CONSTRUCTING MACHINE-PRECISION NEURAL NETWORKS WITH QUASI-INTERPOLANTS

Catherine Deng*
Department of Electrical Engineering
Stanford University

Junmiao Hu* & Jerry Liu
Institute for Computational & Mathematical Engineering
Stanford University

Milan Rohatgi & Christopher Re
Department of Computer Science
Stanford University

ABSTRACT

Neural networks often plateau far above machine precision, limiting their use in scientific computing pipelines. A central question is whether this reflects an expressivity limit or a failure of optimization. In the interpolation setting, we show that optimization is the primary bottleneck by constructing the first explicit MLP interpolant that provably achieves machine-precision accuracy with $\log(1/\varepsilon)$ parameter scaling while remaining implementable in floating-point arithmetic. Our construction, based on quasi-interpolation theory, exposes a dimensionless bandwidth parameter λ that controls the tradeoff between approximation error and numerical stability. Comparing this construction to trained MLPs, we find that optimization drives $\lambda \to 0$, causing the network to collapse to an overly narrow length-scale regime and utilize capacity redundantly, even though the quasi-interpolant itself remains reasonably conditioned. These results provide a principled lens on precision failures in scientific machine learning.

1 INTRODUCTION

Machine learning has the potential to dramatically accelerate scientific workflows by replacing expensive physics-based subroutines with learned surrogates, from neural operators for weather modeling to surrogates for aerodynamics simulations (Pathak et al., 2022; Li et al., 2025; Alkin et al., 2025; Mao et al., 2024). However, numerical precision remains a central barrier to deployment: many scientific pipelines require tight residual tolerances and stable long-horizon rollouts, yet current methods struggle to reliably reach the required fidelity (McGreivy & Hakim, 2024). In fact, neural networks already struggle to reach machine precision in the simple controlled setting of noise-free interpolation of smooth functions from exact samples (Michaud et al., 2023a; Liu et al., 2025a). Even for analytic 1D targets like $f(x) = \sin(2\pi x)$, standard training of single-layer-hidden multilayer perceptrons (MLPs) stalls 6-10 *orders of magnitude* above `fp64` limits ($\sim 10^{-16}$) for (L_2) relative error, plateauing around $L_2 = 10^{-6}$ to 10^{-10} (Figure 1b).

Surprisingly, increasing model size or training time does not reliably yield additional digits of accuracy (Wang et al., 2024; Liu et al., 2025b). Current techniques do not characterize how precision can successfully *scale* with increased model capacity or compute. This stands in contrast to numerical analysis, where decades of work provide provable convergence guarantees (Trefethen & Bau, 2022; Boyd, 1989), and modern language models, where empirical scaling laws predictively relate performance to model size (Hoffmann et al., 2022; Kumar et al., 2024). Understanding whether precision can scale predictably with capacity is therefore a foundational question to unlock broader adoption of machine learning across scientific domains.

*Equal contribution. Corresponding authors: `catherinedeng@stanford.edu`, `junmiaoh@stanford.edu`

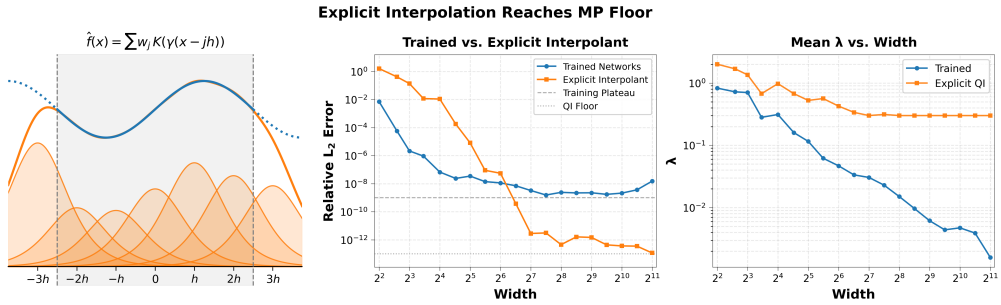

Figure 1: **Left:** Quasi-interpolants represent f as a sum of localized, translated shift-invariant kernels, with additional "halo" nodes outside the interpolation interval $\Omega = [-1, 1]$ (shaded) to reduce boundary error. **Middle:** Relative L_2 error versus width for interpolating $\sin(2\pi x)$: the explicit QI interpolant displays geometric scaling towards the `fp64` floor, while end-to-end trained MLPs plateau three orders of magnitude higher. **Right:** Dimensionless bandwidth λ versus width for interpolating $\sin(2\pi x)$: the QI construction reaches a λ plateau, whereas trained networks drive $\lambda \to 0$, indicating a length-scale mismatch.

Motivated by recent diagnostic work (Michaud et al., 2023a; Wang & Lai, 2023), we study one-hidden-layer MLPs in the noise-free 1D interpolation regime to understand why precision fails to scale with model size and compute. We decompose this problem into two questions:

1. **Expressivity**. What machine-precision representations are numerically realizable as MLPs? Universal approximation theorems (Hornik et al., 1989) guarantee that high-precision interpolating MLPs exist, but classical results either provide vacuous parameter bounds (achieving ε error requires $\text{poly}(1/\varepsilon)$ parameters) or require constructions with exponentially large weight magnitudes (Mhaskar, 1993). We lack explicit representations that are simultaneously (i) machine-precise, (ii) parameter-efficient ($\log(1/\varepsilon)$ scaling), and (iii) numerically realizable in finite-precision arithmetic.

2. **Optimization**. Why does training fail to discover high-precision solutions? Even when numerically stable representations exist in principle, training often fails to find them. We lack understanding of *why* trained networks converge to sub-optimal solutions, and *how* attainable precision depends on the interplay between width, optimizer choice, and training dynamics.

We address both gaps by constructing explicit MLP interpolants using tools from quasi-interpolation theory (Trefethen, 2019; Buhmann & Jäger, 2022). Our contributions are:

- **We provide an explicit construction for high-precision MLP interpolation.** Ours is the first MLP construction that provably achieves machine-precision interpolation with $\log(1/\varepsilon)$ parameter scaling, matching classical polynomial methods. Unlike previous universal approximation results, our construction remains implementable in floating-point arithmetic, yielding a realizable route to machine precision with provable scaling. Critically, our construction also exposes a *dimensionless bandwidth parameter* λ that balances aliasing error against conditioning. We show that λ eventually plateaus as the parameter count scales, yielding deterministic MLP weight scalings.

- **We study why optimization fails to discover high-precision solutions.** Trained MLPs with standard second-order optimizers (Urbán et al., 2025) achieve geometric convergence at small widths, but plateau as capacity increases – the opposite of desirable scaling behavior. Our quasi-interpolant (QI) construction provides a diagnostic lens: we find that (i) trained MLPs learn sparse, non-uniform representations that violate the conditioning requirements exposed by our construction, and (ii) learned weights fail to follow the scalings predicted by our construction.

These results represent a first step towards a principled framework for understanding and achieving machine-precision interpolation with neural networks.

2 Related Work

High-precision scientific ML Precision ceilings in `fp64` training are increasingly recognized as a bottleneck in scientific ML. Across settings ranging from function regression to operator learning, researchers report plateaus far above floating-point limits that persist despite increased model capacity, compute, or data, motivating a range of responses (Michaud et al., 2023b; Wang & Lai, 2023; Cao et al., 2023). Prior attempts to improve the precision floor broadly span architectural changes, optimizer design, and refinement-based training paradigms:

- **Architecture.** Architectural modifications raise attainable precision by mitigating spectral bias (improving high-frequency fitting) or reshaping training geometry to improve conditioning and capacity allocation. Recent works attempt to address spectral bias via sinusoidal activations with principled initialization (Sitzmann et al., 2020), reweight objectives to focus capacity on high-error regions using residual-adaptive mechanisms (Wang et al., 2024), and expose precision-conditioning tradeoffs through barycentric-weight parameterizations (Liu et al., 2025a).
- **Optimizer.** Second-order and geometry-aware optimizers improve precision by providing better-conditioned update directions on ill-conditioned loss landscapes. Recent works attempt to stabilize Shampoo-style preconditioning by operating in the preconditioner's eigenbasis (Vyas et al., 2025; Gupta et al., 2018), leverage natural-gradient structure for faster convergence (Müller & Zeinhofer, 2023), and apply quasi-Newton refinement to push PINN training toward lower residuals (Urbán et al., 2025; Kiyani et al., 2025).
- **Training paradigms.** Other works propose to decompose or reframe the learning problem into better-behaved alternatives. For example, Wang & Lai (2023) attempt to train successive networks on spectrum-normalized residuals (Wang & Lai, 2023), and Bacho et al. (2025) learn surrogates for Cholesky factors of Gauss–Newton operators to amortize solver structure.

Although these advances show that higher precision is achievable, they do not yet explain (i) how attainable precision should scale with width/compute for a single end-to-end MLP in the interpolation setting; and (ii) the training pathologies that yield diminishing or non-monotone returns even when more accurate solutions exist. We address both gaps using a classical constructive approximation framework, quasi-interpolation, that makes the relevant length scales and conditioning constraints explicit and links them to the optimization dynamics seen in practice.

Universal approximation theorems for MLPs Universal approximation theorems guarantee that one-hidden-layer MLPs can approximate continuous functions arbitrarily well (Cybenko, 1989). Classical constructions require width scaling as $O(1/\varepsilon)$ or worse, which is infeasible in practice – this implies requiring $\gtrsim 10^{16}$ parameters for `fp64` machine precision. For analytic targets, Mhaskar (1993) and De Ryck et al. (2021) establish MLP constructions whose precision converges exponentially with width. However, these constructions involve numerically computing W-th order derivatives for W neurons. We show in Table 1 that these constructions demand weight magnitudes that far exceed the range of `fp64`.

Interpolation and quasi-interpolation Classical polynomial interpolation achieves exponential convergence for analytic targets—a rate that is minimax optimal (Trefethen, 2019)—but becomes ill-conditioned as degree increases. Quasi-interpolation sidesteps this by replacing global polynomials with weighted sums of localized kernels, providing explicit stability guarantees through cardinal function constructions (Buhmann & Jäger, 2022). We show that one-hidden-layer MLPs with smooth activations (e.g., tanh, GELU) can exactly represent such quasi-interpolants, with the activation's derivative determining the underlying kernel.

Interpolation in neural networks A related line of work seeks to improve precision by embedding interpolation structure directly into the model architecture. Kolmogorov-Arnold Networks (KANs) replace scalar weights with learnable spline-parameterized univariate functions, turning each layer into an adaptive functional expansion (Liu et al., 2025c). Interpolating Neural Networks (INNs) go further by making interpolation over learned nodes the core representational mechanism (Park et al., 2024). These approaches demonstrate that incorporating interpolation structure can improve accuracy and representation efficiency. However, they do not explain why a standard MLP fails to discover high-precision solutions, nor what representational constraints are required

for it to do so. Rather than replacing the MLP with an interpolation-native architecture, we identify an explicit quasi-interpolant already realizable by a standard one-hidden-layer MLP, and use it to characterize the scaling laws and geometric conditions that training fails to recover.

3 Constructing High-Precision MLPs with Quasi-Interpolants

In this section, we describe our cardinal quasi-interpolant construction for high-precision MLP interpolants. Our main theoretical result is Theorem 1, an explicit error decomposition with geometric parameter convergence. Crucially, our construction is the first to achieve these rates while remaining implementable in floating-point arithmetic (Table 1). We show that one-hidden-layer MLPs with standard activations can implement our constructions. Our analysis identifies a critical parameter λ, which controls a fundamental tradeoff between aliasing error and conditioning. We show that choosing an optimal λ requires the magnitudes of our construction's weights to grow with the number of neurons, contrary to typical initialization schemes.

We proceed by defining the quasi-interpolant and its finite-interval truncations (Section 3.1), showing its realization as a one-hidden-layer MLP (Section 3.2), and finally deriving an error decomposition that makes the geometric convergence rate explicit (Section 3.3).

3.1 Constructing Fourier-normalized cardinal quasi-interpolants

We begin by constructing our quasi-interpolant operator as a cardinal series on $\mathbb{R}$. We reparameterize the cardinal functions to our desired kernels through a Fourier normalization. A boundary-aware truncation is then performed to derive the final interpolant over the target interval $\Omega = [-1, 1]$.

Notation. Fix $N \in \mathbb{N}$ and define the step size $h = 2/N$, nodes $x_k = -1 + kh$, and index $k \in \mathbb{Z}$. Let $R, K_c \in \mathbb{N}$, and define the extended index range

$$I_{R,K_c} = \{-R - K_c, \dots, N + R + K_c\}. \tag{1}$$

Let $K : \mathbb{R} \to \mathbb{R}$ be a base kernel and define its scaled version

$$K_\gamma(x) := \gamma K(\gamma x), \qquad \gamma > 0, \qquad \lambda := \gamma h. \tag{2}$$

The dimensionless bandwidth parameter λ couples the kernel bandwidth γ with the grid scale h. This dimensionless parameter controls a fundamental tradeoff between aliasing and stability by controlling kernel localization, made precise in Theorem 1.

Cardinal Series. Let f be known through uniform samples on a grid of spacing h. Our cardinal quasi-interpolant on $\mathbb{R}$ has the form:

$$(Q_h f)(x) := \sum_{k \in \mathbb{Z}} f(x_k)\, L_h(x - x_k), \qquad x_k := -1 + kh, \tag{3}$$

where L_h is a cardinal function.

Constructing L_h by Fourier normalization. We denote the Fourier transform of g by $\widehat{g}$. Define the Fourier character:

$$\widehat{C}_h(\omega) := \frac{h}{D_h(\omega)} = \frac{h}{\sum_{m \in \mathbb{Z}} \widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)} = \sum_{j \in \mathbb{Z}} c_j e^{-ijh\omega}, \tag{4}$$

where $(c_j)_{j \in \mathbb{Z}}$ denote the Fourier coefficients and we assume the denominator D_h is non-zero on $\mathbb{R}$. We define the stencil truncated cardinal function:

$$L_h^{(K_c)}(x) := \sum_{|j| \le K_c} c_j\, K_\gamma(x - jh), \tag{5}$$

which approaches L_h as $K_c \to \infty$.

Finite operator on Ω. Consider samples $f(x_k)$ on the extended index set I_{R,K_c}. The halo- and stencil-truncated quasi-interpolant is[1]

$$(Q_{h,R}^{(K_c)} f)(x) \;:=\; \sum_{k=-R}^{N+R} f(x_k)\, L_h^{(K_c)}(x - x_k), \qquad x \in \Omega. \tag{6}$$

Re-indexing yields the equivalent single-kernel sum

$$(Q_{h,R}^{(K_c)} f)(x) = \sum_{m=-R}^{N+R} a[m]\; K_\gamma(x - x_m), \qquad a[m] := \sum_{|j|\le K_c} c_j\, f(x_{m-j}), \tag{7}$$

which defines a kernel network in the quasi-interpolant framework.

