{\text{proj}}^{-d/s}})$ even with the optimal band-limited input: *scaling width m to suppress ϵ_{proj} explodes $\mathcal{N}_{net}$ to achieve the desired ϵ_{net} for a generic dense operator.*

Lemma 1 (**Position operator elongates Fourier tail**). *A single position operator, spatial multiplier by x, induces algebraic decay in output Fourier tail when the input is band-limited.*

Sketch of proof. Position operator is kernel $\widehat{x}(\xi) \propto 1/\xi$ in spectral basis. Hence, every mode outside the input spectrum picks up a coefficient of size $\sim 1/|\xi|$, ending up as $|\hat{v}(\xi)| \gtrsim 1/|\xi|$ in the output spectrum. See Appendix C.2 for restatement and full proof. □

Theorem 1 (**FNO fails on position operators**). *Any arbitrary composition of position operators requires FNO to scale super-exponentially on its model size by the desired accuracy.*

Sketch of proof. Iteratively apply Lemma 1, then any arbitrary composition of position operators induce algebraic or slower decay in output Fourier tail even for the optimal band-limited input. This results in super-exponential scaling of latent network size by the desired error as discussed above (Kovachki et al., 2021). See Appendix C.3 for restatement and full proof. □

4 Kolmogorov–Arnold Neural Operator

Motivated by the pure-spectral bottleneck of FNO, we propose the Kolmogorov-Arnold Neural Operator (KANO), an operator network capable of converging closely as desired on a generic position-dependent dynamics with practical model size. We first introduce the KANO architecture, and provide theoretical analysis on its dual-domain expressivity in the following section.

4.1 KANO Architecture

KANO utilizes an iterative structure of KANO layers $\mathcal{L}_{\text{KANO}}$ to learn the unknown operator, akin to FNO. However, KANO excludes lift-up and projection networks to maximize tractability:

$$\mathcal{G}_\theta^{\text{KANO}} \;=\; \mathcal{L}_{\text{KANO}}^{(\ell)} \circ \cdots \circ \mathcal{L}_{\text{KANO}}^{(1)}. \tag{14}$$

$$\mathcal{L}_{\text{KANO}}(\mathbf{a})(\mathbf{x}) \;=\; \boldsymbol{\Phi}_{\theta_1}\Big(\mathcal{F}_m^{-1}\big[\,\boldsymbol{p}_{\theta_2}(\mathbf{x}, \boldsymbol{\xi}) \,*\, \mathcal{F}_m(\mathbf{a})(\boldsymbol{\xi})\,\big](\mathbf{x})\;,\; \mathbf{a}(\mathbf{x})\Big), \tag{15}$$

where $\boldsymbol{\Phi}$ is a KAN sub-network for learnable non-linear activation; $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$ is another KAN sub-network, a pseudo-differential symbol jointly parametrized by both spatial $\mathbf{x}$ and spectral $\boldsymbol{\xi}$ bases[5]. Note the "$*$" notation in $\mathcal{L}_{\text{KANO}}$ (15) instead of the block-diagonal multiplication notation "$\cdot$" in $\mathcal{L}_{\text{FNO}}$ (4). The spatial basis $\mathbf{x}$ of the symbol $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$ is convolution (differential) in spectral domain by the dual relationship (9). Therefore, the pseudo-differential symbol calculus of $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$ needs to be done by quantizing on both spatial and spectral domain (Hörmander, 2007), and we choose Kohn-Nirenberg quantization (Kohn & Nirenberg, 1965) to compute the symbol calculus in $\mathcal{L}_{\text{KANO}}$:

$$\mathcal{F}_m^{-1}\big[\,\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi}) \,*\, \mathcal{F}_m(\mathbf{a})(\boldsymbol{\xi})\,\big](\mathbf{x}) := \Big(\frac{h}{L}\Big)^d \sum_{\boldsymbol{\xi} \in \Xi} \sum_{\mathbf{y} \in \mathcal{Y}} e^{i(\mathbf{x}-\mathbf{y})\cdot\boldsymbol{\xi}}\, \boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})\, \mathbf{a}(\mathbf{y}), \tag{16}$$

[5] Shin et al. (2022) first employed pseudo-differential operator framework for neural operator. They presumed the symbol $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$ to be separable as $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi}) = \boldsymbol{p}_{\mathbf{x}}(\mathbf{x}) \cdot \boldsymbol{p}_{\boldsymbol{\xi}}(\boldsymbol{\xi})$, and used MLP sub-networks while retaining the lift-up and projection networks of a generic neural operator architecture (Kovachki et al., 2023).

where for a periodic domain $D = (-\frac{L}{2}, \frac{L}{2})^d$, $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\} \subset D$ is a uniform discretization with spacing h and $\mathbf{x} \in D$ is an evaluation point. We denote Kohn-Nirenberg quantization (16) as an operator $\mathbf{Op}_m(\boldsymbol{p}) := \boldsymbol{\mathcal{F}}_m^{-1}\big[\, \boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi}) \ast \boldsymbol{\mathcal{F}}_m \big]$ defined by the symbol $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$. In the operator network formulation (1) introduced in Section 2.1, KANO's projection $\boldsymbol{\Pi}_{\text{KANO}}$ is then defined as:

$$\boldsymbol{\Pi}_{\text{KANO}}(\boldsymbol{\mathcal{G}}) \;:=\; \mathbf{Op}_m\big(\boldsymbol{p}_{\boldsymbol{\mathcal{G}}}\big), \qquad \boldsymbol{p}_{\boldsymbol{\mathcal{G}}} \in \arg\min_{\boldsymbol{p}} \big\|\boldsymbol{\mathcal{G}} - \mathbf{Op}_m(\boldsymbol{p})\big\|. \tag{17}$$

Symbolic Interpretability of KANO. By using compact KANs each for the symbol $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$ and non-linear activation $\boldsymbol{\Phi}$ in every KANO layer $\boldsymbol{\mathcal{L}}_{\text{KANO}}$ (15), KANO network $\boldsymbol{\mathcal{G}}_\theta^{\text{KANO}}$ (14) is fully inspectable by visualizing the learned edges of all its KANs, potentially allowing closed-form symbolic formula of the learned operator with the manual provided by Liu et al. (2024b;a).

4.2 KANO's Dual-Domain Expressivity

In contrast to FNO, KANO exploits sparse representations in both the spatial and spectral domains, hence decoupling the scaling of ϵ_{proj} and ϵ_{net} by never letting the latent map be a dense convolution. For instance, for the quantum harmonic oscillator in Eq. (8), a KANO layer $\boldsymbol{\mathcal{L}}_{\text{KANO}}$ (15) can parametrize $\boldsymbol{\mathcal{H}}$ by taking $\boldsymbol{p}(x, \xi) \approx x^2 + \xi^2$, both $-\partial_{xx}$ and x^2 terms are each represented where they are sparse, both leveraging the shift form $\mathbf{S}_n^{(2)}$ (10). By *jointly* parameterizing the operator in both spatial and frequency domains, KANO *cherry-picks* the sparse representation for every term in position-dependent dynamics, building the right inductive bias well-known to be essential for out-of-distribution generalization and model efficiency (Goyal & Bengio, 2022; Trask et al., 2018).

This dual-domain expressivity of KANO first alleviates the input constraint; we first explain that ϵ_{proj} of KANO scales practically by its width m for *any* physical input. Then we provide Theorem 2: as long as the KANO projection (17) of an operator generates smooth symbols KAN can easily approximate, ϵ_{net} scales practically by compact KAN sub-networks independent of ϵ_{proj}. In conclusion, *KANO can converge closely as desired to a generic position-dependent dynamics with practical model size using any physical input, robustly generalizing outside the train subspace.*

KANO practically has no input constraint. According to the quadrature bound from Demanet & Ying (2011, Thm. 1&2), the error estimate of Kohn-Nirenberg quantization (16) obeys

$$\big\|\boldsymbol{\mathcal{G}} - \mathbf{Op}_m(\boldsymbol{p}_{\boldsymbol{\mathcal{G}}})\big\| \;\le\; C\, B\, m^{-s}, \tag{18}$$

given norm-bound (finite-energy)[6] input of $\mathcal{A}_B = \{\mathbf{u} : \|\mathbf{u}\| \le B\}$ where s, C are geometric constants. Hence KANO width m scales polynomially by the desired ϵ_{proj} given any physical data.

