FRITES: Provable Diffusion Posterior Sampling via Forward-Reverse Iterations of Tilted SDEs

Anonymous authors
Paper under double-blind review

Abstract

Score-based diffusion models have emerged as powerful generative frameworks, particularly suited for sampling from high-dimensional distributions. From a Bayesian point of view, the framework provides a potential solution to inverse problems, i.e., combining observations with prior data distributions learned by the diffusion model to obtain posterior sampling — a task of both high practical importance and significant theoretical challenge. Earlier works have achieved high generative quality, but lack evidence on provable posterior sampling. Recent work has sought theoretical guarantees, but faces a persistent dilemma: transport and particle-based methods are either restricted to linear settings or become computationally prohibitive in high dimensions, while annealing-based plug-and-play approaches stall in practice and fail to achieve posterior sampling. To tackle these issues, we introduce Forward-Reverse Iterations of Tilted SDEs (FRITES), a novel algorithm with provable convergence, scalable to high-dimensional data, adaptive to both linear and non-linear inverse problems, and easy to implement. In particular, FRITES introduces a Markov chain built on coupled forward and reverse diffusion dynamics, with a principled tilt so that the target posterior emerges as the unique stationary distribution. We establish in this paper both asymptotic and finite-time theoretical guarantees, and provide empirical evidence that FRITES achieves superior performance spanning synthetic low-dimensional benchmarks and high-dimensional imaging tasks comparing to state-of-the-art baseline methods.

1 Introduction

Inverse problems arise across a broad range of scientific domains, from medical imaging and astronomical observation to geophysics and computational biology. The goal is to recover an unknown signal x^* from noisy, incomplete observations $y = \mathcal{A}(x^*) + n$, where $\mathcal{A}$ is a forward operator and n is observation noise. The Bayesian approach casts this as posterior inference: given a prior p_0 over x^* and a likelihood $p(y|x)$, one seeks to sample from the posterior $p(x|y) \propto p(y|x)\, p_0(x)$, which provides not only point estimates but also principled uncertainty quantification.

The resurgence of score-based diffusion models (Song et al., 2021b; Ho et al., 2020; Vincent, 2011) has shed new light for Bayesian inverse problems: using a pretrained diffusion model that draws high-quality samples from the prior p_0. Diffusion models have demonstrated remarkable ability to capture complex, high-dimensional data distributions in imaging, audio, and scientific applications, making them compelling priors for posterior sampling. A large body of work has since developed methods that leverage pretrained diffusion models to approximately sample from the posterior (Chung et al., 2023; Song et al., 2022; Kawar et al., 2022). These methods achieve impressive empirical performance, but typically introduce uncontrolled errors, i.e., by approximating the intractable conditional score or restricting to a limited variational family, and consequently lack provable guarantees on the quality of their posterior samples. On the other hand, methods that pursue theoretical guarantees (Bruna & Han, 2024; Dou & Song, 2024; Xu & Chi, 2024) face persistent challenges: they are either restricted to linear forward operators, become computationally prohibitive in high dimensions, or cannot achieve posterior sampling in practice due to algorithmic drawbacks, see Appendix A for a detailed review.

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.

In this paper, we introduce **F**orward-**R**everse **I**terations of **T**ilted SD**ES** (FRITES), a novel posterior sampling algorithm that reconciles provable guarantees with practical scalability. FRITES is scalable to high-dimensional data, adaptive to both linear and nonlinear forward operators, and easy to implement on top of any pretrained diffusion model. Our main contributions are as follows.

- **Algorithm.** We introduce FRITES, a Markov chain built on coupled forward-reverse diffusion dynamics with a principled tilt, such that the posterior is its unique stationary distribution for both linear and nonlinear inverse problems (Algorithm 1).
- **Theory.** We establish both asymptotic convergence (Theorem 2, Corollary 1) and finite-step error bounds (Theorem 3, Corollary 2) for FRITES, with the number of iterations scaling only logarithmically in the target accuracy.
- **Experiments.** We validate FRITES on Gaussian mixture models, where ground-truth posteriors are analytically available, and on high-dimensional imaging tasks, demonstrating competitive performance comparing to state-of-the-art baselines.

2 Preliminaries

2.1 Diffusion Models

Score-based diffusion models (Song et al., 2021b; Ho et al., 2020) learn to sample from a data distribution p_0 on $\mathbb{R}^d$ via two coupled processes. The *forward process* is an Ornstein–Uhlenbeck (OU) SDE,

$$dX_t = -X_t\,dt + \sqrt{2}\,dW_t, \qquad X_0 \sim p_0, \tag{1}$$

whose marginal p_t converges to $\mathcal{N}(0, I_d)$ as $t \to \infty$, with the closed-form transition $X_t|X_0 \sim \mathcal{N}(e^{-t}X_0,\, (1-e^{-2t})I_d)$.

The key quantity for reversing this process is the *score function* $s(x,t) := \nabla \log p_t(x)$, which is estimated by a neural network s_θ via denoising score matching (Vincent, 2011):

$$\mathcal{L}(\theta) = \int_0^T \lambda(t)\,\mathbb{E}_{X_0}\mathbb{E}_{X_t|X_0}\Big[\|s_\theta(X_t,t) - \nabla \log p_{0t}(X_t|X_0)\|^2\Big]\,dt, \tag{2}$$

where $\lambda(t)$ is a tuning parameter, and $\nabla \log p_{0t}(X_t|X_0)$ can be easily computed from Eq. (1). Throughout this paper we assume exact score estimation, i.e., we do not distinguish $s(\cdot,t)$ from $s_\theta(\cdot,t)$.

By Anderson's time-reversal theorem (Anderson, 1982), the *reverse process* satisfies

$$d\overleftarrow{X}_t = \big(\overleftarrow{X}_t + 2\,s(\overleftarrow{X}_t, T-t)\big)\,dt + \sqrt{2}\,d\overleftarrow{W}_t, \qquad \overleftarrow{X}_0 \sim p_T. \tag{3}$$

Assuming 0 approximation error, the marginal of $\overleftarrow{X}_t$ equals p_{T-t}, so integrating Eq. (3) from 0 to T yields an exact sample from p_0 (Song et al., 2021b).

2.2 Bayesian inverse problems

An inverse problem seeks to recover an unknown signal $x^* \in \mathbb{R}^d$ from an observation $y \in \mathbb{R}^n$ related to x^* through a *forward operator* $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^n$:

$$y = \mathcal{A}(x^*) + n, \qquad n \sim \mathcal{N}(0, \sigma_y^2 I_n), \tag{4}$$

where $\sigma_y > 0$ is the noise level. Such problems are typically ill-posed, and regularization is essential to obtain meaningful reconstructions.

The Bayesian approach addresses this by placing a *prior* p_0 over x^*, which together with the likelihood $p(y|x) \overset{d}{=} \mathcal{N}(y;\, \mathcal{A}(x),\, \sigma_y^2 I_n)$ yields the *posterior*

$$p(x|y) \propto p(y|x)\,p_0(x) \tag{5}$$

via Bayes' theorem (Stuart, 2010). Sampling from $p(x|y)$ provides both point estimates and uncertainty quantification for the reconstruction.

The expressive power of diffusion models makes them a natural choice of prior p_0, and a growing body of work has sought to leverage a pretrained diffusion model for posterior sampling (Song et al., 2021b; Chung et al., 2023). However, provably correct and computationally tractable methods remain scarce. A key ingredient towards this goal is the tilted transport framework of (Bruna & Han, 2024), which we introduce in the next subsection.

