Functional Multi-Reference Alignment:
A Framework with Fourier Neural Operator and Bispectrum-Guided Diffusion Posteriors

Zixuan Qin

Department of Statistics, University of Chicago

December 8, 2025

1 Introduction

1.1 Problem Overview

In this paper, we study the functional Multi-Reference Alignment (MRA) problem, which seeks to recover an unknown signal f from N independent, noisy, and randomly shifted observations. This problem serves as a simplified mathematical model for Cryo-Electron Microscopy (Cryo-EM), retaining the core challenges of high noise and latent variables while assuming the signal and observations lie in the same domain [3]. As illustrated in Figure 1, the observation process involves a random shift ζ followed by additive noise η.
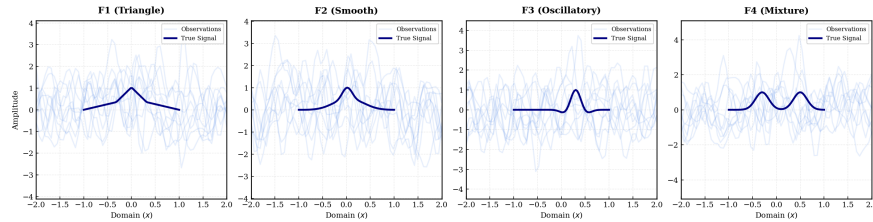

Figure 1 Illustration of the MRA problem.

1.2 Related Work

Standard approaches for MRA include synchronization, which estimates shifts prior to averaging [11], and Expectation-Maximization (EM), which iteratively refines the signal and shift distribution [10]. However, synchronization fails in low signal-to-noise ratio (SNR) regimes, while EM is computationally expensive and sensitive to initialization. Alternatively, invariant feature methods recover the signal from shift-invariant statistics (e.g., bispectrum) [2], but the inversion is often numerically unstable. Our paper builds on this recent paper [1], which framed MRA as a functional deconvolution problem.

The Fourier Neural Operator (FNO) [7] has emerged as a powerful tool for learning operators between infinite-dimensional function spaces, primarily in solving Partial Differential Equations (PDEs). While FNO has been applied to inverse problems involving PDEs [9], its potential for solving the deconvolution operator in the context of functional MRA remains underexplored.

Score-based diffusion models [12] have recently been adopted for sovling inverse problems [4]. In the context of MRA, [6] proposed a diffusion-based solver. However, their method conditions only on the power-spectrum likelihood, which lacks phase information and relies on strong prior assumptions, limiting its effectiveness to simple synthetic settings.

1.3 Contributions

We propose a novel two-stage framework for functional MRA that couples operator learning with diffusion posteriors:

- **Global Structure via Operator Learning:** We interpret the MRA inverse problem as learning a deconvolution operator. We employ FNO to map the first moment of noisy observations directly to a coarse estimate of f. This step effectively aligns and denoises the signal.

- **High-Frequency Recovery via Bispectrum-Guided Diffusion:** To recover fine details lost in the first stage, we introduce a diffusion posterior sampling scheme. We utilize a bispectrum-based gradient to guide the reverse process, which provides phase information invariant to shifts.

- **Efficiency:** The entire pipeline avoids explicit shift estimation or iterative likelihood maximization, and runs in $O(NlogN)$ time through FFT-based operations.

2   Problem Setting

Our problem setting largely follows the functional MRA formulation proposed in [1], with a few modifications described below. Let $D := [-2, 2]^d$ and let $f : D \to \mathbb{R}$ denote a square-integrable signal supported on $[-1, 1]^d$. The goal is to recover f from N independent observations $\{y_n\}_{n=1}^N$ generated as

$$y_n(x) = f(x - \zeta_n) + \eta_n(x), \qquad x \in D, \quad n = 1, \dots, N. \tag{1}$$

Here, $\{\zeta_n\}_{n=1}^N$ are independent copies of a centered random vector ζ with density p_ζ supported on $[-1, 1]^d$, and $\{\eta_n\}_{n=1}^N$ are independent copies of a centered Gaussian process η defined on D with unknown covariance function k_η. Each observation y_n therefore represents a noisy and randomly shifted version of the underlying signal f. We assume that the shifts $\{\zeta_n\}$ and the noise processes $\{\eta_n\}$ are all independent, and we write $y(x) = f(x - \zeta) + \eta(x)$ to denote a generic observation.

We do not assume periodicity of the signal f or cyclic shifts; due to the compact support of both f and the shifts, all observations are supported in the domain D. Unlike [1], our framework does not require access to the covariance function k_η of the noise process.