Theorem 2 (KANO stays practical for smooth symbol). *If the KANO projection of an operator $\boldsymbol{\mathcal{G}}$, $\boldsymbol{\Pi}_{\text{KANO}}(\boldsymbol{\mathcal{G}})$ (17), generates a finite composition of smooth symbols $\boldsymbol{p}_{\boldsymbol{\mathcal{G}}}(\mathbf{x}, \boldsymbol{\xi})$ and finite-degree non-linearities, the model size of KANO scales polynomially by the desired accuracy ε.*

Sketch of proof. Choosing $m \sim (B/\varepsilon)^{1/s}$ scales projection error down to $\varepsilon/2$ by Eq. (18). A fixed-width KAN then approximates the symbols to accuracy $\varepsilon/2$ with $\mathcal{O}(\varepsilon^{-d/(2s_p)})$ parameters (Wang et al., 2024, Corol. 3.4) (s_p is a geometric constant). The finite-degree non-linearities add only constant-size weights by the activation KAN, so the total parameter count is $\mathcal{O}(\varepsilon^{-d/(2s_p)})$. See Appendix C.4 for restatement and full proof. □

Corollary 1 (KANO is practical for generic position-dependent dynamics). *For a finite composition of spatial and spectral multipliers of maximum r-differentiable symbols with finite-degree non-linearity, Theorem 2 yields $|\Theta| = \mathcal{O}\big(\varepsilon^{-d/(2r)}\big)$.*

5 Experimental Results

5.1 Synthetic-Operator Generalization Benchmarks

We benchmark FNO and KANO on three position-dependent operators:

$$\boldsymbol{\mathcal{G}}_1 f = x^2 \cdot f \;-\; \partial_{xx} f, \quad \boldsymbol{\mathcal{G}}_2 f = x \cdot \partial_x f \;+\; \partial_{xx} f, \quad \boldsymbol{\mathcal{G}}_3 f = f^3 + x \cdot \partial_x f + \partial_{xx} f.$$

Our goal is to quantify and compare the generalization of KANO and FNO. We train the models *only* with Group A dataset and evaluate them on the *unseen* Group B dataset.

[6] Norm here and Equation Eq. 18 is the Sobolev norm

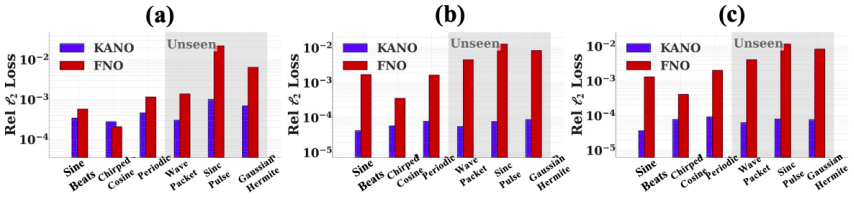

Figure 2: Loss test results. **(a)** $\mathcal{G}_1$ **(b)** $\mathcal{G}_2$ **(c)** $\mathcal{G}_3$. Note the logarithmic scale.

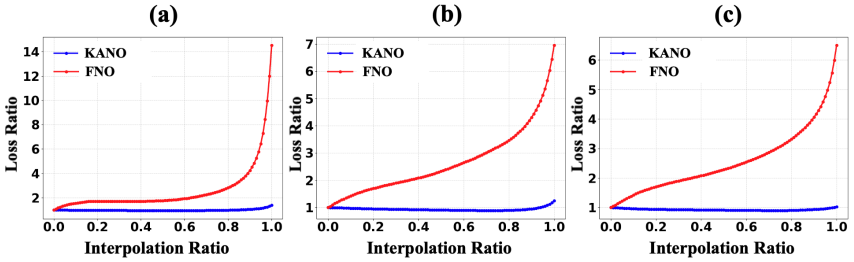

Figure 3: Interpolation test results. **(a)** $\mathcal{G}_1$ **(b)** $\mathcal{G}_2$ **(c)** $\mathcal{G}_3$.

- **Group A (Training families):** Periodic, Chirped Cosine, Sine Beats.
- **Group B (Testing families):** Sinc Pulse, Gaussian×Hermite, Wave Packet.

For each operator, we generate 2000 train pairs from Group A and 400 test pairs each from Group A and B to evaluate the generalization by comparing the average relative ℓ_2 loss (**Loss Test**). We also interpolate the Group A and B function samples in 100 steps, apply ground-truth operators in each step to build the interpolated dataset, and evaluate the loss ratio to that of the Group A samples (**Interpolation Test**). We trained FNO models of 2 layers, 64 width with no mode truncation, and used one-layer KANs of grid 10 cubic B-splines edges for the KANO model. We used Adam optimizer and relative ℓ_2 loss for training. See Appendix B.1 for experiment details.

Results. As shown in Table 1 and Figure 2, KANO shows consistent losses over Group A and Group B. In contrast, FNO shows significant loss increases in Group B dataset. The red curves of the interpolation test in Figure 3 show slow increases on early and mid-interpolation, suggesting that the FNO's learned in-sample mappings are yet close to the ground-truth operators. On the other hand, the red curves abruptly soar up in the latter ratio, suggesting that the interpolated functions are now far outside the train distribution. These results, together with KANO's one-order–of–magnitude lower loss at just 0.03% of FNO's size, are consistent with our claims in previous sections. After convergence, we visualized the embedded KANs (Figure 4). We then froze these learned symbols and continued training, referring to this variant as KANO_symbolic. KANO_symbolic recovered the exact symbolic coefficients of the ground-truth operator to within the fourth decimal place (Table 2). KANO's loss matches KANO_symbolic's loss in Table 1, confirming that KANO converged close to the ground-truth operator.

Table 1: Relative ℓ_2 losses ($\times 10^{-4}$) and parameter counts.

	$\mathcal{G}_1$		$\mathcal{G}_2$		$\mathcal{G}_3$	
Model (params)	A	B	A	B	A	B
FNO (566 k)	6.36	98.8	10.6	87.0	11.4	81.4
KANO (152)	1.04	1.44	0.629	0.749	0.716	0.737
KANO_SYMBOLIC	0.512	0.526	0.498	0.500	0.520	0.536

5.2 Long-Horizon Quantum Dynamics Benchmark

We provide this benchmark on two position-dependent quantum dynamics: the quartic double-well Hamiltonian (DW) and the nonlinear Schrödinger equation with cubic nonlinearity (NLSE):

$$i\partial_t\psi = -\tfrac{1}{2}\partial_{xx}\psi + w(x)\cdot\psi \text{ (DW)}, \quad i\partial_t\psi = -\tfrac{1}{2}\partial_{xx}\psi + w(x)\cdot\psi + |\psi|^2\cdot\psi \text{ (NLSE)},$$

where $w(x) = x^4 - \left(x - \frac{1}{32}\right)^2 + 0.295$.

We generate 200 initial states and yield the state trajectories by the Hamiltonians, sampling momentum/position probability mass functions (PMFs) every 0.1ms for 100 time steps. The first 10 time steps are used for training, and the rest are used to evaluate the long-horizon prediction.

We modify KANO to capture the quantum state evolution: ***Q-KANO***. Symbol $\boldsymbol{p}_\theta$ is parametrized as $\exp\left[-i\Delta T\,\boldsymbol{\phi}_\theta(\mathbf{x},\boldsymbol{\xi})\right]$, where $\Delta T = 0.1$ ms. The adaptive activation is also defined as a complex

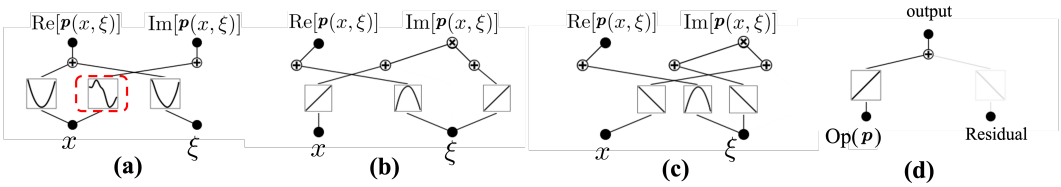

Figure 4: **(a)** $\boldsymbol{p}(x,\xi)$ of $\boldsymbol{\mathcal{G}}_1$. The middle edge does not contribute to the output. **(b)** $\boldsymbol{p}(x,\xi)$ of $\boldsymbol{\mathcal{G}}_2$. **(c)** $\boldsymbol{p}(x,\xi)$ of $\boldsymbol{\mathcal{G}}_3$. **(d)** $\boldsymbol{\Phi}$ of $\boldsymbol{\mathcal{G}}_3$. Edge of the residual in **(d)** looks linear, so we compared two scenarios, linear and cubic, which the latter achieved lower loss and better generalization.

Table 2: Ground-truth vs. learned operators (coefficients rounded to 4^{th} decimal place).

Ground-truth operator	**Learned KANO operator**
$\boldsymbol{\mathcal{G}}_1 f = x^2 \cdot f - \partial_{xx} f$	$\tilde{\boldsymbol{\mathcal{G}}}_1 f = (x^2 + 0.0003) \cdot f - \partial_{xx} f$
$\boldsymbol{\mathcal{G}}_2 f = x \cdot \partial_x f + \partial_{xx} f$	$\tilde{\boldsymbol{\mathcal{G}}}_2 f = 0.9996\, x \cdot \partial_x f + \partial_{xx} f - 0.0003\, f$
$\boldsymbol{\mathcal{G}}_3 f = f^3 + x \cdot \partial_x f + \partial_{xx} f$	$\tilde{\boldsymbol{\mathcal{G}}}_3 f = 1.0001\, f^3 + 0.99997\, x \cdot \partial_x f + 0.99997\, \partial_{xx} f$ $- 0.0002\, f^2 - 0.0003\, f - 0.0001$

exponential with learned phase $\boldsymbol{\vartheta} = \boldsymbol{\Phi}_\theta(|\mathbf{Op}_m(\boldsymbol{p}_\theta)\boldsymbol{\psi}|, \angle\mathbf{Op}_m(\boldsymbol{p}_\theta)\boldsymbol{\psi})$ for input wave function $\boldsymbol{\psi}(\mathbf{x})$:

$$\boldsymbol{\mathcal{G}}_\theta^{\text{Q-KANO}}[\boldsymbol{\psi}] = \mathbf{Op}_m(\exp[-i\Delta T\, \boldsymbol{\phi}_\theta(\mathbf{x}, \boldsymbol{\xi})])\boldsymbol{\psi} \;\cdot\; e^{-i\Delta T\, \boldsymbol{\vartheta}}. \tag{19}$$

We investigate three supervision scenarios: **Full**-type training with full wave function, idealistic yet physically unattainable, **Pos**-type training with only position PMF, physically realistic yet the least informative, and **pos&mom**-type training with both position *and* momentum PMFs, which remains physically attainable while providing richer information although not full. We use Adam optimizer for all trainings. See Appendix B.2 for experiment details.