2.3 Solving linear inverse problems via tilted transport

It is worth asking whether posterior sampling can be achieved by simply running the reverse SDE Eq. (3) with a suitably chosen initialization. In the linear setting $\mathcal{A}(x) = Ax$ where $A \in \mathbb{R}^{n\times d}$, the answer is affirmative. Under the linear setting, the posterior reads

$$q_0(x) \propto p_0(x) \exp\left(-\frac{1}{2}x^\top Q x + x^\top b\right), \tag{6}$$

where $Q = \sigma_y^{-2} A^\top A$ and $b = \sigma_y^{-2} A^\top y$. The key idea of (Bruna & Han, 2024) is that the quadratic structure in the exponent, i.e., the quadratic tilt, can be extended to a time-varying tuple (Q_t, b_t) for $t > 0$, such that the family $q_t \propto p_t \exp(-\frac{1}{2}x^\top Q_t x + x^\top b_t)$ is exactly consistent with the reverse SDE Eq. (3). Under the OU forward process Eq. (1), the log-density $f_t := \log p_t$ solves a viscous Hamilton–Jacobi equation (HJE), see Appendix B.2 for details. Motivated by Eq. (6), we consider the one-parameter family $q_t(x) \propto p_t(x) \exp\left(-\frac{1}{2}x^\top Q_t x + x^\top b_t\right)$, where (Q_t, b_t) are time-varying parameters with $Q_0 = Q$ and $b_0 = b$. Requiring $\log q_t$ to satisfy the same HJE as $\log p_t$ yields the *Riccati ODEs* (Bruna & Han, 2024)

$$\dot{Q}_t = 2(I + Q_t)Q_t, \qquad \dot{b}_t = (I + 2Q_t)b_t, \tag{7}$$

which admit a closed-form solution (see Appendix B.2). The solution blows up at a finite *terminal time* $T^* = \frac{1}{2}\log(1 + \|Q\|^{-1})$, determined by the signal-to-noise ratio of the measurement. For brevity, we do not distinguish throughout this paper T^* and $T^* - \delta$, where $0 < \delta < T^*$ is a constant close to 0.

The solution (Q_t, b_t) defines, for each $t \in [0, T^*]$, a *likelihood path*

$$\mathcal{L}_t(x) := -\frac{1}{2}x^\top Q_t x + x^\top b_t, \tag{8}$$

so that $q_t(x) \propto \exp(\mathcal{L}_t(x))\, p_t(x)$. By (Bruna & Han, 2024), sampling from q_0 is then equivalent to a two-step procedure: first sample from the *boosted posterior* $q_{T^*}(x) \propto \exp(\mathcal{L}_{T^*}(x))\, p_{T^*}(x)$, then run the reverse SDE Eq. (3) from T^* to 0 using the prior score $s(\cdot, t)$. The boosted posterior q_{T^*} is provably easier to sample than q_0: as $t \to T^*$, the prior p_t becomes approximately Gaussian while $\mathcal{L}_t$ simultaneously concentrates, making q_{T^*} more amenable to standard samplers such as Langevin dynamics.

Despite its theoretical appeal, tilted transport has several limitations. First, sampling from q_{T^*} remains challenging in practice: when d is large, as in imaging applications, or when the SNR is low (so that T^* is small and p_{T^*} has not yet approached a Gaussian), adopting standard samplers such as Langevin dynamics are still computationally challenging. Second, tilted transport cannot incorporate the many heuristic posterior sampling methods that, despite lacking theoretical guarantees, have been shown to produce high-quality reconstructions in imaging tasks (Chung et al., 2023; Song et al., 2023). Third, the framework is inherently tied to the linear setting and does not extend directly to nonlinear forward operators $\mathcal{A}$. In the next section, we introduce FRITES, a method built upon tilted transport that addresses all three limitations.

3 Method

3.1 Forward process of tilted transport

We begin by revisiting the unconditional score-based diffusion model (Song et al., 2021b), which underlies our construction. A diffusion model comprises two processes: a forward process, i.e., Eq. (1),

that progressively corrupts data and is used to train the score function, and a reverse process, i.e., Eq. (3) that reconstructs samples from the data distribution by running another diffusion process drifted by the learned score backward in time. Under the assumption of zero approximation errors, the marginal distributions of the forward process p_t and the time-reversed process $\overleftarrow{p_{T-t}}$ coincide exactly.

Tilted transport modifies this framework for posterior sampling by altering the initial condition of the reverse SDE. Specifically, while the reverse SDE remains identical to that of the standard diffusion model, tilted transport initializes the reverse process at the boosted posterior rather than a standard Gaussian. Since the initial condition propagates through the dynamics, the marginal distributions of the resulting reverse process q_t differ from those of $\overleftarrow{p_{T-t}}$ at every time $t \in [0, T^*]$, and in particular q_t is no longer the time-reversal of the standard forward process p_t. This discrepancy raises a natural and important question: what is the corresponding forward process whose time-reversal yields q_t?

Identifying this forward process is not merely of theoretical interest, it is the key structural insight that enables the construction of a Markov chain with the posterior as its invariant measure, as developed in Section 3.2. By applying the theory of time-reversal of diffusion processes, we characterize this forward process in the following theorem.

Theorem 1. *Denote q_0 as the posterior distribution and $\mathcal{L}_t$ as the log-likelihood of measurement y_t in tilted transport. Let $\overrightarrow{q_t}$ be the solution of the following SDE:*

$$dX_t = -X_t\,dt + \sqrt{2}\,dW_t + 2\nabla\mathcal{L}_t(X_t)\,dt, \quad X_0 \sim q_0. \tag{9}$$

Then, the law of $(\overrightarrow{q_t})_{t\in[0,T^*]}$ exactly tracks that of the time-reversal of the tilted transport process $(q_s)_{s\in[0,T^*]}$. In particular, $\overrightarrow{q_t} = q_{T^*-t}$ for all $t \in [0, T^*]$.

The proof is left to Appendix C.1. It writes the score of the tilted marginal as the sum of the untilted score and the likelihood gradient, derives the Fokker–Planck equation satisfying the tilted marginal, and then applies Anderson's time-reversal theorem to read off the corresponding forward SDE. Theorem 1 reveals that the forward process of tilted transport takes the form of a *drifted* diffusion forward process, where the additional drift term $2\nabla\mathcal{L}_t(X_t)dt$ incorporates measurement information through the gradient. Note that this process is initialized at the true posterior q_0. This characterization provides the foundation for the iterative scheme developed in the next section, where we exploit the interplay between this forward process and its time-reversal to construct a Markov chain admitting the posterior as its unique invariant measure.

3.2 Invariant Measure Under the Forward-Reverse Processes

Having characterized the forward process of tilted transport in Theorem 1, we now exploit the interplay between the forward and reverse processes to construct a Markov chain that converges to the posterior q_0. The key idea is that running the forward process from q_0 and then time-reversing it recovers q_0 exactly: a fixed-point property that we will promote into an iterative sampling scheme.

Since both the forward and reverse SDEs are Markovian, their solutions over a fixed time horizon T^* induce well-defined transition kernels. Specifically, we define the forward transition kernel $F_{T^*}(x, \cdot)$ as the conditional distribution of X_{T^*} given $X_0 = x$ under the forward SDE, and the reverse transition kernel $R_{T^*}(x', \cdot)$ as the conditional distribution of X_0 given $X_{T^*} = x'$ under the reverse SDE. Composing these two kernels yields a single Markov transition kernel

$$K_{T^*}(x, x') = \int F_{T^*}(x, y)\, R_{T^*}(y, x')\, dy, \tag{10}$$

which describes one complete forward-reverse iteration: diffuse the current state forward to time T^*, then run the reverse SDE back to time 0. Intuitively, if the current distribution is close to q_0, the forward process brings it to a distribution close to the boosted posterior at time T^*, and the reverse process then reconstructs a sample approximately from q_0. We numerically illustrate the idea by applying the kernel K_{T^*} to the analytically-sampled posterior distribution q_0 in Figure 3, which demonstrates the fact that q_0 is invariant under K_{T^*}. To formalize this intuition, we impose the following mild regularity conditions on the log-likelihood tilt $\mathcal{L}_t$:

Assumption 1 (Regularity)**.** *The log-likelihood tilt $\mathcal{L}_t : \mathbb{R}^d \to \mathbb{R}$ and the prior distribution p_0 satisfies:*

(i) ***(Differentiability)*** *$\mathcal{L}_t(\cdot)$ is differentiable almost everywhere for every $t \in [0, T^*]$.*

(ii) ***(Boundedness from above)*** *$\sup_{x \in \mathbb{R}^d} \mathcal{L}_t(x) < \infty$ for every $t \in [0, T^*]$.*

(iii) ***(Novikov condition)*** *The gradient field satisfies $\mathbb{E}\Big[\exp\Big(\int_0^{T^*} \|\nabla \mathcal{L}_t(X_t)\|^2 \, dt\Big)\Big] < \infty$, where X_t solves the OU process initialized at any $x \in \mathbb{R}^d$.*

(iv) ***(Strict positivity)*** *The prior distribution has full support, meaning $p_0(x) > 0$ for all $x \in \mathbb{R}^d$.*

Conditions (i) and (ii) are standard in the inverse problems literature and are satisfied by most likelihood functions of practical interest, including Gaussian, Poisson, and logistic observation models. Condition (iii) is the classical Novikov condition (Øksendal, 2003), which ensures that the Girsanov exponential martingale associated with the tilt is a true martingale, thereby guaranteeing that F_{T^*} is a well-defined probability kernel. A sufficient condition for (iii) is that $\nabla \mathcal{L}_t$ is bounded uniformly in (t, x), which holds whenever the forward model $\mathcal{A}$ is Lipschitz and the noise level is bounded away from zero. Condition (iv) is imposed to satisfy assumptions of the Jentzsch-Perron theorem (Schaefer et al., 2012) for the convergence analysis.

