ENERGY-GUIDED FUNCTIONAL FLOW MATCHING FOR SOLITON GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study energy-guided Functional Flow Matching for generating solutions of the Korteweg–de Vries equation. A Fourier Neural Operator parameterizes a velocity field that transports a Gaussian reference measure toward a distribution of soliton solutions. We consider both complete space–time trajectories and joint initial–final state representations. Under explicit measure-theoretic assumptions, we derive an exact guidance identity in function space and a tractable covariance-based approximation for sampling. On the reported two- and three-soliton forecasting benchmarks, FFM with initial-condition guidance achieves lower median trajectory errors and PDE residuals than our DiffusionPDE and FunDPS implementations. Additional invariant guidance reduces momentum and energy drift, while its effect on prediction error depends on the soliton regime. These results distinguish conservation from dynamical accuracy and support guided functional generation as a promising approach to dispersive PDEs. The experiments assess performance on a common grid.

1 INTRODUCTION

1.1 GENERATIVE MODELS

Generative models learn complex, high-dimensional distributions. Diffusion models do so through iterative denoising (Ho et al., 2020; Song et al., 2021; Karras et al., 2022), while GANs and VAEs use adversarial and latent-variable objectives, respectively (Goodfellow et al., 2014; Kingma & Welling, 2013). Flow Matching instead learns a time-dependent vector field that transports a simple reference distribution to the target distribution through a deterministic ODE (Lipman et al., 2023; Albergo et al., 2023).

1.2 INFINITE-DIMENSIONAL GENERATIVE MODELS

Most generative models are implemented on a fixed discretization, whereas PDE solutions are naturally functions. Treating observations only as vectors ties the learned representation to the training grid and provides no canonical interpretation under grid refinement. Function-space diffusion models address this issue by defining Gaussian measures and stochastic dynamics on Hilbert spaces (Pidstrigach et al., 2024; Kerrigan et al., 2023). Functional Flow Matching (FFM) similarly extends flow matching to real separable Hilbert spaces (Kerrigan et al., 2024). Parametrizing its velocity field by a Fourier Neural Operator (FNO) allows evaluation on multiple discretizations of the same underlying function (Li et al., 2021), becoming a natural fit for scientific problems with infinite-dimensional data.

1.3 GENERATIVE MODELS APPLIED TO PDES

Generative models have given rise to novel numerical schemes in the study of PDEs. In particular, recent works have employed diffusion models to PDE solving under partial observations, to stabilizing neural solvers (Lippe et al., 2023) and even to discover governing equations from observation data (Xu et al., 2025). In particular, DiffusionPDE (Huang et al., 2024) proposes a diffusion-based scheme to solve forward and inverse problems by learning the joint distribution of coefficients and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

solutions, guiding the denoising process with a modified version of the score function that incorporates observation consistency and PDE-residual terms, and applying its formulation to Darcy Flow, Poisson, Helmholtz, Burgers, and Navier-Stokes equations. However, DiffusionPDE is formulated for a fixed grid resolution rather than in function spaces where the PDE coefficients and solutions belong, tying both the model and its guidance mechanism to a single predefined discretization. FunDPS (Yao et al., 2025) addresses this gap with a diffusion framework that formulates the guidance mechanism directly in function spaces. FunDPS employs a U-shaped neural operator architecture trained under Gaussian Random Fields and provides a rigorous extension of Tweedie's formula to infinite-dimensions, achieving state-of-the-art results over fixed-resolution diffusion baselines. Nevertheless, flow matching schemes have proven to be more efficient during training and sampling, while providing better generalization results compared to diffusion models (Lipman et al., 2023). Hence, we hypothesize that a functional and discretization-invariant flow matching framework, with guidance formulated directly in function spaces may provide a suitable precise and efficient framework for the approximation and generation of PDE solutions by combining the computational advantages of flow matching with the flexibility of function space modeling.

1.4 Contributions

Our contributions are threefold. First, from a theoretical perspective, we formulate energy-guidance within the FFM framework in a Hilbert-space setting and derive Wasserstein-2 distance bounds for general FFM methods. We further develop a covariance-based approximation of the resulting guidance vector field. Second, we study trajectory and joint initial–final representations for KdV soliton generation. Third, we evaluate two guidance strategies within FFM: an observation-guidance, which incorporates general observational data, and invariance-guidance, which enforces the conserved quantities of the KdV equation. We compare both approaches with a DiffusionPDE and FunDPS baselines, assessing prediction accuracy, PDE residuals, and conservation drift.

2 Method

2.1 Korteweg-de Vries equation

We consider the Korteweg-de Vries equation (KdV) on the real line,

$$\partial_\tau u + \partial_x \left(\partial_{xx} u - \mu u^2\right) = 0, \qquad (\tau, x) \in \mathbb{R} \times \mathbb{R}, \tag{1}$$

which describes long-wavelength waves in dispersive media. For $\mu = -1$, equation 1 admits the 1-soliton solutions with scaling/speed parameter $c > 0$

$$u_{c,x_0}(\tau, x) = \frac{3}{2}\, c\, \operatorname{sech}^2\left(\frac{1}{2}\sqrt{c}\,(x - c\tau - x_0)\right), \quad c > 0, \tag{2}$$

representing a localized wave with amplitude proportional to c and propagation speed c.

Integrability and N-soliton. The KdV equation is completely integrable (Gardner et al., 1967). With the Hirota substitution $u = 6\,\partial_x^2 \log f$, the KdV equation becomes a bilinear equation solved exactly by $f = 1 + \sum_{n=1}^{N} \sum_{{}_N C_n} \left(\prod_{k<l}^{(n)} A_{i_k i_l}\right) \exp(s_{i_1} + \cdots + s_{i_n})$, with $s_j = \sqrt{c_j}(x - c_j\tau) - \delta_j$ and interaction coefficients $A_{ij} = \left(\frac{\sqrt{c_i} - \sqrt{c_j}}{\sqrt{c_i} + \sqrt{c_j}}\right)^2$ (Hirota, 1971).These closed form expressions provide exact multi-soliton references for the evaluation of nonlinear interaction dynamics. For further details, see Appendix A.2.

Invariants. The KdV dynamics conserves an infinite hierarchy of quantities. In particular, the first three invariants are

$$M = \int u \,\mathrm{d}x, \quad P = \int u^2 \,\mathrm{d}x, \quad H = \int \left(\tfrac{1}{3}u^3 - \tfrac{1}{2}(\partial_x u)^2\right) \mathrm{d}x, \tag{3}$$

which correspond to mass, momentum, and energy, respectively. Unlike standard image or video domains, these conservation laws offer a quantitative measure of physical fidelity.

2.2 FUNCTIONAL FLOW MATCHING

The FFM framework constructs a path of probability measures $(\mu_t)_{t\in[0,1]}$ that interpolates between a reference Gaussian measure μ_0 and the data distribution ν, so that $\mu_1 \approx \nu$. This path of measures is generated by a time-dependent vector field $v : [0,1] \times \mathcal{F} \to \mathcal{F}$, in the sense that the pair (v_t, μ_t) satisfies the continuity equation in the weak form,

$$\int_0^1 \int_{\mathcal{F}} \left(\partial_t \varphi(g,t) + \langle v_t(g), \nabla_g \varphi(g,t)\rangle\right) \mathrm{d}\mu_t(g)\,\mathrm{d}t = 0, \tag{4}$$

for all smooth cylindrical test functions φ. When a corresponding flow is well posed, the vector field induces a flow ϕ via the abstract ODE

$$\partial_t \phi_t(g) = v_t(\phi_t(g)), \quad \phi_0(g) = g, \tag{5}$$

and for all $t \in [0,1]$, the path of measures is defined by $\mu_t = [\phi_t]_{\#}\mu_0$.

Conditional construction. Each data sample $f \sim \nu$ has an associated conditional path $(\mu_t^f)_{t\in[0,1]}$, with $\mu_0^f = \mu_0$ and μ_1^f a narrow Gaussian centered at f, and defines the marginal measure $\mu_t(A) = \int_{\mathcal{F}} \mu_t^f(A)\,\mathrm{d}\nu(f)$. If each μ_t^f is generated by a conditional vector field v_t^f and $\mu_t^f \ll \mu_t$, then the marginal field

$$v_t(g) = \int_{\mathcal{F}} v_t^f(g)\,\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\nu(f) \tag{6}$$

generates $(\mu_t)_{t\in[0,1]}$ (Kerrigan et al., 2024).

Gaussian paths and OT parameterization. The absolute continuity condition $\mu_t^f \ll \mu_t$ is a fundamental requirement of the FFM framework. In the Gaussian setting, the Feldman–Hájek theorem provides conditions for this property. We therefore consider the parameterization $\mu_t^f = \mathcal{N}(m_t^f, (\sigma_t^f)^2 C_0)$, where C_0 is a fixed self-adjoint, non-negative, and trace-class covariance operator, and adopt the optimal transport (OT) parameterization

$$m_t^f = t\,f, \qquad \sigma_t^f = 1 - (1 - \sigma_{\min})\,t. \tag{7}$$

To construct the desired path of measures, we first define a conditional flow $\phi_t^f(g) = \sigma_t^f g + m_t^f$, designed to push a base distribution $\mathcal{N}(0, C_0)$ onto μ_t^f. Differentiating the conditional flow and applying the flow ODE gives the conditional vector field

$$v_t^f(g) = \frac{(\sigma_t^f)'}{\sigma_t^f}\,(g - m_t^f) + \frac{\mathrm{d}}{\mathrm{d}t} m_t^f. \tag{8}$$

Under equation 7, and the assumption that ν is supported on the Cameron–Martin space of C_0, the conditional measures are mutually absolutely continuous, so equation 6 is well posed.

Training. Directly regressing onto the marginal vector field v_t is computationally infeasible. Instead, we minimize the Conditional Flow Matching (CFM) functional loss, which differs from the marginal one only by a constant independent of the parameters θ (Kerrigan et al., 2024).

$$\mathcal{J}(\theta) = \mathbb{E}_{t\sim\mathcal{U}[0,1],\, f\sim\nu,\, g\sim\mu_t^f} \left\| v_t^f(g) - u_t(g \mid \theta)\right\|_{\mathcal{F}}^2. \tag{9}$$

We parameterize $u_t(\cdot \mid \theta)$ by a Fourier Neural Operator (Li et al., 2021), which acts in the functional space, and allows the learned vector field to be evaluated across discretizations.

Sampling. At inference time, sample generation is performed by solving the flow ODE using the learned vector field. Given an initial sample $g \sim \mu_0$, we solve the ODE $\partial_t \phi_t^\theta(g) = u_t(\phi_t^\theta(g) \mid \theta)$, over the time interval $[0,1]$ using a numerical integrator. The resulting terminal state, $g_1^\theta := \phi_1^\theta(g)$, yields the final generated output.

2.3 SCHEMES OF USE

We consider two different schemes for the PDE generative problem (Figure 1). Each scheme leads to a different formulation based on the chosen data representation.

Full trajectory generation. Samples are full space–time fields $u(\tau, x) \in \mathcal{F}$, where each trajectory represents an entire solution evolution over the time interval $[0, T]$.

Joint distribution generation. Samples represent the joint distribution of the initial condition and its terminal state, namely (u_0, u_T). The generative model learns the relationship between initial and final states, so that a sampled u_0 is accompanied by its predicted evolution u_T. Conditioning on either endpoint allows the model to address forward or inverse problems.

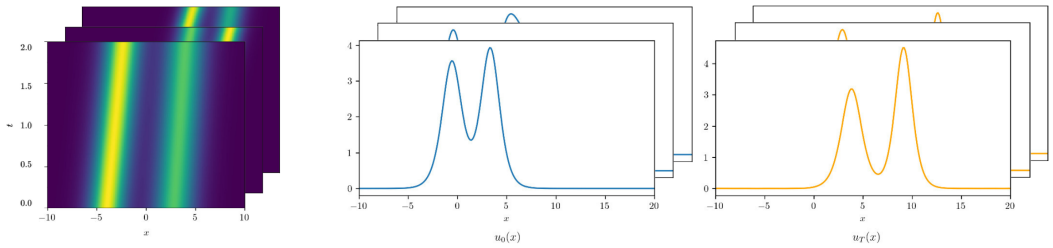

Figure 1: Data representations for KdV. *Left*: a space–time trajectory $u(\tau, x)$. *Right*: the joint distribution over initial-condition and solution pairs (u_0, u_T) used by the generative solver.

2.4 Energy-Guided Generation

Following (Feng et al., 2025), we consider a measurable energy functional $J : \mathcal{F} \to \mathbb{R}$ and define the modified target measure

$$\mathrm{d}\nu'(f) = \frac{1}{Z} e^{-J(f)}\, \mathrm{d}\nu(f), \qquad Z = \int_F e^{-J(f)}\, \mathrm{d}\nu(f) \in (0, \infty). \tag{10}$$

The following result provides an exact guidance field for the FFM framework.

Theorem 1 (General Guidance for FFM)**.** *Assume the hypotheses of Theorem 2. Also assume that the Radon–Nikodym derivatives $d\mu_t^f / d\mu_t$ admit a jointly measurable version in (t, f, g). Assume that $\int_0^1 \int_{\mathcal{F}\times\mathcal{F}} e^{-J(f)} \|v_t^f(g)\|\, \mathrm{d}\mu_t^f(g)\, \mathrm{d}\nu(f)\, \mathrm{d}t < \infty$. Adding the guidance term $\gamma_t(g)$ to $v_t(g)$ defines a modified vector field $v_t'(g)$ that generates the measure path in the weak continuity-equation sense $\mu_t'(A) := \int_{\mathcal{F}} \mu_t^f(A)\, \mathrm{d}\nu'(f)$, $A \in \mathcal{B}(\mathcal{F})$, provided that $\gamma_t(g)$ is given by*

$$\gamma_t(g) := \int_{\mathcal{F}} \left(\frac{e^{-J(f)}}{Z_t(g)} - 1 \right) v_t^f(g) \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}\, \mathrm{d}\nu(f), \tag{11}$$

where $Z_t(g) := \int_{\mathcal{F}} e^{-J(f)} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f)$.

Although this expression is exact, its evaluation requires integration with respect to the posterior measure $\nu_t(\cdot \mid g)$, which is generally intractable. To obtain a computationally tractable approximation, define the conditional mean

$$\hat{g}_1(g) := \mathbb{E}_{f\sim\nu_t(\cdot|g)}[f], \text{ where } \nu_t(\mathrm{d}f \mid g) := \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f). \tag{12}$$

Using equation 8 it can be shown that

$$\hat{g}_1(g) = \sigma_t^f v_t(g) + (1 - \sigma_{\min})g. \tag{13}$$

Assume that J is Fréchet differentiable at $\hat{g}_1(g)$. We use local first-order expansions of both J and e^{-J} around this mean to define a covariance-based guidance approximation. Controlling its error would additionally require bounds on the posterior-weighted Taylor remainders and the normalizing factor. Then,

$$J(f) \approx J(\hat{g}_1(g)) + \langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g) \rangle\,. \tag{14}$$

This gives the covariance-based approximation. Consequently, the exact guidance field can be approximated by

$$\gamma_t^{\mathrm{cov}}(g) = -\frac{1}{\sigma_t^f} \Sigma_{1|t}(g)[\nabla J(\hat{g}_1(g))], \tag{15}$$

where $\Sigma_{1|t}(g)[h] := \mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle h, f-\hat{g}_1(g)\rangle(f-\hat{g}_1(g))\Big]$ denotes the conditional covariance operator. The infinite-dimensional analog of the Jacobian trick relates this covariance operator to the derivative of $\hat{g}_1$. Under the hypotheses of Lemma 9, for $t>0$ and $g\in G_t$, assume additionally that $\hat{g}_1$ is Fréchet differentiable at g. Then,

$$\Sigma_{1|t}(g) = \frac{(\sigma_t^f)^2}{t} D\hat{g}_1(g) C_0. \tag{16}$$

Substituting this identity into the covariance-based guidance, and using the expression of $\hat{g}_1$ in terms of v_t, yields the infinite-dimensional analogue of the $g^{\text{cov-}G}$ guidance (Feng et al., 2025):

$$\gamma_t^{\text{cov-G}}(g) = -\frac{\sigma_t^f}{t}\Big(\sigma_t^f Dv_t(g) + (1-\sigma_{\min})I\Big) C_0[\nabla J(\hat{g}_1(g))]. \tag{17}$$

For a detailed derivation of these results, see Appendix D.

2.5 Guidance strategies

The observation and invariant penalties in this section are evaluated on the finite grids used in the experiments. Their finite-dimensional domains are denoted by $\mathcal{F}_N$ and $\mathcal{Y}_N$. The continuum results instead apply to functionals on $\mathcal{F}$ satisfying the stated measurability and differentiability assumptions. We do not claim that pointwise observations or the KdV energy are differentiable on arbitrary L^2 functions, or that the discretized guided sampler converges to a continuum guided flow.

Observation-Guidance Although the FFM model is trained unconditionally, forecasting can be performed by conditioning the sampling dynamics on partial or noisy observations of the target function $f\sim\nu$. Let $\mathcal{O}:\mathcal{F}\to\mathcal{Y}$ be an observation operator into a Hilbert space $\mathcal{Y}$, and let $y\in\mathcal{Y}$ be the observed data, $y\approx\mathcal{O}(f)$. The observation enters through the energy functional

$$J_{\text{obs}}(f) = \lambda_{\text{obs}}\|\mathcal{O}(f)-y\|_{\mathcal{Y}}^2. \tag{18}$$

A suitable finite-dimensional domain numerical penalization can be quickly implemented based on the previous definition.

Invariance-Guidance Furthermore, prior physical knowledge can be enforced by penalizing deviations from conserved quantities of the underlying dynamics. For each $f\in\mathcal{F}$ representing a full trajectory sample over a temporal grid $\{\tau_k\}_{k=1}^K\subset[0,T]$, write $f_{\tau_k} := f(\tau_k,\cdot)$ for the spatial profile at time τ_k. Considering the first three KdV invariants of equation 3, the deviation may be quantified by the variance of each invariant across the whole trajectory,

$$\text{Var}[I(f)] := \frac{1}{K}\sum_{k=1}^K \big(I(f_{\tau_k}) - \bar{I}(f)\big)^2, \qquad \bar{I(f)} := \frac{1}{K}\sum_{k=1}^K I(f_{\tau_k}), \tag{19}$$

for $I\in\{M,P,H\}$, so that $\text{Var}[I(f)]=0$ when I is conserved along f. The associated energy functional is

$$J_{\text{inv}}(f) = \lambda_M \text{Var}[M(f)] + \lambda_P \text{Var}[P(f)] + \lambda_H \text{Var}[H(f)]. \tag{20}$$

In the numerical experiments, $\lambda_M=\lambda_P=\lambda_H=2$.

Let $g\sim\mu_0$ and denote by $g_t := \phi_t(g)$ the state of the flow at time $t\in[0,1]$. Under the OT parameterization (equation 7), the posterior mean of f given g_t can be expressed as

$$\hat{g}_1(t,g_t) = \sigma_t\, v_t(g_t) + (1-\sigma_{\min})\, g_t, \tag{21}$$

for $\sigma_t := \sigma_t^f = 1-(1-\sigma_{\min})t$. Since v_t is not available at sampling time, we substitute the trained parametric model $u_t(\cdot\mid\theta)\approx v_t$, giving the computable approximation

$$\hat{g}_1(t,g_t) \approx \sigma_t\, u_t(g_t\mid\theta) + (1-\sigma_{\min})\, g_t. \tag{22}$$

Following functional guidance, let $h_t := C_0\nabla J(\hat{g}_1(t,g_t))$, where C_0 is the covariance operator of the Gaussian process prior and $J=J_{\text{obs}}$, $J=J_{\text{inv}}$, or $J=J_{\text{obs}}+J_{\text{inv}}$, depending on the experiment. The guided vector field is computed via

$$u_t^{\text{guide}}(g_t) = u_t(g_t\mid\theta) - \frac{\sigma_t^2}{\max(t,t_{\min})}\left[D_{g_t}u_t(g_t\mid\theta)[h_t] + \frac{1-\sigma_{\min}}{\sigma_t}h_t\right],$$

where $D_{g_t}u_t(g_t\mid\theta)[h_t]$ denotes a Jacobian-vector product.

3 EXPERIMENTS

3.1 NUMERICAL DETAILS

Ground-truth data are generated using a high-resolution finite-difference (Zabusky–Kruskal (Zabusky & Kruskal, 1965)) integrator of equation 1 with $\mu = -1$, seeded by exact single- and multi-soliton initial data defined by the parameters shown in Appendix E, Table 4. The reference measure μ_0 is a centered Gaussian with a Matérn-$\frac{1}{2}$ covariance kernel, with length scale 0.01 and variance 0.1. The vector field is parameterized by a Fourier Neural Operator (FNO) with 16 Fourier modes, 3 layers, and 64–256 hidden channels, depending on the data discretization. We set $\sigma_{\min} = 10^{-4}$ and train the models using Adam with a learning rate of 10^{-3}, and use adaptive Dormand–Prince integration at sampling time. For energy-guidance, we consider $t_{\min} = 10^{-2}$ and $\lambda_{\text{obs}} = 62.5$. Full trajectory models are trained on $20{,}000$ samples and evaluated on 256 test samples, whereas joint schemes are trained on $30{,}000$ and evaluated on $1{,}024$.

The original training and testing trajectories contain 129 temporal and 256 spatial points. For the full trajectory comparisons, each trajectory is represented on a common 64×64 grid. The resampling is performed by bilinear interpolation $\mathbb{R}^{129\times256} \longrightarrow \mathbb{R}^{64\times64}$, using aligned endpoints in both space and time. The full trajectory method comparisons use this common representation. The two- and three soliton datasets are treated as separate learning problems.

All FFM models were trained on a system equipped with an AMD Radeon RX 6600 GPU with 8GB of VRAM and an AMD Ryzen 7 5800H CPU. DiffusionPDE was trained on a 2024 MacBook Pro equipped with an Apple M4 chip and 16GB of unified memory, using the PyTorch MPS backend. All evaluation metrics were computed on the same MacBook Pro. FunDPS was trained on a system equipped with an Intel Core i5-10300H CPU, 16 GB of DDR4 RAM, and an NVIDIA GeForce GTX 1650 Ti GPU with 4 GB of GDDR6 memory.