3.2 MLPs CAN IMPLEMENT QUASI-INTERPOLANTS

Although MLP activation functions often do not satisfy traditional kernel conditions, such as spatial localization, an order derivative of a smooth activation often does ($\mathrm{sech}^2 = \tanh'$)[2]. Assume there exists a smooth activation $\psi : \mathbb{R} \to \mathbb{R}$ and an order $r \in \mathbb{N}$ such that

$$K(u) = \psi^{(r)}(u). \tag{8}$$

Then, by the chain rule,

$$(Q_{h,R}^{(K_c)} f)(x) = \sum_{m=-R}^{N+R} a[m]\; K_\gamma(x - x_m) = \frac{\mathrm{d}^r}{\mathrm{d}x^r} \left(\sum_{m=-R}^{N+R} \frac{a[m]}{\gamma^{r-1}}\, \psi\big(\gamma(x - x_m)\big) \right). \tag{9}$$

Therefore, defining the one-hidden-layer MLP

$$g_{\mathrm{MLP}}(x) := \sum_{m=-R}^{N+R} w[m]\; \psi\big(\gamma(x - x_m)\big), \qquad w[m] := \frac{a[m]}{\gamma^{r-1}}, \tag{10}$$

we obtain the exact identity

$$g_{\mathrm{MLP}}^{(r)}(x) = (Q_{h,R}^{(K_c)} f)(x). \tag{11}$$

Crucially, this shows that standard one-hidden-layer MLPs can exactly represent the quasi-interpolant construction—the expressivity is not the bottleneck. In particular, applying the QI construction to sampled data of an r-th derivative, then g_{MLP} recovers the target up to an integration polynomial of degree at most $r - 1$:

$$\widetilde{f}(x) = p_{r-1}(x) + g_{\mathrm{MLP}}(x), \tag{12}$$

where p_{r-1} can be fixed by boundary conditions. For $\tanh'(x) = \mathrm{sech}^2(x)$ this polynomial reduces to a simple bias. For GeLU and Swish, the second-order derivatives satisfy our kernel conditions (Appendix **??**), and the polynomial reduces to a linear layer.

3.3 ERROR ANALYSIS

We present a numerical error bound on our quasi-interpolant construction. All kernel regularity and non-degeneracy conditions, together with explicit prefactors, are deferred to Appendix **??**; these conditions hold for the kernels used in our experiments.

Theorem 1 (Numerical error bound)**.** *Let $\lambda = \gamma h$. Assume f has exponential Fourier decay, and assume the Fourier normalizer D_h is uniformly bounded away from zero on a fixed*

[1] The extended index set does not require training samples outside the interpolation interval in MLP training. Although the quasi-interpolant utilizes these samples to stabilize boundary error, the MLPs can determine arbitrary function values on the extended set to correct boundary error.

[2] Appendix **??** describes the full set of assumptions we make on our kernels.

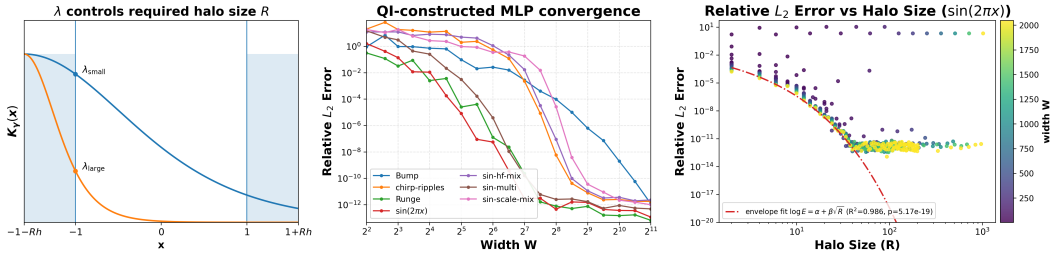

Figure 2: **Left:** λ quantifies the bandwidth of the kernel γ relative to the grid size h. R halo nodes are appended to the boundary of the domain to avoid truncating boundary effects. For wider kernels (smaller λ), wider halo regions are needed for high accuracy. For larger λ, thinner halo regions can sufficiently account for the boundary effects. **Middle:** Convergence plots for QI-constructed MLP against different target functions. Regimes of geometric convergence for the difference target functions (Appendix 8) are clearly visible, before the relative L_2 error plateaus near the conditioning boundary around 10^{-13}. λ is lower bounded by an algebraic blow-up, and we see the scaling depart the rate-theoretic rate. **Right:** Relative L_2 error for the $\sin(2\pi x)$ target across multiple sweeps of (W, R), where the optimal λ is chosen. The error bounds predict a theoretical optimal envelope $\mathcal{O}(e^{-\sqrt{R}})$, and a least squares fit shown through the dashed red line confirms the envelope scaling. The error eventually plateaus around the conditioning floor of the construction, close to numerical precision at larger widths.

*strip (Appendix **??**). Then there exist constants $c_1, c_2, c_3, c_4, C_2 > 0$ and polynomial prefactors $C_1(\lambda), C_3(\lambda), C_4(\lambda) > 0$ such that, for all $x \in \Omega$,*

$$\|f - Q_{h,R}^{(K_c)} f\|_{L^\infty(\Omega)} \le \underbrace{C_1(\lambda) e^{-\frac{c_1}{h}}}_{\text{resolution}} + \underbrace{C_2 e^{-\frac{c_2}{\lambda^p}}}_{\text{aliasing / normalization}} + \underbrace{C_3(\lambda) e^{-c_3 \lambda R}}_{\text{halo truncation}} + \underbrace{C_4(\lambda) e^{-c_4 \lambda K_c}}_{\text{Fourier stencil truncation}} . \tag{13}$$

The exponent p depends on the kernel family: for $K = \operatorname{sech}^2$ one can take $p = 1$.

The full theorem including all the relevant pre-factors can be found in Appendix **??**. Notably, the pre-factors induce a conditioning on the error, as they blow up $\lambda \to 0$. In practice, this then defines a viable operating regime for λ balancing conditioning and convergence.

Equation (13) separates four error terms: the resolution term is controlled by the grid spacing h and the target regularity, the aliasing term represents the aliasing error in our Fourier normalization of the cardinal functions and is controlled by the dimensionless bandwidth λ. Finally, the two truncation terms are due to our projection into a finite sum, and they are controlled by the products λR and λK_c respectively. Hence, λ emerges as the key tradeoff parameter:

Decreasing λ suppresses aliasing error but worsens pre-factor conditioning and truncation errors.

Thus, high accuracy is obtained only in an intermediate regime that balances aliasing, conditioning, and truncation cost. This tradeoff has immediate scaling consequences. By optimizing the error, we discover that optimal λ consistently plateaus as the model width increases (Figure 1c). As such, the bandwidth γ must then scale with MLP width in order to preserve the high-precision regime, increasing as λ plateaus. This is the key structural requirement exposed by the construction.

Numerical validation. We validate our construction numerically in Figure 2b. For optimal λ, the quasi-interpolant achieves `fp64`-level accuracy with predictable scaling in N. This is in contrast to trained MLPs, which fail to maintain the required bandwidth and weight scaling, resulting in a precision plateau above what is achievable by our construction (Section 4).

Comparison to prior constructions. Table 1 compares our construction to prior universal approximation results. Classical constructions (Cybenko, 1989) scale as $\Theta(\varepsilon^{-1})$ in parameters — for `fp64` machine precision, this implies $\gtrsim 10^{16}$ neurons. The construction of Mhaskar (1993) achieves $\widetilde{O}(\log(1/\varepsilon))$ parameters, matching polynomial rates, but computing W-th order finite differences for W neurons yields weights scaling with $\varepsilon^{-1/k}$, which exceeds `fp64` dynamic range at

Method	#params vs. ε	Bit complexity (working precision)	Weight magnitudes
Cybenko (1989)	$\Theta(\varepsilon^{-1})$	$\Theta(\log(1/\varepsilon))$	$\|W\|_{\max} = \Theta(\varepsilon^{-1}\log(1/\varepsilon))$
Mhaskar (1993)	$\tilde{O}(\log(1/\varepsilon)\log\log(1/\varepsilon))$	$\tilde{\Omega}(\log(1/\varepsilon)\log\log(1/\varepsilon))$	$\|W\|_{\max} = \Theta(\varepsilon^{-1/k})$ (order k)
Quasi-interpolant (ours)	$O(\log(1/\varepsilon))$	$O(\log(1/\varepsilon) + \log\log(1/\varepsilon))$	$\|W\|_{\max} = O(\log(1/\varepsilon))$

Table 1: Comparison of parameter count, working-precision bit complexity, and weight magnitudes for prior MLP constructions on analytic targets.

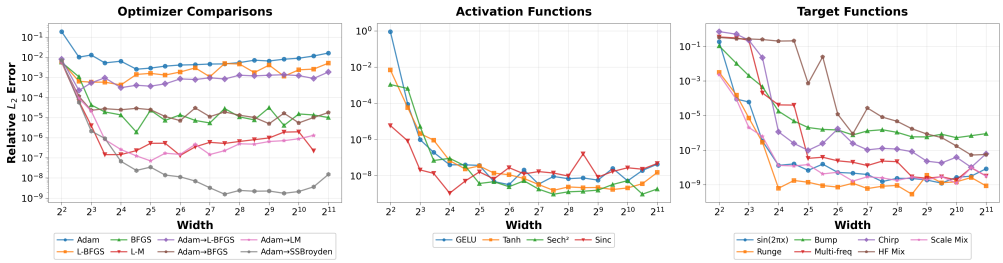

Figure 3: **Direct training plateaus.** Relative ℓ_2 error versus width in `float64`. **Left:** Optimizer comparison on $f(x) = \sin(2\pi x)$ with sech^2 activation ($W = 4$ to 2048). **Middle:** Activations under a fixed two-stage optimizer (Adam→SSBroyden) on the same target. **Right:** Width scaling across a range of target functions trained with $tanh$ activation and 2-stage optimization. Across settings, increasing width yields rapid initial gains followed by saturation well above machine precision.

high precision. In contrast, our construction achieves optimal parameter scaling with weight magnitudes growing only as $O(\log(1/\varepsilon))$. Our construction is also numerically realizable in `fp64` in practice (Figure 2). Refer to Section A for details about these constructions.

4 Experiments

Section 3 shows that machine-precision interpolation is achievable: the QI construction attains near round-off error with one-hidden-layer MLPs once the relevant length scales and normalization are set correctly. Yet in practice, direct training rarely reaches this regime (Figure 3).

We structure the remainder of this section around this gap. We (i) document the plateau phenomenon (Section 4.1), (ii) verify that the QI parameterization is `fp64`-realizable with gradient training and achieves the predicted width–error scaling when the geometry is fixed (Section 4.2), and (iii) analyze how trained networks deviate from the construction (Section 4.3).

Experimental setup. We focus on noiseless 1D interpolation of smooth functions on $[-1, 1]$, training on $n = 256$ uniformly spaced samples and evaluating relative ℓ_2 error on a dense grid of 4096 points. All experiments use `fp64` precision.

4.1 Direct training fails to reach machine precision

Figure 3 reports relative L_2 error versus width for one-hidden-layer MLPs trained on noiseless samples from $f(x) = \sin(2\pi x)$. We quantify the precision ceiling across different optimizers, activations, and target functions. We observe that optimization plateaus across methods: in Figure 3(a), stronger optimizers reduce error by several orders of magnitude compared to Adam, but all tested configurations saturate far above `fp64` machine precision. Notably, widening beyond moderate width yields diminishing or even non-monotone returns. Additionally, plateaus persist across activations and targets: Figures 3(b–c) show the same qualitative behavior under different activations and target functions: geometric convergence at small widths followed by saturation well above machine precision.

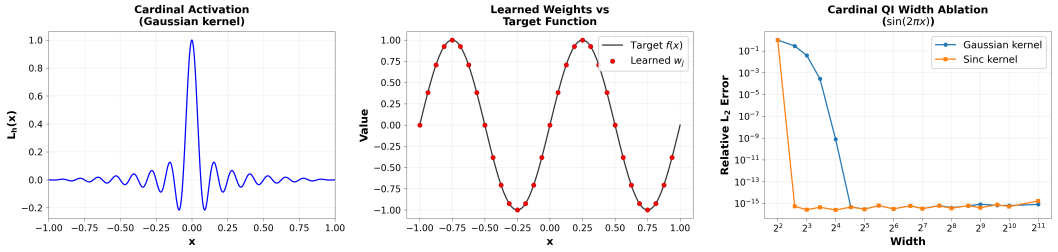

Figure 4: **QI parameterization reaches `fp64` precision.** We fix the QI geometry and fit only the output coefficients in $f(x) \approx \sum_j v_j\, L_h((x - x_j)/h)$. **Left:** Localized cardinal basis function L_h. **Middle:** Fitted coefficients v_j match the samples $f(x_j)$ at grid points. **Right:** Relative ℓ_2 error versus width for Gaussian and sinc kernels, showing geometric convergence to the`fp64` floor.

4.2 MLPs Can Realize Quasi-Interpolants

We verify that the cardinal quasi-interpolants are numerically realizable in `fp64` via gradient descent and exhibit the predicted width-to-error scaling when its geometric assumptions are enforced by altering the activation functions of our MLPs.

We implement the cardinal-kernel parameterization on a uniform grid as described in Equation (3) with a fixed grid and bandwidth γ, so training reduces to estimating only output coefficients $f(x)$. Figure 4 demonstrates that given a localized cardinal activation function $L_h(x)$ (panel a), (i) the learned coefficients γ_j closely match the target samples at grid points, confirming that the coefficient-recovery subproblem is numerically stable when the kernel geometry is specified correctly (panel b) and (ii) the resulting interpolant achieves geometric convergence to the `fp64` floor with increasing width (panel c). Thus, one-hidden-layer MLPs can reach machine precision under our QI construction.

Trained networks place centers outside the domain, qualitatively matching the required halo structure.

4.3 Empirical Diagnosis

Weight magnitude scaling. The QI construction requires width-dependent parameter scaling to keep the dimensionless bandwidth $\lambda = \gamma h$ relatively constant as $h \sim 1/W$ decreases. Figure 5 shows a clear mismatch in magnitude scaling. The explicit QI increases the inverse length-scale γ with width to keep the dimensionless bandwidth $\lambda = \gamma h$ roughly constant as $h \sim 1/W$ decreases. In contrast, end-to-end training keeps $\gamma = \mathcal{O}(1)$ (so $\lambda \to 0$) and uses much larger outer-layer weights than the QI solution across widths. We view this as evidence that training does not recover the QI normalization/geometry, consistent with a less stable (more ill-conditioned) interpolation regime.

This scaling mismatch suggests trained solutions rely more on cancellation among overlapping features, rather than the well-conditioned cardinal regime induced by the QI construction.

Rank saturation and node utilization. MLPs often admit highly compressible subnetworks ((Frankle & Carbin, 2019)). We test whether trained tanh-MLPs exhibit *feature rank saturation* relative to the QI construction.

Consider a MLP and freeze its hidden layer. Let x_i denote the training points and define the feature map $\phi_j(x) = \tanh(\gamma_j(x - x_j))$ for $j = 1, \ldots, N$. The induced linear system for the readout $a \in \mathbb{R}^N$ is

$$\Phi a = b, \qquad \Phi_{ij} = \phi_j(x_i), \qquad b_i = f(x_i) - \overline{f},$$

where $\overline{f}$ denotes the sample mean, so the bias mode is removed. If the learned features become redundant as width grows, then Φ becomes increasingly compressible: a small subset of columns suffices to represent b to high accuracy, and additional neurons do not increase the effective degrees of freedom.

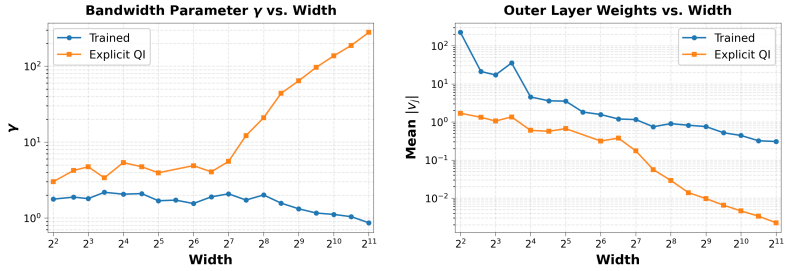

Figure 5: **Weight magnitude scaling** We compare the mean bandwidth γ **(Left)** and mean absolute outer weights $|v_j|$ **(Right)** across network widths, where v_j are the second-layer weights that combine activated features to produce $f(x)$. The blue curve shows trained networks obtained via Adam + SSBroyden optimization, and orange shows explicit quasi-interpolant (QI) constructions computed using theoretical formulas.