Results. We evaluate state infidelity[7] between ground-truth evolution and model prediction at each time step (Table 3, Figure 6). In case of KANO, the **pos & mom**-type training achieves indistinguishable infidelity from the ideal **full**-type training baseline. Meanwhile, the **pos**-type training displays a clear infidelity increase, especially on the NLSE. In contrast, even with **full** type training, FNO fails to maintain low state infidelity after the long-horizon propagation as expected. Iterative time evolution pushes the wave function far outside the train convex hull, and FNO's learned in-sample mapping deviates from the ground-truth evolution rapidly, leading to four orders of infidelity increase compared to KANO.

Table 3: State infidelity after 90 additional time-evolution steps.

Model & Train Type	State Infidelity	
	Double-Well	NLSE
FNO (full)	1.5×10^{-2}	1.6×10^{-2}
Q-KANO (full)	6.3×10^{-6}	6.8×10^{-6}
Q-KANO (pos & mom)	6.3×10^{-6}	5.6×10^{-6}
Q-KANO (pos)	4.7×10^{-3}	6.1×10^{-2}
Q-KANO_SYMBOLIC (full)	2.0×10^{-8}	2.0×10^{-8}
Q-KANO_SYMBOLIC (pos & mom)	2.0×10^{-8}	3.0×10^{-8}
Q-KANO_SYMBOLIC (pos)	5.3×10^{-2}	6.1×10^{-2}

Table 4 juxtaposes the learned symbols with that of the ground-truth Hamiltonians and Figure 5 shows the KAN visualizations from **pos & mom**-type training. With **full**-type training coefficients are recovered to the fourth decimal place, vindicating the ideal capacity of KANO when the information is fully provided. Under the realistic **pos & mom**-type training, the reconstruction remains accurate except for two terms: the constant (global phase) and the NLSE's cubic coefficient. Both discrepancies are predicted by quantum observability: global phases cancel in all PMFs, and the Kerr coefficient enters only through higher-order correlations that become harder to estimate from finite-shot statistics. Q-KANO faithfully reveals what the data support and nothing more.

[7]For predicted state $\tilde{\boldsymbol{\varphi}}$ and ground-truth state $\boldsymbol{\varphi}$, the state fidelity F is defined as the inner product between them ($F := <\tilde{\boldsymbol{\varphi}}, \boldsymbol{\varphi}>$), and the state infidelity is defined as $(1 - F)$, hence shows how distant two states are.

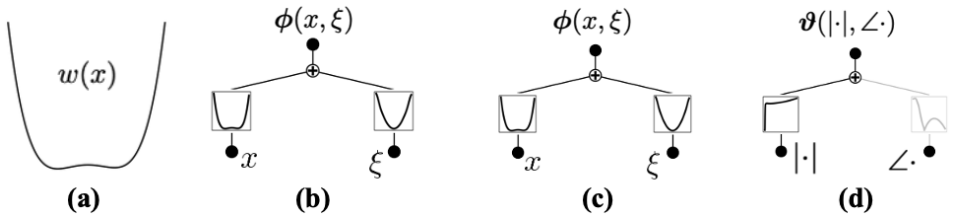

Figure 5: **pos&mom** type training results. **(a)** Structure of the potential $w(x)$ **(b)** $\boldsymbol{p}(x,\xi)$ of DW. **(c)** $\boldsymbol{p}(x,\xi)$ of NLSE. **(d)** $\varphi(|\cdot|,\angle\cdot)$ of NLSE. Potential $w(x)$ structure is clearly reconstructed.

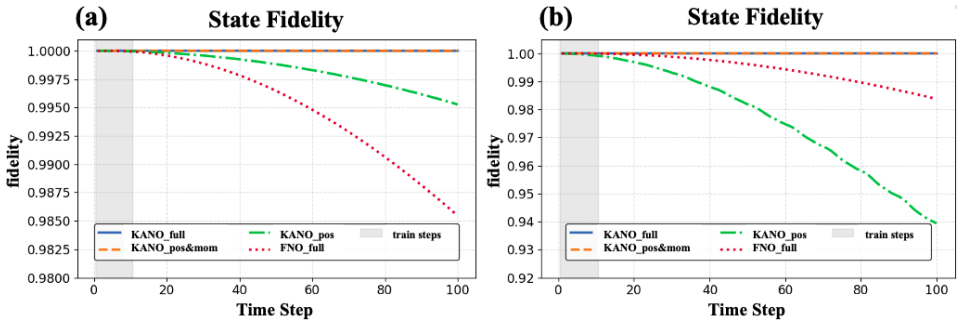

Figure 6: State fidelity over 100 time steps. **(a)** DW **(b)** NLSE

Table 4: Ground truth vs. learned symbols. Coefficients rounded to 4^{th} decimal place.

Hamiltonian	Train Type	Learned symbolic structure
DW	ground truth	$x^4 - x^2 + 0.0625\,x + 0.295 + 0.5\,\xi^2$
	full	$1.0004\,x^4 + 0.0001\,x^3 - 1.0013\,x^2 + 0.0625\,x + 0.2955 + 0.5\,\xi^2$
	pos & mom	$1.0003\,x^4 + 0.0001\,x^3 - 1.0008\,x^2 + 0.0623\,x + 0.0001 + 0.5\,\xi^2$
NLSE	ground truth	$x^4 - x^2 + 0.0625\,x + 0.295 + 0.5\,\xi^2 + \lvert\psi\rvert^2$
	full	$1.0005\,x^4 - 0.0001\,x^3 - 1.0014\,x^2 + 0.0626\,x + 0.2942 + 0.5\,\xi^2 + 0.9815\lvert\psi\rvert^2 + 0.0110\lvert\psi\rvert$
	pos & mom	$0.9999\,x^4 - 0.0003\,x^3 - 1.0001\,x^2 + 0.0630\,x + 0.1141 + 0.5\,\xi^2 + 0.9514\lvert\psi\rvert^2 - 0.5504\lvert\psi\rvert$

6 Conclusion

We have presented the Kolmogorov–Arnold Neural Operator, an interpretable neural operator expressive on a generic position-dependent dynamics. KANO cherry-picks sparse representations of each term via jointly parametrizing on both spectral and spatial bases, and achieves robust generalization outside train distribution while exposing clear tractable representation via its KAN sub-networks. In all our benchmarks, KANO have successfully recovered the ground-truth operators accurately to the fourth decimal place in coefficients. In addition to the superior out-of-distribution generalization, KANO has also achieved orders of magnitude lower losses with less than 0.03% of the model size compared to the FNO baseline. KANO shifts operator learning from an opaque, surrogate-based paradigm towards interpretable data-driven scientific modeling, and provides robust empirical evidence supporting its enhanced dual-domain expressivity and interpretability.

7 Reproducibility Statement

All codes and dataset of the experiments in this work are submitted as a zip file via the Supplementary Material. Details in the data generation and implementation are provided in Appendix B, and the full proofs of Theorems and Lemma are provided in Appendix C along with their mathematical restatements.

8 Ethics Statement

All authors of this work sincerely adhered to the ICLR Code of Ethics. We do not expect any potential violation to the best of our knowledge.

References

Diab W Abueidda, Panos Pantidis, and Mostafa E Mobasher. Deepokan: Deep operator network based on kolmogorov arnold networks for mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 436:117699, 2025.

Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

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.

Lily Chiparova and Vasil Popov. Kolmogorov–arnold networks for system identification of first-and second-order dynamic systems. *Engineering Proceedings*, 100(1):100059, 2025.

Richard Courant, David Hilbert, and Joseph David Harris. Methods of mathematical physics, volume ii: Partial differential equations. *American Journal of Physics*, 31(3):221–221, 1963.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.

Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D Hoffman, et al. Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*, 23(226):1–61, 2022.

Tim De Ryck, Samuel Lanthaler, and Siddhartha Mishra. On the approximation of functions by tanh neural networks. *Neural Networks*, 143:732–750, 2021.

Laurent Demanet and Lexing Ying. Discrete symbol calculus. *SIAM review*, 53(1):71–104, 2011.

Lawrence C Evans. *Partial differential equations*, volume 19. American mathematical society, 2022.

Nart Gashi, Panagiotis Kakosimos, and George Papafotiou. System identification using kolmogorov-arnold networks: A case study on buck converters. In *2025 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–7. IEEE, 2025.

David Gottlieb and Steven A Orszag. *Numerical analysis of spectral methods: theory and applications*. SIAM, 1977.