3   Methods

Our framework recovers the signal f through a coarse-to-fine pipeline. As illustrated in Figure 2, the process begins by aggregating noisy observations into empirical moments, which are mapped by a FNO (Sec. 3.1) to a low-frequency approximation of the signal. This coarse estimate then serves as a structural anchor for a diffusion-based refinement stage (Sec. 3.2), where high-frequency details are restored using bispectrum guidance.

3.1   Coarse Estimation via FNO

The FNO [7] learns an operator between function spaces by parameterizing the integral kernel in the Fourier domain. An FNO block consists of a Fourier transform $\mathcal{F}$, a linear transformation R applied to the lower Fourier modes, and an inverse Fourier transform
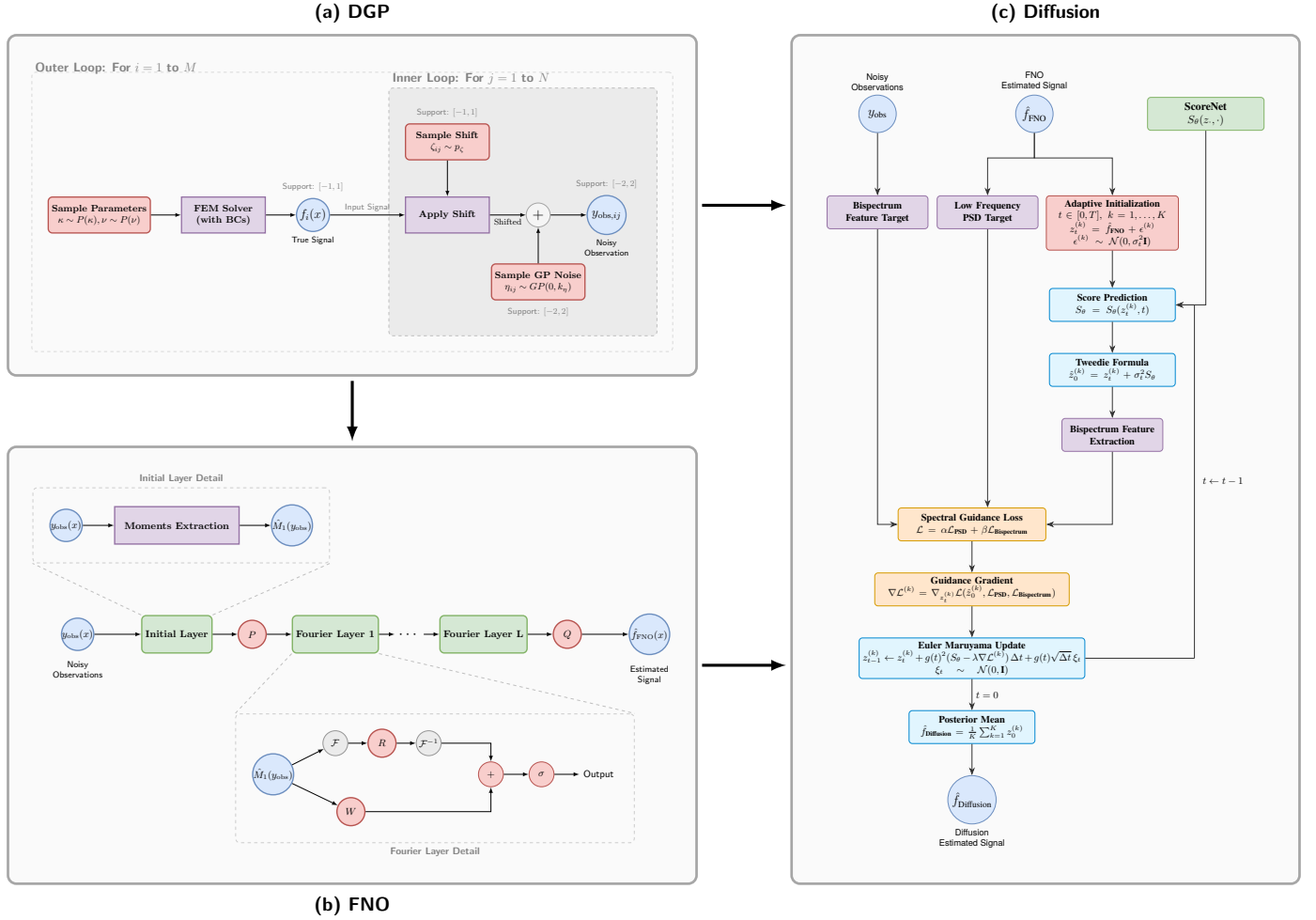
Figure 2 Overview of the proposed two-stage framework. **(a) Data Generation Process (DGP). (b) Stage I: Coarse Estimation. (c) Stage II: High-Frequency Recovery.**

$\mathcal{F}^{-1}$. Letting $v_l(x)$ denote the hidden feature representation at layer l, the update rule is given by:

$$v_{l+1}(x) = \sigma(W v_l(x) + (\mathcal{K} v_l)(x)), \quad (\mathcal{K} v_l)(x) = \mathcal{F}^{-1}(R \cdot \mathcal{F}(v_l))(x), \qquad (2)$$

where σ is a non-linear activation function and W is a linear transformation. A key advantage of FNO is that its spectral parameterization inherently imposes a convolutional structure, providing translation equivariance. This makes it a natural candidate for shift-related problems.