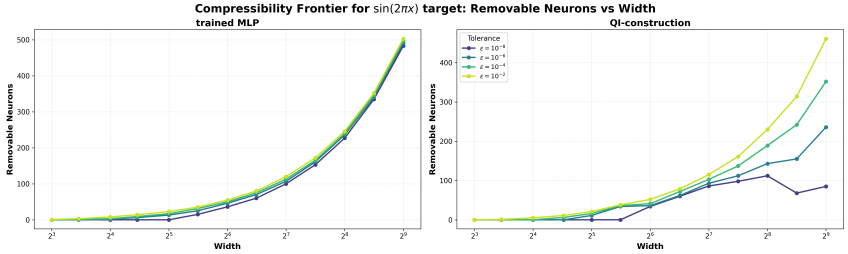

Figure 6: OMP pruning of frozen feature matrices Φ for $f(x) = \sin(2\pi x)$. We report the fraction of removable neurons (columns) while keeping L2RE below a prescribed threshold, as a function of width. **Left:** trained MLP. **Right:** QI-constructed MLP. Trained networks are substantially more compressible, indicating redundant features and rank saturation, whereas the QI-constructed networks distribute approximation across neurons and are less compressible.

Figure 6 quantifies this effect using Orthogonal Matching Pursuit (OMP) (Pati et al., 1993). For each width N we greedily select columns of Φ that minimize the residual, refit a by least squares on the selected columns, and record how many neurons can be removed while maintaining a target relative L_2 error (L2RE) on $f(x) = \sin(2\pi x)$. The trained networks are markedly more compressible: a small subset of neurons attains most of the accuracy as the pruning curves are tightly clustered across error tolerances. In contrast, the QI-constructed networks are less compressible: accuracy degrades sooner under pruning, consistent with a more uniform allocation of approximation power across neurons and reduced redundancy at fixed width. This gap is one manifestation of rank saturation in trained feature maps, and displays how the trained MLPs fall short of the approximation power of our QI construction by learning overly compressible representations.

Endpoint Hessian curvature alone does not explain the gap. A natural hypothesis is that the trained networks prefer compressible features because those representations yield an easier optimization problem. As a first diagnostic, we compare the largest eigenvalue of the empirical-loss Hessian at convergence across widths and target functions. Figure 7 shows that the QI-constructed networks do not exhibit systematically larger top Hessian curvature than the trained networks at comparable widths. This observation does not rule out more subtle conditioning effects, but it suggests that the performance gap is not explained by endpoint curvature alone. Combined with the observed bandwidth-scaling mismatch and the greater compressibility of the trained MLPs, our current evidence is more consistent with a representational mismatch in how width is utilized than with a simple global-conditioning explanation. Further experiments are required to untangle this hypothesis.

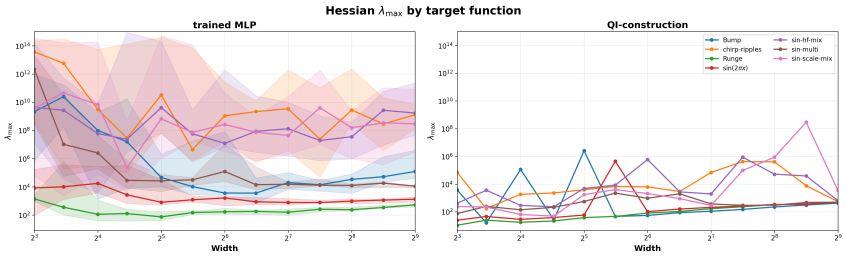

Figure 7: Estimated Hessian largest eigenvalues at convergence. **Left:** trained MLP. **Right:** QI-constructed MLP. The constructed networks do not exhibit worse conditioning than the trained networks.

5 Discussion

Our results suggest that high-precision failures in trained MLPs are not primarily due to expressivity. We construct an explicit quasi-interpolant realizable by a standard one-hidden-layer MLP that i) achieves machine-precision interpolation with ii) logarithmic parameter scaling while iii) remaining realizable in floating-point arithmetic. To our knowledge, ours is the first construction to satisfy all three conditions simultaneously. The construction exposes a dimensionless bandwidth parameter λ governing the tradeoff between aliasing suppression and numerical stability. Empirically, trained MLPs fail to maintain this regime: optimization drives $\lambda \to 0$, leading to length-scale collapse and rank saturation even though the quasi-interpolant construction itself is not prohibitively ill-conditioned. Altogether, these results suggest that optimization, rather than expressivity alone, is the key bottleneck for high-precision neural interpolation. An important next step is to translate these structural insights into optimization methods, for example through λ-aware initialization, bandwidth-preserving parameterizations, or training constraints that maintain the required scaling regime. More broadly, we hope this framework helps motivate a more principled theory of scaling laws for high-precision scientific machine learning.

Acknowledgements

The authors thank Yasa Baig, Roberto Garcia, Sam Layton, and Michael Zhang for their helpful feedback and discussion.

We gratefully acknowledge the support of NIH under Nos. U54EB020405 (Mobilize) and R01GM124443 (SimTK); NSF under Nos. CCF2247015 (Hardware-Aware), CCF1763315 (Beyond Sparsity), CCF1563078 (Volume to Velocity), 1937301 (RTML), 1900638 (Open Federated Virtual Assistants), and 2028602 (PPoSS); DARPA under No. HR00112520038 (Fallingwater); US DEVCOM ARL under Nos. W911NF-23-2-0184 (Long-context) and W911NF-21-2-0251 (Interactive Human-AI Teaming); NSTXL under No. N00164-23-9-G057-01 (Energy-Efficient AI Hardware); ONR under Nos. N000142312633 (Deep Signal Processing), N000142212426 (Misinformation Analysis), N000142012480 (Non-Euclidean Geometry), and N000142012275 (Data Programming); Stanford HAI under Nos. 247183, 215955, and 345077 (Evo); Google DeepMind; Google Research; Google Cloud; NXP, Xilinx, LETI-CEA, Intel, IBM, Microsoft, NEC, Toshiba, TSMC, ARM, Hitachi, BASF, Accenture, Ericsson, Qualcomm, Analog Devices, Salesforce, Total, the Laude Institute, Prime Intellect, Anthropic, the HAI-GCP Cloud Credits for Research program, the Stanford Data Science Initiative (SDSI), the Stanford Marlowe Computing Platform, and members of the Stanford DAWN project: Meta, Google, and VMWare; and members of the Stanford SEAMS project: IBM and Felicis. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views, policies, or endorsements, either expressed or implied, of NIH, ONR, DARPA, DOE, or the U.S. Government. CD is supported by the NSF GRFP and the Stanford EDGE Fellowship. JH is supported by the Aker Scholarship Foundation. JL is supported by the Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0023112.

REFERENCES

Benedikt Alkin, Maurits Bleeker, Richard Kurle, Tobias Kronlachner, Reinhard Sonnleitner, Matthias Dorfer, and Johannes Brandstetter. Ab-upt: Scaling neural cfd surrogates for high-fidelity automotive aerodynamics simulations via anchored-branched universal physics transformers, 2025. URL https://arxiv.org/abs/2502.09692.

Aras Bacho, Aleksei G. Sorokin, Xianjin Yang, Théo Bourdais, Edoardo Calvello, Matthieu Darcy, Alexander Hsu, Bamdad Hosseini, and Houman Owhadi. Operator learning at machine precision, 2025. URL https://arxiv.org/abs/2511.19980.

John P. Boyd. *Chebyshev and Fourier Spectral Methods*. Lecture Notes in Engineering. Springer Berlin, Heidelberg, 1 edition, 1989. ISBN 978-3-540-51487-9.

Martin Buhmann and Janin Jäger. *Quasi-Interpolation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2022.

Lianghao Cao, Thomas O'Leary-Roseberry, Prashant K. Jha, J. Tinsley Oden, and Omar Ghattas. Residual-based error correction for neural operator accelerated infinite-dimensional bayesian inverse problems. *Journal of Computational Physics*, 486:112104, 2023. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2023.112104. URL https://www.sciencedirect.com/science/article/pii/S0021999123001997.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989. doi: 10.1007/BF02551274.

Tim De Ryck, Samuel Lanthaler, and Siddhartha Mishra. On the approximation of functions by tanh neural networks. *Neural Networks*, 143:732–750, November 2021. ISSN 0893-6080. doi: 10.1016/j.neunet.2021.08.015. URL http://dx.doi.org/10.1016/j.neunet.2021.08.015.

Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks, 2019. URL https://arxiv.org/abs/1803.03635.

Vineet Gupta, Tomer Koren, and Yoram Singer. Shampoo: Preconditioned stochastic tensor optimization, 2018. URL https://arxiv.org/abs/1802.09568.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre. Training compute-optimal large language models, 2022. URL https://arxiv.org/abs/2203.15556.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(89)90020-8. URL https://www.sciencedirect.com/science/article/pii/0893608089900208.

Elham Kiyani, Khemraj Shukla, Jorge F. Urbán, Jérôme Darbon, and George Em Karniadakis. Optimizing the optimizer for physics-informed neural networks and kolmogorov-arnold networks, 2025. URL https://arxiv.org/abs/2501.16371.

Tanishq Kumar, Zachary Ankner, Benjamin F. Spector, Blake Bordelon, Niklas Muennighoff, Mansheej Paul, Cengiz Pehlevan, Christopher Ré, and Aditi Raghunathan. Scaling laws for precision, 2024. URL https://arxiv.org/abs/2411.04330.

Zongyi Li, Samuel Lanthaler, Catherine Deng, Michael Chen, Yixuan Wang, Kamyar Azizzadenesheli, and Anima Anandkumar. Scale-consistent learning for partial differential equations, 2025. URL https://arxiv.org/abs/2507.18813.

Jerry Liu, Yasa Baig, Denise Hui Jean Lee, Rajat Vadiraj Dwaraknath, Atri Rudra, and Chris Ré. Bwler: Barycentric weight layer elucidates a precision-conditioning tradeoff for pinns, 2025a. URL https://arxiv.org/abs/2506.23024.

Jerry Liu, Jessica Grogan, Owen Dugan, Ashish Rao, Simran Arora, Atri Rudra, and Christopher Ré. Towards learning high-precision least squares algorithms with sequence models, 2025b. URL https://arxiv.org/abs/2503.12295.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks, 2025c. URL https://arxiv.org/abs/2404.19756.

Chenkai Mao, Robert Lupoiu, Tianxiang Dai, Mingkun Chen, and Jonathan A. Fan. Towards general neural surrogate solvers with specialized neural accelerators, 2024. URL https://arxiv.org/abs/2405.02351.

Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence*, 6 (10):1256–1269, September 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00897-5. URL http://dx.doi.org/10.1038/s42256-024-00897-5.

Hrushikesh Narhar Mhaskar. Approximation properties of a multilayered feedforward artificial neural network. *Advances in Computational Mathematics*, 1(1):61–80, 1993.

Eric J. Michaud, Ziming Liu, and Max Tegmark. Precision machine learning. *Entropy*, 25(1):175, January 2023a. ISSN 1099-4300. doi: 10.3390/e25010175. URL http://dx.doi.org/10.3390/e25010175.

Eric J. Michaud, Ziming Liu, and Max Tegmark. Precision machine learning. *Entropy*, 25(1):175, January 2023b. ISSN 1099-4300. doi: 10.3390/e25010175. URL http://dx.doi.org/10.3390/e25010175.

Johannes Müller and Marius Zeinhofer. Achieving high accuracy with pinns via energy natural gradients, 2023. URL https://arxiv.org/abs/2302.13163.

Chanwook Park, Sourav Saha, Jiachen Guo, Hantao Zhang, Xiaoyu Xie, Miguel A. Bessa, Dong Qian, Wei Chen, Gregory J. Wagner, Jian Cao, and Wing Kam Liu. Interpolating neural network: A novel unification of machine learning and interpolation theory, 2024. URL https://arxiv.org/abs/2404.10296.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators, 2022. URL https://arxiv.org/abs/2202.11214.

Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pp. 40–44 vol.1, 1993. doi: 10.1109/ACSSC.1993.342465.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions, 2020. URL https://arxiv.org/abs/2006.09661.

Lloyd N. Trefethen. *Approximation Theory and Approximation Practice, Extended Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2019. doi: 10.1137/1.9781611975949. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611975949.

Lloyd N. Trefethen and David Bau. *Numerical Linear Algebra, Twenty-fifth Anniversary Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2022. doi: 10.1137/1.9781611977165. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611977165.

Jorge F. Urbán, Petros Stefanou, and José A. Pons. Unveiling the optimization process of physics informed neural networks: How accurate and competitive can pinns be? *Journal of Computational Physics*, 523:113656, 2025. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2024.113656. URL https://www.sciencedirect.com/science/article/pii/S0021999124009045.

Nikhil Vyas, Depen Morwani, Rosie Zhao, Mujin Kwun, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. Soap: Improving and stabilizing shampoo using adam, 2025. URL https://arxiv.org/abs/2409.11321.

Sifan Wang, Bowen Li, Yuhan Chen, and Paris Perdikaris. Piratenets: Physics-informed deep learning with residual adaptive networks, 2024. URL https://arxiv.org/abs/2402.00326.

Yongji Wang and Ching-Yao Lai. Multi-stage neural networks: Function approximator of machine precision, 2023. URL https://arxiv.org/abs/2307.08934.

Marcus Webb, Vincent Coppé, and Daan Huybrechs. Pointwise and uniform convergence of fourier extensions. *Constructive Approximation*, 52(1):139–175, Aug 2020. ISSN 1432-0940. doi: 10.1007/s00365-019-09486-x. URL https://doi.org/10.1007/s00365-019-09486-x.

A DERIVATIONS FOR TABLE 1

Here, we derive the scaling rules reported in Table 1 for analytic targets on $\Omega = [-1, 1]$. We compare three constructions: (i) classical one-hidden-layer constructions based on piecewise approximations (Cybenko, 1989), (ii) Mhaskar's higher-order construction based on polynomial interpolation (Mhaskar, 1993), and (iii) our Fourier-normalized quasi-interpolant (QI) construction (Section 3.1).

For each, we report the following quantities:

- **Number of parameters.** We report the width (number of hidden units) as a function of the target uniform error ε.
- **Bit complexity (working precision).** We report the number of mantissa bits b_{work} sufficient (or, where indicated, necessary due to conditioning) so that the *numerical error* incurred while forming or evaluating the construction does not dominate the *approximation error* ε. Crucially, this is not the same as the number of bits required merely to *store* the final parameters.
- **Weight magnitudes.** We report $\|W\|_{\max}$, the maximum absolute value among the MLP parameters produced by the construction.

Classical piecewise linear construction (Cybenko, 1989). This construction approximates f with a piecewise-constant function on a uniform partition of mesh h. It then implements each jump using a steep sigmoid step $\sigma(\alpha(x - t_i))$ and sums the resulting step increments to obtain a one-hidden-layer network.

Parameter count, weight magnitude, and bit complexity scaling. For Lipschitz f, the discretization error scales as $\|f - f_h\|_\infty = O(h)$, so choosing $h \asymp \varepsilon$ yields $N = \Theta(1/\varepsilon)$ intervals and hence $\Theta(1/\varepsilon)$ hidden units. To approximate a Heaviside step over width h with height error $\le \varepsilon$, one needs slope $\alpha = \Theta(h^{-1} \log(1/\varepsilon))$, yielding $\|W\|_{\max} = \Theta(\varepsilon^{-1} \log(1/\varepsilon))$. Finally, since this construction does not require any high-precision cancellations, the working precision requirement is dominated by resolving ε and representing α. This gives $b_{\text{work}} = \Theta(\log(1/\varepsilon))$.