3.2 UNCONDITIONAL FULL TRAJECTORY GENERATION OF 1, 2, 3-SOLITONS

For unconditional generation, samples are drawn by solving equation 5 using the learned vector field $u_t(\cdot \mid \theta)$, without incorporating additional guidance. We train separate models on full KdV trajectories for the one-, two- and three-soliton cases. Figure 2 presents the generated space–time fields, showing that the models recover clean and coherent soliton structures.

To quantify dynamical consistency, we generate 256 samples for each model—DiffusionPDE, FunDPS, FFM, FFM with energy-guidance—for both the two- and three-soliton cases. The evaluation metrics used in this analysis are described in Appendix G. For each sample, the corresponding initial condition $u_{0,\ell}^{\text{gen}}$ is extracted and used as input for the KdV solver, yielding the reference trajectory $\Phi^{\text{KdV}}(u_{0,\ell}^{\text{gen}})$. Table 1 reports the median value of each metric over the 256 generated samples. Both FFM variants attain lower median values than DiffusionPDE and FunDPS for every reported metric. The differences between FFM with and without energy guidance are smaller and depend on the metric and soliton regime.

Method	Case	$E_{\text{dyn}} \downarrow$	$E_T \downarrow$	$\|\mathcal{R}\|_{\text{RMS}} \downarrow$	$d^M(T) \downarrow$	$d^P(T) \downarrow$	$d^H(T) \downarrow$
DiffusionPDE	2S	1.8092×10^{-1}	2.6982×10^{-1}	4.7473×10^{-1}	5.0019×10^{-2}	1.0176×10^{-2}	1.6968×10^{-1}
DiffusionPDE	3S	1.3736×10^{-1}	2.0446×10^{-1}	7.1654×10^{-1}	5.0329×10^{-2}	1.0283×10^{-1}	1.4957×10^{-1}
FunDPS	2S	7.8586×10^{-1}	8.8044×10^{-1}	4.7807×10^{1}	4.8837×10^{-2}	1.0064×10^{-1}	6.2937×10^{-1}
FunDPS	3S	6.0127×10^{-1}	7.0440×10^{-1}	4.6035×10^{1}	3.2571×10^{-2}	6.6024×10^{-2}	5.3150×10^{-1}
FFM	2S	9.5878×10^{-2}	1.1263×10^{-1}	1.1517×10^{-1}	2.5892×10^{-3}	6.2357×10^{-3}	1.1051×10^{-2}
FFM	3S	1.0345×10^{-1}	1.3419×10^{-1}	2.2988×10^{-1}	4.0313×10^{-3}	6.5129×10^{-3}	1.0434×10^{-2}
FFM + Energy-Guidance	2S	9.1182×10^{-2}	1.1240×10^{-1}	9.4078×10^{-2}	2.2568×10^{-3}	5.4041×10^{-3}	8.5342×10^{-3}
FFM + Energy-Guidance	3S	9.2060×10^{-2}	1.1661×10^{-1}	1.9482×10^{-1}	3.6630×10^{-3}	6.9106×10^{-3}	1.0832×10^{-2}

Table 1: Unconditional full trajectory generation: comparison on 256 generated KdV trajectories per method and soliton regime. Entries are sample medians.

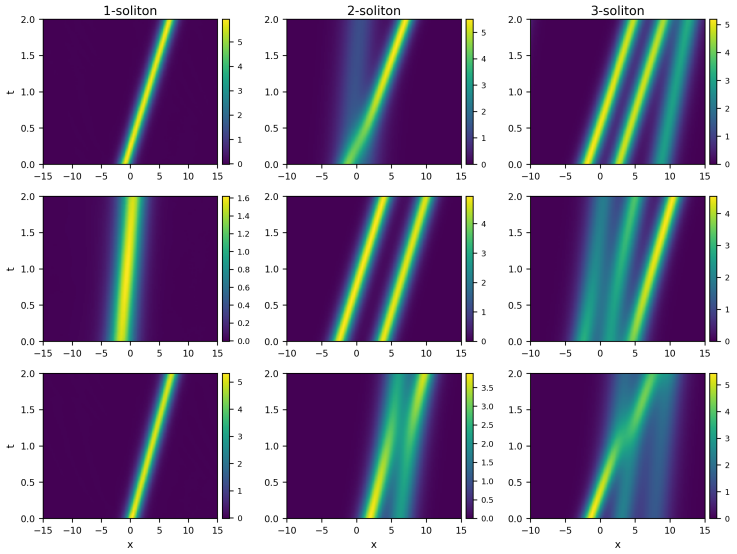

Figure 2: Generated KdV trajectories $u(\tau, x)$ for one-, two- and three-soliton configurations.

3.3 Conditional forecasting from initial data

For full-trajectory forecasting, we compare FFM with our implementations of DiffusionPDE Huang et al. (2024) (Appendix H), and FunDPS Yao et al. (2025), under a common 64×64 protocol. Results are provided in Table 2. From the 256 test trajectories, 8 are used to calibrate FFM observation guidance, leaving 248 for evaluation. Initial-condition (IC) guidance is used for FFM forecasting, and the additional invariant guidance variant isolates the effect of conservation penalties.

FFM with IC guidance has lower median trajectory and final state errors, PDE residuals, and invariant defects than DiffusionPDE and FunDPS in both soliton regimes. Adding invariant guidance reduces the median momentum and energy defects in both regimes. Its effect on forecasting accuracy differs, having an increase in trajectory and final state errors for two solitons, while a slight decrease for three solitons.

Method	Case	$E_{\text{traj}} \downarrow$	$E_{\text{final}} \downarrow$	$\|\mathcal{R}\|_{\text{RMS}} \downarrow$	$d^M(T) \downarrow$	$d^P(T) \downarrow$	$d^H(T) \downarrow$
Test data	2S	---	---	1.0611×10^{-1}	8.9834×10^{-5}	3.3764×10^{-4}	1.3997×10^{-3}
Test data	3S	---	---	2.7962×10^{-1}	9.3532×10^{-4}	3.7315×10^{-4}	2.3332×10^{-3}
DiffusionPDE	2S	3.5121×10^{-1}	5.7962×10^{-1}	1.8513	1.2275×10^{-1}	2.5569×10^{-1}	3.9664×10^{-1}
DiffusionPDE	3S	3.1893×10^{-1}	5.5490×10^{-1}	5.8795	9.9978×10^{-2}	2.5239×10^{-1}	3.7557×10^{-1}
FunDPS	2S	5.2991×10^{-1}	1.0569×10^{0}	4.4124×10^{1}	7.2920×10^{-2}	9.5749×10^{-2}	6.7905×10^{-1}
FunDPS	3S	4.4069×10^{-1}	9.1735×10^{-1}	4.4123×10^{1}	6.5146×10^{-2}	1.1656×10^{-1}	8.4008×10^{-1}
FFM + IC guidance	2S	2.7838×10^{-1}	4.3526×10^{-1}	2.2413×10^{-1}	3.3080×10^{-3}	1.1046×10^{-2}	1.8613×10^{-2}
FFM + IC guidance	3S	2.4227×10^{-1}	3.5292×10^{-1}	4.1800×10^{-1}	5.1094×10^{-3}	7.0658×10^{-3}	1.2984×10^{-2}
FFM + IC + Energy-Guidance	2S	2.9378×10^{-1}	4.5491×10^{-1}	2.1765×10^{-1}	2.9837×10^{-3}	8.5340×10^{-3}	1.3512×10^{-2}
FFM + IC + Energy-Guidance	3S	2.3985×10^{-1}	3.4364×10^{-1}	4.1354×10^{-1}	5.2916×10^{-3}	6.7123×10^{-3}	1.1421×10^{-2}

Table 2: Conditional forecasting on 248 KdV test trajectories under the common protocol. Entries are sample medians.

3.4 Joint Generation and Physical Consistency

We next train the joint model on pairs (u_0, u_T) for one-, two-, and three-soliton data. In the three-soliton setting, we vary the Matérn smoothness parameter $\nu \in \{0.5, 1.5, 2.5\}$ while keeping the architecture and training fixed. Figure 3 overlays the generated initial state u_0^{gen}, the generated terminal state u_T^{gen}, and the reference KdV evolution $\Phi_T^{\text{KdV}}(u_0^{\text{gen}})$ for representative samples. The generated terminal states closely match the reference evolution, including in the three-soliton case. This comparison evaluates whether the generated endpoint pairs are consistent with the KdV flow.

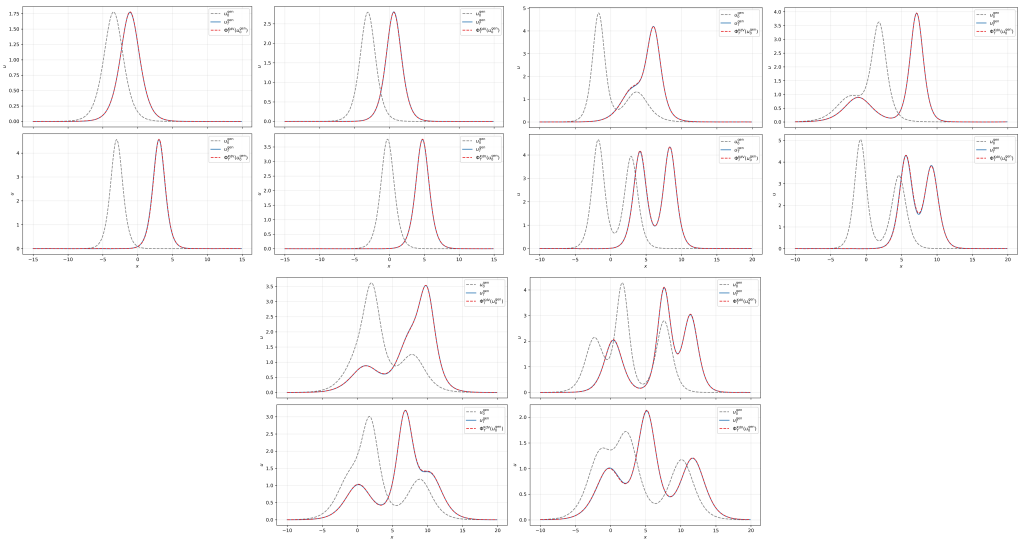

Figure 3: Joint generation for one-, two-, and three-solitons. For each sample: generated initial state u_0^{gen} (gray, dashed), generated solution u_T^{gen} (blue), and the reference KdV evolution of u_0^{gen} (red, dashed). u_T^{gen} matches $\Phi_T^{\text{KdV}}(u_0)$ almost exactly in the displayed samples.

We quantify joint-sample consistency using endpoint invariant defects and the relative L^2 error defined in Appendix G. Table 3 reports medians over 256 unconditional samples. All five experiments produce dynamically consistent pairs: the median L^2 error is below 0.9%. Across the three-soliton ablation, $\nu = 2.5$ gives the smallest median E_T, whereas $\nu = 1.5$ gives the smallest median momentum and energy defects. We also evaluate spatial discretization consistency for each joint model. For a full table see Appendix F.3. Figure 4 shows the relative L^2 discrepancies in the generated initial and terminal channels decrease under spatial refinement. Smoother Matérn priors improve cross-resolution agreement in the three-soliton setting.

Model	ν	$E_T \downarrow$	$d^M(T) \downarrow$	$d^P(T) \downarrow$	$d^H(T) \downarrow$
1–soliton	0.5	7.080×10^{-3}	2.109×10^{-3}	3.425×10^{-3}	5.711×10^{-3}
2–soliton	0.5	8.891×10^{-3}	1.813×10^{-3}	3.383×10^{-3}	5.389×10^{-3}
3–soliton	0.5	8.936×10^{-3}	1.469×10^{-3}	3.045×10^{-3}	4.863×10^{-3}
3–soliton	1.5	8.771×10^{-3}	1.464×10^{-3}	2.882×10^{-3}	4.546×10^{-3}
3–soliton	2.5	8.472×10^{-3}	1.649×10^{-3}	2.980×10^{-3}	5.203×10^{-3}

Table 3: Physical consistency of unconditional joint FFM samples. Entries are sample medians.

3.5 Energy-Guided Targeted Generation

We next evaluate the joint model under endpoint conditioning. Given either an observed initial state u_0 or terminal state u_T, we guide the sampling dynamics using an observation functional acting on the corresponding channel of the joint endpoint estimate. Figure 5 shows forward conditioning from u_0 and inverse conditioning from u_T for one-, two-, and three-soliton configurations. In both directions, the generated pair remains close to the graph of the KdV flow, as verified by independently evolving u_0^{gen} with the reference solver $\Phi_T^{\text{KdV}}(u_0^{\text{gen}})$.

4 Conclusions

We presented energy-guided FFM for generating KdV solitons in function space, using full trajectory and joint initial-final representations. Under the stated assumptions, we derived an exact guidance identity and a covariance-based sampling approximations. The experiments show that joint

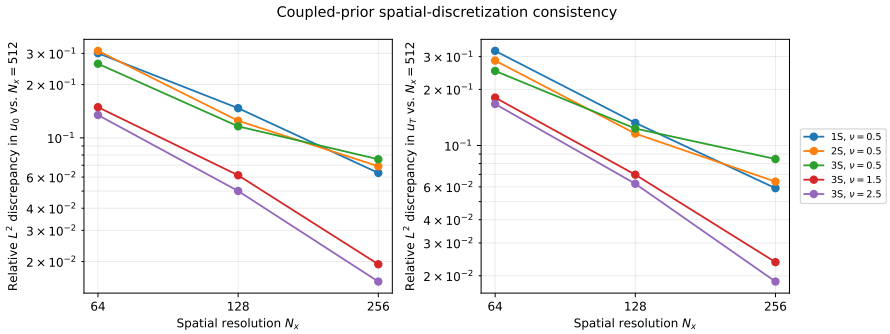

Figure 4: Spatial refinement of joint FFM models with coupled prior samples. Discrepancies in the generated initial and terminal states decrease as the query grid is refined.

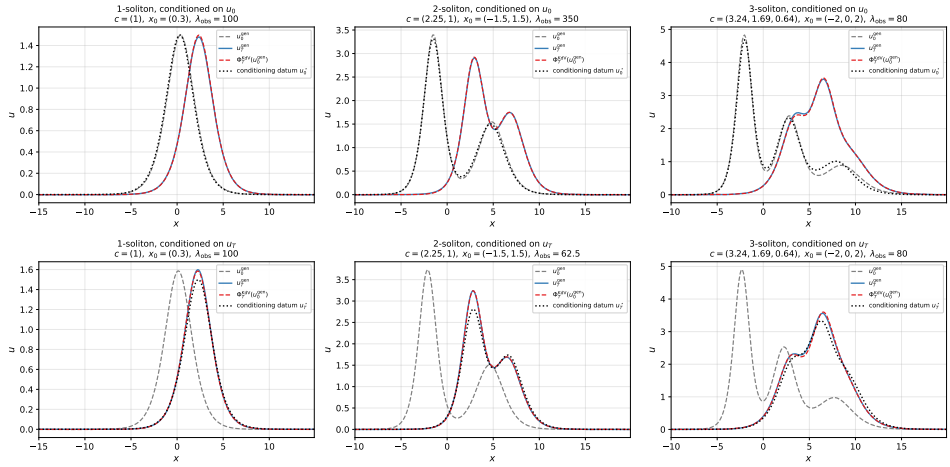

Figure 5: Joint sampling with endpoint conditioning for one-, two-, and three-solitons. The top row conditions on the initial state u_0, while the bottom row conditions on the terminal state u_T.

samples closely follow the KdV flow and that IC-guided FFM improves on our DiffusionPDE and FunDPS baselines in the reported forecasting comparisons. Additional invariant guidance improves momentum and energy conservation, but its effect on forecasting accuracy depends on the soliton regime. These findings motivate evaluating prediction error, PDE residuals, and invariant drift as distinct diagnostics.

Future work includes extending the approach to other dispersive equations, such as mKdV, generalized KdV, wave equations, and the Nonlinear Schrödinger equation (Muñoz & Valenzuela, 2026; Alejo et al., 2025; Freire et al., 2026; Valenzuela et al., 2026). Further theoretical questions include whether dispersives estimates can yield bounds of the form $D_{\mathrm{KL}}(\mu_1 \| \nu) \leq \varepsilon$. Incorporating conservation penalties during training is another direction for investigation.

5 Limitations

Our study is focused on the KdV equation, whose integrable structure provides exact multi-soliton references and conserved quantities; transferring the approach to non-integrable or higher-dimensional dispersive models (e.g. Kadomtsev–Petviashvili) is future work and may require richer priors and architectures. The joint formulation is trained for a fixed horizon T; generating solutions at arbitrary query times or performing long autoregressive rollouts is not yet addressed. Energy guidance biases sampling toward a target set but does not *guarantee* exact conservation; the training-time enforcement of the invariant hierarchy remains open. Finally, our theoretical guarantees concern the FFM framework under explicit assumptions (Appendices B–D).

AI USE STATEMENT

In this work, we used generative AI tools for (i) adapting the previously obtained data towards comparison with DiffusionPDE and FunDPS methods; (ii) obtaining visualizations from the numerical data, and (iii) conducting an internal review of the first draft of this paper. We have not used generative AI tools for fully writing this work, editing, or generating references. We have reviewed all AI-assisted work. We checked all LLM-generated output, including graphics, ideas, texts, etc. References were checked twice to avoid possible bad writing. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The authors of this paper state that all procedures followed in the creation of this work adhere to ICLR 2027 ethical standards and integrity. Ideas were devised by previous research work in the area of dispersive PDEs, along with the interaction with the LLMs and diffusion process community and recent works in the area. All numerical procedures were devised, implemented, and reported following the highest possible standards.

REPRODUCIBILITY STATEMENT

Appendix G defines the evaluation metrics, including dynamical consistency, forecasting error, PDE residuals, and invariant defects. Appendix E describes data generation, and Table 5 reports model configurations and training budgets. Sampling and guidance are described in Sections 2.4 and 2.5, and Appendix H documents the DiffusionPDE implementation. The theoretical appendices state the assumptions and provide the guidance formulation and derivations. The full code implementation and trained checkpoints will be publicly released on a GitHub repository upon publication.

REFERENCES

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic Interpolants: A Unifying Framework for Flows and Diffusions. 2023.

Miguel Á. Alejo, Lucrezia Cossetti, Luca Fanelli, Claudio Muñoz, and Nicolás Valenzuela. Error bounds for Physics Informed Neural Networks in Nonlinear Schrödinger equations placed on unbounded domains, 2025. URL https://arxiv.org/abs/2409.17938.

Vladimir Igorevich Bogachev. *Differentiable measures and the Malliavin calculus*. Number 164. American Mathematical Soc., 2010.

Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, 2nd edition, 2014.

Ruiqi Feng, Chenglei Yu, Wenhao Deng, Peiyan Hu, and Tailin Wu. On the Guidance of Flow Matching. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 16993–17029. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/feng25s.html.

Ricardo Freire, Claudio Muñoz, and Nicolás Valenzuela. Error bounds for Physics Informed Neural Networks in Generalized KdV Equations placed on unbounded domains, 2026. URL https://arxiv.org/abs/2601.00779.

Clifford S. Gardner, John M. Greene, Martin D. Kruskal, and Robert M. Miura. Method for Solving the Korteweg–de Vries Equation. *Physical Review Letters*, 19(19):1095–1097, 1967.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Ryogo Hirota. Exact Solution of the Korteweg–de Vries Equation for Multiple Collisions of Solitons. *Physical Review Letters*, 27(18):1192–1194, 1971.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. DiffusionPDE: Generative PDE-Solving under Partial Observation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the Design Space of Diffusion-Based Generative Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Gavin Kerrigan, Justin Ley, and Padhraic Smyth. Diffusion Generative Models in Infinite Dimensions. In Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent (eds.), *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 9538–9563. PMLR, 25–27 Apr 2023. URL https://proceedings.mlr.press/v206/kerrigan23a.html.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional Flow Matching. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural Operator: Learning Maps Between Function Spaces with Applications to PDEs. *Journal of Machine Learning Research*, 24:1–97, 2023.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier Neural Operator for Parametric Partial Differential Equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Jae Hyun Lim, Nikola B. Kovachki, Ricardo Baptista, Christopher Beckham, Kamyar Azizzadenesheli, Jean Kossaifi, Vikram Voleti, Jiaming Song, Karsten Kreis, Jan Kautz, Christopher Pal, Arash Vahdat, and Anima Anandkumar. Score-based Diffusion Models in Function Space. *arXiv preprint arXiv:2302.07400*, 2023.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Phillip Lippe, Bastiaan S. Veeling, Paris Perdikaris, Richard E. Turner, and Johannes Brandstetter. PDE-Refiner: Achieving Accurate Long Rollouts with Neural PDE Solvers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Claudio Muñoz and Nicolás Valenzuela. Bounds on the approximation error for deep neural networks applied to dispersive models: nonlinear waves. *Partial Differ. Equ. Appl.*, 7(1):Paper No. 8, 60, 2026. ISSN 2662-2963,2662-2971. doi: 10.1007/s42985-026-00369-y. URL https://doi.org/10.1007/s42985-026-00369-y.

Jakiw Pidstrigach, Youssef Marzouk, Sebastian Reich, and Sven Wang. Infinite-dimensional diffusion models. *Journal of Machine Learning Research*, 25(414):1–52, 2024.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Andrew M. Stuart. Inverse Problems: A Bayesian Perspective. *Acta Numerica*, 19:451–559, 2010.

Nicolás Valenzuela, Ricardo Freire, and Claudio Muñoz. Neural discovery of strichartz extremizers, 2026. URL https://arxiv.org/abs/2605.04918.

Hao Xu, Yuntian Chen, Rui Cao, Tianning Tang, Mengge Du, Jian Li, Adrian H Callaghan, and Dongxiao Zhang. Generative discovery of partial differential equations by learning from math handbooks. *Nature Communications*, 16(1):10255, 2025.

Jiachen Yao, Abbas Mammadov, Julius Berner, Gavin Kerrigan, Jong Chul Ye, Kamyar Azizzadenesheli, and Animashree Anandkumar. Guided Diffusion Sampling on Function Spaces with Applications to PDEs. *Advances in Neural Information Processing Systems*, 38:127057–127094, 2025.