The Initialization Layer. Standard FNO architectures typically map a batch of M input functions to M output functions. However, the MRA problem introduces an additional dimension: for each of the M signals, we have a set of N noisy observations.

Consequently, the raw input data forms a tensor of size $M \times N \times G$, where G denotes the spatial grid resolution (i.e., the number of mesh points). Directly processing this tensor incurs a prohibitive $O(M \cdot N \cdot G)$ complexity, becoming intractable in 2D due to quadratic grid growth.

To address this, we introduce a initialization layer that aggregates the observations into their first moment. Consider the expectation of the observation model in Eq. (1):

$$\mathbb{E}[y(x)] = \mathbb{E}[f(x - \zeta)] + \mathbb{E}[\eta(x)] = (f * p_\zeta)(x). \tag{3}$$

The expected observation is the convolution of the signal f with the shift density p_ζ. Consequently, the task of recovering f from $\mathbb{E}[y]$ becomes a deconvolution problem.

We define the FNO input as the mean $\bar{y}(x) = \frac{1}{N} \sum_{n=1}^{N} y_n(x)$. This aggregation collapses the observation dimension, effectively reducing the total dataset size from $M \times N \times G$ to $M \times G$. The FNO is tasked with approximating the deconvolution operator $\mathcal{D}$ defined by $\mathcal{D}(\bar{y}) \approx f$. Furthermore, since $\bar{y}$ can be computed via online updates [13], the peak memory usage is limited to $O(B \cdot G)$, where B is the batch size. As $N \to \infty$, the Law of Large Numbers (LLN) ensures $\bar{y} \to f * p_\zeta$, naturally improving the SNR for $\mathcal{D}$ without incurring additional computational costs.

Limitations. While efficient, this approach has a drawback: convolution with p_ζ acts as a low-pass filter, attenuating high-frequency information. Furthermore, FNO truncate Fourier modes above a threshold ω_{max}. Consequently, the output $\hat{f}_{\text{FNO}}$ captures the global structure and low-frequency components of f but tends to be over-smoothed. This motivates our second stage.

3.2 High-Frequency Recovery via Bispectrum-Guided Diffusion

To recover the missing high-frequency details, we utilize a score-based generative model [12] as a prior. We adopt the Variance Exploding (VE) SDE formulation with a geometric noise schedule. Let $g(t) = \sqrt{\frac{d[\sigma^2(t)]}{dt}}$ be the diffusion coefficient. The forward SDE is $dz_t = g(t)dw_t$, where w_t denotes the standard Wiener process. The reverse-time SDE is given by:

$$dz_t = \left[-g(t)^2 \nabla_{z_t} \log p_t(z_t)\right]dt + g(t)dw_t, \tag{4}$$

where $\nabla_{z_t} \log p_t(z_t)$ is the score function of the perturbed data distribution p_t, approximated by a time-dependent neural network $S_\theta(z, t)$.

Spectral Feature Guidance. We implement posterior sampling by guiding the reverse diffusion process with two shift-invariant spectral features. This strategy enhances high-frequency recovery without the computationally expensive explicit orbit alignment:

1. **Low-Frequency Constraint (PSD):** Leveraging the inherent capability of FNO to resolve global structures, we constrain the diffusion trajectory to match the PSD of $\hat{f}_{\text{FNO}}$ for modes $\omega \leq \omega_{anchor}$. This constraint acts as a spectral regularization, stabilizing the energy profile of the generated samples during the reverse process.