Higher-order construction for analytic targets (Mhaskar, 1993). Mhaskar's analytic construction first uses *order-k* sigmoidal activations (defined as functions with polynomial-rate growth of degree k at $+\infty$) to approximate truncated monomial powers using a deep network. The construction then uses these monomials to implement high-degree polynomial interpolants, whose coefficients are computed explicitly from sampled values of f.

Parameter count scaling. For analytic targets, Mhaskar (1993) shows that exponentially small approximation error is achievable using $O(n \log n)$ neurons arranged in $O(\log n)$ layers, where n plays the role of an effective polynomial degree. In the analytic regime, where polynomial interpolation achieves geometric approximation error $\varepsilon \asymp \rho^n$ for some $\rho \in (0, 1)$, we require $n = \Theta(\log(1/\varepsilon))$. Thus the construction's parameter count is

$$N = O(n \log n) = \widetilde{O}\big(\log(1/\varepsilon) \log\log(1/\varepsilon)\big)$$

as reported in Table 1.

Weight magnitude scaling. To derive the stated $\|W\|_{\max}$ scaling, we study the "steepness" parameter in the key monomial gadget within the construction. An order-k sigmoidal activation behaves like x^k for large $x > 0$, so scaling $\sigma(x/\delta)$ yields an approximation to x_+^k. Achieving uniform error ε requires $\delta = \Theta(\varepsilon^{1/k})$, and therefore the required parameter scaling is $\|W\|_{\max} = 1/\delta = \Theta(\varepsilon^{-1/k})$.

Bit complexity scaling. The numerical bottleneck in implementing the construction from Mhaskar (1993) is in *forming the polynomial interpolant coefficients*. These coefficients are given by high-order finite-difference functionals applied to function samples. Expanding an $(r+1)$-st difference yields an alternating binomial sum with coefficients $\binom{r+1}{j}$, so roundoff is amplified by a factor that is exponential in r. A standard conditioning estimate for an $(r+1)$-st difference at mesh $h \sim 1/n$ yields a cancellation factor $\kappa \sim (2n)^r$, so recovering ε-level accuracy in coefficient formation requires working precision on the order of

$$b_{\text{work}} \gtrsim \log_2(\kappa) + \log_2(1/\varepsilon) \approx r \log(2n) + \log(1/\varepsilon).$$

In the analytic regime, the effective order r grows proportionally to n, and $n \asymp \log(1/\varepsilon)$, so the above conditioning-based estimate implies $b_{\text{work}} \gtrsim \widetilde{\Omega}(\log(1/\varepsilon)\log\log(1/\varepsilon))$. We emphasize that this is a *conditioning-based* bit complexity estimate: it reflects the numerical stability barrier in finite precision implementations, rather than an explicit bit-complexity theorem stated in Mhaskar (1993).

Our quasi-interpolant (QI) construction. Our method constructs a Fourier-normalized cardinal quasi-interpolant on a uniform grid (spacing h) with a bandwidth parameter γ coupled via $\lambda = \gamma h$, and then represents the resulting finite kernel expansion as a one-hidden-layer MLP. We refer to Section 3.1 for the precise definition, truncations, and the MLP projection.

Parameter count scaling. We start from our analytic error decomposition (Theorem 1), which includes a resolution term of the form $e^{-c/h}$. Setting $h = 2/N$, the choice $N = \Theta(\log(1/\varepsilon))$ makes the resolution term $\leq \varepsilon/4$. The halo and stencil truncation terms have the form $e^{-c\lambda R}$ and $e^{-c\lambda K_c}$, so choosing $R, K_c = \Theta(\log(1/\varepsilon)/\lambda)$ makes each $\leq \varepsilon/4$. In the viable regime where λ is bounded away from 0 and the normalizer remains well-conditioned, the resulting width is $O(\log(1/\varepsilon))$ as reported in Table 1.

Weight magnitudes scaling. To derive $\|W\|_{\max}$, we separate the three parameter types in the projected MLP: input scale γ, biases $b_m = -\gamma x_m$, and output weights $w[m]$. Since $\lambda = \gamma h$ and $h = 2/N$, holding λ constant and bounded away from 0 implies

$$\gamma = \lambda/h = \Theta(N) = \Theta(\log(1/\varepsilon)).$$

Because $x_m \in [-1, 1]$, we have $\max_m |b_m| \leq \gamma$. Meanwhile, the output weights satisfy $w[m] = a[m]/\gamma^{r-1}$, where $a[m]$ is obtained by a length-$O(K_c)$ convolution of samples with the character coefficients. Under the mild ℓ^1-summability condition on these coefficients, $\|a\|_\infty \leq \|f\|_\infty \|c\|_{\ell^1}$, hence $\|w\|_\infty \leq \|f\|_\infty \|c\|_{\ell^1} \gamma^{1-r}$. For fixed $r \geq 2$, the dominant contribution to $\|W\|_{\max}$ is therefore γ, yielding $\|W\|_{\max} = O(\log(1/\varepsilon))$.

Bit complexity scaling. Finally, to derive the working-precision bound, we identify the computational bottleneck as forming the coefficients $a[m]$ via a sum of $O(K_c) = O(\log(1/\varepsilon))$ terms. With unit roundoff $u = 2^{-b_{\text{work}}}$, the numerical error is on the order of $O(K_c u)$ times a bound on the weight magnitudes. Thus, taking $b_{\text{work}} = O(\log(1/\varepsilon) + \log\log(1/\varepsilon))$ ensures this arithmetic error is below ε.

Table 2: Target functions used in the interpolation experiments.

Target function	**Definition**
Single-frequency sine	$f(x) = \sin(2\pi x)$
Runge function	$f(x) = \dfrac{1}{1+25x^2}$
Multi-frequency sine	$f(x) = \sin(2\pi x) + \sin(4\pi x) + \sin(6\pi x)$
High-frequency mixture	$f(x) = \dfrac{1}{\|a\|_2} \sum_{k\in\mathcal{K}} a_k \sin(2\pi k x + \phi_k),$ $\mathcal{K} = \{3,5,7,11,13,17,19,23\}, \quad a_k = k^{-1/2}$
Chirp with ripples	$f(x) = \sin\big(2\pi(3t+12t^2)\big) + 0.25\sin(80\pi t), \quad t = \dfrac{x+1}{2}$
Multi-scale sine mixture	$f(x) = 0.6\sin(3\pi x) + 0.3\sin(14\pi x + 0.5) + 0.1\sin(70\pi x)$
Compactly supported bump	$f(x) = \begin{cases} \exp\left(1 - \dfrac{1}{1-u^2}\right), & \|u\| < 1, \\ 0, & \|u\| \geq 1, \end{cases} \quad u = \dfrac{x-0.35}{0.2}$

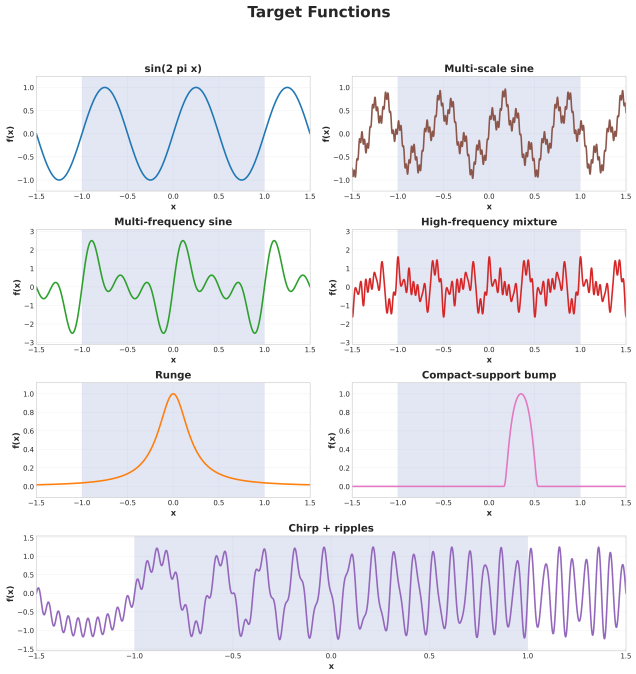

Figure 8: Target functions evaluated throughout the paper. Note the shaded regions denote the interpolation domain $\Omega = [-1, 1]$.

B TARGET FUNCTIONS

Table 2 displays all the target functions used throughout the paper. The target functions are also visualized in Figure 8.

C Theoretical Results

C.1 Uniform-grid setting and finite operator

We work on

$$\Omega := [-1, 1], \qquad h := \frac{2}{N}, \qquad x_k := -1 + kh, \qquad k \in \mathbb{Z}, \tag{14}$$

and, for a halo radius $R \in \mathbb{N}$, we set

$$\Omega_R := [-1 - Rh,\, 1 + Rh]. \tag{15}$$

Let $K : \mathbb{R} \to \mathbb{R}$ be a kernel profile and define the scaled family

$$K_\gamma(x) := \gamma K(\gamma x), \qquad \lambda := \gamma h. \tag{16}$$

For $f \in L^1(\mathbb{R})$, we use the fourier transform convention

$$\widehat{f}(\omega) := \int_{\mathbb{R}} f(x)\, e^{-i\omega x}\, dx, \qquad f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\omega)\, e^{i\omega x}\, d\omega. \tag{17}$$

Define the lattice normalizer and Fourier character by

$$D_h(\omega) := \sum_{m \in \mathbb{Z}} \widehat{K}_\gamma\left(\omega + \frac{2\pi m}{h}\right), \qquad \widehat{C}_h(\omega) := \frac{h}{D_h(\omega)}. \tag{18}$$

Let

$$\widetilde{C}_\lambda(\theta) := \widehat{C}_h(\theta/h), \tag{19}$$

and let $(c_j)_{j \in \mathbb{Z}}$ be the Fourier coefficients of $\widetilde{C}_\lambda$:

$$\widetilde{C}_\lambda(\theta) = \sum_{j \in \mathbb{Z}} c_j e^{-ij\theta}. \tag{20}$$

Define the quasi-cardinal and its Fourier-stencil truncation by

$$L_h(x) := \sum_{j \in \mathbb{Z}} c_j K_\gamma(x - jh), \qquad L_h^{(K_c)}(x) := \sum_{|j| \le K_c} c_j K_\gamma(x - jh). \tag{21}$$

Then the direct finite halo+stencil operator is

$$\big(Q_{h,R}^{(K_c)} f\big)(x) := \sum_{k=-R}^{N+R} f(x_k) L_h^{(K_c)}(x - x_k), \qquad x \in \Omega. \tag{22}$$

C.2 Kernel-side assumptions and structural consequences

Fix $\varepsilon \in (0, 1)$. We introduce five key assumptions on our kernels:

K1 (nonzero mass).

$$\widehat{K}(0) = \int_{\mathbb{R}} K(x)\, dx \neq 0. \tag{23}$$

K2 (interior-band aliasing). Define

$$\omega_c := (1 - \epsilon)\frac{\pi}{h}. \tag{24}$$

Assume there exist constants $C_{\mathrm{alias}}(\epsilon) > 0$, $c_{\mathrm{alias}}(\epsilon) > 0$, and $p_K \ge 1$ such that

$$\sup_{|\omega| \le \omega_c} |1 - M_h(\omega)| \le C_{\mathrm{alias}}(\epsilon) \exp\left(-\frac{c_{\mathrm{alias}}(\epsilon)}{\lambda^{p_K}}\right), \qquad M_h(\omega) := \frac{\widehat{K}_\gamma(\omega)}{D_h(\omega)}. \tag{25}$$

K3 (zero-free strip margin). Assume there exist $\sigma > 0$ and $d_0(\sigma, \lambda) > 0$ such that the rescaled normalizer

$$\widetilde{D}_\lambda(\theta) := D_h(\theta/h) \tag{26}$$

satisfies

$$\inf_{|\Re\theta|\le\pi,\, |\Im\theta|\le\sigma} \left|\widetilde{D}_\lambda(\theta)\right| \ge d_0(\sigma, \lambda). \tag{27}$$

K4 (spatial localization envelope). Assume there exist constants $A_K > 0$ and $a_K > 0$ such that

$$|K_\gamma(x)| \le A_K \gamma e^{-a_K \gamma |x|}, \qquad x \in \mathbb{R}. \tag{28}$$

K5 (strict Fourier positivity). Assume

$$\widehat{K}(\xi) > 0, \qquad \xi \in \mathbb{R}. \tag{29}$$

Stability prefactors Under K3–K4, define

$$C_c(\lambda, \sigma) := \frac{h}{d_0(\sigma, \lambda)}, \tag{30}$$

$$C_Q(\lambda, \sigma) := A_K \gamma\, C_c(\lambda, \sigma)\, \coth\!\left(\frac{a_K \lambda}{2}\right) \coth\!\left(\frac{\sigma}{2}\right), \tag{31}$$

and, assuming $a_K \lambda > \sigma$,

$$C_L(\lambda, \sigma) := A_K \gamma C_c(\lambda, \sigma) \coth\!\left(\frac{a_K \lambda - \sigma}{2}\right). \tag{32}$$

We also abbreviate

$$C_H(\lambda, \sigma) := \frac{2C_L(\lambda, \sigma)}{1 - e^{-\sigma}}, \qquad C_S(\lambda, \sigma) := \frac{2A_K \gamma\, C_c(\lambda, \sigma)}{1 - e^{-\sigma}} \coth\!\left(\frac{a_K \lambda}{2}\right). \tag{33}$$

Lemma 2 (Coefficient decay)**.** *Assume K3. Then*

$$|c_j| \le C_c(\lambda, \sigma) e^{-\sigma |j|}, \qquad j \in \mathbb{Z}. \tag{34}$$

Proof. This is the standard contour-shift argument on the strip cell $\{|\Re\theta| \le \pi,\, |\Im\theta| \le \sigma\}$. Since

$$|\widetilde{C}_\lambda(\theta)| = \left|\frac{h}{\widetilde{D}_\lambda(\theta)}\right| \le C_c(\lambda, \sigma) \tag{35}$$

on that strip, shifting the Fourier-coefficient contour to $\Im\theta = \operatorname{sign}(j)\sigma$ yields the factor $e^{-\sigma|j|}$. □

Lemma 3 (Finite-operator stability)**.** *For every sampled sequence g on the halo nodes,*

$$\|Q_{h,R}^{(K_c)} g\|_{L^\infty(\Omega)} \le C_Q(\lambda, \sigma) \|g\|_{\ell^\infty(\Omega_R \cap h\mathbb{Z})}. \tag{36}$$

In particular, the stability constant is independent of R and K_c.