The following theorem shows that q_0 is the unique invariant measure of K_{T^*}, and that iterating K_{T^*} converges to q_0 from any initial distribution. The proof can be found in Appendix C.2, where it first establishes that the composed kernel is strictly positive everywhere via a Girsanov argument (Øksendal, 2003), which implies irreducibility and aperiodicity of the chain. Detailed balance with respect to the posterior then follows from the time-reversal identity of Theorem 1, and classical Markov chain theory yields convergence (Tierney, 1994).

Theorem 2. *Let $K_{T^*} p(\cdot) = \int p(x) \, K_{T^*}(x, \cdot) \, dx$ denote the action of the transition kernel K_{T^*} on a probability distribution p. Then, for any initial distribution p, the sequence of marginal distributions iteratively updated by K_{T^*} converges to the target posterior q_0 in total variation distance,*

$$\lim_{N \to \infty} \left\| K_{T^*}^{(N)} p - q_0 \right\|_{\mathrm{TV}} = 0. \tag{11}$$

Theorem 2 establishes that iteratively applying K_{T^*}, namely alternately running the forward and reverse SDEs, converges to the target posterior q_0 from any initial distribution. We refer to this iterative procedure as *Forward-Reverse Iterations of Tilted SDEs* (FRITES). However, in practice, we can only run a finite number of iterations N, which raises the question of how quickly the chain mixes. The following theorem provides a quantitative answer by establishing an error bound on the total variation distance. Proof can be found in Appendix C.6 and the graphical illustration of FRITES in Figure 1.

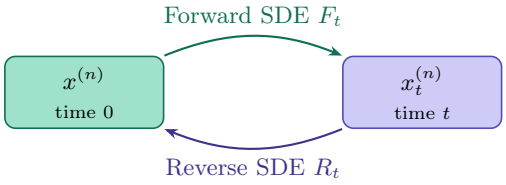


Figure 1: FRITES iterations of the forward kernel F_t and reverse kernel R_t, with $t \in [0, T^*]$.

Theorem 3 (Finite-step error bound)**.** *Under Assumption 1 and the zero score approximation error assumption of Theorem 1, there exists $\lambda_{T^*} = \lambda(q_0, T^*) \in (0, 1)$ such that the following holds. For any initial distribution p with $\chi^2(p \,\|\, q_0) < \infty$ and any target accuracy $\varepsilon_{\mathrm{acc}} > 0$, with $N \asymp \frac{\log(1/\varepsilon_{\mathrm{acc}})}{1 - \lambda_{T^*}}$ iterations we have*

$$\left\| K_{T^*}^{(N)} p - q_0 \right\|_{\mathrm{TV}} \leq \varepsilon_{\mathrm{acc}} \sqrt{\chi^2(p \,\|\, q_0)}. \tag{12}$$

The proof applies the Jentzsch-Perron theorem (Schaefer et al., 2012) to the self-adjoint operator induced by the kernel on the weighted space, obtaining a spectral gap that bounds the operator norm of the centered iterate (Xu & Chi, 2024; De Bortoli, 2022). A Cauchy–Schwarz argument then converts this into a geometric contraction in χ^2-divergence, from which the total variation bound follows via a standard comparison inequality.

Theorem 3 upgrades the asymptotic guarantee of Theorem 2 to a quantitative one: it specifies explicitly how many forward-reverse iterations suffice to reach a target accuracy $\varepsilon_{\mathrm{acc}}$, with the required number

of iterations scaling only logarithmically in $1/\varepsilon_{\text{acc}}$ and inversely in $1 - \lambda_{T^*}$. The factor $\lambda_{T^*} \in (0, 1)$ captures the interplay between the time horizon T^*, the prior p_0, and the likelihood $\mathcal{L}_t$; its precise characterization is given in Section 3.4.

3.3 Nonlinear Extension

In the previous subsection, we constructed the Markov transition kernel K_{T^*} and established its convergence to the posterior q_0. However, the kernel K_{T^*} relies on the FRITES forward process, which requires evaluating the measurement log-likelihood $\mathcal{L}_t$ at intermediate diffusion times $t \in (0, T^*]$. For linear inverse problems, $\mathcal{L}_t$ admits a closed form. For nonlinear measurements, $\mathcal{L}_t$ is intractable in general. This motivates us to seek an alternative formulation that only evaluating $\mathcal{L}_t$ at time $t = 0$.

The key observation is that the convergence property in Theorems 2 and 3 do not rely on the specific choice of time horizon T^*, they hold for any $t \in (0, T^*]$. This is formalized in the following corollaries.

Corollary 1. *Under Assumption 1 and the zero score approximation error assumption of Theorem 1, let $K_t p(\cdot) = \int p(x)\, K_t(x, \cdot)\, dx$ for any $t \in (0, T^*]$. Then for any initial distribution p,*

$$\lim_{N \to \infty} \left\| K_t^{(N)} p - q_0 \right\|_{\text{TV}} = 0 \tag{13}$$

Corollary 2. *Under the hypotheses of Corollary 1, for any $t \in (0, T^*]$ there exists $\lambda_t = \lambda(q_0, t) \in (0, 1)$ such that for any initial distribution p with $\chi^2(p \,\|\, q_0) < \infty$ and any target accuracy $\varepsilon_{\text{acc}} > 0$, with $N \asymp \frac{\log(1/\varepsilon_{\text{acc}})}{1-\lambda_t}$ iterations we have*

$$\left\| K_t^{(N)} p - q_0 \right\|_{\text{TV}} \leq \varepsilon_{\text{acc}} \sqrt{\chi^2(p \,\|\, q_0)}. \tag{14}$$

Corollary 1 and 2 suggest a natural remedy for the nonlinear setting: by choosing t sufficiently close to 0, the forward SDE over the short interval $[0, t]$ can be approximated by its first-order Euler discretization, yielding

$$X_t = (1 - t)X_0 + 2t\nabla\mathcal{L}_0(X_0) + \sqrt{2t}\, w, \quad w \sim \mathcal{N}(0, I_d). \tag{15}$$

Crucially, this discretization involves only $\mathcal{L}_0$, which is the log-likelihood evaluated at the *original* measurement without any diffusion-time perturbation, thereby circumventing the need to evaluate $\mathcal{L}_t$ at intermediate times. Thus for nonlinear inverse problems, though $\mathcal{L}_t$ is not available, this small-t approximation renders the forward transition kernel F_t tractable. Together with the reverse transition kernel R_t, the composed kernel K_t can therefore be applied to sample the posterior distribution in nonlinear settings. We also provide in Appendix C.9 a robustness study concerning score-matching error, in which Theorem 5 gives a finite-step bound when score-matching is not exact.

Computationally, the Euler discretization introduces an approximation error that may violate the detailed balance condition, potentially biasing the stationary distribution away from q_0. To correct for this, we incorporate a Metropolis-Hastings adjustment, which restores exact detailed balance at the cost of occasional rejected proposals. The acceptance probabilities are characterized in Proposition 1 in Appendix D. We summarize the complete FRITES sampling procedure in Algorithm 1 in Appendix E.