Norman J. Zabusky and Martin D. Kruskal. Interaction of Solitons in a Collisionless Plasma and the Recurrence of Initial States. *Physical Review Letters*, 15(6):240–243, 1965.

APPENDIX

A BACKGROUND

A.1 GENERATIVE MODELING IN FUNCTION SPACES

Let $\mathcal{F} \subset \{f : \mathcal{X} \subseteq \mathbb{R}^d \to \mathbb{R}\}$ be a real separable Hilbert space and $\nu \in \mathcal{P}(\mathcal{F})$ the data distribution from which we observe samples $\{f_i\}_{i=1}^N$. A *functional* generative model aims to sample new functions $f \sim \nu$ that can be queried at any $x \in \mathcal{X}$, making the learned distribution independent of the discretization used during training or inference. This is a natural setting for PDEs, and it is the setting used in neural operators (Li et al., 2021; Kovachki et al., 2023; Lu et al., 2021) and function-space diffusion models (Lim et al., 2023; Pidstrigach et al., 2024; Kerrigan et al., 2023). The flow matching framework (Lipman et al., 2023) learns a vector field whose flow transports a reference distribution to ν. Functional Flow Matching (FFM) (Kerrigan et al., 2024) extends this idea to function spaces by formulating the transport in measure theoretic terms since in the infinite-dimensional setting there is no analogue of the Lebesgue measure.

A.2 KORTEWEG–DE VRIES EQUATION

We consider the Korteweg–de Vries equation on the real line,

$$\partial_\tau u + \partial_x\big(\partial_{xx} u - \mu\, u^2\big) = 0, \qquad (\tau, x) \in \mathbb{R} \times \mathbb{R},$$

which describes long-wavelength waves in dispersive media. For $\mu = -1$, equation 1 admits the 1-soliton solutions

$$u_{c,x_0}(\tau, x) = \tfrac{3}{2}\, c\, \operatorname{sech}^2\Big(\tfrac{1}{2}\sqrt{c}\,(x - c\,\tau - x_0)\Big), \quad c > 0,$$

representing a localized wave with amplitude proportional to c and propagation speed c. The KdV equation is completely integrable (Gardner et al., 1967), and exact N-soliton solutions are given by the Hirota bilinear form (Hirota, 1971)

$$u_{c,\delta}(\tau, x) := 6\frac{f\partial_x^2 f - (\partial_x f)^2}{f^2},$$

where $f := f(\tau, x)$ is defined as

$$f = 1 + \sum_{n=1}^{N} \sum_{{}_N C_n} \left(\prod_{k<\ell}^{(n)} \frac{(\sqrt{c_{i_k}} - \sqrt{c_{i_\ell}})^2}{(\sqrt{c_{i_k}} + \sqrt{c_{i_\ell}})^2} \right) \exp\left(s_{i_1} + \ldots + s_{i_n}\right),$$

and for all $j \in \{1, \ldots, N\}$, $s_j = \sqrt{c_j}(x - c_j\tau) - \delta_j$. Here, ${}_N C_n$ indicates all possible combinations of n elements taken from N, and (n) indicates all possible combinations of the n elements (with the specified condition $k < \ell$, as indicated). The dynamics conserves an infinite hierarchy of quantities. In particular, the first three invariants are

$$M = \int u \,\mathrm{d}x, \quad P = \int u^2 \,\mathrm{d}x, \quad H = \int \Big(\tfrac{1}{3}u^3 - \tfrac{1}{2}(\partial_x u)^2\Big)\mathrm{d}x,$$

which correspond to mass, momentum, and energy, respectively. Unlike standard image or video domains, these conservation laws offer a quantitative measure of physical fidelity.

For a multi-soliton solution, previous conserved quantities are well-understood. Particularly, for an N-soliton with speeds $\{c_j\}_{j=1}^N$, the invariants are

$$M_N = 6\sum_{j=1}^{N} c_j^{1/2}, \quad P_N = 6\sum_{j=1}^{N} c_j^{3/2}, \quad H_N = \frac{9}{5}\sum_{j=1}^{N} c_j^{5/2}. \tag{23}$$

Conservation does not determine the dynamics. Let $q_c(x) = \frac{3c}{2}\operatorname{sech}^2(\sqrt{c}x/2)$ and consider $\widetilde{u}(\tau, x) = q_c(x - a(\tau))$ for an arbitrary differentiable shift $a(\tau)$. On the real line, translation invariance implies that $M[\widetilde{u}(\tau)]$, $P[\widetilde{u}(\tau)]$, and $H[\widetilde{u}(\tau)]$ are constant for every choice of a. However, the traveling-wave identity $q_c''' + 2q_c q_c' = cq_c'$ gives

$$\partial_\tau \widetilde{u} + \partial_x^3 \widetilde{u} + 2\widetilde{u}\partial_x \widetilde{u} = \big(c - a'(\tau)\big) q_c'(x - a(\tau)). \tag{24}$$

Thus, all three invariants can be preserved exactly while the propagation speed is incorrect. Taking $a(0) = x_0$ also preserves the prescribed initial condition. This example explains why invariant drift must be complemented by trajectory errors and PDE residuals. On a truncated domain, boundary effects introduce an additional qualification to this argument.

B Functional Flow Matching: Theoretical Framework

This appendix recalls the measure-theoretic foundations of Functional Flow Matching (FFM) used in the main text. Theorems 2–4 are Theorems 1–3 of Kerrigan et al. (2024), restated here in our notation. Standard results concerning Gaussian measures on separable Hilbert spaces, Cameron–Martin spaces, and equivalence of Gaussian measures are taken from Da Prato & Zabczyk (2014).

Setting. Let $\mathcal{F}$ be a separable real Hilbert space of functions $f : X \subseteq \mathbb{R}^d \to \mathbb{R}$ with Borel σ-algebra $\mathcal{B}(\mathcal{F})$, and let $\nu \in \mathcal{P}(\mathcal{F})$ be the data distribution. We seek a generative model producing $f \sim \nu$ as *functions*. Because $\mathcal{F}$ is infinite-dimensional there is no Lebesgue measure and no density w.r.t. it; the theory is therefore phrased in terms of measures.

Flows and the weak continuity equation. A time-dependent velocity field $v : [0,1] \times \mathcal{F} \to \mathcal{F}$ induces a flow $\phi : [0,1] \times \mathcal{F} \to \mathcal{F}$ through the abstract ODE $\partial_t \phi_t(g) = v_t(\phi_t(g))$, $\phi_0(g) = g$, whose well-posedness in infinite dimensions requires v_t to be, e.g., continuously Fréchet differentiable with bounded derivative (Da Prato & Zabczyk, 2014). Given $\mu_0 \in \mathcal{P}(\mathcal{F})$, we set $\mu_t = [\phi_t]_\# \mu_0$. We say (v_t) *generates* (μ_t) when the pair solves the continuity equation weakly,

$$\int_0^1 \int_{\mathcal{F}} \big(\partial_t \varphi(g,t) + \langle v_t(g), \nabla_g \varphi(g,t)\rangle\big)\, \mathrm{d}\mu_t(g)\, \mathrm{d}t = 0, \tag{25}$$

for all smooth cylindrical test functions ϕ with compact support in the generative time interval $(0,1)$. Here, "generates" refers to the weak continuity equation. Whenever a deterministic flow representation is used, we additionally assume that the corresponding ODE is well posed and that $\mu_t = (\phi_t)_\# \mu_0$. The analogous assumptions are required for the guided path.

Conditional and marginal paths. For each datum f we prescribe a conditional path (μ_t^f) with $\mu_0^f = \mu_0$ and μ_1^f a narrow Gaussian centered at f, and define the marginal measure

$$\mu_t(A) = \int_{\mathcal{F}} \mu_t^f(A)\, \mathrm{d}\nu(f), \quad \forall A \in \mathcal{B}(\mathcal{F}), \tag{26}$$

so that $\mu_1 \approx \nu$.

Theorem 2. *Assume $\int_0^1 \int_{\mathcal{F}\times\mathcal{F}} \|v_t^f(g)\|\, \mathrm{d}\mu_t^f(g)\, \mathrm{d}\nu(f)\, \mathrm{d}t < \infty$ and $\mu_t^f \ll \mu_t$ for ν-a.e. f and almost every t. Then the vector field*

$$v_t(g) = \int_{\mathcal{F}} v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f) \tag{27}$$

generates $(\mu_t)_{t\in[0,1]}$ in the weak sense as in equation 25, where $\mathrm{d}\mu_t^f/\mathrm{d}\mu_t$ is the associated Radon–Nikodym derivative.

Theorem 3. *Consider a probability measure ν on $\mathcal{F}$ and a collection of measures μ_t^f parameterized by $f \in \mathcal{F}$. Suppose that the collection of parameterized measures are ν-a.e. mutually absolutely continuous. Define the marginal measure via equation 26. Then $\mu_t^f \ll \mu_t$ for ν-a.e. f.*

The previous theorem asserts that a nontrivial hypothesis about the absolute continuity $\mu_t^f \ll \mu_t$ is implied by the mutual absolute continuity of the family μ_t^f, which we secure via a Gaussian parameterization.

Gaussian measures and the OT parameterization. We take $\mu_t^f = \mathcal{N}(m_t^f, (\sigma_t^f)^2 C_0)$ with a fixed trace-class covariance C_0 (a Matérn kernel), and the optimal-transport schedule $m_t^f = tf$, $\sigma_t^f = 1 - (1 - \sigma_{\min})t$. The conditional flow $\phi_t^f(g) = \sigma_t^f g + m_t^f$ pushes $g \sim \mathcal{N}(0, C_0)$ onto μ_t^f, and the generating conditional field is $v_t^f(g) = \frac{(\sigma_t^f)'}{\sigma_t^f}(g - m_t^f) + \frac{\mathrm{d}}{\mathrm{d}t} m_t^f$.

We assume that C_0 is injective in the Gaussian constructions below. For the posterior differentiation results in Appendix D, we also assume

$$\nu(\mathcal{H}) = 1, \qquad \int_{\mathcal{F}} e^{c\|f\|_{\mathcal{F}}}\, d\nu(f) < \infty \quad \text{for every } c \geq 0, \qquad \mathcal{H} = \mathrm{Im}(C_0^{1/2}).$$

These are standing hypotheses for those results, rather than consequences of Fréchet differentiability of the posterior mean.

Absolute continuity via Cameron–Martin and Feldman–Hájek. For $\mu = \mathcal{N}(0, C)$ the *Cameron–Martin space* is $\mathcal{H} = \mathrm{Im}(C_\mu^{1/2}) = \mathrm{Im}(C_\nu^{1/2})$, with $\langle \cdot, \cdot \rangle_H = \langle C^{-1/2} \cdot, C^{-1/2} \cdot \rangle$. The Feldman–Hájek theorem states that two Gaussian measures μ and ν on $\mathcal{F}$ are either equivalent or mutually singular. In particular, equivalence of μ and ν holds if and only if,

1. μ and ν have the same Cameron–Martin space $\mathcal{H} := C_\mu^{1/2} = C_\nu^{1/2}$;
2. the difference in their means lies in the Cameron–Martin space, i.e. $m_\mu - m_\nu \in \mathcal{H}$; and
3. the operator $(C_\mu^{-1/2} C_\nu^{1/2})(C_\mu^{-1/2} C_\nu^{1/2})^* - I$ is a Hilbert-Schmidt operator on $\bar{\mathcal{H}}$.

For the OT family these statements holds. With $C_t = (\sigma_t^f)^2 C_0$ and $0 < \sigma_t^f \leq M$ one has $\mathrm{Im}(C_t^{1/2}) = \mathrm{Im}(C_0^{1/2})$ (Lemma 6.15 in Stuart (2010)). Under the support assumption $\nu(C_0^{1/2}(\mathcal{F})) = 1$ the difference in their means satisfy $m_t^f - m_t^g = t(f - g) \in C_0^{1/2}(\mathcal{F})$. Finally, equal covariances makes the operator vanish, then is trivially Hilbert–Schmidt. Hence $\mu_t^f \ll \mu_t$ and equation 27 applies.

Training objective. Regressing directly on the marginal v_t is intractable. Instead one minimizes the conditional objective equation 9. The following theorem asserts that the two loss objectives differ by a constant.

Theorem 4 (Loss equivalence)**.** *Assume that the true and model vector fields are square-integrable. i.e. $\int_0^1 \int_{\mathcal{F}} \|v_t(g)\|^2 \,\mathrm{d}\mu_t(g)\,\mathrm{d}t < \infty$ and $\int_0^1 \int_{\mathcal{F}} \|u_t(g|\theta)\|^2 \,\mathrm{d}\mu_t(g)\,\mathrm{d}t < \infty$ $u_t(\cdot|\theta)$. Then, $\mathcal{L}(\theta) = \mathcal{J}(\theta) + C$ with $C \in \mathbb{R}$ independent of θ.*

Thus the tractable conditional regression equation 9 has the same minimizer as the intractable marginal one.

C Distance bounds for FFM

Following the error bound framework developed for infinite-dimensional diffusion models (Pidstrigach et al., 2024), we derive a quantitative bound on the discrepancy between the FFM sampling distribution μ_1 and the target data distribution ν, in terms of the training loss and the numerical integration error.

Wasserstein-2 distance Let $\mathcal{P}_2(\mathcal{F})$ denote the set of Borel probability measures μ on $\mathcal{F}$ with finite second moment, i.e, $\int_{\mathcal{F}} \|x\|_{\mathcal{F}}^2 \, \mathrm{d}\mu(x) < \infty$. For $\mu, \pi \in \mathcal{P}_2(\mathcal{F})$, the *Wasserstein-2 distance* is

$$\mathcal{W}_2(\mu, \pi) = \left(\inf_{\kappa \in \mathcal{Q}(\mu,\pi)} \int_{\mathcal{F}\times\mathcal{F}} \|x - y\|_{\mathcal{F}}^2 \, \mathrm{d}\kappa(x, y) \right)^{1/2}, \tag{28}$$

where $\mathcal{Q}(\mu, \pi)$ is the set of couplings of μ and π, i.e., probability measures on $\mathcal{F} \times \mathcal{F}$ with marginals μ and π.

Since $\mathcal{F}$ is Polish, $(\mathcal{P}_2(\mathcal{F}), \mathcal{W}_2)$ is itself a complete, separable metric space, and $\mathcal{W}_2$-convergence is equivalent to weak convergence together with convergence of second moments.

Theorem 5 (Distance bound for FFM)**.** *Let $\nu \in \mathcal{P}_2(\mathcal{F})$, $\mu_0 = \mathcal{N}(0, C_0)$ be the reference measure with C_0 self-adjoint, non-negative, and trace-class, and consider $\sigma_{\min} > 0$. Suppose $u_t(\cdot \mid \theta)$ is L-Lipschitz in g and K-Lipschitz in t, uniformly in the other variable,*

$$\|u_t(g \mid \theta) - u_s(h \mid \theta)\| \le L\|g - h\| + K|t - s|, \qquad \forall\, g, h \in \mathcal{F},\ t, s \in [0, 1], \tag{29}$$

and v_t is such that

$$M^2 := \sup_{t\in[0,1]} \mathbb{E}_{g\sim\mu_t}\left[\|v_t(g)\|_{\mathcal{F}}^2\right] < \infty. \tag{30}$$

Let $\varepsilon_{\mathrm{Loss}} := \mathcal{L}(\theta) = \int_0^1 \mathbb{E}_{g\sim\mu_t}\left[\|v_t(g) - u_t(g \mid \theta)\|_{\mathcal{F}}^2\right] \mathrm{d}t$ be the flow matching loss. Let μ_{sample} be the terminal law of the explicit Euler approximation in (33), on a partition with maximum step size Δt. Let $\varepsilon_{\mathrm{Num}}$ denote the integrated squared field-freezing defect defined in (47). Under the stated assumptions,

$$\varepsilon_{\mathrm{Num}} \le 2(L^2M^2 + K^2)\Delta t^2.$$

Then

$$\mathcal{W}_2(\nu, \mu_{\mathrm{sample}}) \ \le\ \sigma_{\min}\sqrt{\mathrm{Tr}(C_0)} \ +\ \sqrt{2}\, e^L \Big(\varepsilon_{\mathrm{Loss}}^{1/2} + \varepsilon_{\mathrm{Num}}^{1/2}\Big). \tag{31}$$

Proof. Let $g_t := \phi_t(g)$ denote the solution of

$$\dot{g}_t = v_t(g_t), \qquad g_0 = g, \tag{32}$$

and let $\hat{g}_t$ denote the solution obtained by integrating numerically

$$\dot{\hat{g}}_t = u_{\lfloor t \rfloor}(\hat{g}_{\lfloor t \rfloor} \mid \theta), \qquad \hat{g}_0 = g. \tag{33}$$

Define the error $e_t := g_t - \hat{g}_t$, so that $e_0 = g_0 - \hat{g}_0 = g - g = 0$. Partition $[0, 1]$ into $\tau := \{0 = t_0 < t_1 < \cdots < t_N = 1\}$ with step size Δt. For the given partition we denote

$$\lfloor t \rfloor = \max_{t_i \in \tau}\{t_i \le t\}. \tag{34}$$

Then, for $t \in [t_k, t_{k+1})$ we decompose

$$\begin{aligned} v_t(g_t) - u_{t_k}(\hat{g}_{t_k} \mid \theta) = &\underbrace{v_t(g_t) - u_t(g_t \mid \theta)}_{\text{(I)}} \\ &+ \underbrace{u_t(g_t \mid \theta) - u_{t_k}(g_{t_k} \mid \theta)}_{\text{(II)}} \\ &+ \underbrace{u_{t_k}(g_{t_k} \mid \theta) - u_{t_k}(\hat{g}_{t_k} \mid \theta)}_{\text{(III)}}. \end{aligned}$$

Differential inequality for $\|e_t\|$. Notice that $\dot{e}_t = \dot{g}_t - \dot{\hat{g}}_t = v_t(g_t) - u_{t_k}(\hat{g}_{t_k} \mid \theta) = \text{(I)} + \text{(II)} + \text{(III)}$. For μ_0-almost every initial condition, e_t is absolutely continuous. Hence $t \mapsto \|e_t\|$ is absolutely continuous and satisfies

$$\left|\frac{d}{dt}\|e_t\|\right| \le \|\dot{e}_t\|$$

for almost every t. We obtain

$$\frac{d}{dt}\|e_t\| \le \|\dot{e}_t\| \le \|\text{(I)}\| + \|\text{(II)}\| + \|\text{(III)}\|, \qquad \text{for a.e. } t \in [0, 1]. \tag{35}$$

Assume $u_t(\cdot \mid \theta)$ is L-Lipschitz, uniformly in $t \in [0,1]$. Then

$$\|(\mathrm{III})\| = \|u_{t_k}(g_{t_k} \mid \theta) - u_{t_k}(\hat{g}_{t_k} \mid \theta)\| \le L\|g_{t_k} - \hat{g}_{t_k}\| = L\|e_{t_k}\|. \tag{36}$$

To apply Grönwall's inequality we need the right-hand side to depend on t rather than on the grid point t_k. To this end, define $E(s) := \sup_{r \le s} \|e_r\|$. Since $t_k \le t$, we have $\|e_{t_k}\| \le E(t)$, and therefore

$$\frac{d}{dt}\|e_t\| \le \|(\mathrm{I})\| + \|(\mathrm{II})\| + L\,E(t), \qquad \text{for a.e. } t \in [0,1]. \tag{37}$$

Grönwall's inequality. Integrating from 0 to s and using $e_0 = 0$,

$$\|e_s\| \le \int_0^s \big(\|(\mathrm{I})\| + \|(\mathrm{II})\|\big)\,\mathrm{d}t + L\int_0^s E(t)\,\mathrm{d}t. \tag{38}$$

The right-hand side is nondecreasing in s, so taking the supremum over $r \le s$ on the left-hand side gives

$$E(s) \le \int_0^s \big(\|(\mathrm{I})\| + \|(\mathrm{II})\|\big)\,\mathrm{d}t + L\int_0^s E(t)\,\mathrm{d}t. \tag{39}$$

By Grönwall's inequality, evaluated at $s = 1$,

$$E(1) \le e^L \int_0^1 \big(\|(\mathrm{I})\| + \|(\mathrm{II})\|\big)\,\mathrm{d}t. \tag{40}$$

Squaring both sides and applying the Cauchy–Schwarz inequality on $[0,1]$, followed by the identity $(a+b)^2 \le 2(a^2+b^2)$,

$$\begin{aligned} E(1)^2 &\le e^{2L}\left(\int_0^1 \big(\|(\mathrm{I})\| + \|(\mathrm{II})\|\big)\,\mathrm{d}t\right)^2 \\ &\le e^{2L}\int_0^1 \big(\|(\mathrm{I})\| + \|(\mathrm{II})\|\big)^2\,\mathrm{d}t \\ &\le 2e^{2L}\int_0^1 \Big(\|(\mathrm{I})\|^2 + \|(\mathrm{II})\|^2\Big)\,\mathrm{d}t. \end{aligned}$$

Identifying the loss and numerical error terms. Taking expectation over $g \sim \mu_0$ and using Tonelli's theorem to interchange expectation and time integration,

$$\mathbb{E}_{g\sim\mu_0}[E(1)^2] \le 2e^{2L}\left(\int_0^1 \mathbb{E}_{g\sim\mu_0}\big[\|(\mathrm{I})\|^2\big]\,\mathrm{d}t + \int_0^1 \mathbb{E}_{g\sim\mu_0}\big[\|(\mathrm{II})\|^2\big]\,\mathrm{d}t\right). \tag{41}$$

Since $g_t = \phi_t(g)$, so that $\mu_t = (\phi_t)_{\#}\mu_0$, and $(\mathrm{I}) = v_t(g_t) - u_t(g_t \mid \theta)$ depends on g only through g_t, the change of variables to $g \sim \mu_t$ gives

$$\mathbb{E}_{g\sim\mu_0}\big[\|v_t(g_t) - u_t(g_t \mid \theta)\|^2\big] = \mathbb{E}_{g\sim\mu_t}\big[\|v_t(g) - u_t(g \mid \theta)\|^2\big]. \tag{42}$$