Proof. For $x \in \Omega$,

$$\left|\big(Q_{h,R}^{(K_c)} g\big)(x)\right| \le \|g\|_{\ell^\infty(\Omega_R \cap h\mathbb{Z})} \sum_{k=-R}^{N+R} \left|L_h^{(K_c)}(x - x_k)\right|. \tag{37}$$

Expanding $L_h^{(K_c)}$, taking absolute values, and reindexing gives

$$\sum_{k=-R}^{N+R} \left|L_h^{(K_c)}(x - x_k)\right| \le \left(\sum_{|j|\le K_c} |c_j|\right) \sup_{y\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |K_\gamma(y - mh)|. \tag{38}$$

By Lemma 2,

$$\sum_{|j|\le K_c} |c_j| \le \sum_{j\in\mathbb{Z}} |c_j| \le C_c(\lambda,\sigma) \sum_{j\in\mathbb{Z}} e^{-\sigma|j|} = C_c(\lambda,\sigma)\coth\Big(\frac{\sigma}{2}\Big). \tag{39}$$

For the lattice sum, write $y = ht$ and use K4:

$$|K_\gamma(y - mh)| \le A_K\gamma e^{-a_K\lambda|t-m|}. \tag{40}$$

Hence

$$\sup_{y\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |K_\gamma(y-mh)| \le A_K\gamma \sup_{t\in\mathbb{R}} \sum_{m\in\mathbb{Z}} e^{-a_K\lambda|t-m|} \le A_K\gamma\coth\left(\frac{a_K\lambda}{2}\right). \tag{41}$$

Combining the estimates yields

$$\left|\big(Q_{h,R}^{(K_c)} g\big)(x)\right| \le C_Q(\lambda,\sigma)\,\|g\|_{\ell^\infty(\Omega_R\cap h\mathbb{Z})}, \tag{42}$$

with $C_Q(\lambda,\sigma)$ as defined above. □

Lemma 4 (Spatial decay of the full quasi-cardinal)**.** *Assume Lemma 2 and K4. If $a_K\lambda > \sigma$, then*

$$|L_h(x)| \le C_L(\lambda,\sigma) e^{-\sigma|x|/h}, \qquad x\in\mathbb{R}. \tag{43}$$

Proof. From $L_h(x) = \sum_j c_j K_\gamma(x - jh)$, Lemma 2, and K4,

$$|L_h(x)| \le \sum_{j\in\mathbb{Z}} C_c(\lambda,\sigma) e^{-\sigma|j|}\, A_K\gamma\, e^{-a_K\gamma|x-jh|}.$$

Consider $t := x/h$ so that $\gamma|x - jh| = \lambda|t - j|$. By the inverse triangle inequality:

$$e^{-\sigma|j|} e^{-a_K\lambda|t-j|} \le e^{-\sigma|t|} e^{-(a_K\lambda-\sigma)|t-j|}, \tag{44}$$

Summing the geometric lattice tail gives

$$\sum_{j\in\mathbb{Z}} e^{-(a_K\lambda-\sigma)|t-j|} \le \coth\left(\frac{a_K\lambda-\sigma}{2}\right), \tag{45}$$

which yields the claim. □

Lemma 5 (Fourier transform of the full quasi-cardinal)**.** *Assume Lemma 2. Then $\sum_{j\in\mathbb{Z}} |c_j| < \infty$, and*

$$\widehat{L_h}(\xi) = \widehat{K}_\gamma(\xi) \sum_{j\in\mathbb{Z}} c_j e^{-ijh\xi} = h\, M_h(\xi), \qquad \xi\in\mathbb{R}. \tag{46}$$

In particular,

$$\widehat{L_h}(\xi) = h\, M_h(\xi), \qquad |\xi| \le \frac{\pi}{h}. \tag{47}$$

Proof. By Lemma 2,

$$\sum_{j\in\mathbb{Z}} |c_j| \le C_c(\lambda,\sigma) \sum_{j\in\mathbb{Z}} e^{-\sigma|j|} < \infty.$$

Hence the defining series for L_h converges absolutely in $L^1(\mathbb{R})$, and we may take Fourier transforms term-by-term:

$$\widehat{L_h}(\xi) = \sum_{j\in\mathbb{Z}} c_j e^{-ijh\xi} \widehat{K}_\gamma(\xi) = \widehat{K}_\gamma(\xi) \sum_{j\in\mathbb{Z}} c_j e^{-ijh\xi}. \tag{48}$$

Now $\widetilde{C}_\lambda$ is 2π-periodic with Fourier series

$$\widetilde{C}_\lambda(\theta) = \sum_{j\in\mathbb{Z}} c_j e^{-ij\theta}.$$

Since

$$\widetilde{C}_\lambda(\theta) = \widehat{C}_h(\theta/h),$$

we obtain, after setting $\theta = h\xi$,

$$\sum_{j\in\mathbb{Z}} c_j e^{-ijh\xi} = \widehat{C}_h(\xi) = \frac{h}{D_h(\xi)}. \tag{49}$$

Therefore

$$\widehat{L_h}(\xi) = \widehat{K}_\gamma(\xi)\frac{h}{D_h(\xi)} = h\,\frac{\widehat{K}_\gamma(\xi)}{D_h(\xi)} = h\,M_h(\xi). \tag{50}$$

This proves the claim. □

Lemma 6 (Nyquist-cell multiplier identity)**.** *Define the unshifted infinite-grid operator by*

$$(\widetilde{Q}_h g)(y) := \sum_{k\in\mathbb{Z}} g(kh)\,L_h(y-kh). \tag{51}$$

Let $g \in \mathcal{S}(\mathbb{R})$ satisfy

$$\operatorname{supp}\widehat{g} \subset \left[-\frac{\pi}{h}, \frac{\pi}{h}\right]. \tag{52}$$

Then

$$\widehat{\widetilde{Q}_h g}(\xi) = M_h(\xi)\,\widehat{g}(\xi), \qquad |\xi| \le \frac{\pi}{h}. \tag{53}$$

Equivalently, if Π_h denotes Fourier projection onto the Nyquist cell

$$\widehat{(\Pi_h F)}(\xi) := \mathbf{1}_{\{|\xi|\le\pi/h\}}\,\widehat{F}(\xi), \tag{54}$$

then

$$\Pi_h(\widetilde{Q}_h g) = \mathcal{F}^{-1}\big(M_h\,\widehat{g}\big). \tag{55}$$

Proof. Since $g \in \mathcal{S}(\mathbb{R})$, the sampled sequence $(g(kh))_{k\in\mathbb{Z}}$ decays rapidly, and the defining series for $\widetilde{Q}_h g$ converges absolutely in $L^1_{\mathrm{loc}}(\mathbb{R})$. Therefore we may take Fourier transforms term-by-term:

$$\widehat{\widetilde{Q}_h g}(\xi) = \widehat{L_h}(\xi)\sum_{k\in\mathbb{Z}} g(kh)e^{-i\xi kh}. \tag{56}$$

By Lemma 5,

$$\widehat{L_h}(\xi) = h\,M_h(\xi). \tag{57}$$

It remains to compute the sampled Fourier series. Apply the classical Poisson summation formula to

$$u \longmapsto g(u)e^{-i\xi u} \in \mathcal{S}(\mathbb{R}).$$

Its Fourier transform is

$$\mathcal{F}\big(g(\cdot)e^{-i\xi(\cdot)}\big)(\eta) = \widehat{g}(\eta+\xi).$$

Hence

$$\sum_{k\in\mathbb{Z}} g(kh)e^{-i\xi kh} = \frac{1}{h}\sum_{m\in\mathbb{Z}} \widehat{g}\left(\xi + \frac{2\pi m}{h}\right). \tag{58}$$

Now assume $|\xi| \le \pi/h$. Since $\widehat{g}$ is supported in $[-\pi/h, \pi/h]$, all terms with $m \neq 0$ vanish, and therefore

$$\sum_{k\in\mathbb{Z}} g(kh)e^{-i\xi kh} = \frac{1}{h}\widehat{g}(\xi). \tag{59}$$

Substituting this into the previous display yields

$$\widehat{\widetilde{Q}_h g}(\xi) = h\,M_h(\xi)\cdot\frac{1}{h}\widehat{g}(\xi) = M_h(\xi)\widehat{g}(\xi), \qquad |\xi| \le \frac{\pi}{h}. \tag{60}$$

This proves the Nyquist-cell identity. □

Lemma 7 (Exact full-lattice response on a Fourier mode)**.** *Assume Lemma 4. Define the shifted full-lattice response by*

$$S_h(x,\omega) := \sum_{k\in\mathbb{Z}} e^{i\omega x_k} L_h(x - x_k), \qquad x_k = -1 + kh. \tag{61}$$

Then, for every $|\omega| \le \omega_c$,

$$S_h(x,\omega) = e^{i\omega x} \sum_{m\in\mathbb{Z}} M_h\!\left(\omega + \frac{2\pi m}{h}\right) e^{\,i2\pi m(x+1)/h}. \tag{62}$$

In particular, the Nyquist-cell component of $S_h(\cdot,\omega)$ is

$$\Pi_h\big(S_h(\cdot,\omega)\big) = M_h(\omega) e^{i\omega x}, \qquad |\omega| \le \omega_c. \tag{63}$$

Proof. It is convenient to work first on the unshifted lattice. Define

$$\widetilde{S}_h(y,\omega) := \sum_{k\in\mathbb{Z}} e^{i\omega kh} L_h(y - kh). \tag{64}$$

Set

$$\Phi_\omega(u) := e^{-i\omega u} L_h(u). \tag{65}$$

Then

$$\widetilde{S}_h(y,\omega) = e^{i\omega y} \sum_{k\in\mathbb{Z}} \Phi_\omega(y - kh). \tag{66}$$

By Lemma 4, L_h decays exponentially, hence $\Phi_\omega \in L^1(\mathbb{R})$ and Poisson summation applies:

$$\sum_{k\in\mathbb{Z}} \Phi_\omega(y - kh) = \frac{1}{h} \sum_{m\in\mathbb{Z}} \widehat{\Phi_\omega}\!\left(\frac{2\pi m}{h}\right) e^{\,i2\pi m y/h}. \tag{67}$$

Since multiplication by $e^{-i\omega u}$ shifts Fourier transform,

$$\widehat{\Phi_\omega}(\eta) = \widehat{L_h}(\eta + \omega). \tag{68}$$

Using Lemma 5,

$$\widehat{\Phi_\omega}\!\left(\frac{2\pi m}{h}\right) = h\, M_h\!\left(\omega + \frac{2\pi m}{h}\right). \tag{69}$$

Therefore

$$\widetilde{S}_h(y,\omega) = e^{i\omega y} \sum_{m\in\mathbb{Z}} M_h\!\left(\omega + \frac{2\pi m}{h}\right) e^{\,i2\pi m y/h}. \tag{70}$$

Now return to the shifted lattice $x_k = -1 + kh$. Since

$$x_k = -1 + kh \qquad \text{and} \qquad x - x_k = (x+1) - kh,$$

we have

$$S_h(x,\omega) = e^{-i\omega} \widetilde{S}_h(x+1,\omega). \tag{71}$$

Substituting the previous formula gives

$$S_h(x,\omega) = e^{i\omega x} \sum_{m\in\mathbb{Z}} M_h\!\left(\omega + \frac{2\pi m}{h}\right) e^{\,i2\pi m(x+1)/h}. \tag{72}$$

Finally, if $|\omega| \le \omega_c < \pi/h$, then only the $m = 0$ term lies in the Nyquist cell, which yields

$$\Pi_h\big(S_h(\cdot,\omega)\big) = M_h(\omega) e^{i\omega x}. \tag{73}$$

□

Lemma 8 (Halo response bound)**.** *Assume Lemma 4. Define*

$$S_{h,R}(x,\omega) := \sum_{k=-R}^{N+R} e^{i\omega x_k} L_h(x - x_k). \tag{74}$$

Then, for $x \in \Omega$ and $|\omega| \le \omega_c$,

$$|S_h(x,\omega) - S_{h,R}(x,\omega)| \le C_H(\lambda,\sigma) e^{-\sigma R}. \tag{75}$$

Proof. Since $|e^{i\omega x_k}| = 1$,

$$|S_h(x,\omega) - S_{h,R}(x,\omega)| \le \sum_{k<-R} |L_h(x-x_k)| + \sum_{k>N+R} |L_h(x-x_k)|. \tag{76}$$

Now apply Lemma 4:

$$|L_h(x-x_k)| \le C_L(\lambda,\sigma) e^{-\sigma|x-x_k|/h}.$$

Therefore

$$|S_h(x,\omega) - S_{h,R}(x,\omega)| \le C_L(\lambda,\sigma) \sum_{k<-R} e^{-\sigma|x-x_k|/h} + C_L(\lambda,\sigma) \sum_{k>N+R} e^{-\sigma|x-x_k|/h}. \tag{77}$$

Recall that $x_k = -1 + kh$ and $x_N = 1$. For the left tail, write

$$k = -R - m, \qquad m \ge 1.$$

Then

$$x_k = -1 - (R+m)h,$$

so for $x \in \Omega = [-1,1]$,

$$\frac{|x-x_k|}{h} \ge R + m.$$

Thus

$$\sum_{k<-R} e^{-\sigma|x-x_k|/h} \le \sum_{m=1}^{\infty} e^{-\sigma(R+m)} = e^{-\sigma R} \sum_{m=1}^{\infty} e^{-\sigma m}.$$

Similarly, for the right tail, write

$$k = N + R + m, \qquad m \ge 1.$$

Then

$$x_k = 1 + (R+m)h,$$

and again, for $x \in \Omega$,

$$\frac{|x-x_k|}{h} \ge R + m.$$

Hence

$$\sum_{k>N+R} e^{-\sigma|x-x_k|/h} \le e^{-\sigma R} \sum_{m=1}^{\infty} e^{-\sigma m}.$$

Combining the two estimates,

$$|S_h(x,\omega) - S_{h,R}(x,\omega)| \le 2C_L(\lambda,\sigma) e^{-\sigma R} \sum_{m=1}^{\infty} e^{-\sigma m} = 2C_L(\lambda,\sigma) \frac{e^{-\sigma}}{1-e^{-\sigma}} e^{-\sigma R}.$$

Since

$$2C_L(\lambda,\sigma) \frac{e^{-\sigma}}{1-e^{-\sigma}} \le C_H(\lambda,\sigma),$$

the claimed bound follows. □

Lemma 9 (Fourier-stencil response bound)**.** *Define*

$$S_{h,R}^{(K_c)}(x,\omega) := \sum_{k=-R}^{N+R} e^{i\omega x_k} L_h^{(K_c)}(x-x_k). \tag{78}$$

Then, for $x \in \Omega$ and $|\omega| \le \omega_c$,

$$\left| S_{h,R}(x,\omega) - S_{h,R}^{(K_c)}(x,\omega) \right| \le C_S(\lambda,\sigma) e^{-\sigma K_c}. \tag{79}$$

Proof. Since

$$L_h(x) - L_h^{(K_c)}(x) = \sum_{|j|>K_c} c_j K_\gamma(x - jh), \tag{80}$$

we have

$$\left|S_{h,R}(x,\omega) - S_{h,R}^{(K_c)}(x,\omega)\right| = \left|\sum_{k=-R}^{N+R} e^{i\omega x_k}\left(L_h(x-x_k) - L_h^{(K_c)}(x-x_k)\right)\right| \tag{81}$$

$$\leq \sum_{k=-R}^{N+R}\left|L_h(x-x_k) - L_h^{(K_c)}(x-x_k)\right| \tag{82}$$

$$\leq \sum_{k=-R}^{N+R}\sum_{|j|>K_c} |c_j|\,|K_\gamma(x - x_k - jh)|. \tag{83}$$

Reordering the sums gives

$$\left|S_{h,R}(x,\omega) - S_{h,R}^{(K_c)}(x,\omega)\right| \leq \sum_{|j|>K_c} |c_j| \sum_{k=-R}^{N+R} |K_\gamma(x - x_k - jh)|. \tag{84}$$

Since $x_k = -1 + kh$, we have

$$x - x_k - jh = x - x_{k+j}.$$

Thus, for each fixed j,

$$\sum_{k=-R}^{N+R} |K_\gamma(x - x_k - jh)| = \sum_{k=-R}^{N+R} |K_\gamma(x - x_{k+j})| \leq \sup_{y\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |K_\gamma(y - mh)|. \tag{85}$$

Therefore

$$\left|S_{h,R}(x,\omega) - S_{h,R}^{(K_c)}(x,\omega)\right| \leq \left(\sum_{|j|>K_c} |c_j|\right) \sup_{y\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |K_\gamma(y - mh)|. \tag{86}$$