Anirudh Goyal and Yoshua Bengio. Inductive biases for deep learning of higher-level cognition. *Proceedings of the Royal Society A*, 478(2266):20210068, 2022.

Lars Hörmander. *The analysis of linear partial differential operators III: Pseudo-differential operators*. Springer Science & Business Media, 2007.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

Frank P Incropera, David P DeWitt, Theodore L Bergman, Adrienne S Lavine, et al. *Fundamentals of heat and mass transfer*, volume 1072. New York John Wiley & Sons, Inc., 1990.

John David Jackson and Ronald F Fox. Classical electrodynamics, 1999.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Benjamin C Koenig, Suyong Kim, and Sili Deng. Kan-odes: Kolmogorov–arnold network ordinary differential equations for learning dynamical systems and hidden physics. *Computer Methods in Applied Mechanics and Engineering*, 432:117397, 2024.

Joseph J Kohn and Louis Nirenberg. An algebra of pseudo-differential operators. *Communications on Pure and Applied Mathematics*, 18(1-2):269–305, 1965.

Takeshi Koshizuka, Masahiro Fujisawa, Yusuke Tanaka, and Issei Sato. Understanding the expressivity and trainability of fourier neural operator: A mean-field perspective. *Advances in Neural Information Processing Systems*, 37:11021–11060, 2024.

Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for fourier neural operators. *Journal of Machine Learning Research*, 22(290):1–76, 2021.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Nikola B Kovachki, Samuel Lanthaler, and Hrushikesh Mhaskar. Data complexity estimates for operator learning. *arXiv preprint arXiv:2405.15992*, 2024a.

Nikola B Kovachki, Samuel Lanthaler, and Andrew M Stuart. Operator learning: Algorithms and analysis. *arXiv preprint arXiv:2402.15715*, 2024b.

Pijush K Kundu, Ira M Cohen, David R Dowling, and Jesse Capecelatro. *Fluid mechanics*. Elsevier, 2024.

Samuel Lanthaler, Siddhartha Mishra, and George E Karniadakis. Error estimates for deeponets: A deep learning framework in infinite dimensions. *Transactions of Mathematics and Its Applications*, 6(1):tnac001, 2022.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Chaoyu Liu, Davide Murari, Chris Budd, Lihao Liu, and Carola-Bibiane Schönlieb. Enhancing fourier neural operators with local spatial features. *arXiv preprint arXiv:2503.17797*, 2025.

Ziming Liu, Pingchuan Ma, Yixuan Wang, Wojciech Matusik, and Max Tegmark. Kan 2.0: Kolmogorov-arnold networks meet science. *arXiv preprint arXiv:2408.10205*, 2024a.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*, 2024b.

Miguel Liu-Schiaffini, Julius Berner, Boris Bonev, Thorsten Kurth, Kamyar Azizzadenesheli, and Anima Anandkumar. Neural operators with localized integral and differential kernels. *arXiv preprint arXiv:2402.16845*, 2024.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Kent E Morrison. Spectral approximation of multiplication operators. *New York J. Math*, 1(75):96, 1995.

Raj Kumar Pathria. *Statistical Mechanics: International Series of Monographs in Natural Philosophy*, volume 45. Elsevier, 2017.

Shaoxiang Qin, Fuyuan Lyu, Wenhui Peng, Dingyang Geng, Ju Wang, Xing Tang, Sylvie Leroyer, Naiping Gao, Xue Liu, and Liangzhu Leon Wang. Toward a better understanding of fourier neural operators from a spectral perspective. *arXiv preprint arXiv:2404.07200*, 2024.

Jun John Sakurai and Jim Napolitano. *Modern quantum mechanics*. Cambridge University Press, 2020.

Jin Young Shin, Jae Yong Lee, and Hyung Ju Hwang. Pseudo-differential neural operator: Generalized fourier neural operator for learning solution operators of partial differential equations. *arXiv preprint arXiv:2201.11967*, 2022.

Khemraj Shukla, Juan Diego Toscano, Zhicheng Wang, Zongren Zou, and George Em Karniadakis. A comprehensive and fair comparison between mlp and kan representations for differential equations and operator networks. *Computer Methods in Applied Mechanics and Engineering*, 431:117290, 2024.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. *arXiv preprint arXiv:2111.13802*, 2021.

Andrew Trask, Felix Hill, Scott E Reed, Jack Rae, Chris Dyer, and Phil Blunsom. Neural arithmetic logic units. *Advances in neural information processing systems*, 31, 2018.

Lloyd N Trefethen. *Spectral methods in MATLAB*. SIAM, 2000.

Feng Wang, Hong Qiu, Yingying Huang, Xiaozhe Gu, Renfang Wang, and Bo Yang. Efkan: A kan-integrated neural operator for efficient magnetotelluric forward modeling. *arXiv preprint arXiv:2502.02195*, 2025.

Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

Yixuan Wang, Jonathan W Siegel, Ziming Liu, and Thomas Y Hou. On the expressiveness and spectral bias of kans. *arXiv preprint arXiv:2410.01803*, 2024.

Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.

Zipeng Xiao, Siqi Kou, Hao Zhongkai, Bokai Lin, and Zhijie Deng. Amortized fourier neural operators. *Advances in Neural Information Processing Systems*, 37:115001–115020, 2024.

Dmitry Yarotsky. Error bounds for approximations with deep relu networks. *Neural networks*, 94: 103–114, 2017.

Zhilin You, Zhenli Xu, and Wei Cai. Mscalefno: Multi-scale fourier neural operator learning for oscillatory function spaces. *arXiv preprint arXiv:2412.20183*, 2024.

APPENDIX

A TABLE OF NOTATION

Table 5: Main symbols and notation used in the paper.

Symbol	Meaning / Definition
Domains, spaces, and operators	
$D := (-L/2, L/2)^d$	Periodic d-dimensional spatial box of side length L
$\boldsymbol{\mathcal{G}}$	Ground-truth solution operator to be learned
Spectral & spatial sampling	
$\Xi = \{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m\}$	Retained Fourier modes (truncated spectrum); $m = \lvert\Xi\rvert$
$\boldsymbol{\mathcal{F}}$	Fourier transform
$\boldsymbol{\mathcal{F}}_m,\ \boldsymbol{\mathcal{F}}_m^{-1}$	Truncated Fourier transform by Ξ and its band-limited inverse
$\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_m\} \subset D$	Uniform spatial grid
h	Grid spacing of $\mathcal{Y}$
Fourier Neural Operator (FNO)	
$\boldsymbol{\mathcal{L}}_{\text{FNO}}$	Single FNO layer
$\boldsymbol{\mathcal{G}}_\theta^{\text{FNO}}$	FNO network
$\mathbf{R}_\theta(\boldsymbol{\xi})$	Learnable block-diagonal spectral multiplier
$\mathbf{W}_\theta$	Point-wise learnable linear map
$\sigma(\cdot)$	Point-wise non-linear activation
KANO layer, symbol calculus, and projection	
$\boldsymbol{\mathcal{L}}_{\text{KANO}}$	Single KANO layer
$\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$	Learnable pseudo-differential symbol
$\mathbf{Op}_m(\boldsymbol{p})$	Kohn-Nirenberg quantization of width m defined by $\boldsymbol{p}$
$\boldsymbol{\Pi}_{\text{KANO}}(\boldsymbol{\mathcal{G}})$	KANO projection of $\boldsymbol{\mathcal{G}}$
$\boldsymbol{\mathcal{G}}_\theta^{\text{KANO}}$	KANO network
$\boldsymbol{\Phi}_\theta$	Learnable activation
Kolmogorov–Arnold Network (KAN) primitives	
$\phi_{q,p}^{(\ell)}(\cdot)$	1D edge function on layer ℓ, connecting p^{th} node of layer ℓ to q^{th} node of layer $(\ell+1)$
$b(t),\ \{B_i(t)\}$	Base function and B-spline basis used to parametrize $\phi_{q,p}^{(\ell)}(t)$
Q-KANO (quantum dynamics) notation	
$\psi(x)$	Input wave function
$w(x)$	Quartic double-well potential
ΔT	Time step of propagation
$\boldsymbol{\phi}_\theta(\mathbf{x}, \boldsymbol{\xi})$	Parametrized phase for symbol $\boldsymbol{p}(\mathbf{x}, \boldsymbol{\xi})$ of Q-KANO
$\boldsymbol{\vartheta}_\theta(\lvert\cdot\rvert, \angle\cdot)$	Parametrized phase for non-linear activation of Q-KANO
$\boldsymbol{\mathcal{G}}_\theta^{\text{Q-KANO}}$	Q-KANO network
Function spaces	
$L^2(D)$	Square-integrable function space on domain D
$H^s(D)$	Sobolev function space of order $s \geq 0$ on domain D

B EXPERIMENT DETAILS

B.1 SYNTHETIC OPERATOR BENCHMARK

All experiments are carried out on periodic functions $f : \mathbb{T} \to \mathbb{R}$ with $\mathbb{T} = (-\pi, \pi]$ and a uniform trigonometric grid

$$x_j = x_{\min} + j\,\Delta x, \quad \Delta x = \frac{2\pi}{N}, \quad j = 0, \dots, N-1,$$

with $N = 128$. Unless noted otherwise every random quantity is drawn *independently for every sample*.