Integrating over $[0,1]$ and recalling the flow matching loss,

$$\int_0^1 \mathbb{E}_{g\sim\mu_0}\big[\|(\mathrm{I})\|^2\big]\,\mathrm{d}t = \mathcal{L}(\theta) =: \varepsilon_{\mathrm{Loss}}. \tag{43}$$

Bounding the numerical error term. For $t \in [t_k, t_{k+1})$, the joint Lipschitz condition on $u_t(\cdot \mid \theta)$ gives

$$\|(\mathrm{II})\| = \|u_t(g_t \mid \theta) - u_{t_k}(g_{t_k} \mid \theta)\| \le L\|g_t - g_{t_k}\| + K|t - t_k| \le L\|g_t - g_{t_k}\| + K\Delta t. \tag{44}$$

Since $g_t - g_{t_k} = \int_{t_k}^t v_s(g_s)\,\mathrm{d}s$, Jensen's inequality together with the change of variables $\mu_0 \to \mu_s$ used above give

$$\mathbb{E}_{g\sim\mu_0}\big[\|g_t - g_{t_k}\|^2\big] \le \Delta t \int_{t_k}^{t_{k+1}} \mathbb{E}_{g\sim\mu_s}\big[\|v_s(g)\|^2\big]\,\mathrm{d}s \le M^2\,\Delta t^2. \tag{45}$$

Combining both bounds with $(a+b)^2 \le 2(a^2+b^2)$,

$$\mathbb{E}\big[\|(\mathrm{II})\|^2\big] \le 2\big(L^2M^2+K^2\big)\Delta t^2, \qquad t \in [0,1], \tag{46}$$

and therefore, defining

$$\int_0^1 \mathbb{E}\big[\|(\mathrm{II})\|^2\big]\,\mathrm{d}t =: \varepsilon_{\mathrm{Num}} \le 2\big(L^2M^2+K^2\big)\Delta t^2 = O(\Delta t^2), \tag{47}$$

we obtain

$$\mathbb{E}\big[\|g_1-\hat{g}_1\|^2\big] \le \mathbb{E}_{g\sim\mu_0}[E(1)^2] \le 2e^{2L}\big(\varepsilon_{\mathrm{Loss}}+\varepsilon_{\mathrm{Num}}\big). \tag{48}$$

Since $(g_1,\hat{g}_1)$ is a (not necessarily optimal) coupling of $(\mu_1,\mu_{\mathrm{sample}})$,

$$\mathcal{W}_2(\mu_1,\mu_{\mathrm{sample}}) \le \Big(\mathbb{E}_{g\sim\mu_0}\big[\|g_1-\hat{g}_1\|^2\big]\Big)^{1/2} \le \sqrt{2}\,e^{L}\sqrt{\varepsilon_{\mathrm{Loss}}+\varepsilon_{\mathrm{Num}}}. \tag{49}$$

Using $\sqrt{a+b} \le \sqrt{a}+\sqrt{b}$,

$$\mathcal{W}_2(\mu_1,\mu_{\mathrm{sample}}) \le \sqrt{2}\,e^{L}\left(\varepsilon_{\mathrm{Loss}}^{1/2}+\varepsilon_{\mathrm{Num}}^{1/2}\right). \tag{50}$$

Bounding $\mathcal{W}_2(\nu,\mu_1)$. By construction,

$$\mu_1 = \int \mathcal{N}(f,\sigma_{\min}^2 C_0)\,\mathrm{d}\nu(f) \quad\Longleftrightarrow\quad \mu_1 = \mathrm{Law}(F+\sigma_{\min}Z), \qquad F\sim\nu,\ Z\sim\mathcal{N}(0,C_0), \tag{51}$$

with F and Z independent. Coupling $F\sim\nu$ with $F+\sigma_{\min}Z\sim\mu_1$,

$$\mathcal{W}_2(\nu,\mu_1)^2 \le \mathbb{E}\big[\|F-(F+\sigma_{\min}Z)\|^2\big] = \sigma_{\min}^2\,\mathbb{E}\big[\|Z\|^2\big]. \tag{52}$$

Since $Z\sim\mathcal{N}(0,C_0)$ in the Hilbert space $\mathcal{F}$, let $\{e_i\}_i$ be an orthonormal basis of eigenvectors of C_0 (which exists by the spectral theorem, as C_0 is self-adjoint, non-negative, and trace-class), with eigenvalues $\lambda_i \ge 0$. Writing $Z=\sum_i Z_i e_i$ with $Z_i=\langle Z,e_i\rangle \sim \mathcal{N}(0,\lambda_i)$,

$$\mathbb{E}\big[\|Z\|^2\big] = \mathbb{E}\Big[\sum_i Z_i^2\Big] = \sum_i \mathbb{E}[Z_i^2] = \sum_i \lambda_i = \mathrm{Tr}(C_0). \tag{53}$$

Substituting back,

$$\mathcal{W}_2(\nu,\mu_1) \le \sigma_{\min}\sqrt{\mathrm{Tr}(C_0)}. \tag{54}$$

Since $\mathcal{W}_2$ is a metric on $\mathcal{P}_2(\mathcal{F})$, the triangle inequality gives

$$\begin{aligned}\mathcal{W}_2(\nu,\mu_{\mathrm{sample}}) &\le \mathcal{W}_2(\nu,\mu_1)+\mathcal{W}_2(\mu_1,\mu_{\mathrm{sample}})\\ &\le \sigma_{\min}\sqrt{\mathrm{Tr}(C_0)}+\sqrt{2}\,e^{L}\left(\varepsilon_{\mathrm{Loss}}^{1/2}+\varepsilon_{\mathrm{Num}}^{1/2}\right).\end{aligned}$$ □

This estimate concerns explicit Euler discretization. Since the squared numerical defect enters through a square root, its contribution to the W_2 bound is $O(\Delta t)$. The experiments use adaptive Dormand–Prince integration; no convergence rate for that implementation is established by this proof.

D Energy-Guidance for FFM

Feng et al. (2025) develop the first general guidance framework for finite-dimensional flow matching, deriving a family of techniques for general schemes. However, the extension of these techniques to an infinite-dimensional setting remains unexplored. This section details the extension of energy-based methods to infinite-dimensional function spaces, allowing the well-posedness conditioning of the Functional Flow Matching method.

Let $J:\mathcal{F}\to\mathbb{R}$ be a measurable energy functional. Our goal is to guide the sampling process toward low-energy configurations by sampling from the modified target measure

$$\nu'(\mathrm{d}f) := \frac{e^{-J(f)}}{Z}\,\nu(\mathrm{d}f), \quad Z=\int_{\mathcal{F}} e^{-J(f)}\,\mathrm{d}\nu(f) \in (0,\infty). \tag{55}$$

D.1 Preliminary Results

Lemma 1. *ν' is a probability measure and $\nu \sim \nu'$ with*

$$\frac{\mathrm{d}\nu'}{\mathrm{d}\nu}(f) = \frac{1}{Z} e^{-J(f)}, \tag{56}$$

for every ν-a.e $f \in \mathcal{F}$.

Proof. To verify that ν' is a probability measure, observe that $\nu' \geq 0$ since $e^{-J(f)} > 0$ for $f \in \mathcal{F}$. Moreover,

$$\nu'(\mathcal{F}) = \frac{1}{Z} \int_{\mathcal{F}} e^{-J(f)} \, \mathrm{d}\nu(f) = \frac{Z}{Z} = 1.$$

Hence, ν' is a probability measure. To prove that both measures are equivalent, choose $A \in \mathcal{B}(\mathcal{F})$ such that $\nu(A) = 0$, then $\nu'(A) = \int_A \frac{1}{Z} e^{-J(f)} \, \mathrm{d}\nu(f) = 0$. Conversely, if $\nu(A) > 0$ let $A_n = A \cap \{\frac{e^{-J(f)}}{Z} \geq \frac{1}{n}\}$. Since $A_n \nearrow A$ there exists $n \in \mathbb{N}$ such that $\nu(A_n) > 0$ and then $\nu'(A) \geq \nu'(A_n) = \int_{A_n} \frac{e^{-J(f)}}{Z} \, \mathrm{d}\nu(f) \geq \frac{1}{n} \nu(A_n) > 0$. Thus, ν and ν' are equivalent. □

Lemma 2. *For almost every $t \in [0,1]$, assume that the map*

$$(f, g) \mapsto \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)$$

admits a $\mathcal{B}(\mathcal{F}) \otimes \mathcal{B}(\mathcal{F})$-jointly measurable version. Then, for almost every $t \in [0,1]$ and μ_t-a.e. $g \in \mathcal{F}$, the function $f \mapsto \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)$ defines a probability measure on $\mathcal{F}$ via

$$\nu_t(\mathrm{d}f \mid g) := \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) \, \mathrm{d}\nu(f). \tag{57}$$

Proof. Fix $t \in [0,1]$ such that the assumptions of the lemma hold. By Theorem 3, $\mu_t^f \ll \mu_t$ for ν-a.e. f, so the Radon–Nikodym derivative $\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}$ exists and is measurable for ν-a.e. f. By assumption, choose a jointly measurable version of the Radon–Nikodym derivative.

Since $\mu_t^f \in \mathcal{P}(\mathcal{F})$, for every $A \in \mathcal{B}(\mathcal{F})$

$$\mu_t^f(A) = \int_A \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) \, \mathrm{d}\mu_t(g) \geq 0,$$

which implies that, for ν-a.e. fixed f, $\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) \geq 0$ for μ_t-a.e. g.

Define $N := \{(f,g) : \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) < 0\} \in \mathcal{B}(\mathcal{F}) \otimes \mathcal{B}(\mathcal{F})$ and consider the sets $N^f := \{g : (f,g) \in N\}$ and $N^g := \{f : (f,g) \in N\}$. The above argument states that $\mu_t(N^f) = 0$ for ν-a.e. f. Since $\mathbb{1}_N \geq 0$ is jointly measurable by hypothesis, Tonelli's theorem gives

$$(\nu \otimes \mu_t)(N) = \int_{\mathcal{F}} \mu_t(N^f) \, \mathrm{d}\nu(f) = 0.$$

Reversing the integration order,

$$0 = (\nu \otimes \mu_t)(N) = \int_{\mathcal{F}} \nu(N^g) \, \mathrm{d}\mu_t(g).$$

Therefore, $\nu(N^g) = 0$ for μ_t-a.e. g. Concluding that

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) \geq 0, \quad \text{for } \nu\text{-a.e. } f \text{ and } \mu_t\text{-a.e. } g.$$

Let $B \in \mathcal{B}(\mathcal{F})$ and denote $h_t(g) := \nu_t(\mathcal{F} \mid g)$, by Tonelli's Theorem and the definition equation 26

$$\int_B h_t(g) \, \mathrm{d}\mu_t(g) = \int_{\mathcal{F}} \left(\int_B \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) \, \mathrm{d}\mu_t(g) \right) \mathrm{d}\nu(f) = \int_{\mathcal{F}} \mu_t^f(B) \, \mathrm{d}\nu(f) = \mu_t(B) = \int_B 1 \, \mathrm{d}\mu_t(g).$$

Hence, by uniqueness of the Radon–Nikodym derivative, $h_t(g) = \nu_t(\mathcal{F} \mid g) = 1$ μ_t-a.e. g. Thus, $\nu_t(\mathrm{d}f \mid g) = \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\nu(f)$ is a probability measure on $\mathcal{F}$. □

Following the previous lemma, define for μ_t-a.e. g,

$$Z_t(g) := \int_{\mathcal{F}} e^{-J(f)} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\nu(f) = \mathbb{E}_{f\sim\nu_t(\cdot|g)}[e^{-J(f)}]. \tag{58}$$

Lemma 3. *For almost every $t \in [0,1]$, $\mu_t' \ll \mu_t$ with*

$$\frac{d\mu_t'}{d\mu_t}(g) = \frac{Z_t(g)}{Z} \quad \textit{for } \mu_t\textit{-a.e.}g, \tag{59}$$

where $\mu_t' := \int_{\mathcal{F}} \mu_t^f \,\mathrm{d}\nu'(f)$. Moreover, $\mathbb{E}_{\mu_t}[Z_t] = Z < \infty$.

Proof. Let $A \in \mathcal{B}(\mathcal{F})$. Since $\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, e^{-J(f)} \geq 0$ for ν-a.e. f and μ_t-a.e. g, by Lemma 1 and Tonelli's theorem we have that

$$\begin{aligned}
\mu_t'(A) &= \int_{\mathcal{F}} \mu_t^f(A)\,\mathrm{d}\nu'(f) \\
&= \frac{1}{Z}\int_{\mathcal{F}} \mu_t^f(A)\, e^{-J(f)}\,\mathrm{d}\nu(f) \\
&= \frac{1}{Z}\int_{\mathcal{F}} \left[\int_A \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\mu_t(g)\right] e^{-J(f)}\,\mathrm{d}\nu(f) \\
&= \frac{1}{Z}\int_A \left[\int_{\mathcal{F}} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, e^{-J(f)}\,\mathrm{d}\nu(f)\right] \mathrm{d}\mu_t(g) \\
&= \int_A \frac{Z_t(g)}{Z}\,\mathrm{d}\mu_t(g).
\end{aligned}$$

If $\mu_t(A) = 0$, then $\int_A Z_t(g)\,\mathrm{d}\mu_t(g) = 0$ and therefore $\mu_t'(A) = 0$, proving that $\mu_t' \ll \mu_t$. As this applies for every $A \in \mathcal{B}(\mathcal{F})$, by the uniqueness of the Radon–Nikodym derivative we obtain equation 59.

Finally, taking $A = \mathcal{F}$ we get

$$1 = \mu_t'(\mathcal{F}) = \frac{1}{Z}\int_{\mathcal{F}} Z_t\,\mathrm{d}\mu_t,$$

thus $\mathbb{E}_{\mu_t}[Z_t] = Z < \infty$ since $Z \in (0,\infty)$. □

Lemma 4. *For almost every $t \in [0,1]$ and μ_t-a.e. g, $0 < Z_t(g) < \infty$.*

Proof. By Lemma 3 we have that $\mathbb{E}_{\mu_t}[Z_t] = Z < \infty$, then $Z_t(g) < \infty$ for μ_t-a.e. g. To show that it is positive, fix $g \in \mathcal{F}$ such that $\nu_t(\cdot \mid g)$ is a probability measure and $Z_t(g) < \infty$. Now, since $e^{-J(f)} > 0$ for all $f \in \mathcal{F}$ and $\nu_t(\cdot \mid g) \in \mathcal{P}(\mathcal{F})$, then

$$Z_t(g) = \int_{\mathcal{F}} e^{-J(f)}\,\mathrm{d}\nu_t(f \mid g) \geq 0.$$

To see that $Z_t(g)$ is nonzero, consider the sets $A_n = \{f \in \mathcal{F} : e^{-J(f)} \geq 1/n\}$ and notice that $A_n \nearrow \mathcal{F}$. Suppose that $Z_t(g) = 0$, then

$$0 = Z_t(g) = \int_{\mathcal{F}} e^{-J(f)}\,\mathrm{d}\nu_t(f \mid g) \geq \int_{A_n} e^{-J(f)}\,\mathrm{d}\nu_t(f \mid g) \geq \frac{1}{n}\nu_t(A_n \mid g) \geq 0.$$

Thus $\nu_t(A_n \mid g) = 0$ for all $n \in \mathbb{N}$; so by continuity, we get the contradiction

$$1 = \nu_t(\mathcal{F} \mid g) = \lim_{n\to\infty} \nu_t(A_n \mid g) = 0.$$

Concluding that $Z_t(g) > 0$ μ_t-a.e. g. □

Lemma 5. *$\mu_t^f \ll \mu_t'$ for ν-a.e. f and almost every $t \in [0,1]$ with*

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t'}(g) = \frac{Z}{Z_t(g)} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) \text{ for } \mu_t'\text{-a.e. } g. \tag{60}$$

Proof. Fix $f \in \mathcal{F}$ such that $\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}$ exists. By Lemma 4 we know that $Z_t(g) \in (0,\infty)$ for μ_t-a.e. g and since by Lemma 3 $\mu_t' \ll \mu_t$, it holds for μ_t'-a.e. g, so the right side of 60 is well defined. Now, using Lemma 3 we see that for $A \in \mathcal{B}(\mathcal{F})$

$$\int_A \frac{Z}{Z_t(g)} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\mu_t'(g) = \int_A \frac{Z}{Z_t(g)} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \frac{Z_t(g)}{Z}\, \mathrm{d}\mu_t(g) = \int_A \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\mu_t = \mu_t^f(A).$$

If $\mu_t'(A) = 0$, the previous identity gives $\mu_t^f(A) = 0$. Thus, we conclude that $\mu_t^f \ll \mu_t'$ and by uniqueness of the Radon–Nikodym derivative, it is given by equation 60. □

D.2 Main Results

The following Theorem is the same as the Theorem 1 stated in Section 2.4. We have included here for completeness.

Theorem 6 (General Guidance for FFM)**.** *Assume that the hypothesis from Theorem 2 holds and $\int_0^1 \int_{\mathcal{F}\times\mathcal{F}} e^{-J(f)} \|v_t^f(g)\|\, \mathrm{d}\mu_t^f(g)\, \mathrm{d}\nu(f)\, \mathrm{d}t < \infty$. Adding the guidance vector field $\gamma_t(g)$ to the original vector field $v_t(g)$ will form a modified vector field $v_t'(g)$ that generates the measure path*

$$\mu_t'(A) := \int_{\mathcal{F}} \mu_t^f(A)\, \mathrm{d}\nu'(f),\ A \in \mathcal{B}(\mathcal{F}),$$

as long as $\gamma_t(g)$ follows

$$\gamma_t(g) := \int_{\mathcal{F}} \left(\frac{e^{-J(f)}}{Z_t(g)} - 1 \right) v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}\, \mathrm{d}\nu(f), \tag{61}$$

where $Z_t(g) := \int_{\mathcal{F}} e^{-J(f)}\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f)$.

Proof. Theorem 2 asserts that the vector field $(v_t)_{t\in[0,1]}$ generates the path of measures $(\mu_t)_{t\in[0,1]}$, so in order to prove that $(v_t')_{t\in[0,1]}$ generates $(\mu_t')_{t\in[0,1]}$ we will use the same result for the target measure ν'.

First, by Lemma 5 we have that $\mu_t^f \ll \mu_t'$ for ν-a.e. f and almost every $t \in [0,1]$ and by Lemma 1 we know that $\nu \sim \nu'$, so this also holds for ν'-a.e. f. Moreover, since $\mathrm{d}\nu'(f) = \frac{1}{Z} e^{-J(f)} \mathrm{d}\nu(f)$,

$$\int_0^1 \int_{\mathcal{F}\times\mathcal{F}} \|v_t^f(g)\|\, \mathrm{d}\mu_t^f(g)\, \mathrm{d}\nu'(f)\, \mathrm{d}t = \frac{1}{Z} \int_0^1 \int_{\mathcal{F}\times\mathcal{F}} e^{-J(f)} \|v_t^f(g)\|\, \mathrm{d}\mu_t^f(g)\, \mathrm{d}\nu(f)\, \mathrm{d}t < \infty,$$

where the last term is finite because $Z > 0$ and $\int_0^1 \int_{\mathcal{F}\times\mathcal{F}} e^{-J(f)} \|v_t^f(g)\|\, \mathrm{d}\mu_t^f(g)\, \mathrm{d}\nu(f)\, \mathrm{d}t < \infty$. Then, the vector field

$$v_t'(g) = \int_{\mathcal{F}} v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t'}(g)\, \mathrm{d}\nu'(f)$$

generates the path of measures $(\mu_t')_{t\in[0,1]}$. By Lemma 5 the Radon–Nikodym derivative is given by

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t'}(g) = \frac{Z}{Z_t(g)} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g),$$

and by Lemma 1 $\mathrm{d}\nu'(f) = \frac{1}{Z} e^{-J(f)} \mathrm{d}\nu(f)$. Substituting these expressions we get

$$\begin{aligned} v_t'(g) &= \int_{\mathcal{F}} v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t'}(g)\, \mathrm{d}\nu'(f) \\ &= \int_{\mathcal{F}} v_t^f(g) \left[\frac{Z}{Z_t(g)} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) \right] \left(\frac{1}{Z} e^{-J(f)} \right) \mathrm{d}\nu(f) \\ &= \int_{\mathcal{F}} \frac{e^{-J(f)}}{Z_t(g)}\, v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f). \end{aligned}$$

Now define the vector field $\gamma_t(g) = v_t'(g) - v_t(g)$ which by construction is given by

$$\begin{aligned}\gamma_t(g) &= \int_{\mathcal{F}} \frac{e^{-J(f)}}{Z_t(g)}\, v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f) - \int_{\mathcal{F}} v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f) \\ &= \int_{\mathcal{F}} \left(\frac{e^{-J(f)}}{Z_t(g)} - 1 \right) v_t^f(g)\, \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\, \mathrm{d}\nu(f).\end{aligned}$$

Finally, we conclude that $v_t' = v_t + \gamma_t$ generates $(\mu_t')_{t\in[0,1]}$. □

D.2.1 Approximated Guidance

Although Theorem 6 provides an explicit formula for the guidance vector field, the marginalization over ν makes it computationally intractable. To address this issue, the following sections develop a rigorous framework for deriving a computationally feasible approximation of the guidance vector field.