2. **Phase Recovery (Bispectrum):** To recover the phase information lost in the PSD, we utilize the bispectrum [2]. Let $\tilde{z}$ denote the Fourier transform of z, and let $\tilde{z}^*$ denote the complex conjugate of $\tilde{z}$. The bispectrum is defined as $\mathcal{B}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \tilde{z}(\boldsymbol{\omega}_1)\tilde{z}(\boldsymbol{\omega}_2)\tilde{z}^*(\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2)$. For high-dimensional signals (e.g., 2D images), the full bispectrum scales as $O(G^2)$, leading to prohibitive memory and computational costs due to the dimensionality explosion. Instead, we compute only two specific subsets enough for phase retrieval:

- **Chain**: $\mathcal{B}_{chain}(\boldsymbol{\omega}) = \tilde{z}(\mathbf{1})\tilde{z}(\boldsymbol{\omega})\tilde{z}^*(\boldsymbol{\omega}+\mathbf{1})$. This term recursively links the phase of frequency $\boldsymbol{\omega} + \mathbf{1}$ to $\boldsymbol{\omega}$ via the fundamental frequency.
- **Diagonal**: $\mathcal{B}_{diag}(\boldsymbol{\omega}) = \tilde{z}(\boldsymbol{\omega})^2\tilde{z}^*(2\boldsymbol{\omega})$. This term constrains the harmonic relationships by linking frequency $2\boldsymbol{\omega}$ to $\boldsymbol{\omega}$.

Restricting computation to these subsets reduces both the computational and memory complexity from $O(G^2)$ to $O(G)$.

Tweedie-Stabilized Guidance. Evaluating bispectrum gradients directly on the noisy state z_t is numerically unstable. We instead apply Tweedie's formula [5] to estimate the denoised signal expectation $\hat{z}_0(z_t) = z_t + \sigma_t^2 S_\theta(z_t, t)$. Crucially, since $\hat{z}_0$ is fully differentiable with respect to z_t (backpropagating through S_θ), the gradients of the spectral loss can be transmitted to guide the sampling trajectory. To avoid scaling discrepancies, we define the total guidance loss $\mathcal{L}(z_t)$ as a weighted combination of a low-frequency PSD constraint and a normalized bispectrum phase loss:

$$\mathcal{L}(z_t) = \alpha \mathcal{L}_{\text{PSD}} + \beta \mathcal{L}_{\text{Bispectrum}}, \tag{5}$$

where the components are explicitly defined as:

$$\mathcal{L}_{\text{PSD}} = \|\text{PSD}(\hat{z}_0)_{\text{low}} - \text{PSD}(\hat{f}_{\text{FNO}})_{\text{low}}\|^2, \tag{6}$$

$$\mathcal{L}_{\text{Bispectrum}} = \sum_{j \in \{\text{chain}, \text{diag}\}} \left\| \frac{\mathcal{B}_j(\hat{z}_0)}{|\mathcal{B}_j(\hat{z}_0)|} - \frac{\mathcal{B}_j(\mathcal{Y})}{|\mathcal{B}_j(\mathcal{Y})|} \right\|^2. \tag{7}$$

Here, $\mathcal{Y}$ denotes the observed data statistics, and α, β are hyperparameters balancing the spectral constraints. Note that this refinement stage is performed during inference. Posterior samples are drawn one signal at a time, restricting the active memory requirement for $\mathcal{Y}$ to $O(N \cdot G)$ and effectively avoiding the dimensionality bottleneck.

Adaptive Initialization. To optimize efficiency, we adapt the refinement intensity based on the spectral quality of $\hat{f}_{\text{FNO}}$. We quantify this by fitting a power-law decay rate γ to its PSD (i.e., $S(\omega) \propto \omega^{-\gamma}$). The starting time is derived via a linear mapping:

$$t_{start}(\gamma) = t_{max} + \frac{\gamma - \gamma_{min}}{\gamma_{max} - \gamma_{min}}(t_{min} - t_{max}), \tag{8}$$

clipped to $[t_{min}, t_{max}]$. A larger γ (smoother signal) maps to a smaller t_{start} to preserve the FNO's global structure, while smaller γ allows for deeper refinement. Finally, we initialize $z_{t_{start}} \sim \mathcal{N}(\hat{f}_{\text{FNO}}, \sigma_{t_{start}}^2 \mathbf{I})$ and solve the reverse SDE.

Algorithm 3.1 Coarse-to-Fine Reconstruction

1: **Input**: Noisy observations $\mathcal{Y} = \{y_n\}_{n=1}^{N}$, trained FNO $\mathcal{A}_\varphi$, ScoreNet S_θ, particles K, anchors ω_{anchor}, guidance scale λ.