3.4 Mixing Time

Having established convergence of FRITES for any $t \in (0, T^*]$, a natural question arises in the linear setting where all choices of t are theoretically valid: which value of t yields the fastest convergence? From the proof of Corollary 2 that the χ^2-divergence contracts geometrically at rate λ_t^2 per iteration, so minimizing λ_t directly minimizes the number of forward-reverse iterations required to reach a given approximation accuracy. This motivates a careful analysis of how the spectral gap of K_t depends on t.

Intuitively, a larger t corresponds to a longer forward diffusion, which brings the distribution more blurry and thereby makes the subsequent reverse process more informative. One might therefore expect the mixing time to decrease monotonically as t increases. The following theorem confirms this intuition and identifies $t = T^*$ as the optimal choice.

		ULA		Tilted		DPnP		FRITES	
d	d'	mean	std	mean	std	mean	std	mean	std
20	18	6.1255	3.1006	3.4254	2.5637	3.2761	**2.0837**	**1.9400**	2.4910
40	36	6.0435	2.7606	3.3099	**1.5663**	4.3013	3.7264	**1.9463**	2.9799
80	72	6.3772	3.1447	4.0864	**1.5164**	3.8659	2.7486	**1.5998**	2.7929

Table 1: SWD between reconstructions and ground truth ($d' = 0.9d$). Best values per task are in **bold**

		NUTS		Tilted		DPnP		FRITES	
d	d'	mean	std	mean	std	mean	std	mean	std
20	1	1.5870	1.0734	1.8263	1.3039	3.5652	1.5248	**0.9060**	**0.8158**
40	1	2.7049	1.5029	3.2645	2.3101	3.8803	1.3220	**0.7048**	**0.5883**
80	1	3.8742	2.2717	4.3990	2.7055	4.1526	1.3479	**0.9336**	**1.2365**

Table 2: SWD between reconstructions and ground truth ($d' = 1$). Best values per task are in **bold**

Theorem 4 (Mixing time monotonicity)**.** *Under Assumption 1, let $\lambda_t \in (0, 1)$ be the contraction factor of K_t from Corollary 2, defined as the second largest eigenvalue of the operator $\mathbf{K}_t$ on $L^2(q_0)$. Then for any $0 < t < t' \leq T^*$, $\lambda_{t'} < \lambda_t$. In other words, the map $t \mapsto \lambda_t$ is strictly decreasing on $(0, T^*]$. Equivalently, the mixing time of K_t, defined as the number of iterations N needed to achieve $\|K_t^{(N)} p - q_0\|_{\mathrm{TV}} \leq \varepsilon$, is strictly decreasing in the diffusion horizon t: running the forward-reverse iteration over a longer time horizon yields strictly faster convergence to the posterior.*

Theorem 4 establishes that $t = T^*$ uniquely minimizes the spectral gap of K_t, and therefore yields the fastest mixing among all valid choices of t. This provides a principled justification for the choice of $t = T^*$ in the linear setting, precisely quantifying the trade-off between tractability and mixing speed that underlies the design of FRITES. Proof is left to Appendix C.10. It uses the variational characterization of the spectral gap and a conditional Jensen inequality: the conditional expectation of the test function contracts along the Markov chain as t increases, so the spectral gap can only improve. Strictness follows from the non-degenerate Gaussian noise in the forward SDE, which ensures the conditional variance is always strictly positive between two distinct time points.

4 Experiments

4.1 Gaussian mixture prior with linear measurement

We follow Bruna & Han (2024) in our Gaussian mixture model (GMM) experiments. For a given signal dimension $d \in \{20, 40, 80\}$, the prior is a mixture of 25 Gaussians whose means tile the planar grid point $(8i, 8j)$ across the d coordinates, i.e. $(8i, 8j, \ldots, 8i, 8j) \in \mathbb{R}^d$ for $i, j \in \{-2, -1, 0, 1, 2\}$, each with identity covariance. The mixture weights are generated randomly.

We use linear measurements so that the true posterior is available analytically. The measurement forward operator is regenerated for each experiment, with singular values drawn from Unif[1/20, 1], and the observation noise standard deviation is set to 1. We consider two settings: the observation dimension equals 90% of the signal dimension, $d' = 0.9d$, and $d' = 1$. We evaluate four algorithms: FRITES and three baselines, and use the true analytical score to circumvent approximation errors. For FRITES we use a small-t implementation in both settings, for consistency with the later nonlinear imaging experiments; we set $N = 200$ and $t = 0.01$. For $d' = 0.9d$ we initialize FRITES with the pseudo-inverse of the observation, and for $d' = 1$ we initialize with annealed Langevin dynamics (Song & Ermon, 2019).

For comparison, we implement tilted transport Bruna & Han (2024) and Diffusion Plug-and-Play (DPnP) (Xu & Chi, 2024) in both experimental settings. Tilted transport follows (Bruna & Han, 2024) exactly. DPnP uses the same initialization as FRITES, with hyperparameters tuned accordingly. The DDS part of DPnP is implemented using DDPM (Ho et al., 2020). As the fourth baseline, we use the unadjusted Langevin algorithm (ULA) (Durmus & Moulines, 2017) for $d' = 0.9d$, and the No-U-Turn

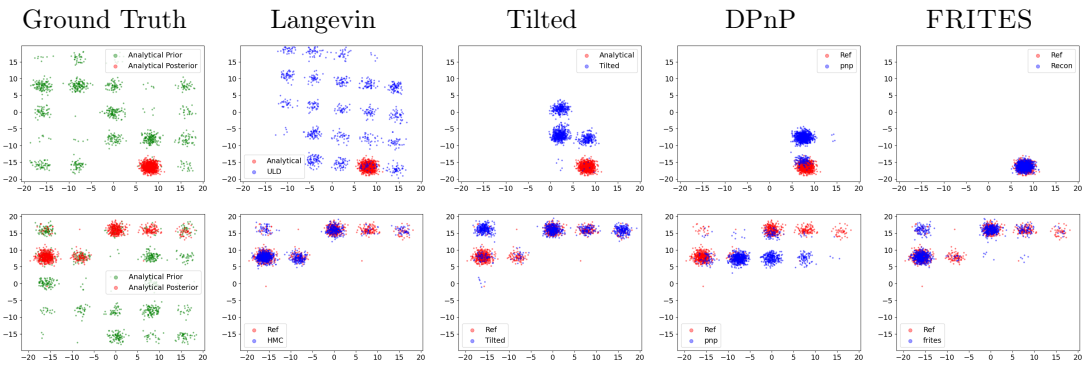


Figure 2: Qualitative results on the Gaussian mixture model experiments. Each panel shows posterior samples of a 20 dimension GMM projected onto two coordinates. The top row corresponds to the $d' = 0.9d$ setting and the bottom row to $d' = 1$. From left to right: ground-truth posterior, Langevin baseline (ULA for $d' = 0.9d$ and NUTS for $d' = 1$), tilted transport, DPnP, and our proposed FRITES method. Throughout, red represents the ground truth posterior, blue the reconstruction, and green the ground truth prior GMM.

Sampler (NUTS) (Hoffman et al., 2014), a Hamiltonian Monte Carlo (HMC) method, for $d' = 1$. Both follow the implementations in Bruna & Han (2024).

Since the posterior is available in closed form, we draw ground-truth samples directly from it. The output of each experiment is compared to these samples using the sliced-Wasserstein distance (SWD) (Kolouri et al., 2019). For each setting and algorithm we run a batch of 50 experiments and report the mean and standard deviation of the SWD. Quantitative results are reported in Table 1 and Table 2, and visual results in Figure 2.

The results suggest that in both settings FRITES obtains higher posterior sampling quality. It attains the lowest mean SWD across every dimension and both observation regimes, and its advantage is most pronounced in the harder $d' = 1$ setting, where the baselines degrade markedly as d grows while FRITES remains accurate. Notably, the mean SWD of FRITES stays roughly stable as the signal dimension increases from 20 to 80, whereas the competing samplers exhibit a clear upward trend, indicating that FRITES scales more gracefully with dimension.

4.2 Imaging Problems

We provide numerical evidence that FRITES obtains promising results in solving both linear and nonlinear inverse problems on high-resolution imaging.