Posterior Mean. Recall that the conditional flow is defined by

$$\phi_t^f(g_0) = \sigma_t^f g_0 + m_t^f, \quad g_0 \sim \mu_0 = \mathcal{N}(0, C_0),$$

where $m_t^f = tf$ and $\sigma_t^f = 1 - (1-\sigma_{\min})t$. Moreover, as σ_t^f does not depend on f it will be denoted as σ_t in the following sections. Using the abstract ODE it can be shown that, for all $g \in \mathcal{F}$, the conditional vector field is given by

$$\begin{aligned}v_t^f(g) &= \frac{(\sigma_t)'}{\sigma_t}(g - m_t^f) + \frac{\mathrm{d}}{\mathrm{d}t} m_t^f \\ &= \frac{(\sigma_t)'}{\sigma_t} g + \left(\frac{\mathrm{d}}{\mathrm{d}t} m_t^f - \frac{(\sigma_t)'}{\sigma_t} m_t^f \right) \\ &= \frac{-(1-\sigma_{\min})}{\sigma_t} g + \frac{f}{\sigma_t}.\end{aligned}$$

For fixed $g \in \mathcal{F}$ and taking $\mathbb{E}_{f\sim\nu_t(\cdot|g)}$ in the previous expression,

$$\mathbb{E}_{f\sim\nu_t(\cdot|g)}[v_t^f(g)] = \mathbb{E}_{f\sim\nu_t(\cdot|g)} \left[\frac{-(1-\sigma_{\min})}{\sigma_t} g + \frac{f}{\sigma_t} \right].$$

Since $v_t(g) = \int_{\mathcal{F}} v_t^f(g)\, \mathrm{d}\nu_t(f \mid g) = \mathbb{E}_{f\sim\nu_t(\cdot|g)}[v_t^f(g)]$, by Lemma 2

$$\begin{aligned}v_t(g) &= \mathbb{E}_{f\sim\nu_t(\cdot|g)} \left[\frac{-(1-\sigma_{\min})}{\sigma_t} g + \frac{f}{\sigma_t} \right] \\ &= \frac{-(1-\sigma_{\min})}{\sigma_t} g + \frac{1}{\sigma_t} \mathbb{E}_{f\sim\nu_t(\cdot|g)}[f].\end{aligned}$$

Thus, defining $\hat{g}_1(g) := \mathbb{E}_{f\sim\nu_t(\cdot|g)}[f]$ as the posterior mean of the endpoint f given a noisy intermediate state g at time t, we obtain

$$\hat{g}_1(g) = \sigma_t v_t(g) + (1-\sigma_{\min}) g. \tag{62}$$

Localization. Assume that J is Fréchet differentiable at $\hat{g}_1(g)$. Since $\mathcal{F}$ is a Hilbert space, the Riesz representation theorem states that the Fréchet derivative $DJ(\hat{g}_1(g)) \in \mathcal{F}^*$ is identified with a unique element $\nabla J(\hat{g}_1(g)) \in \mathcal{F}$ such that

$$DJ(\hat{g}_1(g))[h] = \langle \nabla J(\hat{g}_1(g)), h \rangle, \text{ for every } h \in \mathcal{F}.$$

Hence, by the Fréchet differentiability of J at $\hat{g}_1(g)$,

$$J(f) = J(\hat{g}_1(g)) + \langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g) \rangle + o(\|f - \hat{g}_1(g)\|).$$

Using the first order Fréchet expansion of J around $\hat{g}_1(g)$, followed by the first order Taylor expansion of the exponential function, we obtain the approximation

$$e^{-J(f)} \approx e^{-J(\hat{g}_1(g))} e^{-\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g) \rangle} \approx e^{-J(\hat{g}_1(g))} (1 - \langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g) \rangle).$$

Taking $\mathbb{E}_{f\sim\nu_t(\cdot|g)}$ in the previous expression and using the definition of $\hat{g}_1(g)$,

$$Z_t(g) = \mathbb{E}_{f\sim\nu_t(\cdot|g)}[e^{-J(f)}] \approx e^{-J(\hat{g}_1(g))}(1 - \langle \nabla J(\hat{g}_1(g)), \mathbb{E}_{f\sim\nu_t(\cdot|g)}[f] - \hat{g}_1(g)\rangle) = e^{-J(\hat{g}_1(g))}.$$

Therefore,

$$Z_t(g) \approx e^{-J(\hat{g}_1(g))}. \tag{63}$$

Substituting the above approximation into the guidance term 61 gives

$$\frac{e^{-J(f)}}{Z_t(g)} - 1 \approx \frac{e^{-J(\hat{g}_1(g))}(1 - \langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle)}{e^{-J(\hat{g}_1(g))}} - 1 = -\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle.$$

We define the localized approximation of the guidance term as $\gamma_t^{\text{loc}}(g)$. By Lemma 2, it can be written as

$$\gamma_t^{\text{loc}}(g) = -\mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle\, v_t^f(g)\Big]. \tag{64}$$

Using the expression for the conditional vector field, $v_t^f(g) = \frac{-(1-\sigma_{\min})}{\sigma_t} g + \frac{f}{\sigma_t}$, and replacing it into the localized guidance term, we obtain

$$\begin{aligned}
\gamma_t^{\text{loc}}(g) &= -\mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle\, v_t^f(g)\Big] \\
&= \frac{(1-\sigma_{\min})}{\sigma_t}\, g\, \mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle\Big] \\
&\quad - \frac{1}{\sigma_t}\mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle\, f\Big] \\
&= \frac{(1-\sigma_{\min})}{\sigma_t}\, g\, \langle \nabla J(\hat{g}_1(g)), \mathbb{E}_{f\sim\nu_t(\cdot|g)}[f] - \hat{g}_1(g)\rangle \\
&\quad - \frac{1}{\sigma_t}\mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle(\hat{g}_1(g) + (f - \hat{g}_1(g)))\Big] \\
&= -\frac{1}{\sigma_t}\mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle \nabla J(\hat{g}_1(g)), f - \hat{g}_1(g)\rangle(f - \hat{g}_1(g))\Big].
\end{aligned}$$

Defining $\Sigma_{1|t}(g)[h] := \mathbb{E}_{f\sim\nu_t(\cdot|g)}\Big[\langle h, f - \hat{g}_1(g)\rangle(f - \hat{g}_1(g))\Big]$ the localized guidance term can be written in terms of the conditional covariance operator $\Sigma_{1|t}(g)$,

$$\gamma_t^{\text{cov}}(g) = -\frac{1}{\sigma_t}\Sigma_{1|t}(g)[\nabla J(\hat{g}_1(g))]. \tag{65}$$

To ensure that the posterior mean $\hat{g}_1$ and the conditional covariance operator $\Sigma_{1|t}$ are well defined, we require the corresponding Bochner integrals to be finite. A natural sufficient condition is to assume that $\nu \in \mathcal{P}_2(\mathcal{F})$, i.e., ν is a probability measure on $\mathcal{F}$ with finite second moment. In the soliton setting, this condition follows provided the family of solutions is uniformly bounded in the norm of $\mathcal{F}$.

Since μ_t^f is a probability measure for almost every $t \in [0,1]$ and ν-a.e. f, by Tonelli's Theorem,

$$\begin{aligned}
\int_{\mathcal{F}} \mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f\|_{\mathcal{F}}^2\right] \mathrm{d}\mu_t(g) &= \int_{\mathcal{F}}\int_{\mathcal{F}} \|f\|_{\mathcal{F}}^2 \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\nu(f)\,\mathrm{d}\mu_t(g) \\
&= \int_{\mathcal{F}} \|f\|_{\mathcal{F}}^2 \left(\int_{\mathcal{F}} \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\mu_t(g)\right) \mathrm{d}\nu(f) \\
&= \int_{\mathcal{F}} \|f\|_{\mathcal{F}}^2\,\mathrm{d}\nu(f) < \infty.
\end{aligned}$$

Hence, $\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f\|_{\mathcal{F}}^2\right] < \infty$ for μ_t-a.e. g. Then, using the Cauchy–Schwarz inequality,

$$\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f\|_{\mathcal{F}}\right] \le \left(\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f\|_{\mathcal{F}}^2\right]\right)^{1/2} < \infty,$$

which guarantees that the posterior mean $\hat{g}_1(g)$ is well defined for μ_t-a.e. g.

Moreover,

$$\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f-\hat{g}_1(g)\|_{\mathcal{F}}^2\right]=\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f\|_{\mathcal{F}}^2\right]-\|\hat{g}_1(g)\|_{\mathcal{F}}^2\leq\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f\|_{\mathcal{F}}^2\right]<\infty.$$

Therefore, for every $h\in\mathcal{F}$,

$$\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|\langle h,f-\hat{g}_1(g)\rangle(f-\hat{g}_1(g))\|_{\mathcal{F}}\right]\leq\|h\|_{\mathcal{F}}\mathbb{E}_{f\sim\nu_t(\cdot|g)}\left[\|f-\hat{g}_1(g)\|_{\mathcal{F}}^2\right]<\infty.$$

Thus, the Bochner integral defining the conditional covariance operator $\Sigma_{1|t}(g)[h]$ is well defined for every $h\in\mathcal{F}$ and μ_t-a.e. g.

Infinite-dimensional Jacobian Trick. We now derive an infinite-dimensional analogue of the Jacobian Trick used in Feng et al. (2025) to conclude the covariance-guided approximation that can be computationally implemented.

Lemma 6 (Logarithmic derivative of μ_t^f)**.** *Consider a non-negative, self-adjoint and trace-class covariance operator C_0. Let $k\in\mathcal{F}$ and set $h=C_0k$. For every $f\in\mathcal{F}$ the measure $\mu_t^f=\mathcal{N}(tf,\sigma_t^2C_0)$ is Fomin differentiable in the direction h. Its logarithmic derivative is given by*

$$\beta_h^f(g)=-\frac{1}{(\sigma_t)^2}\langle g-tf,k\rangle. \tag{66}$$

Moreover, $\beta_h^f\in L^p(\mu_t^f)$ for every $0<p<\infty$ and

$$\|d_h\mu_t^f\|_{TV}=\int_{\mathcal{F}}|\beta_h^f|\,\mathrm{d}\mu_t^f=\sqrt{\frac{2}{\pi}}\frac{\langle C_0k,k\rangle^{1/2}}{\sigma_t}. \tag{67}$$

Proof. Let $\rho_t:=\mathcal{N}(0,\sigma_t^2C_0)$ with covariance operator $R_{\rho_t}=\sigma_t^2C_0$. Since $h=C_0k$, we can write

$$h=R_{\rho_t}\left(\frac{k}{\sigma_t^2}\right).$$

Thus, h belongs to the Cameron–Martin space of ρ_t.

By definition of the measurable linear functional associated with a Cameron–Martin direction Bogachev (2010), if

$$h=R_{\rho_t}l,\quad l=\frac{k}{\sigma_t^2}\in\mathcal{F},$$

then the corresponding measurable linear functional can be represented by the continuous linear functional l. Hence,

$$\hat{h}_{\rho_t}(g)=\frac{1}{\sigma_t^2}\langle k,g\rangle.$$

Therefore, Theorem 3.1.9 of Bogachev (2010) gives

$$d_h\rho_t=-\hat{h}_{\rho_t}\rho_t,$$

and in consequence

$$\beta_h^{\rho_t}(g)=-\frac{1}{\sigma_t^2}\langle k,g\rangle. \tag{68}$$

Now, for a measure μ and $a\in\mathcal{F}$, let

$$\mu_a(A):=\mu(A-a),\quad A\in\mathcal{B}(\mathcal{F}).$$

Moreover, by definition of the Fomin derivative

$$d_h\mu_a(A)=\lim_{s\to0}\frac{\mu_a(A+sh)-\mu_a(A)}{s}=\lim_{s\to0}\frac{\mu(A+sh-a)-\mu(A-a)}{s}=d_h\mu(A-a).$$

Thus

$$d_h\mu_a=(d_h\mu)_a.$$

Furthermore, as $d_h\mu(\mathrm{d}g)=\beta_h^\mu(g)\mu(\mathrm{d}g)$,

$$d_h\mu_a(A)=\int_A\beta_h^\mu(g-a)\,\mathrm{d}\mu_a(g),$$

so $\beta_h^{\mu_a}(g) = \beta_h^{\mu}(g-a)$.

Since $\mu_t^f = (\rho_t)_{tf}$, expression 68 and the previous identity gives

$$\beta_h^f(g) = \beta_h^{\rho_t}(g - tf) = -\frac{1}{\sigma_t^2}\langle g - tf, k\rangle.$$

To prove the integrability property notice that, under μ_t^f, the element $g - tf$ is centered and has covariance operator $\sigma_t^2 C_0$. Then, by Fernique's theorem, all moments are finite. Hence, $\beta_h^f \in L^p(\mu_t^f)$ for every $0 < p < \infty$.

Finally, the definition of total variation and the first moment formula for a Gaussian measure gives

$$\begin{aligned}
\|d_h\mu_t^f\|_{\mathrm{TV}} &= \int_{\mathcal{F}} |\beta_h^f(g)| \,\mathrm{d}\mu_t^f(g) \\
&= \frac{1}{\sigma_t^2}\int_{\mathcal{F}} |\langle g - tf, k\rangle| \,\mathrm{d}\mu_t^f(g) \\
&= \frac{1}{\sigma_t^2}\mathbb{E}_{g\sim\mu_t^f}[|\langle g - tf, k\rangle|] \\
&= \frac{1}{\sigma_t^2}\mathbb{E}[|X|], \quad X \sim \mathcal{N}(0, \sigma_t^2\langle C_0 k, k\rangle) \\
&= \frac{1}{\sigma_t^2}\sqrt{\frac{2}{\pi}}\sigma_t\langle C_0 k, k\rangle^{1/2} \\
&= \sqrt{\frac{2}{\pi}}\frac{\langle C_0 k, k\rangle^{1/2}}{\sigma_t}.
\end{aligned}$$

□

Lemma 7 (Logarithmic derivative of μ_t)**.** *Let $k \in \mathcal{F}$ and $h = C_0 k$. Under the posterior differentiation hypotheses of Appendix B, μ_t is Fomin differentiable along h with $d_h\mu_t = \int_{\mathcal{F}} d_h\mu_t^f \,\mathrm{d}\nu(f)$ and $\|d_h\mu_t\|_{TV} \le \sqrt{\frac{2}{\pi}}\frac{\langle C_0 k,k\rangle^{1/2}}{\sigma_t}$. Moreover, its logarithmic derivative satisfies*

$$\beta_h^{\mu_t}(g) = \mathbb{E}_{f\sim\nu_t(\cdot|g)}[\beta_h^f(g)] = -\frac{1}{\sigma_t^2}\langle g - t\hat{g}_1(g), k\rangle, \tag{69}$$

for μ_t-a.e. g.

Proof. By Lemma 6, for every $f \in \mathcal{F}$, μ_t^f is Fomin differentiable along h and

$$\beta_h^f(g) = -\frac{1}{\sigma_t^2}\langle g - tf, k\rangle,$$

with

$$\|d_h\mu_t^f\|_{\mathrm{TV}} = M := \sqrt{\frac{2}{\pi}}\frac{\langle C_0 k, k\rangle^{1/2}}{\sigma_t}.$$

Define

$$\eta(A) := \int_{\mathcal{F}} d_h\mu_t^f(A) \,\mathrm{d}\nu(f), \quad A \in \mathcal{B}(\mathcal{F}).$$

The map $f \mapsto d_h\mu_t^f(A)$ is Borel measurable since

$$d_h\mu_t^f(A) = -\frac{1}{\sigma_t^2}\int_{\mathcal{F}} \mathbb{1}_A(x + tf)\langle x, k\rangle \,\mathrm{d}\rho_t(x),$$

where $\rho_t = \mathcal{N}(0, \sigma_t^2 C_0)$. Moreover,

$$|d_h\mu_t^f(A)| \le \|d_h\mu_t^f\|_{\mathrm{TV}} \le M.$$

Thus η is a finite signed measure, and $\|\eta\|_{\mathrm{TV}} \leq M$. Proposition 3.3.4 of Bogachev (2010) and the Fomin differentiability of μ_t^f give

$$\mu_t^f(A + sh) - \mu_t^f(A) = \int_0^s d_h\mu_t^f(A + uh)\,\mathrm{d}u.$$

Then, integrating with respect to ν, the integrand is bounded, and Fubini's theorem yields

$$\begin{aligned}\mu_t(A + sh) - \mu_t(A) &= \int_{\mathcal{F}}\int_0^s d_h\mu_t^f(A + uh)\,\mathrm{d}u\,\mathrm{d}\nu(f)\\ &= \int_0^s\int_{\mathcal{F}} d_h\mu_t^f(A + uh)\,\mathrm{d}\nu(f)\,\mathrm{d}u\\ &= \int_0^s \eta(A + uh)\,\mathrm{d}u.\end{aligned}$$

Furthermore, Corollary 3.3.3 of Bogachev (2010) asserts that the Fomin differentiability of μ_t^f implies that $d_h\mu_t^f$ is continuous along h,

$$\|(d_h\mu_t^f)_{sh} - d_h\mu_t^f\|_{\mathrm{TV}} \to 0 \quad \text{as } s \to 0,$$

We need to show that $\|\eta_{sh} - \eta\|_{\mathrm{TV}} \to 0 \quad$ as $s \to 0$. By the definition of η,

$$\begin{aligned}|(\eta_{sh} - \eta)(A)| &= \left|\int_{\mathcal{F}} (d_h\mu_t^f)_{sh}(A) - d_h\mu_t^f(A)\,\mathrm{d}\nu(f)\right|\\ &\leq \int_{\mathcal{F}} |(d_h\mu_t^f)_{sh}(A) - d_h\mu_t^f(A)|\,\mathrm{d}\nu(f)\\ &\leq \int_{\mathcal{F}} \|(d_h\mu_t^f)_{sh} - d_h\mu_t^f\|_{\mathrm{TV}}\,\mathrm{d}\nu(f).\end{aligned}$$

Taking the supremum, we obtain that

$$\|\eta_{sh} - \eta\|_{\mathrm{TV}} \leq \int_{\mathcal{F}} \|(d_h\mu_t^f)_{sh} - d_h\mu_t^f\|_{\mathrm{TV}}\,\mathrm{d}\nu(f),$$

and since for every $f \in \mathcal{F}$ $\|(d_h\mu_t^f)_{sh} - d_h\mu_t^f\|_{\mathrm{TV}} \leq 2M$, the dominated convergence theorem concludes the continuity of η along h.

Hence, Proposition 3.3.4 of Bogachev (2010) implies that μ_t is Fomin differentiable along h and

$$d_h\mu_t = \eta = \int_{\mathcal{F}} d_h\mu_t^f\,\mathrm{d}\nu(f).$$

The bound of its derivative follows from

$$\|d_h\mu_t\|_{\mathrm{TV}} = \|\eta\|_{\mathrm{TV}} \leq M = \sqrt{\frac{2}{\pi}}\frac{\langle C_0 k, k\rangle^{1/2}}{\sigma_t}.$$

To prove that the logarithmic derivative satisfies 69, recall from Lemma 2 that there exists a jointly Borel measurable version of $\frac{d\mu_t^f}{d\mu_t}(g)$.

Since $\mu_t^f(dg) = \frac{d\mu_t^f}{d\mu_t}(g)\mu_t(dg)$, the Fomin derivative can be written in the form $d_h\mu_t^f(dg) = \beta_h^f(g)\frac{d\mu_t^f}{d\mu_t}(g)\mu_t(dg)$.

Moreover,

$$\int_{\mathcal{F}}\int_{\mathcal{F}} |\beta_h^f(g)|\frac{d\mu_t^f}{d\mu_t}(g)\,\mathrm{d}\mu_t(g)\,\mathrm{d}\nu(f) = \int_{\mathcal{F}} \|d_h\mu_t^f\|_{\mathrm{TV}}\,\mathrm{d}\nu(f) = M < \infty.$$

Hence, Fubini's theorem applies, and for every $A \in \mathcal{B}(\mathcal{F})$,

$$d_h\mu_t(A) = \int_{\mathcal{F}} d_h\mu_t^f(A)\,\mathrm{d}\nu(f) = \int_A\int_{\mathcal{F}} \beta_h^f(g)\frac{d\mu_t^f}{d\mu_t}(g)\,\mathrm{d}\nu(f)\,\mathrm{d}\mu_t(g).$$

By the uniqueness of the Radon–Nikodym derivative,

$$\beta_h^{\mu_t}(g) = \int_{\mathcal{F}} \beta_h^f(g) \frac{d\mu_t^f}{d\mu_t}(g)\, \mathrm{d}\nu(f) = \mathbb{E}_{f\sim\nu_t(\cdot|g)}[\beta_h^f(g)],$$

for μ_t-a.e. g.

Replacing expression 66 from Lemma 6 we obtain

$$\begin{aligned}
\beta_h^{\mu_t}(g) &= -\frac{1}{\sigma_t^2}\langle g, k\rangle + \frac{t}{\sigma_t^2}\mathbb{E}_{f\sim\nu_t(\cdot|g)}[\langle f, k\rangle] \\
&= -\frac{1}{\sigma_t^2}\langle g, k\rangle + \frac{t}{\sigma_t^2}\langle \mathbb{E}_{f\sim\nu_t(\cdot|g)}[f], k\rangle \\
&= -\frac{1}{\sigma_t^2}\langle g, k\rangle + \frac{t}{\sigma_t^2}\langle \hat{g}_1(g), k\rangle \\
&= -\frac{1}{\sigma_t^2}\langle g - t\hat{g}_1(g), k\rangle,
\end{aligned}$$

for μ_t-a.e. g. □

Lemma 8. *Let $\mathcal{H} := C_0^{1/2}(\mathcal{F})$ be the Cameron–Martin space of $\mathcal{N}(0, C_0)$, and let $(e_i)_{i\geq 1}$ be an orthonormal eigenbasis of C_0, with $C_0 e_i = \lambda_i e_i$. For $u \in \mathcal{H}$, define*

$$S_n^u(g) := \sum_{i=1}^{n} \lambda_i^{-1}\langle u, e_i\rangle\langle g, e_i\rangle,$$

and $\mathcal{D}_u := \{g \in \mathcal{F} : (S_n^u(g))_{n\geq 1}$ converges$\}$. For $g \in \mathcal{D}_u$, define $\hat{u}(g) := \lim_{n\to\infty} S_n^u(g)$, and set $\hat{u}(g) = 0$ for $g \notin \mathcal{D}_u$. Then,

1. *$\mathcal{D} := \{(u, g) \in \mathcal{H} \times \mathcal{F} : g \in \mathcal{D}_u\}$ is Borel and $(u, g) \mapsto \hat{u}(g)$ is jointly measurable.*
2. *For every $u \in \mathcal{H}$, $\rho_t(\mathcal{D}_u) = 1$.*
3. *For every $u \in \mathcal{H}, g \in \mathcal{D}_u, k \in \mathcal{F}$ and $s \in \mathbb{R}$,*

$$g + sC_0k \in \mathcal{D}_u, \quad \hat{u}(g + sC_0k) = \hat{u}(g) + s\langle u, k\rangle.$$

Proof. For fixed n, the map $(u, g) \mapsto \lambda_i^{-1}\langle u, e_i\rangle\langle g, e_i\rangle$ is jointly continuous since it is a finite sum of continuous linear functionals. On the other hand, in $\mathbb{R}$, $(S_n^u(g))_{n\geq 1}$ converges if and only if $(S_n^u(g))_{n\geq 1}$ is Cauchy. By definition, for $g \in \mathcal{D}_u$,

$$\mathcal{D} = \bigcap_{r\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{n,m\geq N} \left\{(u, g) \in \mathcal{H} \times \mathcal{F} : |S_n^u(g) - S_m^u(g)| < \frac{1}{r}\right\}.$$

Hence, $\mathcal{D} \in \mathcal{B}(\mathcal{H} \times \mathcal{F})$. Furthermore, $\mathcal{D}_u = \{g \in \mathcal{F} : (u, g) \in \mathcal{D}\} \in \mathcal{B}(\mathcal{F})$. Since each S_n^u is Borel and $\mathcal{D}$ is Borel, the function $\hat{u}(g) = \mathbb{1}_D \lim_{n\to\infty} S_n^u(g)$ is jointly Borel.