Outer envelope. To avoid the Gibbs phenomenon all basis functions are multiplied by a smooth taper that decays to zero in a $\pi/6$-wide buffer near the periodic boundary:

$$A(x) = \begin{cases} 1, & |x| \le 5\pi/6, \\ \cos^4\Big[\dfrac{|x| - 5\pi/6}{\pi/6}\,\dfrac{\pi}{2}\Big], & 5\pi/6 < |x| < \pi, \\ 0, & |x| \ge \pi. \end{cases}$$

The full "base" function is always $f_{\text{base}}(x) = A(x)\, g(x)$.

Spectral derivatives and ground-truth operator. Derivatives are computed with an exact Fourier stencil:

$$f'(x) \;=\; \boldsymbol{\mathcal{F}}^{-1}\big[\, i\xi\, \widehat{f}(\xi)\big], \qquad f''(x) \;=\; \boldsymbol{\mathcal{F}}^{-1}\big[-\xi^2 \widehat{f}(\xi)\big].$$

$U[a, b]$ denotes random digit drawn from range $[a, b]$.

Training families (Group A)

A1. sine_beats:

$$g(x) = \sin(\omega_1 x + \phi_1)\, \sin(\omega_2 x + \phi_2), \quad \omega_i = 8\, U[0.5, 3],\ \phi_i = U[0, 2\pi].$$

A2. chirped_cosine:

$$g(x) = \cos\big(\alpha x^2\big), \qquad \alpha = 12\, U[0.5, 2].$$

A3. periodic (random harmonic series):

$$g(x) = \sin(\omega x + \phi_1) + \cos(\omega x + \phi_2), \qquad \omega = 8\, U[0.5, 3],\ \phi_{1,2} = U[0, 2\pi].$$

Unseen families (Group B)

B1. wave_packet:

$$g(x) = \exp\Big[-\tfrac{(x-\mu)^2}{2\sigma^2}\Big]\, \sin\big(\omega x + \phi\big),\ \mu = U[-2, 2],\ \sigma = \tfrac{1}{12}\, U[0.5, 2],\ \omega = 12\, U[2, 6],\ \phi = U[0, 2\pi].$$

B2. sinc_pulse:

$$g(x) = \begin{cases} \dfrac{\sin(\alpha x)}{\alpha x}, & |x| > 10^{-12}, \\ 1, & |x| \le 10^{-12}, \end{cases} \qquad \alpha = 12\, U[0.5, 3].$$

B3. gaussian_hermite:

$$g(x) = H_n\Big(\frac{x - \mu}{\sigma}\Big)\, \exp\Big[-\tfrac{(x-\mu)^2}{2\sigma^2}\Big],\ n \in \{1, 2, 3\} \text{ uniform},\ \mu = U[-2, 2],\ \sigma = \tfrac{1}{8}\, U[0.5, 2],$$

where H_n is the degree-n Hermite polynomial.

NORMALIZATION

Each realization is divided by its maximum absolute value, $\|f\|_\infty$, to obtain $\|f\|_\infty = 1$. The envelope guarantees periodicity and keeps the numerical spectrum sharply band-limited.

Sample counts. #train $= 2000$ samples from the three Group A families for train data and #test $= 400$ samples each from the Group A and Group B families for generalization tests..

B.2 Quantum Dynamics Benchmark

We model a quantum apparatus with 200 state-preparation protocols each with perfect reproducibility, capable of generating an identical initial state whose wave function is drawn from one of the three families: *Periodic*, *Gaussian wave-packet*, and *Gaussian×Hermite*. The prepared initial states evolve under one of two unknown, time-independent Hamiltonians, and two arrays of 128 detectors measure position and momentum on uniform grids, yielding probability mass functions (PMFs) every 0.1ms for 100 time steps. PMFs collected from the first 10 time steps are used for training the models, and the rest of the PMFs collected from the remaining 90 time steps are used to evaluate the long-horizon fidelity drop beyond the train steps.

B.2.1 Quantum Apparatus Assumptions

1. **State preparation.** A *collection of calibrated protocols* can each prepare a designated initial wave-function each of one of three real-valued families *Periodic*, *Gaussian wave-packet*, or *Gaussian–Hermite*: $\boldsymbol{\psi}_0^{(m)}(x) \in L^2(\mathbb{T})^8$, $m = 1, \ldots, 200$. Repeated shots under the *same* protocol start from **exactly** the same $\boldsymbol{\psi}_0^{(m)}$, enabling trajectory-level reproducibility for every member of the ensemble.
2. **Hamiltonian stability.** The (unknown) Hamiltonian is *time–independent*, so trajectories are perfectly repeatable once $\boldsymbol{\psi}_0$ is fixed.
3. **Dual-basis detection.** Two 128-grid projective detectors measure the position basis $\{|x_i\rangle\}$ and the momentum basis $\{|\xi_j\rangle\}$, yielding empirical probability mass functions (PMFs) $\hat{p}_x(i) = |\psi(x_i)|^2$ and $\hat{p}_\xi(j) = |\widehat{\psi}(\xi_j)|^2$ on a common torus grid $\mathbb{T}_L$, $L = 4$.

B.2.2 Data Generation Details

For each of 200 distinct sample trajectories we

1. draw the initial wave function and propagate on the Hamiltonian with a high-resolution Strang split: $\delta t = 1\,\mu$s for 10 000 micro-steps, producing coarse snapshots every 100 steps ($\Delta T = 0.1$ ms, $T = 1, \ldots, 100$);
2. store $(\boldsymbol{\psi}_T,\ \mathbf{p}_x^T,\ \mathbf{p}_\xi^T)$ where $p_x^T(i) = |\psi_T(x_i)|^2$ and $p_\xi^T(j) = |\widehat{\psi}_T(\xi_j)|^2$.

Only the first 10 coarse steps are used for training; the remaining 90 steps test fidelity drop on long-horizon. All simulations employ an $n = 128$-point FFT grid to match the detectors.

Spatial discretization. We place the problem on a periodic box of length $L = 4$ with $N = 128$ grid points $x_j = x_{\min} + j\Delta x,\ \Delta x = L/N$. Periods suppress wrap–around artifacts because every initial state is tapered by the smooth envelope $A(x)$ defined in Appendix B.1. Spatial derivatives are taken spectrally: let $\xi_m = 2\pi m/L$ for $m = -N/2, \ldots, N/2 - 1$. Writing $\widehat{\psi}_m = \boldsymbol{\mathcal{F}}[\psi](\xi_m)$,

$$\partial_x \psi = \boldsymbol{\mathcal{F}}^{-1}\big[i\,\xi_m \widehat{\psi}_m\big], \qquad \partial_{xx}\psi = \boldsymbol{\mathcal{F}}^{-1}\big[-\xi_m^2 \widehat{\psi}_m\big].$$

Strang–splitting time integrator. Let $\boldsymbol{\mathcal{K}} := -\frac{1}{2}\partial_{xx}$ (kinetic), $\boldsymbol{\mathcal{V}} : \psi \mapsto w(x)\cdot\psi$ (potential) and $\boldsymbol{\mathcal{N}} : \psi \mapsto |\psi|^2 \cdot \psi$ (cubic nonlinearity). With time step Δt the second–order Strang factorization reads

$$e^{(\boldsymbol{\mathcal{K}}+\boldsymbol{\mathcal{V}}+\boldsymbol{\mathcal{N}})\Delta t} \;=\; e^{\frac{\Delta t}{2}(\boldsymbol{\mathcal{K}}+\boldsymbol{\mathcal{V}})}\, e^{\Delta t \boldsymbol{\mathcal{N}}}\, e^{\frac{\Delta t}{2}(\boldsymbol{\mathcal{K}}+\boldsymbol{\mathcal{V}})} \;+\; \mathcal{O}(\Delta t^3).$$

Because $\boldsymbol{\mathcal{K}}$ is diagonal in Fourier space and $\boldsymbol{\mathcal{V}}$ in real space we implement each half step explicitly:

$$\begin{aligned}
\psi \;&\xleftarrow{\boldsymbol{\mathcal{V}}/2}\; e^{-\frac{i}{2}\Delta t\, w(x)}\psi, \\
\widehat{\psi} \;&\xleftarrow{\boldsymbol{\mathcal{K}}/2}\; \boldsymbol{\mathcal{F}}[\psi], \quad \widehat{\psi}_m \leftarrow e^{-\frac{i}{4}\Delta t\, \xi_m^2}\widehat{\psi}_m, \\
\psi \;&\xleftarrow{\boldsymbol{\mathcal{N}}}\; \boldsymbol{\mathcal{F}}^{-1}[\widehat{\psi}], \quad \psi \leftarrow e^{-i\Delta t|\psi|^2}\psi,
\end{aligned}$$

repeat $\boldsymbol{\mathcal{K}}/2$ and $\boldsymbol{\mathcal{V}}/2$.

[8] Square-integrable function space.

The L^2 norm is renormalized after every macro step to compensate numerical drift. For DW Hamiltonian, non-linearity time evolution is omitted.

Integrator parameters. We use an inner step $\delta t = 10^{-6}$. A *macro* step of $\Delta T = 100\,\delta t = 10^{-4}$ is recorded and the sequence is propagated for $T_{\max} = 100\,\Delta T = 10^{-2}$.