Two points about evaluation are worth noting up front. First, because FRITES is a posterior sampling algorithm, the standard imaging metrics may understate its quality: these metrics tend to favor sharp reconstructions, whereas posterior samples need not be sharp. We therefore modify FRITES in practice to produce sharper images. Second, FRITES depends on its initialization, and different initializations call for different tuning parameters. To enable a controlled comparison, we fix a single initialization—a reasonable but deliberately far-from-optimal reconstruction—and compare FRITES against state-of-the-art baseline DPnP (Xu & Chi, 2024) under the same initialization. We do not aim for SOTA performance; rather, these experiments are designed to show that, given the same task and initialization, FRITES outperforms DPnP.

We consider the following experiments to cover both linear and nonlinear settings. Both inverse problems are corrupted by independent Gaussian noise of standard deviation $\sigma_y = 0.05$. We consider phase retrieval with a coded diffraction mask, a classical nonlinear inverse problem (Candes et al., 2015). We first draw a random mask $M \in \mathbb{R}^{256\times256}$, shared across the three color channels. The mask is applied to the image x via the Hadamard product $M \odot x$, the result is mapped through the Fourier transform $\mathcal{F}$, and only the magnitudes of the transformed signal are retained. The forward operator is therefore $\mathcal{A}(x) = \mathrm{mag}\big(\mathcal{F}(M \odot x)\big)$, where $\mathrm{mag}(\cdot)$ takes the entrywise modulus of a complex-valued

matrix. For super-resolution the forward model is the bicubic downsampling operator (Keys, 1981), which is linear. We use a downsampling factor of 4 in all experiments.

We implement FRITES (see Algorithm 2 in Appendix E) and DPnP on the FFHQ validation dataset (Karras et al., 2019).We use pre-trained score functions as in Chung et al. (2023) and we normalize all images into the range $[-1, 1]$. For quantitative comparison of the reconstruction results, we use two standard metrics: peak signal-to-noise-ratio (PSNR), structural similarity index (SSIM) together with the perceptual metric: Learned Perceptual Image Patch Similarity (LPIPS) distance (Zhang et al., 2018). We obtain the shared initialization from DPS (Chung et al., 2023). A full DPS run uses 1000 inference steps; we instead run only 500, yielding a reconstruction that is sensible but noticeably short of what the full procedure achieves. This gives us exactly the kind of starting point our setup calls for: good enough to be meaningful, yet leaving clear room for improvement. We also report quantitative metrics for this initialization, so that it serves as a common reference point against which both FRITES and DPnP can be measured.

For the reverse process, we find that the SDE-type implementation (Eq. (3)), which injects noise at each step, blurs the reconstruction and degrades quality as noise accumulates over iterations. We therefore adopt a DDIM reverse process in Eq. (16) instead. We fix the number of DDIM steps to 20 and set $t = 0.01$ across all tasks. We also find that the Metropolis-adjustment is not needed in practice. A summary of the modified algorithm is given in Appendix F.2.

Results Table 3 reports reconstruction quality on phase retrieval and super-resolution. Across both tasks, FRITES improves on DPnP on every metric (PSNR, SSIM, and LPIPS) when started from the same DPS initialization. The gains are largest on phase retrieval, the harder nonlinear problem, where FRITES raises PSNR by roughly 2.6 dB over DPnP and cuts LPIPS by about 30%; on super-resolution the improvement is smaller but consistent across all three metrics. In both cases FRITES also moves well beyond the 500-step DPS starting point, confirming that the gains come from the method rather than the initialization. We emphasize that these results are obtained without tuning toward state-of-the-art performance: the comparison is controlled to isolate the effect of the algorithm given a fixed task and initialization. Visual results are shown in Figure 5.

Table 3: Imaging results. Reconstruction quality for phase retrieval and super-resolution, averaged over all runs. All methods share the same 500-step DPS initialization. Best values per task are in **bold**.

Task	Algorithm	PSNR ↑	SSIM ↑	LPIPS ↓
Phase retrieval	DPS (init)	18.4741	0.5280	0.3479
	DPnP	20.5101	0.5855	0.2802
	FRITES	**23.1483**	**0.6536**	**0.1968**
Super-resolution	DPS (init)	25.1331	0.7173	0.1654
	DPnP	25.5581	0.7281	0.1485
	FRITES	**27.7269**	**0.8045**	**0.1199**

5 Conclusion

We presented FRITES, a posterior sampling method for inverse problems based on a Markov chain that couples forward and reverse diffusion dynamics through a principled tilt. By construction, the target posterior is the unique stationary distribution of this chain, and the same formulation applies to both linear and nonlinear measurement models. On the theoretical side, we showed that FRITES converges asymptotically to the posterior and admits finite-step error bounds, with the iteration count growing only logarithmically in the desired accuracy. Empirically, we validated the method on Gaussian mixture models, where the true posterior is available in closed form, and on high-dimensional imaging problems, where FRITES compares favorably with strong baselines under controlled conditions.

These results suggest that coupling diffusion dynamics with an explicit tilt is a promising route to provable and practical posterior sampling. A limitation of the approach is structural: because FRITES relies on a forward-reverse update, it enforces measurement consistency and prior consistency in two separate stages: the forward step pulling the iterate toward the measurements, and the reverse step pulling it back toward the prior. When the measurement is highly ill-posed, this separation can steer the chain toward spurious modes. The same difficulty is shared by Plug-and-Play methods more broadly, which alternate between analogous likelihood and prior steps, and we leave a thorough treatment of it to future work.

AI use statement

In this work, we used generative AI tools to assist in the writing of proofs and provide some critical ingredients for proving some mathematical claims. We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, support qualitative and thematic data analysis, implement methods, and interpret results, and assisting with translation and cleaning and reformating dataset are not applicable to this work. Additionally, we used generative AI tools for editing a research paper to improve readability and identifying relevant literature. We have reviewed all AI-assisted work. LLM-generated wording was checked for correctness against original draft version. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Reproducibility Statement

We have taken several steps to make our results reproducible. All theoretical claims state their assumptions and are proven in full in Appendix C. The algorithm is detailed in Appendix E and the experimental implementations are detailed Section 4 and Appendix F. Code will be released to reproduce all results.

References

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Joan Bruna and Jiequn Han. Provable posterior sampling with denoising oracles via tilted transport. In *Advances in Neural Information Processing Systems*, volume 37, pp. 82863–82894, 2024.

Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

Gabriel Cardoso, Sylvain Le Corff, and Eric Moulines. Monte carlo guided denoising diffusion models for bayesian linear inverse problems. In *International Conference on Learning Representations*, volume 2024, pp. 44001–44037, 2024.

Hyungjin Chung, Jeongsol Kim, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *11th International Conference on Learning Representations, ICLR 2023*, 2023.

Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *arXiv preprint arXiv:2208.05314*, 2022.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *The Twelfth International Conference on Learning Representations*, 2024.

Alain Durmus and Eric Moulines. Nonasymptotic convergence analysis for the unadjusted langevin algorithm. *Annals of Applied Probability*, 27(3):1551–1587, 2017.

Wendell H. Fleming and H. Mete Soner. *Controlled Markov processes and viscosity solutions*. Springer, 2006.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in neural information processing systems*, volume 33, pp. 6840–6851, 2020.

Matthew D Hoffman, Andrew Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989.

Alicia A Johnson, Miles Q Ott, and Mine Dogucu. *Bayes rules!: An introduction to applied Bayesian modeling*. Chapman and Hall/CRC, 2022.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in neural information processing systems*, volume 35, pp. 23593–23606, 2022.

Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. Generalized sliced wasserstein distances. In *Advances in neural information processing systems*, volume 32, 2019.

Xiang Li, Soo Min Kwon, Shijun Liang, Ismail R Alkhouri, Saiprasad Ravishankar, and Qing Qu. Decoupled data consistency with diffusion purification for image restoration. *arXiv preprint arXiv:2403.06054*, 2024.

Xiangming Meng and Yoshiyuki Kabashima. Diffusion model based posterior sampling for noisy linear inverse problems. *arXiv preprint arXiv:2211.12343*, 2022.

Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations: an introduction with applications*, pp. 38–50. Springer, 2003.

Yury Polyanskiy and Yihong Wu. *Information theory: From coding to learning*. Cambridge university press, 2025.

Yinuo Ren, Wenhao Gao, Lexing Ying, Grant M Rotskoff, and Jiequn Han. Driftlite: Lightweight drift control for inference-time scaling of diffusion models. *arXiv preprint arXiv:2509.21655*, 2025.

Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning*, pp. 5546–5557. PMLR, 2019.

Laurent Saloff-Coste. Lectures on finite markov chains. In *Lectures on probability theory and statistics: École d'été de Probabilités de Saint-Flour XXVI-1996*, pp. 301–413. Springer, 2006.

Helmut H Schaefer et al. Banach lattices and positive operators [electronic resource]. *Springer Science and Business Media*, 2012.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International conference on learning representations*, 2023.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in neural information processing systems*, volume 32, 2019.

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*, 2021b.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022.

Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta numerica*, 19:451–559, 2010.

Luke Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, pp. 1701–1728, 1994.

Brian L Trippe, Luhuan Wu, Christian A Naesseth, David Blei, and John Patrick Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *ICML 2023 Workshop on Structured Probabilistic Inference & Generative Modeling*, 2023.

Alexandre B Tsybakov and Vladimir Zaiats. *Introduction to nonparametric estimation*, volume 11. Springer, 2009.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 31372–31403, 2023.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. In *Advances in Neural Information Processing Systems*, volume 37, pp. 118389–118427, 2024.

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. In *Advances in Neural Information Processing Systems*, volume 37, pp. 36148–36184, 2024.

Yici Yan, Yichi Zhang, Xiangming Meng, and Zhizhen Zhao. Fig: Flow with interpolant guidance for linear inverse problems. In *The Thirteenth International Conference on Learning Representations*, 2025.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 20895–20905, 2025.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1219–1229, 2023.

Appendix: FRITES: Provable Diffusion Posterior Sampling via Forward-Reverse Iterations of Tilted SDEs

This appendix is structured as follows: in Appendix A, we cover the related work; in Appendix B, we provide additional background on the diffusion models sampling process and tilted transport; in Appendix C, we provide proofs for claims in the main paper; in Appendix D, we detail the Metropolis Hasting correction; in Appendix E, we summarize algorithms; in Appendix F, we provide additional implementation details and results.

A Related works

A large body of work has leveraged pretrained diffusion models for posterior sampling by approximating the intractable conditional score $\nabla \log p_t(x_t|y)$. Among all, DPS (Chung et al., 2023) uses a Tweedie-based correction at each reverse step, and ΠGDM (Song et al., 2023), improves upon DPS via a pseudoinverse-guided correction under a Gaussian approximation. Song et al. (2022); Yan et al. (2025) explores interpolated measurements for conditional score approximation, and DMPS (Meng & Kabashima, 2022), assumes an uninformative prior over intermediate states. More recently, DAPS (Zhang et al., 2025) introduced a decoupled noise annealing process to handle highly nonlinear problems. While these methods achieve impressive empirical results, the approximations they rely on are essentially uncontrolled, and lack guarantees that the output distribution converges to the target posterior.

A second line of work seeks provable guarantees by casting posterior sampling as a Sequential Monte Carlo (SMC) problem, using particles to approximate the conditional distribution of intermediate diffusion states. Cardoso et al. (2024); Trippe et al. (2023); Wu et al. (2023); Dou & Song (2024) develop SMC samplers that use the diffusion reverse process as a proposal with importance reweighting, providing consistency as the particle count grows. DriftLite (Ren et al., 2025) addresses weight degeneracy by exploiting a degree of freedom in the Fokker–Planck equation to actively control the sampling drift, improving stability with modest overhead. These methods offer rigorous theoretical guarantees and are consistent in the limit of many particles, but their computational cost scales poorly with the data dimension due to the many-particle system, making them impractical for high-dimensional imaging tasks.

A third group draws on the classical plug-and-play (PnP) framework, alternating between a data-consistency update and a diffusion denoising step used as an implicit prior. Xu & Chi (2024); Wu et al. (2024); Li et al. (2024); Zhu et al. (2023); Ryu et al. (2019) introduce auxiliary variables to decouple the likelihood from the diffusion prior, establishing provable robustness guarantees for both linear and nonlinear forward operators. A common feature of PnP-based methods is the use of an annealing schedule on the noise level injected during the diffusion step. While annealing helps the algorithm explore the posterior at high noise levels, it creates a practical failure mode: when the noise level becomes small, the proximal regularizer dominates the update, causing the chain to stagnate near its current iterate and preventing convergence in practice.

B Preliminaries and Backgrounds

B.1 DDIM

In practice, the stochastic reverse process Eq. (3) can be replaced by a deterministic sampler without retraining. Denoising diffusion implicit models (DDIM, (Song et al., 2021a)) generalize the forward process to a non-Markovian family that shares the same marginals p_t and hence the same training objective Eq. (2). Setting the stochasticity to zero yields the deterministic update

$$X_{t-\Delta t} = \sqrt{\alpha_{t-\Delta t}} \underbrace{\frac{X_t - \sqrt{1-\alpha_t}\,\epsilon_\theta(X_t, t)}{\sqrt{\alpha_t}}}_{\text{predicted } X_0} + \sqrt{1-\alpha_{t-\Delta t}}\,\epsilon_\theta(X_t, t), \tag{16}$$

where $\alpha_t \coloneqq e^{-2t}$ and $\epsilon_\theta(X_t, t) \coloneqq -s_\theta(X_t, t)\sqrt{1-\alpha_t}$ is the predicted noise. The resulting trajectory is fully deterministic given the initial $X_T \sim p_T$, which enables faithful inversion of the generative process.

B.2 Tilted transport background

This appendix provides the technical details and derivations underlying Section 2.3. Throughout this section, we work with the Ornstein–Uhlenbeck (OU) forward process described in Eq. (1) and use the notation introduced in Section 2.3.

Hamilton–Jacobi equation. Let $f_t \coloneqq \log p_t$ denote the log-density of the forward process marginal p_t. By differentiating the Fokker-Planck equation, $\partial_t p_t = \nabla \cdot (x p_t) + \Delta p_t$, with respect to x and rearranging the terms, we obtain the following PDE for f_t:

$$\partial_t f_t = \Delta f_t + \|\nabla f_t\|^2 + x \cdot \nabla f_t + d, \qquad f_0 = \log p_0. \tag{17}$$

This formulation takes the form of a viscous Hamilton–Jacobi equation (HJE) (Fleming & Soner, 2006).

Riccati equations. Motivated by the form of the posterior in Eq. (6), we seek a family q_t of the form

$$\log q_t(x) = \log p_t(x) - \frac{1}{2} x^\top Q_t x + x^\top b_t + F(t), \tag{18}$$

where $F(t)$ is a normalizing term depending only on t, and (Q_t, b_t) satisfy $Q_0 = Q$, $b_0 = b$. We require $\log q_t$ to satisfy the same HJE in Eq. (17). Since $\log p_t$ already satisfies Eq. (17), substituting the ansatz in Eq. (18) and taking the difference gives the condition

$$-\frac{1}{2} x^\top \dot{Q}_t x + x^\top \dot{b}_t + \dot{F}(t) = (-Q_t x + b_t) \cdot (x + 2\nabla f_t) + \operatorname{tr}(Q_t) + \|\nabla f_t\|^2 - \|\nabla f_t - Q_t x + b_t\|^2. \tag{19}$$

By expanding this expression and independently matching the coefficients of the quadratic, linear, and scalar terms in x, we derive the *Riccati ordinary differential equations (ODEs)*,

$$\begin{cases} \dot{Q}_t = 2(I + Q_t)Q_t, & \text{with } Q_0 = Q, \\ \dot{b}_t = (I + 2Q_t)b_t, & \text{with } b_0 = b. \end{cases} \tag{20}$$

Note that the coupled ODEs are closed in (Q_t, b_t) and do not depend on p_t, which makes the tilted transport approach computationally tractable.