For the second statement, let $G \sim \rho_t = \mathcal{N}(0, \sigma_t^2 C_0)$. Since ρ_t is a centered Gaussian measure, each coordinate $\langle G, e_i\rangle$ is a centered Gaussian random variable. Moreover,

$$\mathrm{Var}(\langle G, e_i\rangle) = \mathbb{E}[\langle G, e_i\rangle^2] = \langle \sigma_t^2 C_0 e_i, e_i\rangle = \sigma_t^2 \lambda_i.$$

Consider the random variables $X_i := \lambda_i^{-1}\langle u, e_i\rangle\langle G, e_i\rangle$. Since G is Gaussian, the vector $(X_1, \ldots, X_n)$ is jointly Gaussian and for $i \neq j$,

$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}[\langle G, e_i\rangle\langle G, e_j\rangle] = \langle \sigma_t^2 C_0 e_i, e_j\rangle = \sigma_t^2 \lambda_i \langle e_i, e_j\rangle = 0.$$

Thus, X_i are independent, centered and satisfy, for $u \in \mathcal{H}$,

$$\sum_{i=1}^{\infty} \mathrm{Var}(X_i) = \sigma_t^2 \sum_{i=1}^{\infty} \lambda_i^{-1}\langle u, e_i\rangle^2 = \sigma_t^2 \|u\|_{\mathcal{H}}^2 < \infty.$$

Therefore, by Kolmogorov's two-series theorem, $\sum_{i=1}^{\infty} X_i$ converges almost surely, proving that $\rho_t(\mathcal{D}_u) = 1$.

Finally, for $h = C_0 k$,

$$S_n^u(g + sh) = S_n^u(g) + s\sum_{i=1}^{n} \lambda_i^{-1}\langle u, e_i\rangle\langle C_0 k, e_i\rangle$$
$$= S_n^u(g) + s\sum_{i=1}^{n}\langle u, e_i\rangle\langle k, e_i\rangle.$$

Since $u, k \in \mathcal{F}$, Parseval's identity gives $\sum_{i=1}^{\infty}\langle u, e_i\rangle\langle k, e_i\rangle = \langle u, k\rangle$. Hence, for $g \in \mathcal{D}_u$,

$$S_n^u(g + sC_0 k) \longrightarrow \hat{u}(g) + s\langle u, k\rangle.$$

Therefore, $g + sC_0 k \in \mathcal{D}_u$ and $\hat{u}(g + sC_0 k) = \hat{u}(g) + s\langle u, k\rangle$. □

Lemma 9. *For $t \in (0, 1]$ consider $\rho_t := \mathcal{N}(0, \sigma_t^2 C_0)$. Denote by $\mathcal{H} := \mathcal{H}(C_0)$ the Cameron–Martin space of $\mathcal{N}(0, C_0)$. Assume that $\nu(\mathcal{H}) = 1$ and for every $c \geq 0$, $\int_{\mathcal{F}} e^{c\|f\|}\,\mathrm{d}\nu(f) < \infty$. Fix $t \in (0, 1]$, for $f \in \mathcal{H}$ and $g \in \mathcal{D}_f$, define*

$$\Psi_f(g) := \exp\left(\frac{t}{\sigma_t^2}\hat{f}(g) - \frac{t^2}{2\sigma_t^2}\|f\|_{\mathcal{H}}^2\right), \tag{70}$$

and set $\Psi_f(g) = 0$ otherwise. For $c \geq 0$, define

$$I_c(g) := \int_{\mathcal{F}} \Psi_f(g) e^{c\|f\|}\,\mathrm{d}\nu(f), \tag{71}$$

and $\mathcal{G}_t := \{g \in \mathcal{F} : 0 < I_c(g) < \infty, \forall c \geq 0\}$. Then,

1. *$\mathcal{G}_t \in \mathcal{B}(\mathcal{F})$ and $\rho_t(\mathcal{G}_t) = \mu_t(\mathcal{G}_t) = 1$. Moreover, $\mathcal{G}_t + C_0(\mathcal{F}) = \mathcal{G}_t$.*

2. *For every $f \in \mathcal{H}, \mu_t^f \sim \rho_t$, $\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\rho_t}(g) = \Psi_f(g)$ for ρ_t-a.e. g. Furthermore, $\mu_t \sim \rho_t$, $\frac{\mathrm{d}\mu_t}{\mathrm{d}\rho_t}(g) = I_0(g)$ for ρ_t-a.e. g. Consequently, the function*

$$r_t(f, g) := \mathbb{1}_{\mathcal{G}_t}(g)\frac{\Psi_f(g)}{I_0(g)},$$

   defines a jointly Borel measurable version of $\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}$ for ν-a.e. f and for every $f \in \mathcal{H}$.

3. *For every $c \geq 0, \mathbb{E}_{f\sim\nu_t(\cdot|g)}[e^{c\|f\|}] = \frac{I_c(g)}{I_0(g)} < \infty$.*

4. *For every $g \in \mathcal{G}_t, k \in \mathcal{F}, s \in \mathbb{R}$ and $f \in \mathcal{H}$,*

$$\Psi_f(g + sC_0 k) = \Psi_f(g)e^{a_f s}, \tag{72}$$

   where $a_f := \frac{t}{\sigma_t^2}\langle f, k\rangle$, and $g + sC_0 k \in \mathcal{G}_t$. Consequently,

$$I_0(g + sC_0 k) = \int_{\mathcal{F}} \Psi_f(g) e^{a_f s}\,\mathrm{d}\nu(f) \tag{73}$$

Proof. By Lemma 8 (1), the map $(f, g) \mapsto \Psi_f(g)$ is jointly Borel measurable. Hence, by Tonelli's theorem, I_c is Borel measurable for every $c \geq 0$.

Fix $c \geq 0$ and $f \in \mathcal{H}$. Cameron–Martin theorem gives

$$\int_{\mathcal{F}} \Psi_f(g)\,\mathrm{d}\rho_t(g) = 1. \tag{74}$$

Moreover, since $\nu(\mathcal{H}) = 1$ it follows that

$$\begin{aligned}\int_{\mathcal{F}} I_c(g)\,\mathrm{d}\rho_t(g) &= \int_{\mathcal{F}}\int_{\mathcal{F}} \Psi_f(g) e^{c\|f\|}\,\mathrm{d}\nu(f)\,\mathrm{d}\rho_t(g)\\ &= \int_{\mathcal{F}} e^{c\|f\|}\left(\int_{\mathcal{F}} \Psi_f(g)\,\mathrm{d}\rho_t(g)\right)\mathrm{d}\nu(f)\\ &= \int_{\mathcal{F}} e^{c\|f\|}\,\mathrm{d}\nu(f) < \infty.\end{aligned}$$

Then, $I_c(g) < \infty$ for ρ_t-a.e. g. Since I_c is Borel measurable and nondecreasing, $\mathcal{G}_t$ can be written as

$$\mathcal{G}_t = \{I_0 > 0\} \cap \bigcap_{n=1}^{\infty} \{I_n < \infty\}.$$

Thus, $\mathcal{G}_t \in \mathcal{B}(\mathcal{F})$.

By Lemma 8 (2), for every fixed $f \in \mathcal{H}$, $\rho_t(\mathcal{D}_f) = 1$, i.e., $\rho_t(\mathcal{D}_f^c) = 0$. Furthermore, Lemma 8 (1) asserts that

$$\mathcal{D} = \{(f, g) \in \mathcal{H} \times \mathcal{F} : g \in \mathcal{D}_f\}$$

is Borel measurable. Define, $E := \mathcal{D}^c$, which is also Borel measurable. The definition of the product measure gives

$$(\nu \otimes \rho_t)(E) = \int_{\mathcal{F}} \rho_t(\{g \in \mathcal{F} : (f, g) \in E\})\,\mathrm{d}\nu(f) = \int_{\mathcal{F}} \rho_t(\mathcal{D}_f^c)\,\mathrm{d}\nu(f) = 0.$$

Then,

$$0 = (\nu \otimes \rho_t)(E) = \int_{\mathcal{F}} \nu(\{f \in \mathcal{H} : g \notin \mathcal{D}_f\})\,\mathrm{d}\rho_t(g),$$

therefore, $\nu(\{f \in \mathcal{H} : g \notin \mathcal{D}_f\}) = 0$ for ρ_t-a.e. g. This means that for ρ_t-a.e. g, $g \in \mathcal{D}_f$ for ν-a.e. f, where by definition $\Psi_f(g) > 0$. Since, ν is a probability measure,

$$I_0(g) = \int_{\mathcal{F}} \Psi_f(g)\,\mathrm{d}\nu(f) > 0.$$

Consider the sets

$$N_1 := \bigcup_{n=1}^{\infty} \{g \in \mathcal{F} : I_n(g) = \infty\} \text{ and } N_2 := \{g \in \mathcal{F} : I_0(g) = 0\}. \tag{75}$$

By the previous results we have that $\rho_t(N_1) \le \sum_{n=1}^{\infty} \rho_t(\{I_n = \infty\}) = 0$ and $\rho_t(N_2) = 0$. Therefore, if $g \notin N_1 \cup N_2$, $0 < I_0(g) \le I_c(g) < \infty$ for every $n \in \mathbb{N}$. Let $c \ge 0$ and choose $n \in \mathbb{N}$ such that $n \ge c$. As $c \mapsto I_c$ is increasing, $I_c(g) \le I_n(g) < \infty$, and $I_c(g) \ge I_0(g) > 0$. Thus, $0 < I_c(g) < \infty$ for every $c \ge 0$ and then $g \in \mathcal{G}_t$. Moreover, $\mathcal{G}_t^c \subset N_1 \cup N_2$ and $\rho_t(\mathcal{G}_t^c) \le \rho_t(N_1) + \rho_t(N_2) = 0$, concluding that $\rho_t(\mathcal{G}_t) = 1$.

Since $f \in \mathcal{H}$, the Cameron–Martin theorem gives $\mu_t^f \sim \rho_t$ with

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\rho_t}(g) = \exp\left(\frac{t}{\sigma_t^2}\hat{f}(g) - \frac{t^2}{2\sigma_t^2}\|f\|_{\mathcal{H}}^2\right) = \Psi_f(g) \quad \text{for } \rho_t\text{-a.e. } g.$$

Now, for $A \in \mathcal{B}(\mathcal{F})$,

$$\begin{aligned}\mu_t(A) &= \int_{\mathcal{F}} \mu_t^f(A)\,\mathrm{d}\nu(f)\\ &= \int_{\mathcal{F}} \left(\int_A \Psi_f(g)\,\mathrm{d}\rho_t(g)\right)\mathrm{d}\nu(f)\\ &= \int_A \left(\int_{\mathcal{F}} \Psi_f(g)\,\mathrm{d}\nu(f)\right)\mathrm{d}\rho_t(g)\\ &= \int_A I_0(g)\,\mathrm{d}\rho_t(g).\end{aligned}$$

Thus, $\frac{\mathrm{d}\mu_t}{\mathrm{d}\rho_t} = I_0$. Since $I_0(g) > 0$ for ρ_t-a.e., we obtain $\mu_t \sim \rho_t$. For $g \in \mathcal{G}_t, I_0(g) > 0$ and $\frac{\Psi_f(g)}{I_0(g)}$ is well defined. Therefore, the function $r_t(f,g) := \mathbb{1}_{\mathcal{G}_t}(g)\frac{\Psi_f(g)}{I_0(g)}$ is jointly Borel measurable by the measurability of $\mathcal{G}_t$ and the joint measurability of Ψ_f and I_0. Moreover, for every $A \in \mathcal{B}(\mathcal{F})$,

$$\int_A r_t(f,g)\,\mathrm{d}\mu_t(g) = \int_A \frac{\Psi_f(g)}{I_0(g)} I_0(g)\,\mathrm{d}\rho_t(g) = \int_A \Psi_f(g)\,\mathrm{d}\rho_t(g) = \mu_t^f(A),$$

hence r_t is a version of $\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}$.

Now, by definition the probability measure $\nu_t(\cdot|g)$ can be written in terms of r_t as $\nu_t(df|g) = r_t(f,g)\,\mathrm{d}\nu(f)$. Furthermore,

$$\begin{aligned}\mathbb{E}_{f\sim\nu_t(\cdot|g)}[e^{c\|f\|}] &= \int_{\mathcal{F}} e^{c\|f\|} r_t(f,g)\,\mathrm{d}\nu(f)\\ &= \frac{1}{I_0(g)}\int_{\mathcal{F}} \Psi_f(g) e^{c\|f\|}\,\mathrm{d}\nu(f)\\ &= \frac{I_c(g)}{I_0(g)} < \infty.\end{aligned}$$

For the last statement, let $h = C_0 k$. By Lemma 8 (3), $\hat{f}(g + sC_0k) = \hat{f}(g) + s\langle f, k\rangle$ for $g \in \mathcal{D}_f$. Hence, defining $a_f := \frac{t}{\sigma_t^2}\langle f, k\rangle$, we obtain

$$\Psi_f(g + sC_0k) = \Psi_f(g)\exp\left(\frac{st}{\sigma_t^2}\langle f, k\rangle\right) = \Psi_f(g)e^{a_f s}.$$

Setting $\alpha := \frac{t\|k\|}{\sigma_t^2}$, $|a_f| \le \frac{t\|k\|}{\sigma_t^2}\|f\| = \alpha\|f\|$. Therefore, for every $c \ge 0$,

$$\begin{aligned}I_c(g + sC_0k) &= \int_F \Psi_f(g) e^{a_f s} e^{c\|f\|}\,\mathrm{d}\nu(f)\\ &\le \int_F \Psi_f(g) e^{(c+\alpha|s|)\|f\|}\,\mathrm{d}\nu(f)\\ &= I_{c+\alpha|s|}(g) < \infty.\end{aligned}$$

Applying Lemma 8 (3) with s and $-s$ gives $g \in \mathcal{D}_f \iff g + sC_0k \in \mathcal{D}_f$ for every $f \in \mathcal{H}$. Hence, the identity $\Psi_f(g + sC_0k) = \Psi_f(g)e^{a_f s}$ holds for all $f \in \mathcal{H}$, and since $e^{a_f s} > 0$, the functions $f \mapsto \Psi_f(g + sC_0k)$ and $f \mapsto \Psi_f(g)$ have the same zero set. As $I_0(g) > 0$,

$$\nu(Q) := \nu(\{f \in \mathcal{H} : \Psi_f(g) > 0\}) > 0,$$

and therefore

$$I_0(g + sC_0k) = \int_{\mathcal{F}} \Psi_f(g) e^{a_f s}\,\mathrm{d}\nu(f) \ge \int_Q \Psi_f(g) e^{a_f s}\,\mathrm{d}\nu(f) > 0$$

Thus,

$$g + sC_0k \in \mathcal{G}_t.$$

Finally,

$$I_0(g + sC_0k) = \int_{\mathcal{F}} \Psi_f(g) e^{a_f s}\,\mathrm{d}\nu(f).$$

□

Lemma 10 (Directional derivative of $\mathrm{d}\mu_t^f/\mathrm{d}\mu_t$)**.** *Fix $t \in (0,1]$. Let $g \in \mathcal{G}_t, k \in \mathcal{F}$ and $h = C_0k$. Then, for ν-a.e. $f \in \mathcal{F}$, the directional derivative $D_h\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)$ exists and*

$$D_h\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) = \frac{t}{\sigma_t^2}\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\langle f - \hat{g}_1(g), k\rangle. \tag{76}$$

Equivalently,

$$D_h\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) = \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\left(\beta_h^f(g) - \beta_h^{\mu_t}(g)\right). \tag{77}$$

Proof. Set $a_f := \frac{t}{\sigma_t^2}\langle f, k\rangle$. Lemma 9 provides two useful results. First, the Radon–Nikodym derivative can be written as

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) = \frac{\Psi_f(g)}{I_0(g)}.$$

Second, Ψ_f satisfies the translation identity,

$$\Psi_f(g + sh) = \Psi_f(g)e^{a_f s}.$$

Then, defining $N_g(s) := \int_{\mathcal{F}} \Psi_{f'}(g)e^{a_{f'}s}\,\mathrm{d}\nu(f')$, we obtain

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g + sh) = \frac{\Psi_f(g + sh)}{I_0(g + sh)} = \frac{\Psi_f(g)e^{a_f s}}{N_g(s)}.$$

Since $h = C_0 k$ and $\mathcal{G}_t + C_0(\mathcal{F}) = \mathcal{G}_t$, the previous expression is well defined over the line $g + \mathbb{R}h$. The next step is to show that

$$N_g'(0) = \int_{\mathcal{F}} a_f \Psi_f(g)\,\mathrm{d}\nu(f). \tag{78}$$

By definition,

$$\begin{aligned}
N_g'(0) &= \lim_{s\to 0} \frac{N_g(s) - N_g(0)}{s} \\
&= \lim_{s\to 0} \frac{1}{s}\left(\int_{\mathcal{F}} \Psi_{f'}(g)e^{a_{f'}s}\,\mathrm{d}\nu(f') - \int_{\mathcal{F}} \Psi_{f'}(g)\,\mathrm{d}\nu(f')\right) \\
&= \lim_{s\to 0} \int_{\mathcal{F}} \Psi_{f'}(g)\frac{e^{a_{f'}s} - 1}{s}\,\mathrm{d}\nu(f') \\
&= \int_{\mathcal{F}} \Psi_{f'}(g) \lim_{s\to 0} \frac{e^{a_{f'}s} - 1}{s}\,\mathrm{d}\nu(f') \\
&= \int_{\mathcal{F}} a_{f'}\Psi_{f'}(g)\,\mathrm{d}\nu(f').
\end{aligned}$$

The interchange between the limit and the integral is justified by the dominated convergence theorem.

Indeed, set $\alpha := \frac{t\|k\|}{\sigma_t^2}$. The Cauchy-Schwarz inequality gives

$$|a_f| = \frac{t}{\sigma_t^2}|\langle f, k\rangle| \le \alpha\|f\|.$$

Consider $0 < |s| \le 1$. By the mean value theorem,

$$\left|\frac{e^{a_f s} - 1}{s}\right| \le |a_f|e^{|a_f s|}.$$

Since $|s| \le 1$, $|a_f s| \le |a_f| \le \alpha\|f\|$, hence

$$\left|\frac{e^{a_f s} - 1}{s}\right| \le \alpha\|f\|e^{\alpha\|f\|} \le \alpha e^{(\alpha+1)\|f\|},$$

where the last inequality follows from the fact that $x \le e^x$ for every $x \in \mathbb{R}$. Thus,

$$\left|\Psi_f(g)\frac{e^{a_f s} - 1}{s}\right| \le \alpha\Psi_f(g)e^{(\alpha+1)\|f\|},$$

and since $g \in \mathcal{G}_t$

$$\int_{\mathcal{F}} \Psi_f(g)e^{(\alpha+1)\|f\|}\,\mathrm{d}\nu(f) < \infty.$$

Moreover, $N_g(0) = \int_{\mathcal{F}} \Psi_f(g)\,\mathrm{d}\nu(f) = I_0(g) > 0$.