2: Compute empirical mean: $\bar{y} \leftarrow \frac{1}{N} \sum_{n=1}^{N} y_n$

3: FNO Prediction: $\widehat{f}_{\text{FNO}} \leftarrow \mathcal{A}_\varphi(\bar{y})$

4: Target Extraction:

5:     $\mathcal{L}_{\text{PSD}} \leftarrow \text{PSD}(\widehat{f}_{\text{FNO}})[0 : \omega_{anchor}]$

6:     $\mathcal{L}_{\text{Bispectrum}} \leftarrow \text{Bispec}(\mathcal{Y})_{\text{chain, diag}}$

7: Adaptive Initialization:

8:     Fit spectral decay rate γ where $\text{PSD}(\widehat{f}_{\text{FNO}}) \propto \omega^{-\gamma}$

9:     Compute $t_{start}(\gamma)$ via linear mapping

10:     Initialize particles $z_{t_{start}}^{(k)} \sim \mathcal{N}(\widehat{f}_{\text{FNO}}, \sigma_{t_{start}}^2 \mathbf{I})$ for $k = 1 \dots K$

11: Refinement Loop:

12: **for** t from t_{start} down to 0 **do**

13:     **for** $k = 1$ to K **do**

14:         Estimate clean signal (Tweedie): $\widehat{z}_0^{(k)} \leftarrow z_t^{(k)} + \sigma_t^2 S_\theta(z_t^{(k)}, t)$

15:         Compute guidance gradient: $\nabla\mathcal{L}^{(k)} \leftarrow \nabla_{z_t^{(k)}} \mathcal{L}(\widehat{z}_0^{(k)}, \mathcal{L}_{\text{PSD}}, \mathcal{L}_{\text{Bispectrum}})$

16:         Update via Euler-Maruyama:

17:         $z_{t-1}^{(k)} \leftarrow z_t^{(k)} + g(t)^2 \Big(S_\theta(z_t^{(k)}, t) - \lambda\nabla\mathcal{L}^{(k)} \Big)\Delta t + g(t)\sqrt{\Delta t}\xi_t$

18:     **end for**

19: **end for**

20: Aggregate: $\widehat{f}_{\text{Diffusion}} \leftarrow \frac{1}{K} \sum_{k=1}^{K} z_0^{(k)}$

21: **Output**: Posterior Mean $\widehat{f}_{\text{Diffusion}}$.

Posterior Sampling. The refinement process executes the guided reverse SDE for K parallel particles $\{z_t^{(k)}\}_{k=1}^{K}$ from t_{start} to 0. We employ the Euler-Maruyama discretization, updating each particle via:

$$z_{t-1}^{(k)} \leftarrow z_t^{(k)} + g(t)^2 \Big(S_\theta(z_t^{(k)}, t) - \nabla\mathcal{L}(z_t^{(k)}) \Big)\Delta t + g(t)\sqrt{\Delta t}\xi_t. \tag{9}$$

Explicit alignment is unnecessary as the particles inherit the aligned orientation from the initialization $\widehat{f}_{\text{FNO}}$. Consequently, we directly aggregate the samples via the posterior mean. Additionally, the empirical variance across the particles naturally quantifies the estimation uncertainty. Integrating this refinement process with the coarse estimation from Sec. 3.1, the complete two-stage framework is summarized in Algorithm 3.1.

3.3 Data Generation via SPDEs

To train our models and evaluate performance, we require a diverse set of ground-truth signals with controllable geometric properties. Instead of relying on fixed datasets, we adopt the Stochastic Partial Differential Equation (SPDE) approach [8]. We generate the signal $f(x)$ as a stationary solution to the following equation:

$$(\kappa^2 - \Delta)^{\nu/2}(\tau f(x)) = \mathcal{W}(x), \quad x \in [-1, 1]^d, \tag{10}$$

Algorithm 3.2 SPDE-based Data Generation Process

1: **Input**: Number of signals M, observations per signal N, spatial grid resolution G, hyperparameter priors $P(\kappa), P(\tau), P(\nu)$, shift density p_ζ, noise covariance k_η.
2: Initialize FEM triangulation mesh $\mathcal{T}$ on domain $[-1, 1]^d$ with resolution G.
3: Initialize dataset $\mathcal{Y} \leftarrow \emptyset$.
4: **for** $i = 1$ to M **do**
5:     Sample hyperparameters: $\kappa_i \sim P(\kappa), \tau_i \sim P(\tau), \nu_i \sim P(\nu)$.
6:     Construct sparse precision matrix $\mathbf{Q}_i$ via FEM discretization of $(\kappa_i^2 - \Delta)^{\nu_i/2}$.
7:     Sample ground truth signal: $f_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i^{-1})$.
8:     **for** $j = 1$ to N **do**
9:         Sample random shift: $\zeta_{i,j} \sim p_\zeta$.
10:        Apply shift: $f_{i,j}^{\text{shifted}}(x) \leftarrow f_i(x - \zeta_{i,j})$.
11:        Sample noise: $\eta_{i,j} \sim \mathcal{N}(\mathbf{0}, k_\eta \mathbf{I})$.
12:        Generate noisy observation: $y_{i,j} \leftarrow f_{i,j}^{\text{shifted}} + \eta_{i,j}$.
13:        Store observation: $\mathcal{Y} \leftarrow \mathcal{Y} \cup \{y_{i,j}\}$.
14:     **end for**
15: **end for**
16: **Output**: Dataset $\mathcal{Y}$.