Initial–state families (real wave-functions). Each sample draws a real profile $f(x)$ from $\{A_1, A_2, A_3\}$ below, multiplies it by the envelope A and normalizes it in L^2:

$$\psi_0(x) = \frac{A(x)\, f(x)}{\|Af\|_2}.$$

A1. random harmonic series

$$f(x) = \sum_{m=0}^{5} \tfrac{1}{2}\,\xi_m^{(c)} \cos(2\pi m\xi) + \tfrac{1}{2}\,\xi_m^{(s)} \sin(2\pi m\xi), \quad \xi = \frac{x+2}{4},\ \xi_m^{(\cdot)} \sim \mathcal{N}(0,1).$$

A2. Gaussian wave packet $f(x) = \exp[-(x-x_0)^2/(2\sigma^2)]$ with $x_0 \sim U[-0.4\pi, 0.4\pi]$ and $\sigma \sim U[0.1, 0.3]$.

A3. Gaussian–Hermite mode $f(x) = H_n\big(\frac{x-x_0}{\sigma}\big) \exp[-(x-x_0)^2/(2\sigma^2)]$ with $n \in \{0,1,2\}$ uniform, x_0, σ as above, and H_n the Hermite polynomial.

Stored quantities. For every sample index s and every snapshot $T \in \{0, 1, \ldots, 100\}$ we save

$$\big(\mathbf{x},\ \boldsymbol{\psi}(x,T),\ |\boldsymbol{\psi}(x,T)|^2,\ |\widehat{\boldsymbol{\psi}}(\xi,T)|^2\big) \quad \longrightarrow \quad \texttt{wavefunc},\ \texttt{pos\_pdf},\ \texttt{mom\_pdf}.$$

All arrays are written in `float32` except the complex wave-function, stored as `complex64`. Altogether one call to `generate_dataset(num_samples=200)` produces $200 \times 101 \times 4 = 80{,}800$ labeled records.

B.2.3 Train Type Details

Let $\boldsymbol{\psi}_\Theta^{(T)}$ be the T-step prediction of Q-KANO given $\boldsymbol{\psi}_0$.

Train Type	Train Dataset	Loss function
full	complex $\boldsymbol{\psi}^{(T)}$	$\mathcal{L}_{full} = \dfrac{\|\boldsymbol{\psi}_\Theta^{(T)} - \boldsymbol{\psi}^{(T)}\|_2}{\|\boldsymbol{\psi}^{(T)}\|_2}$
pos	$\mathbf{p}_x^{(T)}$	$\mathcal{L}_{pos} = D_{\mathrm{KL}}\big(\mathbf{p}_x^{(T)} \parallel \|\boldsymbol{\psi}_\Theta^{(T)}|^2\big)$
pos & mom	$\mathbf{p}_x^{(T)}$, $\mathbf{p}_\xi^{(T)}$	$\mathcal{L}_{pos\&mom} = D_{\mathrm{KL}}\big(\mathbf{p}_x^{(T)} \parallel \|\boldsymbol{\psi}_\Theta^{(T)}|^2\big) + D_{\mathrm{KL}}\big(\mathbf{p}_\xi^{(T)} \parallel \|\widehat{\boldsymbol{\psi}}_\Theta^{(T)}|^2\big)$

C Proofs

C.1 Notation and preliminaries

Throughout, $\mathbb{T}^d := [-\pi, \pi]^d$ denotes the flat d–torus and $\mathbb{Z}^d$ the lattice of Fourier indices. For $\boldsymbol{\xi} \in \mathbb{Z}^d$ let $e_{\boldsymbol{\xi}}(\mathbf{x}) := e^{i\boldsymbol{\xi}\cdot\mathbf{x}}$. The Fourier coefficient of a square integrable function $\mathbf{f}$ is

$$\widehat{\mathbf{f}}(\boldsymbol{\xi}) \ := \ (2\pi)^{-d} \int_{\mathbb{T}^d} \mathbf{f}(\mathbf{x})\, e^{-i\boldsymbol{\xi}\cdot\mathbf{x}}\, \mathrm{d}\mathbf{x}.$$

In a Sobolev space H^s with an order of smoothness $s \in \mathbb{R}$, the Sobolev norm of function $\mathbf{f}$ is

$$\|\mathbf{f}\|_{H^s}^2 \ := \frac{(2\pi)^d}{2} \sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} (1 + |\boldsymbol{\xi}|^{2s})\, |\widehat{\mathbf{f}}(\boldsymbol{\xi})|^2.$$

Asymptotics. Write $A \lesssim B$ if $A \le C\,B$ for a constant C depending only on fixed parameters (dimension, regularity exponents, etc.).

Vector Notation Fix a spatial dimension $d \geq 1$ and an index $j \in \{1, \dots, d\}$. For $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d) \in \mathbb{Z}^d$ we write

$$\boldsymbol{\xi}_{-j} := (\xi_1, \dots, \xi_{j-1}, \xi_{j+1}, \dots, \xi_d) \in \mathbb{Z}^{d-1}$$

for the vector obtained by *removing* the j-th coordinate of $\boldsymbol{\xi}$. Conversely, for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{d-1}) \in \mathbb{Z}^{d-1}$ and $n \in \mathbb{Z}$ we define the *insertion* map

$$(\boldsymbol{\alpha}, n)_j := (\alpha_1, \dots, \alpha_{j-1},\, n,\, \alpha_j, \dots, \alpha_{d-1}) \in \mathbb{Z}^d.$$

We use $|\cdot|_\infty$ for the max–norm on $\mathbb{Z}^d$, i.e. $|\boldsymbol{\xi}|_\infty = \max_{1 \leq i \leq d} |\xi_i|$. When we write $\widehat{\mathbf{u}}(\boldsymbol{\alpha}, r)$, this is shorthand for the d–dimensional coefficient $\widehat{\mathbf{u}}\big((\boldsymbol{\alpha}, r)_j\big)$.

C.2 Proof of Lemma 1

Restatement of Lemma 1 (with explicit notation). Let $\mathbf{u} \in H^s(\mathbb{T}^d)$ with $s > 0$, and assume that its Fourier coefficients are compactly supported:

$$\widehat{\mathbf{u}}(\boldsymbol{\xi}) = \mathbf{0} \quad \text{for all } |\boldsymbol{\xi}|_\infty > N_0, \qquad \text{and} \qquad \widehat{\mathbf{u}} \not\equiv \mathbf{0}.$$

Fix $j \in \{1, \dots, d\}$ and set $\mathbf{v}(\mathbf{x}) := x_j\, \mathbf{u}(\mathbf{x})$. Then there exist

$$\boldsymbol{\alpha} \in \mathbb{Z}^{d-1}, \qquad m \in \{1, 2, \dots, 2N_0 + 1\}, \qquad c > 0, \quad R \in \mathbb{N},$$

and an infinite set of the fiber at j^{th} coordinate,

$$\Xi_{\boldsymbol{\alpha}, R} \subset \big\{\boldsymbol{\xi} \in \mathbb{Z}^d : \ \boldsymbol{\xi}_{-j} = \boldsymbol{\alpha},\ |\xi_j| \geq R\big\}$$

such that

$$\big|\widehat{\mathbf{v}}(\boldsymbol{\xi})\big| \ \geq \ \frac{c}{\big(1 + |\xi_j|\big)^m} \qquad \forall\, \boldsymbol{\xi} \in \Xi_{\boldsymbol{\alpha}, R}. \tag{20}$$

In particular, if for some $\boldsymbol{\alpha} \in \mathbb{Z}^{d-1}$ one has

$$\sum_{r=-N_0}^{N_0} \widehat{\mathbf{u}}\big((\boldsymbol{\alpha}, r)_j\big) \neq \mathbf{0} \quad \text{(equivalently, } \sum_{r=-N_0}^{N_0} \widehat{\mathbf{u}}(\boldsymbol{\alpha}, r) \neq \mathbf{0} \text{ in the shorthand above)},$$

then the bound equation 20 holds with the sharper exponent $m = 1$.

Proof. We work with square-integrable function $\mathbf{v} \in L^2(\mathbb{T}^d)$ where $\mathbb{T}^d = [0, 2\pi]^d$.