By Lemma 2,

$$\sum_{|j|>K_c} |c_j| \leq 2C_c(\lambda,\sigma) \sum_{j=K_c+1}^{\infty} e^{-\sigma j} = \frac{2C_c(\lambda,\sigma)}{1-e^{-\sigma}}\, e^{-\sigma(K_c+1)}. \tag{87}$$

For the lattice sum, write $y = ht$ and use K4:

$$|K_\gamma(y - mh)| \leq A_K \gamma e^{-a_K \lambda |t-m|}. \tag{88}$$

Hence

$$\sup_{y\in\mathbb{R}} \sum_{m\in\mathbb{Z}} |K_\gamma(y - mh)| \leq A_K \gamma \coth\left(\frac{a_K \lambda}{2}\right). \tag{89}$$

Combining the estimates,

$$\left|S_{h,R}(x,\omega) - S_{h,R}^{(K_c)}(x,\omega)\right| \leq \frac{2A_K \gamma C_c(\lambda,\sigma)}{1-e^{-\sigma}} \coth\left(\frac{a_K \lambda}{2}\right) e^{-\sigma(K_c+1)}. \tag{90}$$

Since $e^{-\sigma(K_c+1)} \leq e^{-\sigma K_c}$, the stated bound follows with the previously defined constant $C_S(\lambda,\sigma)$. □

Lemma 10 (Reduction of the full mode-response alias term under Fourier positivity)**.** *Assume K2 and K5. Define the full mode-response aliasing factor by*

$$A_{\text{alias}}^{\text{full}}(\lambda,\epsilon) := \sup_{x\in\Omega,\ |\omega|\leq\omega_c} \left|1 - \sum_{m\in\mathbb{Z}} M_h\left(\omega + \frac{2\pi m}{h}\right) e^{\,i2\pi m(x+1)/h}\right|. \tag{91}$$

Then

$$A_{\text{alias}}^{\text{full}}(\lambda,\epsilon) \leq 2 \sup_{|\omega|\leq\omega_c} \left|1 - M_h(\omega)\right|. \tag{92}$$

Consequently,

$$A_{\text{alias}}^{\text{full}}(\lambda,\epsilon) \leq 2C_{\text{alias}}(\epsilon) \exp\left(-\frac{c_{\text{alias}}(\epsilon)}{\lambda^{p_K}}\right). \tag{93}$$

Proof. Since K5 implies $\widehat{K}_\gamma(\xi) = \widehat{K}(\xi/\gamma) > 0$ for all $\xi \in \mathbb{R}$, we have

$$D_h(\omega) = \sum_{m\in\mathbb{Z}} \widehat{K}_\gamma\Big(\omega + \frac{2\pi m}{h}\Big) > 0, \qquad \omega \in \mathbb{R}. \tag{94}$$

Hence

$$M_h\Big(\omega + \frac{2\pi m}{h}\Big) = \frac{\widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)}{D_h(\omega)} \geq 0. \tag{95}$$

Using the $2\pi/h$-periodicity of D_h, we have

$$\sum_{m\in\mathbb{Z}} M_h\Big(\omega + \frac{2\pi m}{h}\Big) = \frac{\sum_{m\in\mathbb{Z}} \widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)}{D_h(\omega)} = 1. \tag{96}$$

Thus, for every $x \in \Omega$ and $|\omega| \leq \omega_c$,

$$\begin{aligned}
&\left|1 - \sum_{m\in\mathbb{Z}} M_h\Big(\omega + \frac{2\pi m}{h}\Big)\, e^{\,i2\pi m(x+1)/h}\right| \\
&\quad = \left|\sum_{m\neq 0} M_h\Big(\omega + \frac{2\pi m}{h}\Big)\Big(1 - e^{\,i2\pi m(x+1)/h}\Big)\right| \\
&\quad \leq 2\sum_{m\neq 0} M_h\Big(\omega + \frac{2\pi m}{h}\Big) = 2\big(1 - M_h(\omega)\big).
\end{aligned} \tag{97}$$

Since $0 < M_h(\omega) \leq 1$ on $|\omega| \leq \omega_c$, taking the supremum yields

$$A^{\text{full}}_{\text{alias}}(\lambda, \epsilon) \leq 2 \sup_{|\omega|\leq\omega_c} \big|1 - M_h(\omega)\big|. \tag{98}$$

The final estimate follows immediately from K2. □

C.3 Target function assumption and bandlimited approximants

Let

$$L := 1 + Rh, \qquad \Omega_R = [-L, L], \qquad g(y) := f(Ly), \quad y \in [-1, 1]. \tag{99}$$

For $T > 1$, define the Fourier-extension map

$$m_T(y) := 2\,\frac{\cos\big(\frac{\pi y}{T}\big) - \cos\big(\frac{\pi}{T}\big)}{1 - \cos\big(\frac{\pi}{T}\big)} - 1, \tag{100}$$

the Bernstein ellipse of index $\varrho > 1$ by

$$\mathcal{B}(\varrho) := \left\{\frac{1}{2}\left(z + z^{-1}\right) : 1 \leq |z| \leq \varrho\right\}, \tag{101}$$

and the mapped Bernstein ellipse by

$$\mathcal{D}_T(\varrho) := m_T^{-1}\big(\mathcal{B}(\varrho)\big). \tag{102}$$

Also set

$$E(T) := \cot^2\Big(\frac{\pi}{4T}\Big). \tag{103}$$

We make the following analyticity assumption on our target function:

T1 (mapped-ellipse analyticity). There exist $T > 1$ and $\varrho_\star > E(T)$ such that g extends holomorphically to an open neighborhood of $\mathcal{D}_T(\varrho_\star)$ and continuously to $\overline{\mathcal{D}_T(\varrho_\star)}$. We write

$$M_{\mathrm{FE}}(f) := \sup_{z \in \mathcal{D}_T(\varrho_\star)} |g(z)| = \sup_{z \in L\mathcal{D}_T(\varrho_\star)} |f(z)|. \tag{104}$$

Remark 11. Assumption T1 is local and compact-domain based, and covers a broad range of functions. It includes all polynomials, all entire functions, and more generally every function that is analytic on an open set containing $L\mathcal{D}_T(\varrho_\star)$.

The purpose of writing the assumption this way is to leverage uniform convergence in Fourier extensions. Hence, for the exact continuous Fourier extension, define the approximation space

$$H_m^T := \operatorname{span}\left\{ e^{i\pi k y/T} : |k| \le m \right\}. \tag{105}$$

Let $G_m \in H_m^T$ denote the $L^2(-1,1)$-orthogonal projection of g onto H_m^T.

We use the standard uniform convergence theorem for exact Fourier extensions together with one FE-side coefficient estimate:

Proposition 1 (Fourier extension uniform convergence (Webb et al., 2020, Theorem 3.1))**.** *Assume T1. Fix any q with $1 < q < E(T)$. Then there exists a constant $C_{\mathrm{FE}}(T,q) > 0$ such that*

$$\|g - G_m\|_{L^\infty([-1,1])} \le C_{\mathrm{FE}}(T,q) M_{\mathrm{FE}}(f) q^{-m}, \qquad m \ge 0. \tag{106}$$

Proof. This is exactly the standard uniform convergence theorem for exact Fourier extensions of analytic functions. For the proof, see (Webb et al., 2020). □

Proposition 2 (Fourier extension coefficient-growth input in the good regime)**.** *Assume T1. Write the exact FE in the exponential basis as*

$$G_m(y) = \sum_{|k| \le m} b_k e^{i\pi k y/T}. \tag{107}$$

Then there exists a constant $C_{\mathrm{coef}}(T, \varrho_\star) > 0$ such that

$$\|b\|_{\ell^2} \le C_{\mathrm{coef}}(T, \varrho_\star) M_{\mathrm{FE}}(f). \tag{108}$$

Proof. Let

$$g(y) := f(Ly), \qquad y \in [-1,1],$$

and decompose

$$g = g_{\mathrm{e}} + g_{\mathrm{o}}, \qquad g_{\mathrm{e}}(y) := \frac{g(y) + g(-y)}{2}, \qquad g_{\mathrm{o}}(y) := \frac{g(y) - g(-y)}{2}.$$

By the standard polynomialization of Fourier extension, there exist functions

$$h_1,\ h_2$$

analytic on the Bernstein ellipse $\mathcal{B}(\varrho_\star)$ such that

$$g_{\mathrm{e}}(y) = h_1(m_T(y)), \qquad g_{\mathrm{o}}(y) = \sin\Big(\frac{\pi y}{T}\Big) h_2(m_T(y)),$$

with

$$\|h_1\|_{L^\infty(\mathcal{B}(\varrho_\star))} + \|h_2\|_{L^\infty(\mathcal{B}(\varrho_\star))} \le C_0(T)\, M_{\mathrm{FE}}(f).$$

Let $\{\phi_n\}_{n \ge 0}$ and $\{\psi_n\}_{n \ge 0}$ be the orthonormal polynomial families in the two weighted FE polynomial spaces, and define

$$\Phi_n(y) := \phi_n(m_T(y)), \qquad \Psi_n(y) := \sin\Big(\frac{\pi y}{T}\Big) \psi_n(m_T(y)).$$

Then

$$G_m(y) = \sum_{n=0}^{m} \alpha_n \Phi_n(y) + \sum_{n=0}^{m-1} \beta_n \Psi_n(y),$$

where α_n, β_n are the corresponding orthogonal projection coefficients of h_1, h_2.

Since h_1 is analytic on $\mathcal{B}(\varrho_\star)$, Bernstein–Walsh approximation gives, for each $n \geq 0$, a polynomial $p_{n-1} \in \mathbb{P}_{n-1}$ such that

$$\|h_1 - p_{n-1}\|_{L^\infty([-1,1])} \leq C_1(\varrho_\star)\|h_1\|_{L^\infty(\mathcal{B}(\varrho_\star))}\varrho_\star^{-n}.$$

Using the orthogonality of ϕ_n to $\mathbb{P}_{n-1}$,

$$\alpha_n = \langle h_1 - p_{n-1}, \phi_n \rangle_{w_1},$$

and hence

$$|\alpha_n| \leq \|h_1 - p_{n-1}\|_{L^2(w_1)} \leq C_A(T, \varrho_\star)\, M_{\mathrm{FE}}(f)\, \varrho_\star^{-n}.$$

The same argument applied to h_2 yields

$$|\beta_n| \leq C_A(T, \varrho_\star)\, M_{\mathrm{FE}}(f)\, \varrho_\star^{-n}.$$

Therefore,

$$|\alpha_n| + |\beta_n| \leq C_A(T, \varrho_\star)\, M_{\mathrm{FE}}(f)\, \varrho_\star^{-n}. \tag{109}$$

Next, write the FE basis functions in the exponential basis:

$$\Phi_n(y) = \sum_{|k|\leq n} d_k^{(n)} e^{i\pi k y/T}, \qquad \Psi_n(y) = \sum_{|k|\leq n+1} e_k^{(n)} e^{i\pi k y/T}.$$

Since $\Phi_n, \Psi_n \in H_{n+1}^T$, the standard FE continuation estimate implies that there exist constants $C_B(T) > 0$, $p \geq 0$, and $E(T) > 1$ such that

$$\|\Phi_n\|_{L^\infty([-T,T])} + \|\Psi_n\|_{L^\infty([-T,T])} \leq C_B(T)(1+n)^p E(T)^n.$$

By Parseval on $[-T, T]$,

$$\|d^{(n)}\|_{\ell^2} = (2T)^{-1/2}\|\Phi_n\|_{L^2([-T,T])} \leq \|\Phi_n\|_{L^\infty([-T,T])},$$

and similarly for $e^{(n)}$. Hence

$$\|d^{(n)}\|_{\ell^2} + \|e^{(n)}\|_{\ell^2} \leq C_B(T)(1+n)^p E(T)^n. \tag{110}$$

Let

$$G_m(y) = \sum_{|k|\leq m} b_k e^{i\pi k y/T},$$

and denote by b its exponential coefficient vector. Since

$$b = \sum_{n=0}^{m} \alpha_n d^{(n)} + \sum_{n=0}^{m-1} \beta_n e^{(n)},$$

the triangle inequality together with Equation (109) and Equation (110) gives

$$\|b\|_{\ell^2} \leq C_A(T, \varrho_\star) C_B(T) M_{\mathrm{FE}}(f) \sum_{n=0}^{m} (1+n)^p \left(\frac{E(T)}{\varrho_\star}\right)^n.$$

By T1, $\varrho_\star > E(T)$, so the series

$$\sum_{n=0}^{\infty} (1+n)^p \left(\frac{E(T)}{\varrho_\star}\right)^n$$

converges. Therefore the partial sums are uniformly bounded in m, and we obtain

$$\|b\|_{\ell^2} \leq C_{\mathrm{coef}}(T, \varrho_\star)\, M_{\mathrm{FE}}(f),$$

for some constant $C_{\mathrm{coef}}(T, \varrho_\star) > 0$ independent of m.

□

Remark 12. Proposition 2 is the only additional FE-side input beyond the standard uniform convergence theorem. With the uniform ℓ^2 estimate above, the later ℓ^2-to-ℓ^1 passage contributes only a $(1+\omega_c)^{1/2}$ prefactor. A sharper target-side construction with an $O(1)$ total-variation bound would remove even that loss.

Theorem 13 (Bandlimited approximant from FE-A)**.** *Assume T1 and fix any cutoff $\omega_c > 0$. Then there exists a finite complex measure μ_{ω_c} supported in $[-\omega_c, \omega_c]$ such that the bandlimited function*

$$P_{\omega_c}(x) := \int_{-\omega_c}^{\omega_c} e^{i\omega x}\, d\mu_{\omega_c}(\omega) \tag{111}$$

satisfies

$$\|f - P_{\omega_c}\|_{L^\infty(\Omega_R)} \le C_{\mathrm{BL}} M_{\mathrm{FE}}(f) e^{-c_{\mathrm{BL}}\omega_c}, \tag{112}$$

and

$$\|\mu_{\omega_c}\|_{\mathrm{TV}} \le C_{\mathrm{spec}} M_{\mathrm{FE}}(f)(1+\omega_c)^{1/2}. \tag{113}$$

Here $C_{\mathrm{BL}}, c_{\mathrm{BL}}, C_{\mathrm{spec}} > 0$ depend only on the chosen FE geometry $(T, \varrho_\star, q, L)$.