Closed-form solution. Let $A = U\Sigma V^\top$ be the SVD of A with singular values $\lambda_1 \geq \cdots \geq \lambda_n > 0$, so that $Q = \sigma_y^{-2} V \operatorname{diag}(\lambda_1^2, \ldots, \lambda_n^2, 0, \ldots, 0) V^\top$. Since Q is diagonalized by V, so is Q_t for all t, and the first ODE in Eq. (20) reduces to the scalar equation $\dot{q} = 2(1+q)q$ for each non-zero diagonal entry. This ODE has explicit solution

$$Q_t = V \operatorname{diag}\left(\frac{e^{2t}}{1 + \sigma_y^2/\lambda_1^2 - e^{2t}}, \ldots, \frac{e^{2t}}{1 + \sigma_y^2/\lambda_n^2 - e^{2t}}, 0, \ldots, 0 \right) V^\top, \tag{21}$$

which is well-defined up to the critical blowup time $T^* = \frac{1}{2}\log(1 + \sigma_y^2/\lambda_1^2)$. To solve for the linear term b_t, we project it onto the eigenvectors of Q_t as $b_t = \sum_i \xi_i(t) v_i$. By observing that the ratio $\xi_i(t)/(Q_t)_{ii}$ satisfies the decay equation $\frac{d}{dt}[\xi_i/(Q_t)_{ii}] = -\xi_i/(Q_t)_{ii}$, we obtain the closed-form expression:

$$\xi_i(t) = \frac{e^t (U^\top y)_i}{\lambda_i (1 + \sigma_y^2/\lambda_i^2 - e^{2t})}, \qquad i = 1, \ldots, n, \tag{22}$$

where $\xi_i(t) = 0$ for all indices $i > n$.

The tilted family and its exactness. Combining the derivations above, the family

$$q_t(x) \propto p_t(x) \exp\left(-\frac{1}{2} x^\top Q_t x + x^\top b_t\right) \tag{23}$$

is explicitly constructed to satisfy the same HJE Eq. (17) as p_t, with the initial condition q_0 equal to the target posterior in Eq. (6). By Anderson's time-reversal theorem, it follows that if we initialize $X_{T^*} \sim q_{T^*}$ and simulate the reverse SDE in Eq. (3) from T^* to 0 using the prior score $s(\cdot, t)$, the intermediate states will perfectly track the tilted marginals, $X_s \sim q_{T^*-s}$ for all $s \in [0, T^*]$. In particular, this guarantees that the final state precisely recovers the target posterior, yielding $X_{T^*} \sim q_0$.

C Proof of theoretical results

C.1 Proof of Theorem 1

Proof. Let $s_t(\cdot) \coloneqq \nabla \log p_t(\cdot)$ denote the score function of the untilted marginal p_t, which need not have a closed form. Since $q_t \propto e^{\mathcal{L}_t} p_t$, the score of q_t decomposes as

$$\nabla \log q_t(x) = \nabla \mathcal{L}_t(x) + s(x, t). \tag{24}$$

Step 1: Fokker–Planck equation for q_t.

The untilted process $dX_t = -X_t\, dt + \sqrt{2}\, dW_t$ has marginal p_t satisfying

$$\partial_t p_t = \nabla \cdot (x\, p_t) + \Delta p_t. \tag{25}$$

Thus the time-reverse Fokker–Planck equation reads

$$\partial_t \overleftarrow{p_t} = -\nabla \cdot [(x + 2 s_{T-t})\, \overleftarrow{p_t}] + \Delta \overleftarrow{p_t}. \tag{26}$$

Since q_t differs from $\overleftarrow{p_t}$ only by the initial condition, the Fokker–Planck equation of q_t is the same as Eq. (26). For clarity we denote q_t, $t \in [0, T^*]$ as the solution of

$$\partial_t \tilde{v}_t = -\nabla \cdot [(x + 2 s_{T^*-t})\, \tilde{v}_t] + \Delta \tilde{v}_t, \quad \tilde{v}_0 = q_{T^*}. \tag{27}$$

Step 2: Breaking down q_t.

By definition $q_t = e^{\mathcal{L}_t} p_t$ we have

$$\nabla \log q_t = \nabla \log p_{T^*-t} + \nabla \mathcal{L}_t = s_{T^*-t} + \nabla \mathcal{L}_t. \tag{28}$$

Step 3: Time-reversal. By time-reversing Eq. (27), we have

$$\begin{aligned} \partial_t v_t &= \nabla \cdot [(x + 2 s_{T^*-t})\, v_t] - \Delta v_t \\ &= \nabla \cdot [(x + 2 s_{T^*-t})\, v_t] - 2\nabla \cdot (\nabla \log v_t \cdot v_t) + \Delta v_t \\ &= \nabla \cdot [(x + 2 s_{T^*-t} - 2\nabla \log v_t)\, v_t] + \Delta v_t. \end{aligned}$$

Since q_t is the solution, we substitute $\nabla \log v_t$ by Eq. (28) and obtain

$$\partial_t v_t = -\nabla \cdot [(2\nabla \mathcal{L}_t - x)\, v_t] + \Delta v_t, \quad v_0 = q_0. \tag{29}$$

Denote the solution of Eq. (29) as $\overrightarrow{q_t}$, the corresponding SDE is

$$dX_t = -X_t\, dt + \sqrt{2}\, dW_t + 2\nabla \mathcal{L}_t(X_t)\, dt, \quad X_0 \sim q_0.$$

□

C.2 Proof of Theorem 2

To prove Theorem 2 we first state the following lemmas. Proofs of the lemmas are delayed to Appendix C.3 and Appendix C.4.

Lemma 1 (Explicit transition kernel and strict positivity)**.** *Under Assumption 1, the one-step transition kernel K_{T^*} of the forward–reverse iteration admits the explicit form*

$$K_{T^*}(x, x') = p_0(x') \int_{\mathbb{R}^d} \frac{p^{\mathrm{OU}}_{T^*}(z \mid x)\, p^{\mathrm{OU}}_{T^*}(z \mid x')\, e^{\mathcal{L}_{T^*}(z)}}{p_{T^*}(z)\, Z_{\mathrm{tilt}}(x)}\, dz, \tag{30}$$

where $p^{\mathrm{OU}}_{T^*}(z \mid x) = \mathcal{N}\big(z;\, xe^{-T^*},\, (1 - e^{-2T^*})I\big)$ is the OU transition density, p_{T^*} is the marginal of the (unconditional) forward process at time T^*, and $Z_{\mathrm{tilt}}(x) = \int p^{\mathrm{OU}}_{T^*}(z \mid x)\, e^{\mathcal{L}_{T^*}(z)}\, p_{T^*}(z)^{-1}\, dz$ is a finite normalizing factor. In particular,

$$K_{T^*}(x, x') > 0 \qquad \text{for all } x, x' \in \mathbb{R}^d \tag{31}$$

Lemma 2 (Reversibility and Harris recurrence)**.** *Under Assumption 1 and the hypotheses of Theorem 1, the following hold.*

(i) ***(Reversibility.)*** *The kernel K_{T^*} satisfies the detailed balance condition with respect to q_0:*

$$q_0(x)\, K_{T^*}(x, x') = q_0(x')\, K_{T^*}(x', x) \qquad \forall\, x, x' \in \mathbb{R}^d. \tag{32}$$

In particular, q_0 is a stationary distribution of K_{T^*}.

(ii) ***(Harris recurrence.)*** *The Markov chain with kernel K_{T^*} is Harris recurrent with invariant measure q_0.*

Proof of Theorem 2. The result follows immediately from Lemma 2(ii). Specifically, the Harris recurrence and aperiodicity of K_{T^*} established therein, together with q_0 as its unique stationary distribution from Lemma 2(i), yield

$$\lim_{N \to \infty} \left\| K^{(N)}_{T^*} p - q_0 \right\|_{\mathrm{TV}} = 0, \tag{33}$$

for any initial distribution p, under Assumption 1 and the zero score approximation error assumption of Theorem 1. □

C.3 Proof of Lemma 1

Proof. **Step 1: Forward kernel F_{T^*}.** The tilted forward SDE in Eq. (9) may be written as a Girsanov perturbation of the standard OU process

$$dX_t = -X_t\, dt + \sqrt{2}\, dW_t.$$

Let $\mathbb{P}^{\mathrm{OU}}_x$ denote the law of the OU process started at x. Under Assumption 1(iii), the Novikov condition is satisfied, so by Girsanov's theorem, the tilted forward measure $\mathbb{Q}_x$ is absolutely continuous with respect to $\mathbb{P}^{\mathrm{OU}}_x$, with Radon-Nikodým density

$$\frac{d\mathbb{Q}_x}{d\mathbb{P}^{\mathrm{OU}}_x} = \exp\left(\sqrt{2} \int_0^{T^*} \nabla \mathcal{L}_t(X_t)^\top dW_t - \int_0^{T^*} \|\nabla \mathcal{L}_t(X_t)\|^2\, dt \right) =: \mathcal{E}_{T^*}.$$

Marginalizing $\mathcal{E}_{T^*}$ over all paths that end at z and using the tower property, the forward transition kernel density is shown in the following lemma.