Step 1 — The periodic "coordinate" and its Fourier coefficients. The function $\mathbf{x} \mapsto x_j$ is not periodic. Introduce the zero-mean, periodic 1D sawtooth

$$\psi_j(\mathbf{x}) \ := \ x_j - \pi, \qquad \mathbf{x} \in [0, 2\pi]^d,$$

extended periodically to $\mathbb{T}^d$. A direct computation (factorization of the integral and one-dimensional integration by parts) shows that its Fourier coefficients are supported on the j^{th}-coordinate: for $\boldsymbol{\xi} \in \mathbb{Z}^d$,

$$\widehat{\psi_j}(\boldsymbol{\xi}) = \begin{cases} -\dfrac{1}{i\,\xi_j}, & \text{if } \boldsymbol{\xi}_{-j} = \mathbf{0} \text{ and } \xi_j \neq 0, \\ 0, & \text{if } \boldsymbol{\xi} = \mathbf{0} \text{ or } \boldsymbol{\xi}_{-j} \neq \mathbf{0}. \end{cases} \tag{21}$$

Moreover $x_j = \psi_j + \pi$, hence

$$\widehat{\mathbf{v}}(\boldsymbol{\xi}) \ = \ \boldsymbol{\mathcal{F}}[\psi_j \mathbf{u}](\boldsymbol{\xi}) + \pi\, \widehat{\mathbf{u}}(\boldsymbol{\xi}). \tag{22}$$

Step 2 — Exact coefficient formula outside the support of $\widehat{\mathbf{u}}$. Since $\widehat{\mathbf{u}}(k) = \mathbf{0}$ for $|\boldsymbol{\xi}|_\infty > N_0$, the second term in equation 22 vanishes whenever $|\boldsymbol{\xi}|_\infty > N_0$. Using equation 21 and the convolution theorem, we obtain for any $\boldsymbol{\xi} \in \mathbb{Z}^d$ with $|\boldsymbol{\xi}|_\infty > N_0$:

$$\widehat{\mathbf{v}}(\boldsymbol{\xi}) = \sum_{\boldsymbol{\ell} \in \mathbb{Z}^d} \widehat{\psi_j}(\boldsymbol{\xi} - \boldsymbol{\ell})\, \widehat{\mathbf{u}}(\boldsymbol{\ell}) = -\frac{1}{i} \sum_{\substack{\boldsymbol{\ell}_{-j} = \boldsymbol{\xi}_{-j} \\ |\ell_j| \leq N_0}} \frac{\widehat{\mathbf{u}}(\boldsymbol{\ell})}{\xi_j - \ell_j}. \tag{23}$$

Thus, along any fixed transverse index $\boldsymbol{\alpha} := \boldsymbol{\xi}_{-j} \in \mathbb{Z}^{d-1}$, the tail $\widehat{\mathbf{v}}(\boldsymbol{\alpha}, n)$ for large $|n|$ is a finite sum of simple fractions in the single variable n.

Step 3 — Moments on a nontrivial fiber and the first non-vanishing moment. Because $\widehat{\mathbf{u}} \not\equiv \mathbf{0}$, there exists at least one $\boldsymbol{\alpha} \in \mathbb{Z}^{d-1}$ for which the fiber

$$\mathcal{F}_{\boldsymbol{\alpha}} := \{r \in \mathbb{Z} : \ \widehat{\mathbf{u}}(\boldsymbol{\alpha}, r) \neq \mathbf{0}\}$$

is nonempty. Define the (finite) coefficients $\mathbf{c}_r := \widehat{\mathbf{u}}(\boldsymbol{\alpha}, r)$ for $r \in [-N_0, N_0]$, and their moments

$$\boldsymbol{\mu}_p := \sum_{r=-N_0}^{N_0} r^p \mathbf{c}_r \qquad (p \geq 0).$$

Let $m \in \{1, \ldots, 2N_0 + 1\}$ be the smallest index for which $\mu_{m-1} \neq \mathbf{0}$. Such an m exists since not all $\mathbf{c}_r$ vanish.

Step 4 — Asymptotics along a line and a polynomial lower bound. For $n \in \mathbb{Z}$ with $|n| > N_0$, formula equation 23 along the line $\boldsymbol{\xi}_{-j} = \boldsymbol{\alpha}$ reads

$$\widehat{\mathbf{v}}(\boldsymbol{\alpha}, n) = -\frac{1}{i} \sum_{r=-N_0}^{N_0} \frac{\mathbf{c}_r}{n - r}.$$

Expanding $\frac{1}{n-r} = \frac{1}{n} \sum_{q \geq 0} \left(\frac{r}{n}\right)^q$ for $|n| > 2N_0$ and collecting terms yields the asymptotic expansion

$$\widehat{\mathbf{v}}(\boldsymbol{\alpha}, n) = -\frac{1}{i} \left(\frac{\boldsymbol{\mu}_0}{n} + \frac{\boldsymbol{\mu}_1}{n^2} + \cdots + \frac{\boldsymbol{\mu}_{m-1}}{n^m} + O\left(\frac{1}{|n|^{m+1}}\right)\right), \qquad |n| \to \infty.$$

By the choice of m, the first nonzero term is $\boldsymbol{\mu}_{m-1}/n^m$. Consequently, there exist $R \in \mathbb{N}$ and $c > 0$ such that

$$\left|\widehat{\mathbf{v}}(\boldsymbol{\alpha}, n)\right| \ \geq \ \frac{c}{|n|^m} \qquad \text{for all } |n| \geq R, \ n \in \mathbb{Z}.$$

Step 5 — Conclusion and the special case $m = 1$. Then equation 20 holds for all $\boldsymbol{\xi} \in \Xi_{\boldsymbol{\alpha}, R}$ with the exponent m determined in Step 3, by the definition of $\Xi_{\boldsymbol{\alpha}, R}$ and that $|n| \geq R$. If $\boldsymbol{\mu}_0 = \sum_r \mathbf{c}_r \neq \mathbf{0}$ for the chosen fiber (equivalently, $\sum_{r=-N_0}^{N_0} \widehat{\mathbf{u}}(\boldsymbol{\alpha}, r) \neq \mathbf{0}$), then $m = 1$ and we obtain the sharper $|\widehat{\mathbf{v}}(\boldsymbol{\xi})| \gtrsim (1 + |\xi_j|)^{-1}$ along the line $\boldsymbol{\xi}_{-j} = \boldsymbol{\alpha}$. □

Remark 1. *The explicit one-dimensional tail equation 23 shows that multiplying a band-limited field by the coordinate x_j produces a* polynomial *Fourier tail decay along lines parallel to the j^{th} coordinate, with rate $(1 + |\xi_j|)^{-m}$ where m is the first non-vanishing moment of the finitely many coefficients on the relevant fiber. In particular, when $m = 1$, the decay is exactly $(1 + |\xi_j|)^{-1}$. Such algebraic tails are consistent with the pseudo-spectral projection error estimate quoted in Kovachki et al. (2021, Thm. 40).*

C.3 Proof of Theorem 1

Restate of Theorem 1. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ with total degree $M := |\boldsymbol{\alpha}| \geq 1$ and define the position-multiplier

$$\boldsymbol{\mathcal{M}}(\mathbf{x}) := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}.$$

For inputs band-limited to radius N_0 and lying in $H^s(\mathbb{T}^d)$ with $s > \frac{d}{2}$, any Fourier Neural Operator $\boldsymbol{\mathcal{G}}_\theta^{\text{FNO}}$ that achieves $\|\boldsymbol{\mathcal{M}}(\mathbf{x}) - \boldsymbol{\mathcal{G}}_\theta^{\text{FNO}}\|_{H^s \to H^{s-\delta}} \leq \varepsilon$ $(0 < \delta < 1)$ must employ a spectral bandwidth (FNO width) $m \gtrsim \varepsilon^{-M/s}$ and a parameter count $|\theta| \ \geq \ \exp\big(c\, \varepsilon^{-Md/s}\big)$ for some $c > 0$ depending only on (d, s, δ, N_0).

Proof. Set $s' := s - \delta$ with $0 < \delta < 1$.

Step 1 – Algebraic tail produced by $\boldsymbol{\mathcal{M}}(\mathbf{x})$. Applying Lemma 1 once per factor of x_j shows that for some constant $C_0 > 0$ and an infinite set $\Xi_\infty \subset \mathbb{Z}^d$,

$$|\boldsymbol{\mathcal{F}}[\boldsymbol{\mathcal{M}}(\mathbf{x})\mathbf{u}](\boldsymbol{\xi})| \ \geq \ \frac{C_0}{(1 + |\boldsymbol{\xi}|)^{M+1}} \qquad \forall\, \boldsymbol{\xi} \in \Xi_\infty. \tag{24}$$

Step 2 – Pseudo-spectral projection lower bound. For any $\mathbf{f} \in H^s$, pseudo-spectral projection error estimate gives (Gottlieb & Orszag, 1977)

$$\|(\boldsymbol{\mathcal{I}} - \boldsymbol{\Pi}_{\text{FNO}})\mathbf{f}\|_{H^{s'}} \;\geq\; C_1\Big[\sum_{|\boldsymbol{\xi}|_\infty > N} (1 + |\boldsymbol{\xi}|^{2s'})\,|\widehat{\mathbf{f}}(\boldsymbol{\xi})|^2\Big]^{1/2},$$

where $\boldsymbol{\mathcal{I}}$ is an identity operator. Insert $\mathbf{f} = \boldsymbol{\mathcal{M}}(\mathbf{x})\mathbf{u}$ and the tail bound equation 24; summing over $\Xi_\infty \cap \{|\boldsymbol{\xi}|_\infty > m\}$ yields

$$\|(\boldsymbol{\mathcal{I}} - \boldsymbol{\Pi}_{\text{FNO}})\boldsymbol{\mathcal{M}}(\mathbf{x})\mathbf{u}\|_{H^{s'}} \;\gtrsim\; m^{-(M-\delta)}.$$

Imposing this residual $\leq \frac{1}{2}\varepsilon$ forces

$$m \;\geq\; C_2\,\varepsilon^{-M/s}, \qquad C_2 = C_2(d, s, \delta, M, N_0) > 0. \tag{25}$$

Step 3 – Canonical neural scaling in the latent map. An FNO with spectral radius (width) m manipulates a latent vector of dimension $(2m+1)^d \sim m^d$. Approximating a generic Lipschitz map $\mathbf{G} : \mathbb{C}^{m^d} \to \mathbb{C}^{m^d}$ to accuracy $\varepsilon/2$ with a fully connected network requires (Yarotsky, 2017; De Ryck et al., 2021) *at least*

$$\text{parameters} \;\gtrsim\; \varepsilon^{-m^d}.$$

(Kovachki et al., 2021, Remark 22)

Step 4 – Substitute the bandwidth lower bound. Using equation 25, $m^d \sim \varepsilon^{-Md/s}$. Hence the latent network must have at least

$$|\theta| \;\gtrsim\; \varepsilon^{-\varepsilon^{-Md/s}},$$

a *super-exponential* curse of dimensionality in the target accuracy ε.