Proof. Fix $q \in (1, E(T))$ once and for all. Given $\omega_c > 0$, choose

$$m := \left\lfloor \frac{TL}{\pi}\, \omega_c \right\rfloor. \tag{114}$$

Let $G_m \in H_m^T$ be the exact FE of g, and define the scaled approximant on Ω_R by

$$P_{\omega_c}(x) := G_m(x/L). \tag{115}$$

Since

$$G_m(y) = \sum_{|k|\le m} b_k e^{i\pi k y/T}, \tag{116}$$

we may write

$$P_{\omega_c}(x) = \sum_{|k|\le m} b_k e^{i\pi k x/(TL)} = \int e^{i\omega x}\, d\mu_{\omega_c}(\omega), \tag{117}$$

with the discrete spectral measure

$$\mu_{\omega_c} := \sum_{|k|\le m} b_k\, \delta_{\pi k/(TL)}. \tag{118}$$

By construction,

$$\operatorname{supp}\mu_{\omega_c} \subset \left[-\frac{\pi m}{TL}, \frac{\pi m}{TL}\right] \subset [-\omega_c, \omega_c]. \tag{119}$$

Moreover, because $x \in \Omega_R$ if and only if $x/L \in [-1,1]$,

$$\|f - P_{\omega_c}\|_{L^\infty(\Omega_R)} = \|g - G_m\|_{L^\infty([-1,1])} \le C_{\mathrm{FE}}(T,q) M_{\mathrm{FE}}(f) q^{-m} \tag{120}$$

by Proposition 1. Since

$$m \ge \frac{TL}{\pi}\, \omega_c - 1, \tag{121}$$

we get

$$q^{-m} \le q\, \exp\!\left(-\frac{TL\log q}{\pi}\, \omega_c\right). \tag{122}$$

Thus the interval error bound follows with

$$C_{\mathrm{BL}} := q\, C_{\mathrm{FE}}(T,q), \qquad c_{\mathrm{BL}} := \frac{TL\log q}{\pi}. \tag{123}$$

For the spectral mass, Proposition 2 gives

$$\|b\|_{\ell^2} \le C_{\mathrm{coef}}(T, \varrho_\star) M_{\mathrm{FE}}(f). \tag{124}$$

By Cauchy–Schwarz,

$$\|\mu_{\omega_c}\|_{\mathrm{TV}} = \sum_{|k|\le m} |b_k| \le \sqrt{2m+1}\, \|b\|_{\ell^2} \le C\, M_{\mathrm{FE}}(f)(1+m)^{1/2}, \tag{125}$$

for a constant C depending only on T and $\varrho_\star$. Since

$$1 + m \le 1 + \frac{TL}{\pi}\, \omega_c, \tag{126}$$

this yields the desired total-variation bound after enlarging the constant once more. □

C.4 FINITE-INTERVAL QUASI-INTERPOLATION THEOREM

We now present the full finite-interval quasi-interpolation theorem.

Theorem 14 (Final truncated Quasi-Interpolant Bound)**.** *Fix $\epsilon \in (0,1)$ and set*

$$\omega_c := (1-\epsilon)\frac{\pi}{h}. \tag{127}$$

Assume K1–K5 and T1, and suppose additionally that

$$a_K \lambda > \sigma. \tag{128}$$

Then

$$\begin{aligned}\|f - Q_{h,R}^{(K_c)} f\|_{L^\infty(\Omega)} \le{}& C_{\mathrm{BL}}\big(1 + C_Q(\lambda,\sigma)\big) M_{\mathrm{FE}}(f) e^{-c_{\mathrm{BL}}\omega_c} \\ &+ C_{\mathrm{spec}} M_{\mathrm{FE}}(f)(1+\omega_c)^{1/2}\Big[2A_{\mathrm{alias}}(\lambda,\epsilon) + C_H(\lambda,\sigma)e^{-\sigma R} + C_S(\lambda,\sigma)e^{-\sigma K_c}\Big],\end{aligned} \tag{129}$$

where

$$A_{\mathrm{alias}}(\lambda,\epsilon) := \sup_{|\omega|\le\omega_c} |1 - M_h(\omega)|. \tag{130}$$

In particular, K2 implies

$$2A_{\mathrm{alias}}(\lambda,\epsilon) \le 2C_{\mathrm{alias}}(\epsilon)\exp\left(-\frac{c_{\mathrm{alias}}(\epsilon)}{\lambda^{p_K}}\right). \tag{131}$$

Equivalently, since $\omega_c = (1-\epsilon)\pi/h$, the first term is geometric in $1/h$, while the remaining three terms are the usual aliasing, halo, and Fourier-stencil contributions multiplied by the polynomial spectral-mass factor $(1+\omega_c)^{1/2}$.

Proof. Let

$$Q := Q_{h,R}^{(K_c)}, \qquad P := P_{\omega_c}, \tag{132}$$

where P_{ω_c} is the bandlimited approximant from Theorem 13. Consider the error decomposition:

$$f - Qf = (f - P) + (P - QP) + Q(P - f). \tag{133}$$

Step 1: resolution term. Since $\Omega \subset \Omega_R$,

$$\|f - P\|_{L^\infty(\Omega)} \le \|f - P\|_{L^\infty(\Omega_R)} \le C_{\mathrm{BL}} M_{\mathrm{FE}}(f) e^{-c_{\mathrm{BL}}\omega_c}. \tag{134}$$

Step 2: filtered tail. By Lemma 3,

$$\|Q(P-f)\|_{L^\infty(\Omega)} \le C_Q(\lambda,\sigma)\|P - f\|_{\ell^\infty(\Omega_R \cap h\mathbb{Z})} \le C_Q(\lambda,\sigma)\|P - f\|_{L^\infty(\Omega_R)}. \tag{135}$$

Using Theorem 13 once more gives

$$\|Q(P-f)\|_{L^\infty(\Omega)} \le C_Q(\lambda,\sigma) C_{\mathrm{BL}} M_{\mathrm{FE}}(f) e^{-c_{\mathrm{BL}}\omega_c}. \tag{136}$$

Combining Steps 1 and 2 yields the first resolution bound term in the theorem.

Step 3: modewise representation of $P - QP$. Write

$$P(x) = \int_{-\omega_c}^{\omega_c} e^{i\omega x}\, d\mu_{\omega_c}(\omega). \tag{137}$$

Define

$$S_{h,R}^{(K_c)}(x,\omega) := \sum_{k=-R}^{N+R} e^{i\omega x_k} L_h^{(K_c)}(x - x_k). \tag{138}$$

Then

$$P(x) - QP(x) = \int_{-\omega_c}^{\omega_c} \Big(e^{i\omega x} - S_{h,R}^{(K_c)}(x,\omega)\Big)\, d\mu_{\omega_c}(\omega). \tag{139}$$

Therefore,

$$|P(x) - QP(x)| \le \|\mu_{\omega_c}\|_{\mathrm{TV}} \sup_{|\omega|\le\omega_c} \Big|e^{i\omega x} - S_{h,R}^{(K_c)}(x,\omega)\Big|. \tag{140}$$

Step 4: split the mode defect. Define the intermediate full-cardinal finite response

$$S_{h,R}(x,\omega) := \sum_{k=-R}^{N+R} e^{i\omega x_k} L_h(x - x_k). \tag{141}$$

Then

$$e^{i\omega x} - S_{h,R}^{(K_c)}(x,\omega) = \underbrace{\big(e^{i\omega x} - S_h(x,\omega)\big)}_{\text{full-lattice aliasing / mode-response defect}} + \underbrace{\big(S_h(x,\omega) - S_{h,R}(x,\omega)\big)}_{\text{halo truncation}} + \underbrace{\big(S_{h,R}(x,\omega) - S_{h,R}^{(K_c)}(x,\omega)\big)}_{\text{Fourier-stencil truncation}}. \tag{142}$$

By Lemma 7,

$$S_h(x,\omega) = e^{i\omega x} \sum_{m\in\mathbb{Z}} M_h\left(\omega + \frac{2\pi m}{h}\right) e^{\,i2\pi m(x+1)/h}. \tag{143}$$

Therefore

$$\left|e^{i\omega x} - S_h(x,\omega)\right| = \left|1 - \sum_{m\in\mathbb{Z}} M_h\left(\omega + \frac{2\pi m}{h}\right) e^{\,i2\pi m(x+1)/h}\right|. \tag{144}$$

By the definition of $A_{\text{alias}}^{\text{full}}(\lambda,\epsilon)$ from Lemma 10,

$$\left|e^{i\omega x} - S_h(x,\omega)\right| \le A_{\text{alias}}^{\text{full}}(\lambda,\epsilon). \tag{145}$$

By Lemma 8,

$$|S_h(x,\omega) - S_{h,R}(x,\omega)| \le C_H(\lambda,\sigma)e^{-\sigma R}. \tag{146}$$

By Lemma 9,

$$\left|S_{h,R}(x,\omega) - S_{h,R}^{(K_c)}(x,\omega)\right| \le C_S(\lambda,\sigma)e^{-\sigma K_c}. \tag{147}$$

Therefore, uniformly for $x \in \Omega$ and $|\omega| \le \omega_c$,

$$\left|e^{i\omega x} - S_{h,R}^{(K_c)}(x,\omega)\right| \le A_{\text{alias}}^{\text{full}}(\lambda,\epsilon) + C_H(\lambda,\sigma)e^{-\sigma R} + C_S(\lambda,\sigma)e^{-\sigma K_c}. \tag{148}$$

Finally, by Lemma 10,

$$A_{\text{alias}}^{\text{full}}(\lambda,\epsilon) \le 2A_{\text{alias}}(\lambda,\epsilon). \tag{149}$$

Therefore, uniformly for $x \in \Omega$ and $|\omega| \le \omega_c$,

$$\left|e^{i\omega x} - S_{h,R}^{(K_c)}(x,\omega)\right| \le 2A_{\text{alias}}(\lambda,\epsilon) + C_H(\lambda,\sigma)e^{-\sigma R} + C_S(\lambda,\sigma)e^{-\sigma K_c}. \tag{150}$$

Step 5: integrate the mode bound against the spectral measure. Using Theorem 13,

$$\|\mu_{\omega_c}\|_{\text{TV}} \le C_{\text{spec}} M_{\text{FE}}(f)(1+\omega_c)^{1/2}. \tag{151}$$

Therefore,

$$\|P - QP\|_{L^\infty(\Omega)} \le C_{\text{spec}} M_{\text{FE}}(f)(1+\omega_c)^{1/2}\Big[2A_{\text{alias}}(\lambda,\epsilon) + C_H(\lambda,\sigma)e^{-\sigma R} + C_S(\lambda,\sigma)e^{-\sigma K_c}\Big]. \tag{152}$$

Combining this with the decomposition equation 133 and Steps 1–2 proves the theorem. □

C.5 MLP REALIZATION

Corollary 15 (Derivative-trick realization)**.** *Assume there exists a smooth activation ψ and an integer $r \geq 1$ such that*

$$K(u) = \psi^{(r)}(u). \tag{153}$$

Then, for the finite kernel network

$$\big(Q_{h,R}^{(K_c)} f\big)(x) = \sum_{m=-R}^{N+R} a[m] K_\gamma(x - x_m), \tag{154}$$

with

$$a[m] := \sum_{|j| \leq K_c} c_j f(x_{m-j}), \tag{155}$$

the one-hidden-layer MLP

$$g_{\mathrm{MLP}}(x) := \sum_{m=-R}^{N+R} \frac{a[m]}{\gamma^{r-1}} \psi\big(\gamma(x - x_m)\big) \tag{156}$$

satisfies

$$g_{\mathrm{MLP}}^{(r)}(x) = \big(Q_{h,R}^{(K_c)} f\big)(x). \tag{157}$$

Consequently, if $f = \widetilde{f}^{(r)}$, then g_{MLP} recovers $\widetilde{f}$ up to a polynomial of degree at most $r-1$, fixed by boundary conditions.

C.6 KERNEL VERIFICATIONS

We verify Assumptions K1 –K5 for differnet kernels.

C.6.1 KERNEL VERIFICATION FOR THE NORMALIZED sech^2 FAMILY

We verify Assumptions K1 –K5 for the normalized profile

$$K(x) := \frac{1}{2}\,\mathrm{sech}^2(x), \qquad K_\gamma(x) = \frac{\gamma}{2}\,\mathrm{sech}^2(\gamma x). \tag{158}$$

K1 (mass). Since $\int_{\mathbb{R}} \mathrm{sech}^2(x)\,dx = 2$, we have

$$\widehat{K}(0) = \int_{\mathbb{R}} K(x)\,dx = 1. \tag{159}$$

Closed-form Fourier transform. We have

$$\widehat{K}(\xi) = \frac{\frac{\pi}{2}\,\xi}{\sinh\big(\frac{\pi}{2}\,\xi\big)}, \qquad \xi \in \mathbb{R}, \tag{160}$$

with continuous extension $\widehat{K}(0) = 1$. Consequently $\widehat{K}(\xi) > 0$ for all $\xi \in \mathbb{R}$.

K5 (Fourier positivity). By Equation 160,

$$\widehat{K}(\xi) = \frac{\frac{\pi}{2}\,\xi}{\sinh\big(\frac{\pi}{2}\,\xi\big)}.$$

For $\xi \neq 0$, the numerator and denominator have the same sign because $\sinh$ is odd and strictly increasing, hence $\widehat{K}(\xi) > 0$. At $\xi = 0$, the continuous extension gives $\widehat{K}(0) = 1$. Therefore

$$\widehat{K}(\xi) \geq 0, \qquad \xi \in \mathbb{R}, \tag{161}$$

so K5 holds.

K2 (interior-band aliasing) with $p_K = 1$.

Proposition 3 (Interior-band aliasing for sech^2)**.** *Fix $\varepsilon \in (0,1)$. There exists $C_{\mathrm{alias}}(\varepsilon) > 0$ such that for all $h > 0$, $\gamma > 0$,*

$$\sup_{|\omega| \le (1-\varepsilon)\pi/h} \big|1 - M_h(\omega)\big| \le C_{\mathrm{alias}}(\varepsilon) \exp\!\left(-\frac{\varepsilon\pi^2}{\lambda}\right), \qquad \lambda = \gamma h. \tag{162}$$

In particular, K2 holds with $p_K = 1$ and one may take $c_{\mathrm{alias}}(\varepsilon) = \varepsilon\pi^2$.

Proof. By Equation 160, $\widehat{K}_\gamma(\omega) \ge 0$, hence $D_h(\omega) \ge \widehat{K}_\gamma(\omega)$ and

$$|1 - M_h(\omega)| = \frac{\sum_{m\neq 0} \widehat{K}_\gamma(\omega + 2\pi m/h)}{D_h(\omega)} \le \sum_{m\neq 0} \frac{\widehat{K}_\gamma(\omega + 2\pi m/h)}{\widehat{K}_\gamma(\omega)}.$$

Write $\widehat{K}_\gamma(\omega) = \widehat{K}(\omega/\gamma)$ and set $F(t) := \widehat{K}(t)$ for $t \ge 0$. On the interior band, $|\omega| \le (1-\varepsilon)\pi/h$ implies $t := |\omega|/\gamma \le (1-\varepsilon)\pi/\lambda$. For $m \ge 1$, the closest approach is at $\omega = -(1-\varepsilon)\pi/h$, giving $|\omega + 2\pi m/h|/\gamma \ge (2m - 1 + \varepsilon)\pi/\lambda$, hence the gap $\delta_m \ge 2(m - 1 + \varepsilon)\pi/\lambda$. Using $F(t) = \frac{(\pi/2)t}{\sinh((\pi/2)t)}$ and the inequality $\sinh(a+b) \ge \frac{1}{2}e^b \sinh(a)$ for $a, b \ge 0$, one obtains $F(t + \delta)/F(t) \le C(\varepsilon)e^{-(\pi/2)\delta}$ uniformly for $t \in [0, (1-\varepsilon)\pi/\lambda]$ (after absorbing the mild $(t+\delta)/t$ factor into $C(\varepsilon)$). Thus

$$\frac{\widehat{K}_\gamma(\omega + 2\pi m/h)}{\widehat{K}_\gamma(\omega)} \le C(\varepsilon) \exp\!\Big(-\frac{\pi}{2}\delta_m\Big) \le C(\varepsilon) \exp\!\left(-\frac{(m - 1 + \varepsilon)\pi^2}{\lambda}\right).$$

Summing the geometric series over $m \ge 1$ and using symmetry for $m \le -1$ yields the claimed bound. □

K4 (spatial localization envelope). Using $\mathrm{sech}(t) \le 2e^{-|t|}$,

$$|K_\gamma(x)| = \frac{\gamma}{2}\,\mathrm{sech}^2(\gamma x) \le 2\gamma\, e^{-2\gamma|x|}, \tag{163}$$

so K4 holds with $(A_K, a_K) = (2, 2)$.

K3 (zero-free strip margin) via diagonal dominance. We include a convenient sufficient condition that makes the λ–σ feasibility explicit.