---

where Δ is the Laplacian operator, $\mathcal{W}$ is Gaussian white noise, and the hyperparameters κ, τ, ν control the correlation length, variance, and smoothness, respectively.

A key theoretical advantage of this formulation is that the parameter ν explicitly determines the regularity of the generated signals. Specifically, the sample paths of f belong to the Sobolev space $H^s([-1, 1]^d)$ for $s < \nu - d/2$, allowing us to rigorously control the smoothness of the prior distribution.

Numerically, we solve this SPDE using the Finite Element Method (FEM). This discretization approximates the continuous field by a Gaussian Random Field (GRF) defined on a triangulation of the domain. The resulting precision matrix is sparse, which enables computationally efficient sampling and storage even for fine meshes, avoiding the prohibitive costs associated with dense covariance matrices.

Finally, we integrate this signal generator with the MRA observation model described in Sec. 2. For each sampled signal f, we simulate N noisy, shifted observations to form the training set. The complete data generation pipeline is detailed in Algorithm 3.2.

4 Numerical Experiments

We evaluate our framework on a 1D functional MRA problem. We assess both the convergence of the coarse FNO estimator with respect to sample size N and the ability of the diffusion posterior sampling to refine high-frequency details and quantify uncertainty.

4.1 Experimental Setup

Data Generation and Baselines. We adopt the domain $D = [-2, 2]$ with signals supported on $[-1, 1]$ with Dirichlet boundary condition, discretized on a grid of size $G = 64$. Observations are generated via Algorithm 3.2, where shifts $\zeta_n \sim \mathcal{U}[-1, 1]$ and the noise η_n is drawn from a Gaussian Process with a Gaussian kernel ($\sigma_\eta = 1.0, \lambda_\eta = 0.1$).

To train the FNO, we generate a dataset of 1,024 signals using Algorithm 3.2. The SPDE parameters are sampled from a mixture distribution to ensure diversity: smoothness $\nu \in \{0.5, 1.5, 2.5\}$ with probabilities $\{0.4, 0.3, 0.3\}$, and correlation length parameter κ scaled to $[1, 20]$ from a Beta$(5, 2)$ distribution. The test set consists of 256 signals generated from the same distribution.

Out-of-Distribution Evaluation. To rigorously test generalization, we additionally define four specific Compare Signals (f_1 to f_4) with distinct geometric properties, defined as follows (where h^{*n} denotes the n-fold convolution of h):

- **Triangle (f_1):** $h_1(x) = \mathbf{1}^{*2}_{[-.5,.5]}(x)$; $f_1(x) = h_1(x) + h_1(\pi x)$.

- **Smooth (f_2):** $h_2(x) = \mathbf{1}^{*4}_{[-.25,.25]}(x)$; $f_2(x) = h_2(x) + h_2(\pi x)$.

- **Oscillatory (f_3):** $h_3(x) = e^{-20x^2}\cos(8x)$; $f_3(x) = h_3(x - 0.3)$.

- **Mixture (f_4):** PDF of a Gaussian mixture $0.5\mathcal{N}(-.3, .02) + 0.5\mathcal{N}(.5, .02)$.

Model Architectures and Training. The **FNO** consists of 3 Fourier layers with $\omega_{max} = 16$ and a width of 32, and is optimized using the relative L_2 loss. It inputs the empirical mean of N observations. The **ScoreNet** adopts an 8-layer CNN architecture following [6] and is trained via the standard Mean Squared Error (MSE) objective. To prevent the generative model from biasing towards rougher samples, we train the ScoreNet on a dataset of 16,384 samples generated via Algorithm 3.2 (steps 1-7), but with $\kappa \sim \mathcal{U}[1, 20]$. For posterior sampling, we use guidance scale $\lambda = 5.0$ and simulate $K = 16$ particles.

4.2 Stage I: Coarse Estimation Scaling

We first investigate the asymptotic performance of the FNO (Algorithm 3.1 (steps 1-3)). We vary the number of observations N from 2^{10} to 2^{18}. Figure 3 illustrates the reconstruction of the four Compare Signals at $N = 2^{10}$ and $N = 2^{18}$. The FNO successfully recovers the global structure and supports. Figure 4 quantifies the error decay. Both relative L_2 and L_∞ errors decrease monotonically as N increases, validating that the empirical moment input effectively aggregates information via the LLN.