□

C.4 Proof of Theorem 2

Restate of Theorem 2. Let $s \geq s' \geq 0$ and $s_p > d/2$. Consider any finite composition

$$\boldsymbol{\mathcal{G}} \;=\; (\varsigma_\ell \circ \boldsymbol{\mathcal{L}}_\ell) \circ \cdots \circ (\varsigma_1 \circ \boldsymbol{\mathcal{L}}_1), \qquad \boldsymbol{\mathcal{L}}_i = \mathbf{Op}_m(\boldsymbol{p}_i),$$

where each Kohn–Nirenberg symbol $\boldsymbol{p}_i$ belongs to $W^{s_p,2}\big(\mathbb{T}^d_{\mathbf{x}} \times \mathbb{T}^d_{\boldsymbol{\xi}}\big)$[9], and each ς_i is a point-wise nonlinearity of uniformly bounded degree (so the number of such nonlinearities is $O(\ell)$ and independent of ε). For every $\varepsilon > 0$ there exists a *single-layer* KANO $\boldsymbol{\mathcal{G}}^{\text{KANO}}_\theta$ such that

$$\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{G}}^{\text{KANO}}_\theta\|_{H^s \to H^{s'}} \leq \varepsilon, \qquad |\theta| \;=\; O\big(\varepsilon^{-d/(2s_p)}\big).$$

Proof. **Step 1 – Kohn–Nirenberg quantization error.** By the quadrature bound from Demanet & Ying (2011), for each $\boldsymbol{\mathcal{L}}_i$ we have

$$\big\|\boldsymbol{\mathcal{L}}_i - \mathbf{Op}_m(\boldsymbol{p}_i)\big\|_{H^s \to H^{s'}} \;\leq\; C'\,B\,m^{-(s-s')},$$

where $\mathbf{Op}_m(\boldsymbol{p}_i)$ keeps only frequencies $|\boldsymbol{\xi}|_\infty \leq m$. Since each ς_i is a bounded-degree pointwise map, its Nemytskii operator is Lipschitz on bounded sets; write $L_i := \text{Lip}(\varsigma_i)$[10] on the relevant range and set $L_{\max} := \max_i L_i$. Let $M := \max_j\big\{\|\boldsymbol{\mathcal{L}}_j\|_{H^{s'} \to H^{s'}},\, \|\mathbf{Op}_m(\boldsymbol{p}_j)\|_{H^{s'} \to H^{s'}}\big\}$. Because $s_p > \frac{d}{2}$ implies $\|\boldsymbol{p}_j\|_{L^\infty} \lesssim \|\boldsymbol{p}_j\|_{W^{s_p,2}}$, both $\boldsymbol{\mathcal{L}}_j$ and $\mathbf{Op}_m(\boldsymbol{p}_j)$ are bounded on $H^{s'}$ with a bound independent of m. A telescoping estimate for the interleaved composition then yields

$$\big\|\boldsymbol{\mathcal{G}} - \boldsymbol{\Pi}_{\text{KANO}}(\boldsymbol{\mathcal{G}})\big\|_{H^s \to H^{s'}} \;\leq\; C_*\,C'\,B\,m^{-(s-s')}, \qquad C_* \;\leq\; \ell\,(L_{\max}M)^{\ell-1}L_{\max}.$$

[9] Standard square integrable periodic Sobolev space on the product torus $\mathbb{T}^{2d}$.

[10] Lipschitz constant of the point-wise nonlinearity ς_i on the relevant value range, i.e. $L_i := \sup_{a \neq b} \frac{|\varsigma_i(a) - \varsigma_i(b)|}{|a-b|}$ with a, b restricted to the compact interval attained by the i-th preactivations.

Pick $m := \big(2\,C_*\,C'B/\varepsilon\big)^{1/(s-s')}$. Then the full composition $\boldsymbol{\mathcal{G}}$ deviates from its projected counterpart

$$\boldsymbol{\Pi}_{\text{KANO}}(\boldsymbol{\mathcal{G}}) := (\varsigma_\ell \circ \mathbf{Op}_m(\boldsymbol{p}_\ell)) \circ \cdots \circ (\varsigma_1 \circ \mathbf{Op}_m(\boldsymbol{p}_1))$$

by at most $\varepsilon/2$ in $H^s \to H^{s'}$ operator norm.

Step 2 – KAN approximation of both symbols and pointwise nonlinearities. By the width-fixed KAN scaling law from Wang et al. (2024), for any $\eta > 0$ there exists a Kolmogorov–Arnold Network $\boldsymbol{\Phi}_{i,\eta}$ with $|\boldsymbol{\Phi}_{i,\eta}| = O(\eta^{-d/(2s_p)})$ such that $\|\boldsymbol{p}_i - \boldsymbol{\Phi}_{i,\eta}\|_{L^\infty} \le \eta$ on $\mathbb{T}^d_{\mathbf{x}} \times \mathbb{T}^d_{\boldsymbol{\xi}}$. Likewise, since each ς_i is a fixed-degree pointwise map, its action over the compact value range visited by the projected flow can be uniformly approximated by a width-fixed KAN $\boldsymbol{\Psi}_{i,\eta}$ with size $|\boldsymbol{\Psi}_{i,\eta}| = O\big(\eta^{-1/(2s_p)}\big)$ and $\|\varsigma_i - \boldsymbol{\Psi}_{i,\eta}\|_{L^\infty} \le \eta$. Choose

$$\eta_{\text{sym}} := \frac{\varepsilon}{4\ell}, \qquad \eta_{\text{nl}} := \frac{\varepsilon}{4\ell}.$$

Define the single-layer KANO

$$\boldsymbol{\mathcal{G}}_\theta^{\text{KANO}} := (\boldsymbol{\Psi}_{\ell,\eta_{\text{nl}}} \circ \mathbf{Op}_m(\boldsymbol{\Phi}_{\ell,\eta_{\text{sym}}})) \circ \cdots \circ (\boldsymbol{\Psi}_{1,\eta_{\text{nl}}} \circ \mathbf{Op}_m(\boldsymbol{\Phi}_{1,\eta_{\text{sym}}})),$$

where the same spectral radius m from Step 1 is used in every $\mathbf{Op}_m(\,\cdot\,)$.

Step 3 – Error accumulation beyond projection. Using linearity of the symbol-to-operator map and stability of Nemytskii (pointwise) maps under uniform approximation, the post-projection error splits into a sum of the symbol parts and the nonlinearity parts:

$$\begin{aligned}\big\|\boldsymbol{\Pi}_{\text{KANO}}(\boldsymbol{\mathcal{G}}) - \boldsymbol{\mathcal{G}}_\theta^{\text{KANO}}\big\|_{H^s \to H^{s'}} &\le \sum_{i=1}^{\ell} \Big\|\mathbf{Op}_m(\boldsymbol{p}_i) - \mathbf{Op}_m(\boldsymbol{\Phi}_{i,\eta_{\text{sym}}})\Big\|_{H^s \to H^{s'}} \;+\; \sum_{i=1}^{\ell} \|\varsigma_i - \boldsymbol{\Psi}_{i,\eta_{\text{nl}}}\|_{L^\infty} \\ &\le \ell\,\eta_{\text{sym}} + \ell\,\eta_{\text{nl}} \;=\; \varepsilon/2.\end{aligned}$$

Combining with Step 1 yields $\|\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{G}}_\theta^{\text{KANO}}\|_{H^s \to H^{s'}} \le \varepsilon$.

Step 4 – Parameter complexity. Summing the sizes of all KAN blocks gives

$$|\theta| \;=\; \sum_{i=1}^{\ell} O\big(\eta_{\text{sym}}^{-d/(2s_p)}\big) \;+\; \sum_{i=1}^{\ell} O\big(\eta_{\text{nl}}^{-1/(2s_p)}\big) \;+\; O(1) \;=\; O\big((\varepsilon/\ell)^{-d/(2s_p)}\big) + O\big((\varepsilon/\ell)^{-1/(2s_p)}\big).$$

Since ℓ and the number/degree of the ς_i are fixed (do not scale with ε), the dominating term is $O(\varepsilon^{-d/(2s_p)})$, establishing the claimed complexity. □

D USE OF LARGE LANGUAGE MODEL

Large Language Model (LLM) is used to polish the writing in this paper, such as detecting grammar errors and typos. LLM is also used to search for appropriate references for correct citations, and all the proposed references are fully inspected and verified before citing in the paper.