By the definition of the directional derivative,

$$\begin{aligned} D_h \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) &= \frac{d}{ds}\left(\frac{\Psi_f(g)e^{a_f s}}{N_g(s)}\right)\Big|_{s=0} \\ &= \frac{a_f \Psi_f(g) N_g(0) - \Psi_f(g) N_g'(0)}{N_g(0)^2} \\ &= \frac{\Psi_f(g)}{N_g(0)}\left(a_f - \frac{N_g'(0)}{N_g(0)}\right) \\ &= \frac{\Psi_f(g)}{I_0(g)}\left(a_f - \frac{N_g'(0)}{N_g(0)}\right) \\ &= \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\left(a_f - \frac{N_g'(0)}{N_g(0)}\right). \end{aligned}$$

Furthermore,

$$\frac{N_g'(0)}{N_g(0)} = \int_{\mathcal{F}} a_f \frac{\Psi_f(g)}{I_0(g)}\,\mathrm{d}\nu(f) = \int_{\mathcal{F}} a_f \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\nu(f) = \mathbb{E}_{f\sim\nu_t(\cdot|g)}[a_f].$$

Replacing $a_f = \frac{t}{\sigma_t^2}\langle f, k\rangle$,

$$\frac{N_g'(0)}{N_g(0)} = \mathbb{E}_{f\sim\nu_t(\cdot|g)}[a_f] = \frac{t}{\sigma_t^2}\mathbb{E}_{f\sim\nu_t(\cdot|g)}[\langle f, k\rangle] = \frac{t}{\sigma_t^2}\langle \hat{g}_1(g), k\rangle.$$

Finally, we conclude that

$$D_h \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g) = \frac{t}{\sigma_t^2}\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\langle f - \hat{g}_1(g), k\rangle.$$

Equation 77 follows by replacing the logarithmic derivative expressions from Lemmas 6 and 7. □

Lemma 11 (Differentiation of the posterior mean)**.** *Let $g \in \mathcal{G}_t, k \in \mathcal{F}$ and $h = C_0 k$. Then , $\hat{g}_1(g)$ is differentiable along h, and*

$$D_h \hat{g}_1(g) = \int_{\mathcal{F}} f D_h \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\nu(f) = \mathrm{Cov}_{f\sim\nu_t(\cdot|g)}(f, \beta_h^f(g)). \tag{79}$$

Proof. By definition, the directional derivative of $\hat{g}_1(g)$ is given by

$$D_h \hat{g}_1(g) = \lim_{s\to 0}\frac{\hat{g}_1(g+sh) - \hat{g}_1(g)}{s} = \lim_{s\to 0}\int_{\mathcal{F}} f\left(\frac{\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+sh) - \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)}{s}\right)\mathrm{d}\nu(f)$$

At this point, we would like to use the dominated convergence theorem to conclude the first equality of 79,

$$D_h \hat{g}_1(g) = \int_{\mathcal{F}} f\left(\lim_{s\to 0}\frac{\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+sh) - \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)}{s}\right)\mathrm{d}\nu(f) = \int_{\mathcal{F}} f D_h \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\,\mathrm{d}\nu(f).$$

To this end, by the mean value theorem, we obtain

$$\left|\frac{\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+sh) - \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)}{s}\right| \le \sup_{|u|\le 1}\left|\frac{d}{du}\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+uh)\right|$$

Then, by an analogous procedure as in Lemma 9, the integrand can be bounded by

$$\begin{aligned} \left|f\frac{\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+sh) - \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)}{s}\right| &\le \|f\| \sup_{|u|\le 1}\left|\frac{d}{du}\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+uh)\right| \\ &\le \|f\| \sup_{|u|\le 1}\left|\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+uh)\left(a_f - \frac{N_g'(u)}{N_g(u)}\right)\right| \\ &\le \|f\| \sup_{|u|\le 1}\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+uh)\left(|a_f| + \left|\frac{N_g'(u)}{N_g(u)}\right|\right), \end{aligned}$$

where $a_f = \frac{t}{\sigma_t^2}\langle f, k\rangle$. Defining $\alpha := \frac{t\|k\|}{\sigma_t^2}$, $|a_f| \leq \alpha\|f\|$. Furthermore, since

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+uh) = \frac{\Psi_f(g)e^{a_f u}}{N_g(u)},$$

and $g + uh \in \mathcal{G}_t$, for every $u \in [-1,1]$, we have $N_g(u) = I_0(g+uh) > 0$. Moreover, N_g is continuous on $[-1,1]$ by dominated convergence. Hence,

$$m := \inf_{|u|\leq 1} N_g(u) > 0.$$

By the same dominated convergence argument it is possible to prove that N_g' is continuous on $[-1,1]$, thus,

$$C_g := \sup_{|u|\leq 1}\left|\frac{N_g'(u)}{N_g(u)}\right| < \infty.$$

Then, since $0 < |s| \leq 1$,

$$\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+sh) = \frac{\Psi_f(g)e^{a_f s}}{N_g(s)} \leq \frac{\Psi_f(g)e^{a_f s}}{m} \leq \frac{\Psi_f(g)e^{\alpha\|f\|}}{m}.$$

In consequence,

$$\left\| f\frac{\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g+sh) - \frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)}{s}\right\| \leq \frac{\|f\|\Psi_f(g)}{m}e^{\alpha\|f\|}(\alpha\|f\| + C_g).$$

Furthermore, using that $\|f\| \leq e^{\|f\|}$, there exists a constant $\tilde{C} > 0$ independent of s and f such that

$$\frac{\|f\|\Psi_f(g)}{m}e^{\alpha\|f\|}(\alpha\|f\| + C_g) \leq \tilde{C}\Psi_f(g)e^{(\alpha+2)\|f\|}$$

Define $G(f) := \tilde{C}\Psi_f(g)e^{(\alpha+2)\|f\|}$. Since $g \in \mathcal{G}_t$,

$$\int_{\mathcal{F}} \Psi_f(g)e^{(\alpha+2)\|f\|}\,\mathrm{d}\nu(f) < \infty.$$

Hence, $G \in L^1(\nu)$ and the interchange between limit and integral is justified.

Substituting expression 76 and the definition of $\hat{g}_1(g)$,

$$\begin{aligned}
D_h\hat{g}_1(g) &= \frac{t}{\sigma_t^2}\int_{\mathcal{F}} f\frac{\mathrm{d}\mu_t^f}{\mathrm{d}\mu_t}(g)\langle f - \hat{g}_1(g), k\rangle\,\mathrm{d}\nu(f)\\
&= \frac{t}{\sigma_t^2}\mathbb{E}_{f\sim\nu_t(\cdot|g)}[f\langle f - \hat{g}_1(g), k\rangle]\\
&= \frac{t}{\sigma_t^2}\mathbb{E}_{f\sim\nu_t(\cdot|g)}[(f - \hat{g}_1(g))\langle f - \hat{g}_1(g), k\rangle + \hat{g}_1(g)\langle f - \hat{g}_1(g), k\rangle]\\
&= \frac{t}{\sigma_t^2}\mathbb{E}_{f\sim\nu_t(\cdot|g)}[(f - \hat{g}_1(g))\langle f - \hat{g}_1(g), k\rangle].
\end{aligned}$$

Finally, to show that it coincides with $\mathrm{Cov}_{f\sim\nu_t(\cdot|g)}(f, \beta_h^f(g))$, by 66,

$$\begin{aligned}
\mathrm{Cov}_{f\sim\nu_t(\cdot|g)}(f, \beta_h^f(g)) &= \mathrm{Cov}_{f\sim\nu_t(\cdot|g)}\left(f, -\frac{1}{\sigma_t^2}\langle g, k\rangle + \frac{t}{\sigma_t^2}\langle f, k\rangle\right)\\
&= \mathrm{Cov}_{f\sim\nu_t(\cdot|g)}\left(f, \frac{t}{\sigma_t^2}\langle f, k\rangle\right)\\
&= \frac{t}{\sigma_t^2}\mathbb{E}_{f\sim\nu_t(\cdot|g)}[(f - \hat{g}_1(g))\langle f - \hat{g}_1(g), k\rangle],
\end{aligned}$$

concluding the last equality of 79. □

Lemma 12 (Jacobian Trick for FFM). *For $g \in \mathcal{G}_t, t \in (0,1]$ and $k \in \mathcal{F}$ the Gâteaux derivative of $\hat{g}_1(g)$ in the direction $h = C_0 k$ of the Cameron–Martin space satisfies*

$$D_h \hat{g}_1(g) = \frac{t}{(\sigma_t)^2} \Sigma_{1|t}(g)[k]. \tag{80}$$

In particular, if $\hat{g}_1(g)$ is Fréchet differentiable at g, then

$$\Sigma_{1|t}(g) = \frac{(\sigma_t)^2}{t} D\hat{g}_1(g) C_0. \tag{81}$$

Proof. By Lemma 11

$$D_h \hat{g}_1(g) = \mathrm{Cov}_{\nu_t(\cdot|g)}(f, \beta_h^f(g)).$$

Replacing the expression for $\beta_h^f(g)$ of Lemma 6

$$D_h \hat{g}_1(g) = -\frac{1}{\sigma_t^2} \mathrm{Cov}_{f \sim \nu_t(\cdot|g)} \left(f, \langle g - tf, k \rangle\right).$$

As g is fixed when taking the conditional covariance

$$\mathrm{Cov}_{f \sim \nu_t(\cdot|g)} \left(f, \langle g, k \rangle\right) = 0.$$

Then,

$$D_h \hat{g}_1(g) = \frac{t}{\sigma_t^2} \mathrm{Cov}_{f \sim \nu_t(\cdot|g)} \left(f, \langle f, k \rangle\right).$$

By definition of $\Sigma_{1|t}$,

$$\mathrm{Cov}_{f \sim \nu_t(\cdot|g)} \left(f, \langle f, k \rangle\right) = \Sigma_{1|t}(g)[k].$$

Thus,

$$D_h \hat{g}_1(g) = \frac{t}{\sigma_t^2} \Sigma_{1|t}(g)[k],$$

for every $k \in \mathcal{F}$. Now, if $\hat{g}_1$ is Fréchet differentiable, then

$$D_h \hat{g}_1(g) = D\hat{g}_1(g)[h] = D\hat{g}_1(g)[C_0 k].$$

Hence,

$$D\hat{g}_1(g)[C_0 k] = \frac{t}{\sigma_t^2} \Sigma_{1|t}(g)[k].$$

Since this holds for every $k \in \mathcal{F}$,

$$D\hat{g}_1(g) C_0 = \frac{t}{\sigma_t^2} \Sigma_{1|t}(g).$$

Finally,

$$\Sigma_{1|t}(g) = \frac{\sigma_t^2}{t} D\hat{g}_1(g) C_0.$$

□

Theorem 7 (Infinite-dimensional covariance-based guidance). *Let $g \in \mathcal{G}_t$, suppose that J is Fréchet differentiable at $\hat{g}_1(g)$ and $\hat{g}_1(g)$ is Fréchet differentiable at g. Then,*

$$\gamma_t^{\text{cov-G}}(g) = -\frac{\sigma_t}{t} \Big(\sigma_t D v_t(g) + (1 - \sigma_{\min}) I \Big) C_0 [\nabla J(\hat{g}_1(g))]. \tag{82}$$

Proof. Recall that $\gamma_t^{\text{cov}}(g) = -\frac{1}{\sigma_t} \Sigma_{1|t}(g)[\nabla J(\hat{g}_1(g))]$. Substituting the expression for $\Sigma_{1|t}$ of Lemma 12, the localized guidance term can be written as

$$\gamma_t^{\text{cov-G}}(g) = -\frac{\sigma_t}{t} D\hat{g}_1(g) C_0 [\nabla J(\hat{g}_1(g))].$$

By 62, the Fréchet derivative $D\hat{g}_1(g)$ is given by

$$D\hat{g}_1(g) = \sigma_t D v_t(g) + (1 - \sigma_{\min}) I.$$

Thus, substituting into the $\gamma_t^{\text{cov-G}}$ expression, we conclude that

$$\gamma_t^{\text{cov-G}}(g) = -\frac{\sigma_t}{t} \Big(\sigma_t D v_t(g) + (1 - \sigma_{\min}) I \Big) C_0 [\nabla J(\hat{g}_1(g))].$$

□

Invariant guidance for ideal data. If ν is supported on exact KdV trajectories and the invariants are evaluated exactly, then $J_{\mathrm{inv}} = 0$ ν-almost surely. Its exact reweighting therefore leaves ν unchanged, even when combined with observation guidance. The improvements obtained using the approximate invariant guidance concern the learned, discretized sampler; they do not follow from the exact reweighting theorem.

E Data Generation and Architecture details

Ground-truth trajectories are produced with a Zabusky–Kruskal finite-difference integrator (Zabusky & Kruskal, 1965) of equation 1 ($\mu = -1$), using a leap-frog time step with the stability-controlled parameters of Table 4. Initial data are exact single- or multi-soliton profiles with speed/amplitude parameters and randomized phases x_0 in the ranges of Table 4. We build two datasets: (i) *trajectory*, storing the full $u(\tau, x)$ field on the space–time grid; and (ii) *joint*, storing the pair (u_0, u_T) as a two-channel function. Trajectory datasets uses $20{,}000$ training and 256 test samples, and joint datasets use $30{,}000$ and $1{,}024$, respectively.

Symbol	Meaning	Value
$\Delta\tau$	solver time step	3×10^{-5}
$[x_a, x_b]$	spatial domain (1/2/3-soliton)	$[-15, 15]$ / $[-10, 20]$ / $[-10, 20]$
T	horizon	2
c	soliton velocity range (1/2/3-soliton)	$[0.3, 4.0]$ / $[0.5, 4.0]$ / $[0.5, 4.0]$
x_0	soliton phase range (1/2/3-soliton)	$[-3.5, 3.5]$ / $[-2.5, 2.5]$ / $[-3.0, 2.0]$
shoots	saved time slices	128 (129 considering the initial state)
μ	KdV nonlinearity sign	-1

Table 4: KdV solver and dataset parameters.

The velocity field is a Fourier Neural Operator (Li et al., 2021) with the per-experiment configuration of Table 5. Wall-clock times are on a single GPU.

Model	Repr.	Matérn ν	Modes	Layers	Hidden	Ch.	Epochs	s/epoch
1-sol	joint 1D ($N_x = 256$)	0.5	16	3	256	2	100	$\sim$37.5
2-sol	joint 1D ($N_x = 256$)	0.5	16	3	256	2	100	$\sim$46.6
3-sol	joint 1D ($N_x = 256$)	0.5	16	3	256	2	100	$\sim$37.4
3-sol	joint 1D ($N_x = 256$)	1.5	16	3	256	2	100	$\sim$37.4
3-sol	joint 1D ($N_x = 256$)	2.5	16	3	256	2	100	$\sim$37.5
2-sol	traj 2D (64×64)	0.5	16	3	128	1	100	$\sim$259.7
3-sol	traj 2D (64×64)	0.5	16	3	128	1	100	$\sim$260.0

Table 5: FNO architectures and training budget per experiment. "Ch." is the number of output channels (2 for the joint (u_0, u_T) representation, 1 for the trajectory field). Times are measured wall-clock on a single GPU.

F Additional Qualitative Results

Figure 6 contrasts unguided and guided single-soliton generation: unguided samples cover the data distribution, while guidance concentrates them on the prescribed target. Additional multi-soliton samples and animated space–time rollouts are included in the released material.

F.1 Unconditional full trajectory generation

Figures 7 and 8 provide additional unconditional sampling for the two- and three-soliton trajectory models. Figure 7 compares representative temporal snapshots generated by FFM, Energy-Guided FFM, and DiffusionPDE, while Figure 8 shows the corresponding full space-time trajectories. These qualitative results complement the dynamical consistency metrics reported in Table 1.

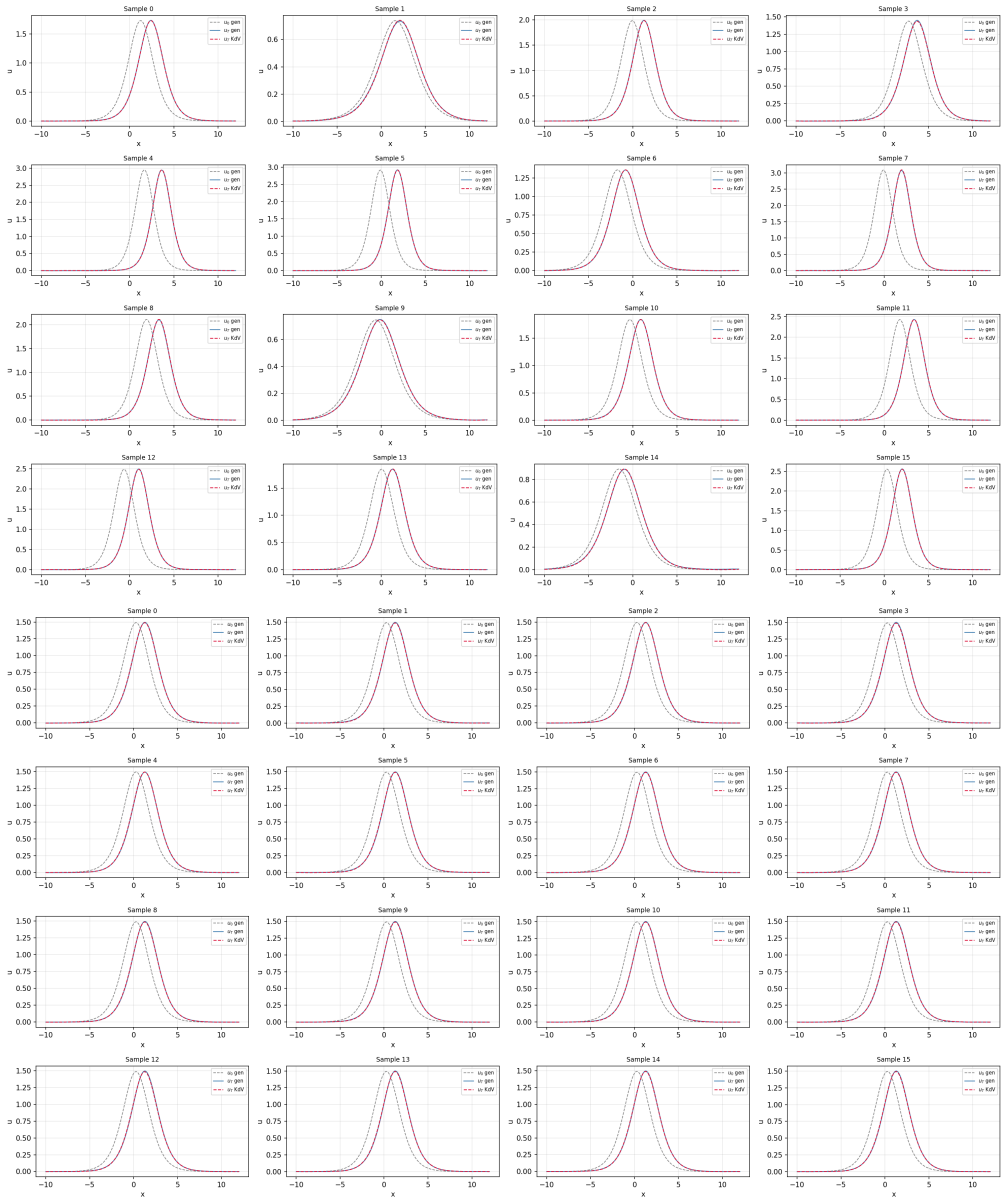

Figure 6: Single-soliton generation *without* guidance (first four rows) and *with* energy-guidance toward a target soliton (last four rows).

F.2 TRAJECTORY DENOISING

Denoising is evaluated separately from main experiments. In this task, each method receives a corrupted version of the complete trajectory.

For FFM, the corrupted trajectory is sampled from its native interpolation,

$$g_t = tf + \sigma_t Z, \qquad Z \sim \mathcal{N}(0, C_0),$$

and the clean endpoint is estimated as

$$\hat{g}_1(t, g_t) = \sigma_t u_t(g_t \mid \theta) + (1 - \sigma_{\min}) g_t.$$

For DiffusionPDE, the clean trajectory is corrupted with independent Gaussian noise and evaluated directly with the EDM denoiser. The physical pointwise noise RMS is matched to that of the FFM

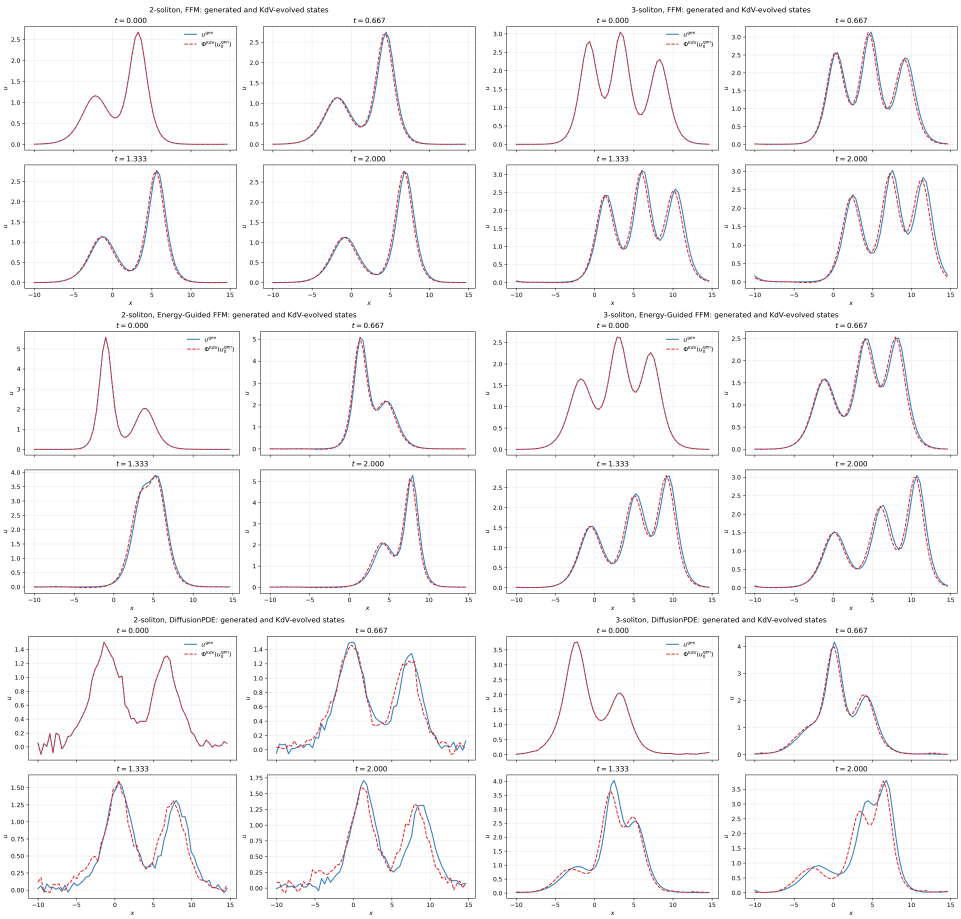

Figure 7: Generation of two- (first two columns) and three-solitons (last two columns) using FFM (first two rows) FFM + Energy-Guidance (middle two rows) and DiffusionPDE (last two rows), in different time snapshots.

corruption. We evaluate the flow times $s \in \{0.25, 0.5, 0.75\}$, corresponding to physical noise RMS values approximately equal to 0.949, 0.316, and 0.105.

This experiment matches the pointwise noise amplitude, but not its covariance: FFM uses Gaussian process correlated noise, whereas DiffusionPDE uses i.i.d. Gaussian noise. No initial condition or invariant guidance is used in the denoising experiment. The results of this experiment are settled in Table 6.

F.3 Spatial discretization consistency for joint FFM models

For the joint experiments, Table equation 7 shows spatial-discretization consistency, with four different grids, $N_x \in \{64, 128, 256, 512\}$.

G Evaluation Metrics

We employed several metrics to compare the generated trajectories with the reference test data and with the baseline methods. In this section, we describe their numerical computation. Unless we say the opposite, all the tables entries correspond to the median value over all the sampling.