4.3 Stage II: Refinement and Uncertainty Quantification

We then apply the bispectrum-guided diffusion (Algorithm 3.1 steps 4-21) to refine the FNO estimates obtained at $N = 2^{18}$. Figure 5 compares the FNO initialization, the Diffusion posterior mean, and the Ground Truth for selected samples. The refinement stage enhances high-frequency details and corrects amplitude discrepancies.
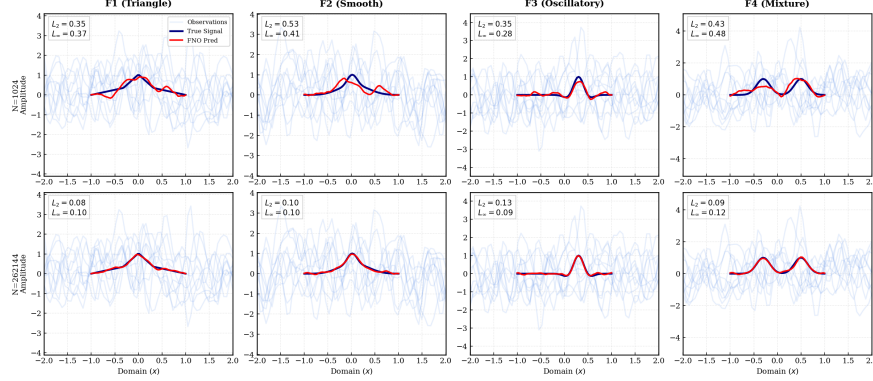
Figure 3 Stage I Results: FNO estimation vs. Ground Truth for the four specific compare signals at $N = 2^{10}$ and $N = 2^{18}$.
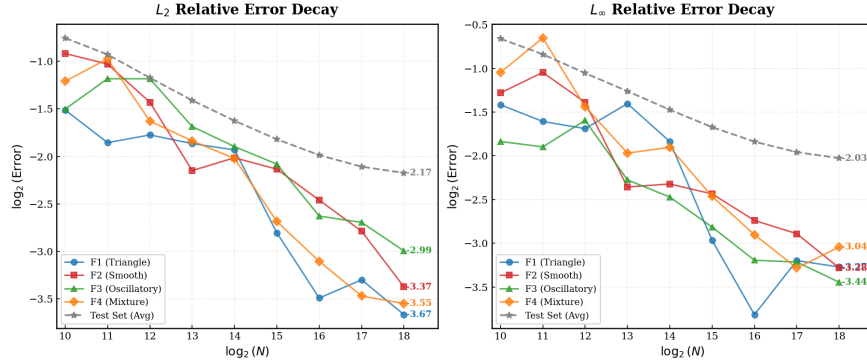

Figure 4 Error convergence of the FNO estimator with respect to sample size $N \in [2^{10}, 2^{18}]$. (a) Relative L_2 Error. (b) Relative L_∞ Error.

Table 4.1 presents the quantitative comparison. On the Compare Set, the diffusion refinement reduces the average L_2 error from 0.0966 to 0.0939 and L_∞ error from 0.1050 to 0.0884. A similar improvement is observed on the Test Set (32 samples)[1].

Finally, Figure 6 visualizes the Uncertainty Quantification (UQ) provided by the particle variance. We calculate the empirical coverage of the 95% Credible Interval across the evaluated samples. The actual coverage is 85.63%. While slightly under-covered compared to the target 95%, the uncertainty bands effectively capture the signal variability and potential reconstruction errors.

5 Discussion

Due to space limitations, we present only a subset of our results focusing on the 1D setting. Preliminary experiments indicate that the FNO initialization remains robust in

---

[1]Due to the high memory of storing observations at $N = 2^{18}$, we restrict evaluation to a subset (the first 32 samples) of the test set to ensure feasibility on a PC.

Table 4.1 Quantitative comparison of reconstruction errors (Relative L_2 and L_∞).