Proposition 4 (Poisson form of the normalizer)**.** *Assume K_γ decays sufficiently fast for Poisson summation. Then*

$$D_h(\omega) = h \sum_{k\in\mathbb{Z}} K_\gamma(kh)\, e^{-i\omega kh}, \qquad \widetilde{D}_\lambda(\theta) = D_h(\theta/h) = \lambda \sum_{k\in\mathbb{Z}} K(\lambda k)\, e^{-ik\theta}. \tag{164}$$

Proof. This is the standard Poisson summation identity applied to $x \mapsto K_\gamma(x)e^{-i\omega x}$. □

Proposition 5 (Diagonal dominance sufficient condition for K3: sech^2)**.** *Let $K(x) = \frac{1}{2}\mathrm{sech}^2(x)$. Fix $\sigma > 0$ and assume $2\lambda > \sigma$. Then for all $\theta \in \mathbb{C}$ with $|\Im\theta| \le \sigma$,*

$$\big|\widetilde{D}_\lambda(\theta)\big| \ge \frac{\lambda}{2}\Big(1 - q(\sigma,\lambda)\Big), \qquad q(\sigma,\lambda) := \frac{8e^{-(2\lambda-\sigma)}}{1 - e^{-(2\lambda-\sigma)}}. \tag{165}$$

In particular, if $q(\sigma,\lambda) < 1$ then K3 holds with $d_0(\sigma,\lambda) \ge \frac{\lambda}{2}(1 - q(\sigma,\lambda)) > 0$.

Proof. By Proposition 4, $\widetilde{D}_\lambda(\theta) = \lambda\sum_{k\in\mathbb{Z}} K(\lambda k)e^{-ik\theta}$. Write $\theta = \theta_r + i\theta_i$ with $|\theta_i| \le \sigma$. Then $|e^{-ik\theta}| = e^{k\theta_i} \le e^{\sigma|k|}$. Separating the $k = 0$ term and applying the triangle inequality,

$$|\widetilde{D}_\lambda(\theta)| \ge \lambda K(0) - \lambda \sum_{k\neq 0} |K(\lambda k)| e^{\sigma|k|}.$$

Using $|K(\lambda k)| \le 2e^{-2\lambda|k|}$ gives $|K(\lambda k)|e^{\sigma|k|} \le 2e^{-(2\lambda-\sigma)|k|}$, so $\sum_{k\neq 0}|K(\lambda k)|e^{\sigma|k|} \le 4\sum_{k\ge 1} e^{-(2\lambda-\sigma)k} = 4\frac{e^{-(2\lambda-\sigma)}}{1-e^{-(2\lambda-\sigma)}}$. Since $K(0) = 1/2$, the claim follows. □

C.6.2 KERNEL VERIFICATION FOR THE NORMALIZED GELU″ FAMILY

We verify Assumptions K1 –K5 for the GELU″ profile

$$K(x) := (2 - x^2)\,\phi(x), \qquad \phi(x) := (2\pi)^{-1/2} e^{-x^2/2}. \tag{166}$$

Since $\int_{\mathbb{R}} K(x)\,dx = 1$ below, no additional normalization is required.

K1 (mass). Since ϕ is entire and $x \mapsto (2-x^2)$ is a polynomial, K is entire. Moreover, $K \in L^1(\mathbb{R})$ because it is a polynomial times a Gaussian. Using $\int_{\mathbb{R}} \phi(x)\,dx = 1$ and $\int_{\mathbb{R}} x^2\phi(x)\,dx = 1$,

$$\widehat{K}(0) = \int_{\mathbb{R}} K(x)\,dx = 2\int_{\mathbb{R}} \phi(x)\,dx - \int_{\mathbb{R}} x^2\phi(x)\,dx = 1, \tag{167}$$

so K1 holds.

Closed-form Fourier transform. One has

$$\widehat{K}(\omega) = 2\widehat{\phi}(\omega) - \widehat{x^2\phi}(\omega) = (1+\omega^2)e^{-\omega^2/2}, \qquad \omega \in \mathbb{R}. \tag{168}$$

K5 (Fourier positivity). Equation equation 168 gives

$$\widehat{K}(\omega) = (1+\omega^2)e^{-\omega^2/2}.$$

Both factors are strictly positive for every $\omega \in \mathbb{R}$, so

$$\widehat{K}(\omega) > 0, \qquad \omega \in \mathbb{R}, \tag{169}$$

and hence K5 holds.

K2 (interior-band aliasing) with $p_K = 2$. Since $\widehat{K} \geq 0$, the same positivity reduction used in the sech^2 proof applies: for $|\omega| \leq (1-\varepsilon)\pi/h$,

$$|1 - M_h(\omega)| = \frac{\sum_{m\in\mathbb{Z}\setminus\{0\}} \widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)}{D_h(\omega)} \leq \sum_{m\in\mathbb{Z}\setminus\{0\}} \frac{\widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)}{\widehat{K}_\gamma(\omega)}. \tag{170}$$

Write $\widehat{K}_\gamma(\omega) = \widehat{K}(\omega/\gamma)$ and set

$$s := \frac{|\omega|}{\gamma}, \qquad s_m(\omega) := \frac{1}{\gamma}\left|\omega + \frac{2\pi m}{h}\right|.$$

On the interior band, $s \leq (1-\varepsilon)\pi/\lambda$. For $m \geq 1$, the closest approach occurs at $\omega = -(1-\varepsilon)\pi/h$, giving

$$s_m(\omega) \geq \frac{(2m - 1 + \varepsilon)\pi}{\lambda}.$$

Using equation 168, define

$$F(t) := (1+t^2)e^{-t^2/2}, \qquad t \geq 0.$$

Then

$$\frac{\widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)}{\widehat{K}_\gamma(\omega)} = \frac{F\big(s_m(\omega)\big)}{F(s)} = \frac{1 + s_m(\omega)^2}{1+s^2}\exp\Big(-\frac{s_m(\omega)^2 - s^2}{2}\Big). \tag{171}$$

For $\Delta := s_m(\omega)^2 - s^2 \geq 0$, the prefactor obeys

$$\frac{1 + s_m(\omega)^2}{1+s^2} \leq 1 + \Delta,$$

so $(1+\Delta)e^{-\Delta/2} \leq Ce^{-\Delta/4}$ for a numerical constant C. Hence equation 171 yields

$$\frac{F\big(s_m(\omega)\big)}{F(s)} \leq C\exp\Big(-\frac{s_m(\omega)^2 - s^2}{4}\Big).$$

In particular, for $m = 1$ one has $s_1(\omega) \geq (1+\varepsilon)\pi/\lambda$ and $s \leq (1-\varepsilon)\pi/\lambda$, so

$$s_1(\omega)^2 - s^2 \geq \Big((1+\varepsilon)^2 - (1-\varepsilon)^2\Big)\frac{\pi^2}{\lambda^2} = \frac{4\varepsilon\pi^2}{\lambda^2}.$$

Summing equation 170 over $m \neq 0$ and absorbing the geometric remainder into a constant gives

$$\sup_{|\omega|\leq(1-\varepsilon)\pi/h} |1 - M_h(\omega)| \leq C_{\mathrm{alias}}(\varepsilon)\exp\Big(-\frac{c_{\mathrm{alias}}(\varepsilon)}{\lambda^2}\Big), \qquad c_{\mathrm{alias}}(\varepsilon) \asymp \varepsilon\pi^2, \tag{172}$$

so K2 holds with $p_K = 2$.

K4 (spatial localization envelope). Since $|K(x)| \le (2+x^2)\phi(x)$ and ϕ decays faster than any exponential, there exist constants $A_G > 0$ and $a_G > 0$ such that

$$|K(x)| \le A_G e^{-a_G|x|}, \qquad x \in \mathbb{R}. \tag{173}$$

For instance, one may take $a_G = 1$ and

$$A_G := \sup_{x\in\mathbb{R}} |K(x)|e^{|x|} < \infty.$$

After scaling,

$$|K_\gamma(x)| = \gamma|K(\gamma x)| \le A_G\,\gamma e^{-a_G\gamma|x|},$$

so K4 holds.

K3 (zero-free strip margin) via diagonal dominance. Using the spatial representation of the normalizer,

$$\widetilde{D}_\lambda(\theta) = \lambda\sum_{k\in\mathbb{Z}} K(\lambda k)e^{-ik\theta},$$

write $\theta = \theta_r + i\theta_i$. For $|\theta_i| \le \sigma$, one has $|e^{-ik\theta}| \le e^{\sigma|k|}$, hence

$$|\widetilde{D}_\lambda(\theta)| \ge \lambda|K(0)| - \lambda\sum_{k\in\mathbb{Z}\setminus\{0\}} |K(\lambda k)|e^{\sigma|k|} = \lambda K(0) - 2\lambda\sum_{k=1}^{\infty}|K(\lambda k)|e^{\sigma k}. \tag{174}$$

Using equation 173, $|K(\lambda k)|e^{\sigma k} \le A_G e^{-(a_G\lambda-\sigma)k}$, so whenever $a_G\lambda > \sigma$,

$$|\widetilde{D}_\lambda(\theta)| \ge \lambda K(0) - 2\lambda A_G\,\frac{e^{-(a_G\lambda-\sigma)}}{1-e^{-(a_G\lambda-\sigma)}} = \lambda K(0)\big(1-q_G(\sigma,\lambda)\big), \tag{175}$$

where

$$q_G(\sigma,\lambda) := \frac{2A_G}{K(0)}\,\frac{e^{-(a_G\lambda-\sigma)}}{1-e^{-(a_G\lambda-\sigma)}}. \tag{176}$$

In particular, if $q_G(\sigma,\lambda) < 1$, then K3 holds with

$$d_0(\sigma,\lambda) \ge \lambda K(0)\big(1-q_G(\sigma,\lambda)\big) > 0. \tag{177}$$

C.6.3 Kernel verification for the normalized Swish″ family

We verify Assumptions K1 –K5 for the Swish″ profile. Let

$$\varsigma(x) := (1+e^{-x})^{-1}, \qquad \mathrm{swish}(x) := x\,\varsigma(x), \tag{178}$$

and define

$$K(x) := \mathrm{swish}''(x) = \varsigma(x)\big(1-\varsigma(x)\big)\Big(2 + x\big(1-2\varsigma(x)\big)\Big). \tag{179}$$

Using

$$\varsigma(x)\big(1-\varsigma(x)\big) = \tfrac{1}{4}\operatorname{sech}^2(x/2), \qquad 1-2\varsigma(x) = -\tanh(x/2),$$

this admits the even closed form

$$K(x) = \frac{1}{4}\Big(2 - x\tanh(x/2)\Big)\operatorname{sech}^2(x/2), \qquad K(x) = K(-x). \tag{180}$$

K1 (mass). Since $K = \mathrm{swish}''$ and

$$\mathrm{swish}'(x) = \varsigma(x) + x\varsigma(x)\big(1-\varsigma(x)\big)$$

satisfies

$$\lim_{x\to+\infty}\mathrm{swish}'(x) = 1, \qquad \lim_{x\to-\infty}\mathrm{swish}'(x) = 0,$$

we have

$$\widehat{K}(0) = \int_{\mathbb{R}} K(x)\,dx = \mathrm{swish}'(+\infty) - \mathrm{swish}'(-\infty) = 1, \tag{181}$$

so K1 holds.

Closed-form Fourier transform.

$$\widehat{K}(\omega) = (\pi\omega)^2 \coth(\pi\omega)\,\operatorname{csch}(\pi\omega) = (\pi\omega)^2\,\frac{\cosh(\pi\omega)}{\sinh^2(\pi\omega)}, \tag{182}$$

with the continuous extension $\widehat{K}(0) = 1$. Moreover, $\widehat{K}$ has exponential tails:

$$\widehat{K}(\omega) \lesssim (1+\omega^2)e^{-\pi|\omega|} \qquad \text{as } |\omega| \to \infty.$$

K5 (Fourier positivity). From equation 182, for $\omega \neq 0$ we may rewrite

$$\widehat{K}(\omega) = (\pi\omega)^2\,\frac{\cosh(\pi\omega)}{\sinh^2(\pi\omega)}.$$

Here $(\pi\omega)^2 > 0$, $\cosh(\pi\omega) > 0$, and $\sinh^2(\pi\omega) > 0$, so $\widehat{K}(\omega) > 0$ for every $\omega \neq 0$. At $\omega = 0$, the continuous extension gives $\widehat{K}(0) = 1$. Therefore

$$\widehat{K}(\omega) \geq 0, \qquad \omega \in \mathbb{R}, \tag{183}$$

and K5 holds.

K2 (interior-band aliasing) with $p_K = 1$. Since K5 holds,

$$|1 - M_h(\omega)| = \frac{\sum_{m\in\mathbb{Z}\setminus\{0\}} \widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)}{D_h(\omega)} \leq \sum_{m\in\mathbb{Z}\setminus\{0\}} \frac{\widehat{K}_\gamma\big(\omega + \frac{2\pi m}{h}\big)}{\widehat{K}_\gamma(\omega)}. \tag{184}$$

Using the exponential tail in equation 182, one obtains an interior-band alias-gap bound of the form

$$\sup_{|\omega|\leq(1-\varepsilon)\pi/h} |1 - M_h(\omega)| \leq C_{\mathrm{alias}}(\varepsilon)\exp\Big(-\frac{c_{\mathrm{alias}}(\varepsilon)}{\lambda}\Big), \qquad \lambda = \gamma h, \tag{185}$$

so K2 holds with $p_K = 1$. One convenient way to see equation 185 is to bound the ratio

$$\frac{\widehat{K}(t+\delta)}{\widehat{K}(t)} \leq Ce^{-\pi\delta} \qquad (t, \delta \geq 0),$$

using

$$\sinh(u+\delta) \geq \tfrac{1}{2}e^{\delta}\sinh(u), \qquad \cosh(u+\delta) \leq e^{\delta}\cosh(u), \qquad u, \delta \geq 0,$$

and then proceed exactly as in the sech^2 proof.

K4 (spatial localization envelope). From equation 180, $|\tanh(x/2)| \leq 1$ and $\operatorname{sech}^2(x/2) \leq 4e^{-|x|}$ give

$$|K(x)| \leq \frac{1}{4}(2+|x|)\,4e^{-|x|} = (2+|x|)e^{-|x|} \leq 2e^{-|x|/2},$$

so $|K(x)| \leq A_S e^{-a_S|x|}$ holds with $(A_S, a_S) = (2, 1/2)$. After scaling,

$$|K_\gamma(x)| = \gamma|K(\gamma x)| \leq A_S\gamma e^{-a_S\gamma|x|},$$

so K4 holds.

K3 (zero-free strip margin) via diagonal dominance. Using again

$$\widetilde{D}_\lambda(\theta) = \lambda\sum_{k\in\mathbb{Z}} K(\lambda k)e^{-ik\theta}$$

and $|e^{-ik\theta}| \leq e^{\sigma|k|}$ for $|\Im\theta| \leq \sigma$, we have

$$|\widetilde{D}_\lambda(\theta)| \geq \lambda K(0) - 2\lambda\sum_{k=1}^{\infty} |K(\lambda k)|e^{\sigma k}.$$

With the exponential envelope $|K(x)| \leq A_S e^{-a_S|x|}$, $|K(\lambda k)|e^{\sigma k} \leq A_S e^{-(a_S\lambda-\sigma)k}$, hence for $a_S\lambda > \sigma$,

$$|\widetilde{D}_\lambda(\theta)| \geq \lambda K(0) - 2\lambda A_S\,\frac{e^{-(a_S\lambda-\sigma)}}{1-e^{-(a_S\lambda-\sigma)}}. \tag{186}$$

Since $K(0) = \operatorname{swish}''(0) = 1/2$, equation 186 yields an explicit sufficient condition for K3.