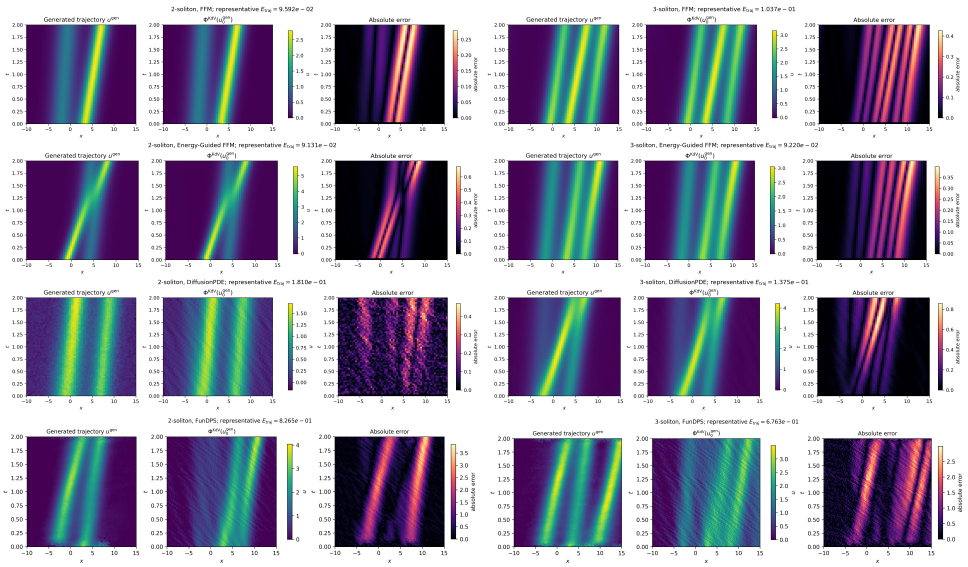

Figure 8: Generation of two- (first three columns) and three-solitons (last three columns) using FFM (first row), FFM + Energy-Guidance (second row), DiffusionPDE (third row), and FunDPS (last row), for the whole trajectory.

Method	Case	Noise RMS	$E_{\text{traj}} \downarrow$	$E_{\text{final}} \downarrow$	$\|\mathcal{R}\|_{\text{RMS}} \downarrow$	$d^M(T) \downarrow$	$d^P(T) \downarrow$	$d^H(T) \downarrow$
Test data	2S	—	—	—	1.068×10^{-1}	8.853×10^{-5}	3.378×10^{-4}	1.400×10^{-3}
Test data	3S	—	—	—	2.796×10^{-1}	9.353×10^{-4}	3.686×10^{-4}	2.273×10^{-3}
DiffusionPDE	2S	9.487×10^{-1}	3.150×10^{-2}	4.371×10^{-2}	2.373×10^{0}	2.112×10^{-2}	3.363×10^{-2}	4.878×10^{-2}
DiffusionPDE	3S	9.487×10^{-1}	3.035×10^{-2}	4.166×10^{-2}	2.630×10^{0}	1.615×10^{-2}	2.740×10^{-2}	4.372×10^{-2}
DiffusionPDE	2S	3.163×10^{-1}	1.397×10^{-2}	2.311×10^{-2}	1.261×10^{0}	1.003×10^{-2}	1.708×10^{-2}	3.034×10^{-2}
DiffusionPDE	3S	3.163×10^{-1}	1.233×10^{-2}	2.012×10^{-2}	1.303×10^{0}	9.120×10^{-3}	1.462×10^{-2}	2.417×10^{-2}
DiffusionPDE	2S	1.054×10^{-1}	6.217×10^{-3}	1.060×10^{-2}	6.637×10^{-1}	4.952×10^{-3}	6.833×10^{-3}	1.115×10^{-2}
DiffusionPDE	3S	1.054×10^{-1}	5.493×10^{-3}	9.759×10^{-3}	6.177×10^{-1}	4.638×10^{-3}	6.064×10^{-3}	1.039×10^{-2}
FFM	2S	9.487×10^{-1}	3.475×10^{-2}	3.918×10^{-2}	9.326×10^{-1}	4.526×10^{-3}	8.505×10^{-3}	1.500×10^{-2}
FFM	3S	9.487×10^{-1}	3.238×10^{-2}	3.683×10^{-2}	1.222×10^{0}	5.563×10^{-3}	9.416×10^{-3}	1.418×10^{-2}
FFM	2S	3.163×10^{-1}	1.182×10^{-2}	1.514×10^{-2}	5.357×10^{-1}	2.794×10^{-3}	6.159×10^{-3}	9.886×10^{-3}
FFM	3S	3.163×10^{-1}	1.279×10^{-2}	1.542×10^{-2}	7.138×10^{-1}	4.478×10^{-3}	7.351×10^{-3}	1.102×10^{-2}
FFM	2S	1.054×10^{-1}	4.885×10^{-3}	6.434×10^{-3}	2.830×10^{-1}	2.399×10^{-3}	4.134×10^{-3}	7.103×10^{-3}
FFM	3S	1.054×10^{-1}	5.331×10^{-3}	7.411×10^{-3}	4.273×10^{-1}	3.636×10^{-3}	5.119×10^{-3}	8.338×10^{-3}

Table 6: KdV trajectory denoising at matched pointwise physical noise RMS on 256 test trajectories.

G.1 Relative L^2 errors

In the case unconditional sampling, we obtain either complete trajectories $\{u_\ell^{\text{gen}}(\tau_j, x_k)\}_{j,k=0}^{N_\tau-1,N_x-1}$, or the joint distribution $(u_{\ell,0}^{\text{gen}}, u_{\ell,T}^{\text{gen}})$ at a sample ℓ. For both experiments we are able to obtain the generated initial condition $u_{\ell,0}^{\text{gen}}$, and therefore evolve this initial data with a KdV solver (The Zabusky-Kruskal method in our experiments), obtaining the complete trajectory $\Phi_\tau^{\text{KdV}}(u_{\ell,0}^{\text{gen}})$, or the final state $\Phi_T^{\text{KdV}}(u_{\ell,0}^{\text{gen}})$. Then, it is natural to compute the L^2-relative error between the generated trajectories (or final states), with the solver evolution. For this, we introduce the two metrics:

$$E_{\text{dyn},\ell} := \frac{\|u_\ell^{\text{gen}} - \Phi_{(\cdot)}^{\text{KdV}}(u_{\ell,0}^{\text{gen}})\|_{\ell^2_{\tau,x}}}{\|\Phi_{(\cdot)}^{\text{KdV}}(u_{\ell,0}^{\text{gen}})\|_{\ell^2_{\tau,x}}},$$

and

$$E_{T,\ell} := \frac{\|u_{\ell,T}^{\text{gen}} - \Phi_T^{\text{KdV}}(u_{\ell,0}^{\text{gen}})\|_{\ell^2_x}}{\|\Phi_T^{\text{KdV}}(u_{\ell,0}^{\text{gen}})\|_{\ell^2_x}}.$$

Model	N_x	$E_{u_0}^{512} \downarrow$	$E_{u_T}^{512} \downarrow$	$E_T \downarrow$	$d^P(T) \downarrow$	$d^H(T) \downarrow$
1S, $\nu = 0.5$	64	3.008×10^{-1}	3.219×10^{-1}	7.359×10^{-3}	3.544×10^{-3}	5.925×10^{-3}
1S, $\nu = 0.5$	128	1.471×10^{-1}	1.323×10^{-1}	6.721×10^{-3}	3.502×10^{-3}	5.822×10^{-3}
1S, $\nu = 0.5$	256	6.334×10^{-2}	5.903×10^{-2}	7.740×10^{-3}	3.652×10^{-3}	6.090×10^{-3}
1S, $\nu = 0.5$	512	0	0	7.593×10^{-3}	3.627×10^{-3}	6.015×10^{-3}
2S, $\nu = 0.5$	64	3.101×10^{-1}	2.855×10^{-1}	9.847×10^{-3}	4.066×10^{-3}	6.472×10^{-3}
2S, $\nu = 0.5$	128	1.249×10^{-1}	1.159×10^{-1}	9.726×10^{-3}	4.410×10^{-3}	6.714×10^{-3}
2S, $\nu = 0.5$	256	6.906×10^{-2}	6.399×10^{-2}	9.406×10^{-3}	3.228×10^{-3}	5.216×10^{-3}
2S, $\nu = 0.5$	512	0	0	9.059×10^{-3}	3.192×10^{-3}	4.652×10^{-3}
3S, $\nu = 0.5$	64	2.622×10^{-1}	2.513×10^{-1}	8.223×10^{-3}	2.814×10^{-3}	4.331×10^{-3}
3S, $\nu = 0.5$	128	1.161×10^{-1}	1.233×10^{-1}	8.206×10^{-3}	2.003×10^{-3}	3.550×10^{-3}
3S, $\nu = 0.5$	256	7.568×10^{-2}	8.463×10^{-2}	9.085×10^{-3}	2.185×10^{-3}	3.653×10^{-3}
3S, $\nu = 0.5$	512	0	0	8.702×10^{-3}	1.902×10^{-3}	3.264×10^{-3}
3S, $\nu = 1.5$	64	1.490×10^{-1}	1.808×10^{-1}	9.202×10^{-3}	2.493×10^{-3}	3.835×10^{-3}
3S, $\nu = 1.5$	128	6.144×10^{-2}	6.975×10^{-2}	9.227×10^{-3}	2.394×10^{-3}	4.070×10^{-3}
3S, $\nu = 1.5$	256	1.929×10^{-2}	2.367×10^{-2}	9.154×10^{-3}	2.308×10^{-3}	3.281×10^{-3}
3S, $\nu = 1.5$	512	0	0	8.970×10^{-3}	2.350×10^{-3}	3.457×10^{-3}
3S, $\nu = 2.5$	64	1.346×10^{-1}	1.670×10^{-1}	7.960×10^{-3}	3.065×10^{-3}	5.073×10^{-3}
3S, $\nu = 2.5$	128	5.007×10^{-2}	6.235×10^{-2}	6.936×10^{-3}	2.924×10^{-3}	4.682×10^{-3}
3S, $\nu = 2.5$	256	1.541×10^{-2}	1.862×10^{-2}	6.737×10^{-3}	3.004×10^{-3}	4.887×10^{-3}
3S, $\nu = 2.5$	512	0	0	6.783×10^{-3}	2.957×10^{-3}	4.729×10^{-3}

Table 7: Spatial-discretization consistency of the joint FFM models. $E_{u_0}^{512}$ and $E_{u_T}^{512}$ compare each output with the coupled $N_x = 512$ output. E_T compares the generated terminal channel with the independent KdV evolution of its generated initial channel. Entries are sample medians over 64 coupled latent realizations.

On the other hand, for the forecasting experiments we conditioned the sampling to the initial datum $u_{\ell,0}^{\mathrm{ref}}$ from the test data. Therefore we can compare the generated trajectory u_ℓ^{gen} (conditioned on $u_{\ell,0}^{\mathrm{ref}}$), with the corresponding reference test trajectory $\left\{u_\ell^{\mathrm{ref}}(\tau_j, x_k)\right\}_{j,k=0}^{N_\tau-1,N_x-1}$. Thus we have the L^2 relative errors

$$E_{\mathrm{traj},\ell} := \frac{\|u_\ell^{\mathrm{gen}} - u_\ell^{\mathrm{ref}}\|_{\ell^2_{\tau,x}}}{\|u_\ell^{\mathrm{ref}}\|_{\ell^2_{\tau,x}}}, \tag{83}$$

and

$$E_{\mathrm{final},\ell} := \frac{\|u_{\ell,T}^{\mathrm{gen}} - u_{\ell,T}^{\mathrm{ref}}\|_{\ell^2_x}}{\|u_{\ell,T}^{\mathrm{ref}}\|_{\ell^2_x}}. \tag{84}$$

G.2 Root Mean Square error

In the main tables we have provided the root mean square (RMS) error for the residual of KdV. More precisely, for each sample ℓ, we compute the residual term as

$$\mathcal{R}_\ell(\tau_j, x_k) = \partial_\tau u_\ell(\tau_j, x_k) + \partial_{xxx} u_\ell(\tau_j, x_k) + 2u_\ell(\tau_j, x_k)\partial_x u_\ell(\tau_j, x_k),$$

and

$$\|\mathcal{R}_\ell\|_{\mathrm{RMS}} = \left(\frac{1}{(N_\tau - 2)N_x} \sum_{j=1}^{N_\tau-2} \sum_{k=0}^{N_x-1} |\mathcal{R}_\ell(\tau_j, x_k)|^2 \right)^{1/2}.$$

The indexes $j = 1, \ldots, N_\tau - 2$ were chosen given the approximation of the temporal derivative via centered finite differences:

$$\partial_\tau u_\ell(\tau_j, x_k) \approx \frac{u_\ell(\tau_{j+1}, x_k) - u_\ell(\tau_{j-1}, x_k)}{2\Delta\tau},$$

while the spatial points were taken from all the indexes by assuming periodicity.

G.3 INVARIANTS

The invariants in equation 3 are evaluated numerically at every time step. For each sample ℓ and each field $X \in \{\mathrm{ref}, \mathrm{gen}\}$, we compute

$$M_\ell^X(\tau_j) \approx \Delta x \sum_{k=0}^{N_x-1} u_\ell^X(\tau_j, x_k), \quad P_\ell^X(\tau_j) \approx \Delta x \sum_{k=0}^{N_x-1} \left(u_\ell^X(\tau_j, x_k)\right)^2.$$

To evaluate H, the spatial derivative is computed spectrally. In particular,

$$\left(D_x u_\ell^X\right)(\tau_j, x_k) = \left[\mathcal{F}^{-1}_{\xi\to x}\left(i\xi\, \mathcal{F}_{x\to\xi}\left[u_\ell^X(\tau_j, \cdot)\right](\xi)\right)\right](x_k),$$

where

$$\xi_m = 2\pi\ \mathrm{fftfreq}(N_x, \Delta x)_m$$

is the angular wave-number grid associated with the periodic spatial grid.

For the KdV parametrization considered in this work, the energy is approximated in physical space as

$$H_\ell^X(\tau_j) \approx \Delta x \sum_{k=0}^{N_x-1} \left[\frac{1}{3}\left(u_\ell^X(\tau_j, x_k)\right)^3 - \frac{1}{2}\left(D_x u_\ell^X(\tau_j, x_k)\right)^2\right].$$

For each invariant $I \in \{M, P, H\}$, its absolute relative defect is computed for each sample ℓ as

$$d^I_{\ell,X}(\tau_j) = \frac{\left|I_\ell^X(\tau_j) - I_\ell^X(0)\right|}{\left|I_\ell^X(0)\right|}.$$

The same computation is performed independently across all the experiments.

H PROTOCOL FOR THE DIFFUSION HEAD-TO-HEAD

The baseline is an EDM denoiser (Karras et al., 2022) with a U-Net backbone trained on the *same* soliton dataset and grid, using a white-noise prior and the standard stochastic (Heun/SDE) sampler; for PDE-conditioned generation we follow the observation-plus-residual guidance of DiffusionPDE (Huang et al., 2024) (ζ_{obs}, ζ_{pde}).

DiffusionPDE represents each complete space-time trajectory as a two-dimensional field. In our implementation, the temporal and spatial coordinates play the role of the two image coordinates, whereas the solution value $u(\tau, x)$ is represented by a single channel.

Let $U \in \mathbb{R}^{N_\tau \times N_x}$ denote a clean trajectory. The diffusion model is trained as an EDM preconditioned denoiser D_θ, using corrupted trajectories of the form

$$U_\sigma = U + \sigma\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I),$$

and the weighted denoising objective

$$\mathcal{L}_{\mathrm{EDM}}(\theta) = \mathbb{E}_{U,\sigma,\varepsilon}\left[\lambda(\sigma)\|D_\theta(U_\sigma; \sigma) - U\|^2_{L^2}\right].$$

The denoiser uses the EDM preconditioning wrapper around a SongUNet. The same architecture is used for the two- and three-soliton datasets, and only the training data, normalization statistics, and resulting checkpoints differ. Its complete configuration is summarized in Table 8.

We use $\sigma_{\mathrm{data}} = 0.5$. The model is trained for 6000 optimization steps with a batch size of 16, Adam learning rate 2×10^{-4}, gradient clipping at 10, and exponential moving average decay 0.999. Before training, the data are normalized through an affine transformation computed exclusively from the training set. The normalization parameters are $(0.68465, 2.20860)$ for the two soliton dataset and $(1.01804, 2.40705)$ for the three soliton dataset, where the first value is the offset and the second one is the scale.

At inference, the reverse diffusion dynamics are integrated with the second-order EDM/Heun sampler using 75 steps, $\sigma_{\min} = 0.002$, $\sigma_{\max} = 80$, and $\rho = 7$. Forecasting from an initial state u_0 is obtained by introducing the observation functional

$$\mathcal{J}_{\mathrm{obs}}(U) = \|U(0, \cdot) - u_0\|^2_{L^2}.$$

Component	Configuration
Backbone	EDMPrecond/SongUNet. Standard encoder and decoder
Input/output and levels	one channel; $64 \times 64 \to 32 \times 32 \to 16 \times 16$
Channel widths	$32 \to 64 \to 64$
Residual blocks/attention	4 per resolution/ one head at 16×16
Noise embedding	positional, 32 noise/128 mapped channels
Resampling/dropout	filter $[1, 1]$/ probability 0.05
Trainable parameters	2,740,193

Table 8: DiffusionPDE architecture used for both KdV baselines. The channel schedule uses 32 model channels and channel mult. $[1, 2, 2]$.

Its gradient is used to guide the reverse diffusion process, and the initial temporal row is projected onto the observation at the end of sampling. The observation guidance is selected on validation data, resulting in $\zeta_{\text{obs}} = 450$ for two solitons and $\zeta_{\text{obs}} = 550$ for three solitons. Positive values of the PDE-residual guidance ζ_{PDE} were also tested. However, the validation error was minimized at $\zeta_{\text{PDE}} = 0$ in both cases. Therefore, the reported DiffusionPDE forecasts use initial condition guidance but no PDE guidance.

The reported DiffusionPDE forecasts use initial-condition guidance without PDE-residual guidance, since validation selected $\zeta_{\text{PDE}} = 0$. The methods also use different training budgets: the diffusion baseline receives 96,000 example presentations over 6,000 updates, whereas trajectory FFM is trained for 100 epochs on 20,000 examples. These experiments therefore compare the reported implementations and budgets, rather than isolating the effect of the model family under matched training exposure.

I Protocol for the FunDPS Baseline

We adapt FunDPS (Yao et al., 2025) to model complete KdV trajectories. Each sample is a space-time field $u(\tau, x)$: time and space form the two grid coordinates, and the solution value occupies one channel. Separate unconditional diffusion priors are specified for the two- and three-soliton datasets.

Let $U \in \mathbb{R}^{N_\tau \times N_x}$ denote a clean trajectory, resampled to a 64×64 grid. Before training, we normalize it as

$$Z = \frac{0.5}{s_{\text{train}}}(U - \mu_{\text{train}}),$$

where μ_{train} and s_{train} are computed after resampling, using only the training split. Each soliton dataset has its own statistics, which are then applied unchanged to its validation and test splits.

FunDPS trains an EDM-preconditioned denoiser D_θ using Gaussian random-field noise. A corrupted trajectory has the form

$$Z_\sigma = Z + \sigma\xi, \qquad \xi \sim \mathcal{N}(0, C_{\text{RBF}}),$$

where C_{RBF} is the covariance produced by the implementation's RBFKernel sampler. The training objective is

$$\mathcal{L}_{\text{FunDPS}}(\theta) = \mathbb{E}_{Z,\sigma,\xi}\left[\lambda(\sigma)\|D_\theta(Z_\sigma;\sigma) - Z\|_{L^2}^2\right], \qquad \lambda(\sigma) = \frac{\sigma^2 + \sigma_{\text{data}}^2}{(\sigma\sigma_{\text{data}})^2}.$$

We use $\sigma_{\text{data}} = 0.5$ and $\log\sigma \sim \mathcal{N}(-1.2, 1.2^2)$. The denoiser consists of the EDMPrecond wrapper around a SongUNO neural operator. The specified architecture is the same for both soliton datasets; their training data, normalization statistics, and checkpoints are separate. Its configuration is summarized in Table 9.

The current training configuration uses Adam with learning rate 10^{-4} and a global batch size of 8. Its target budget is 80 kimg, equivalent to 80,000 trajectory presentations and 10,000 optimization updates if the run is completed. An exponential moving average of the network weights provides the denoiser used at inference. Training and validation are split by complete trajectory.

At inference, sampling begins from a Gaussian random field generated with the same covariance sampler used for training. The reverse dynamics use the second-order EDM/Heun integrator with

Component	Configuration
Backbone	EDMPrecond/SongUNO (`ddpmpp-uno`)
Input/output and levels	one channel; $64^2 \to 32^2 \to 16^2 \to 8^2$
Channel widths	$32 \to 64 \to 128 \to 128$
Operator blocks	four encoder and five decoder blocks per resolution
Attention	one head at 8×8
Noise embedding	positional, 32 noise/128 mapped channels
Operator settings	Fourier-mode fraction 0.5; rank 0.1
Resampling/dropout	filter $[1, 1]$/probability 0.13
Gaussian-field sampler	RBFKernel, scale 0.05
Trainable parameters	13,419,703

Table 9: Specified SongUNO architecture for the 64×64 KdV experiments. Channel widths correspond to 32 base channels and multipliers $[1, 2, 4, 4]$.

500 steps, $\sigma_{\min} = 0.002$, $\sigma_{\max} = 80$, and $\rho = 7$. Forecasting from an initial state u_0 uses the observation functional

$$\mathcal{J}_{\mathrm{obs}}(U) = \frac{1}{N_x} \sum_{j=0}^{N_x-1} \left[\frac{0.5}{s_{\mathrm{train}}} \big(U(0, x_j) - u_0(x_j) \big) \right]^2 .$$

Diffusion posterior sampling (DPS) differentiates this loss through the denoised trajectory estimate and applies the resulting guidance during reverse sampling. The implementation scales observation guidance by $\min\{1, \sigma\}$ at noise level σ. The current configuration sets $\zeta_{\mathrm{obs}} = 1000$ for both soliton datasets; this is a starting value that has not yet been selected by validation.

The forecasting profile observes the complete initial temporal row and contains no PDE-residual guidance term. It does not impose the initial condition through a final hard projection. A separate sensor profile observes spatial locations throughout the trajectory; because those observations include later times, that profile represents trajectory reconstruction rather than forecasting from u_0 alone.