Method	Compare Set (4 samples)		Test Set (32 subset)	
	Avg L_2	Avg L_∞	Avg L_2	Avg L_∞
Stage I (FNO)	0.0966	0.1050	0.3174	0.3700
Stage II (Posterior)	**0.0939**	**0.0884**	**0.2879**	**0.3403**
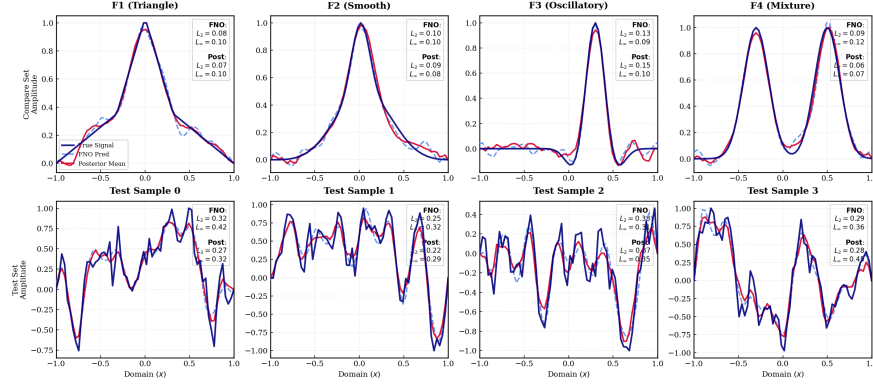
Figure 5 Visual comparison of Ground Truth, Stage I (FNO), and Stage II (Diffusion Posterior Mean) on selected samples from the Compare and Test sets.
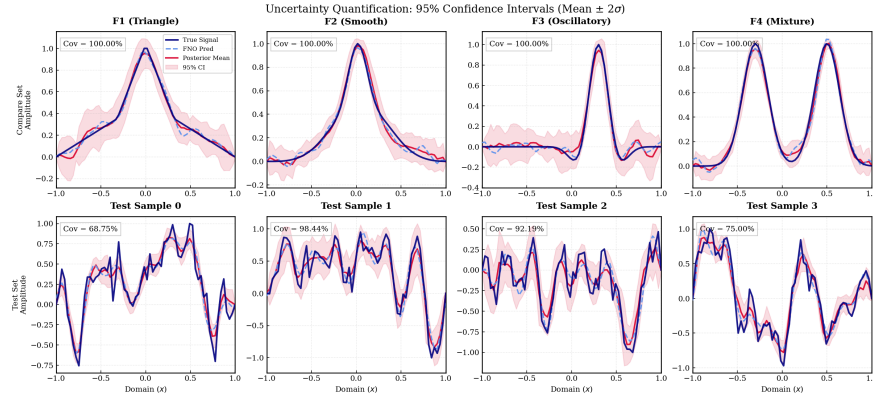
Figure 6 Uncertainty Quantification. The shaded regions represent the ± 2 std deviation intervals derived from $K = 16$ particles.

2D, while the extension of the full diffusion refinement pipeline to 2D is currently under active development. Regarding noise sensitivity, our empirical analysis suggests that the FNO estimator requires a sample size scaling of $N = O(\sigma_\eta^2)$ to maintain constant error, aligning with the theoretical convergence rate of the first moment.

This manuscript is currently a draft and will be continuously updated. The code and full experimental suite are available[2]. If citation is needed, a recommended BibTeX entry is provided in the README file of the repository.

---

[2] https://github.com/Zixuanqin1897/MRA_FNO_Diffusion.git

References

[1] O. Al-Ghattas, A. Little, D. Sanz-Alonso, and M. Sweeney. Functional multi-reference alignment via deconvolution, 2025. URL https://arxiv.org/abs/2506.12201.

[2] T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on signal processing*, 66(4):1037–1050, 2017.

[3] T. Bendory, A. Bartesaghi, and A. Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Processing Magazine*, 37(2):58–76, 2020. doi: 10.1109/MSP.2019.2957822.

[4] G. Daras, H. Chung, C.-H. Lai, Y. Mitsufuji, J. C. Ye, P. Milanfar, A. G. Dimakis, and M. Delbracio. A survey on diffusion models for inverse problems, 2024. URL https://arxiv.org/abs/2410.00083.

[5] B. Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

[6] A. Janson and J. Andén. Moment-based posterior sampling for multi-reference alignment, 2025. URL https://arxiv.org/abs/2510.12651.

[7] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

[8] F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, Sept. 2011. doi: 10.1111/j.1467-9868.2011.00777.x. URL https://doi.org/10.1111/j.1467-9868.2011.00777.x.

[9] D. Long, Z. Xu, Q. Yuan, Y. Yang, and S. Zhe. Invertible fourier neural operators for tackling both forward and inverse problems, 2025. URL https://arxiv.org/abs/2402.11722.

[10] S. H. Scheres. Relion: Implementation of a bayesian approach to cryo-em structure determination. *Journal of Structural Biology*, 180(3):519–530, 2012. ISSN 1047-8477. doi: https://doi.org/10.1016/j.jsb.2012.09.006. URL https://www.sciencedirect.com/science/article/pii/S1047847712002481.

[11] A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2010.02.001. URL https://www.sciencedirect.com/science/article/pii/S1063520310000205.

[12] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations, 2021. URL https://arxiv.org/abs/2011.13456.

[13] B. P. Welford. Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4(3):419–420, 1962. doi: 10.1080/00401706.1962. 10490022. URL https://www.tandfonline.com/doi/abs/10.1080/00401706. 1962.10490022.