Lemma 3 (Forward transition kernel)**.** *Under Assumption 1, the forward transition kernel $F_{T^*}(x, \cdot)$ admits the explicit form*

$$F_{T^*}(x, z) = \frac{p^{\mathrm{OU}}_{T^*}(z \mid x)\, e^{\mathcal{L}_{T^*}(z)}}{Z_{\mathrm{tilt}}(x)}, \tag{34}$$

where $p^{\mathrm{OU}}_{T^*}(z \mid x) = \mathcal{N}\big(z;\, xe^{-T^*},\, (1 - e^{-2T^*})I\big)$ is the OU transition density, and

$$Z_{\mathrm{tilt}}(x) \coloneqq \int_{\mathbb{R}^d} p^{\mathrm{OU}}_{T^*}(z \mid x)\, e^{\mathcal{L}_{T^*}(z)}\, dz \;<\; \infty. \tag{35}$$

A detailed proof is presented in Appendix C.5.

Step 2: Reverse kernel R_{T^*}. The reverse SDE uses the unconditional score $\nabla \log p_t$, so it is the exact time-reversal of the standard OU process. By Bayes' rule applied to the OU joint density, the conditional law of X_0 given $X_{T^*} = z$ under the OU process is

$$R_{T^*}(z, x') = \frac{p_0(x')\, p_{T^*}^{\mathrm{OU}}(z \mid x')}{p_{T^*}(z)}.$$

Step 3: Composed kernel. Composing the two kernels gives

$$\begin{aligned} K_{T^*}(x, x') &= \int_{\mathbb{R}^d} F_{T^*}(x, z)\, R_{T^*}(z, x')\, dz \\ &= \int_{\mathbb{R}^d} \frac{p_{T^*}^{\mathrm{OU}}(z \mid x)\, e^{\mathcal{L}_{T^*}(z)}}{Z_{\mathrm{tilt}}(x)} \cdot \frac{p_0(x')\, p_{T^*}^{\mathrm{OU}}(z \mid x')}{p_{T^*}(z)}\, dz, \end{aligned}$$

which is exactly Eq. (30).

Step 4: Strict positivity. Fix any $x, x' \in \mathbb{R}^d$. The integrand in Eq. (30) is non-negative, so it suffices to show it is positive on a set of positive measure. Since $p_{T^*}^{\mathrm{OU}}(\cdot \mid x)$ and $p_{T^*}^{\mathrm{OU}}(\cdot \mid x')$ are both full-support Gaussian densities on $\mathbb{R}^d$, their product is strictly positive everywhere. Furthermore $e^{\mathcal{L}_{T^*}(z)} > 0$ everywhere, and $p_{T^*}(z) > 0$ everywhere. Hence the integrand is strictly positive for all z, so $K_{T^*}(x, x') > 0$. □

C.4 Proof of Lemma 2

Proof. **Proof of (i): detailed balance.** We verify $q_0(x)\, K_{T^*}(x, x') = q_0(x')\, K_{T^*}(x', x)$ directly using the time-reversal identity from Theorem 1.

Denote the joint density of (X_0, X_{T^*}) under the tilted forward process Eq. (9) initialized at $X_0 \sim q_0$ by

$$p^{\rightarrow}(x, z) \coloneqq q_0(x)\, F_{T^*}(x, z). \tag{36}$$

By Theorem 1, the tilted forward process $\overrightarrow{q_t}$ is the time-reversal of the reverse process q_t, and in particular the two processes share the same joint distribution over their endpoints. This yields the fundamental identity

$$q_0(x)\, F_{T^*}(x, z) = \overrightarrow{q_{T^*}}(z)\, R_{T^*}(z, x), \tag{37}$$

which states that the probability of the forward chain going from x to z (weighted by the initial distribution q_0) equals the probability of the reverse chain going from z to x (weighted by the terminal distribution $\overrightarrow{q_{T^*}}$).

Now compute $q_0(x)\, K_{T^*}(x, x')$ using the composition $K_{T^*} = F_{T^*} \circ R_{T^*}$:

$$\begin{aligned} q_0(x)\, K_{T^*}(x, x') &= \int_{\mathbb{R}^d} q_0(x)\, F_{T^*}(x, z)\, R_{T^*}(z, x')\, dz \\ &= \int_{\mathbb{R}^d} \overrightarrow{q_{T^*}}(z)\, R_{T^*}(z, x)\, R_{T^*}(z, x')\, dz, \end{aligned} \tag{38}$$

where the second line substitutes Eq. (37). The right-hand side of Eq. (38) is manifestly symmetric in $x \leftrightarrow x'$, since $R_{T^*}(z, x)\, R_{T^*}(z, x')$ is symmetric in its two arguments. Therefore

$$q_0(x)\, K_{T^*}(x, x') = q_0(x')\, K_{T^*}(x', x) \qquad \forall\, x, x' \in \mathbb{R}^d, \tag{39}$$

establishing Eq. (32). Since $K_{T^*}(x, x') > 0$ for all x, x' (Lemma 1), uniqueness of the stationary distribution follows from the Harris recurrence established in part (ii).

Proof of (ii): Harris recurrence. We verify the three conditions for Harris recurrence of K_{T^*} with respect to q_0.

Irreducibility. Fix any Borel set $A \subset \mathbb{R}^d$ with $q_0(A) > 0$ and any starting point $x \in \mathbb{R}^d$. From Lemma 1, the kernel satisfies $K_{T^*}(x, x') > 0$ for all x, x'. Therefore

$$K_{T^*}(x, A) = \int_A K_{T^*}(x, x')\, dx' > 0, \tag{40}$$

so the chain can reach any positive-measure set in a single step from anywhere. This establishes q_0-irreducibility.

Aperiodicity. Since $K_{T^*}(x, x) > 0$ for all x (again by strict positivity from Lemma 1), the chain has a positive probability of remaining at its current location, which rules out periodicity. Hence the chain is aperiodic.

Harris recurrence. By q_0-irreducibility together with q_0 being a stationary distribution (established in part (i)), a standard result in the theory of Markov chains (Tierney, 1994) implies that the chain is positive Harris recurrent with invariant measure q_0.

□

C.5 Proof of Lemma 3

Proof. The proof proceeds in three steps: (i) we derive a PDE that $\mathcal{L}_t$ satisfies by enforcing consistency of the density ratio q_t/p_t under the respective Fokker–Planck equations; (ii) we use this PDE together with Itô's formula to show that the Girsanov weight $\mathcal{E}_{T^*}$ depends only on the endpoints of the trajectory; (iii) we marginalize over paths to obtain the kernel.

Step 1: PDE for $\mathcal{L}_t$. Define the density ratio

$$r_t(x) \coloneqq \frac{q_t(x)}{p_t(x)} = \frac{e^{\mathcal{L}_t(x)}}{Z_t}, \qquad Z_t \coloneqq \int_{\mathbb{R}^d} e^{\mathcal{L}_t(x)}\, p_t(x)\, dx, \tag{41}$$

where p_t is the marginal of the standard OU forward process and q_t is the marginal of the tilted forward process Eq. (9). Both marginals satisfy Fokker–Planck equations. For p_t (standard OU):

$$\partial_t p_t = \nabla \cdot (x\, p_t) + \Delta p_t, \tag{42}$$

and for q_t (tilted forward SDE with additional drift $2\nabla\mathcal{L}_t$):

$$\partial_t q_t = \nabla \cdot (x\, q_t) + \Delta q_t - 2\nabla \cdot (q_t\, \nabla\mathcal{L}_t). \tag{43}$$

Writing $q_t = r_t p_t$, expanding $\partial_t(r_t p_t) = p_t \partial_t r_t + r_t \partial_t p_t$, and substituting Eq. (42) to eliminate $r_t \partial_t p_t$ gives

$$\begin{aligned} p_t\, \partial_t r_t &= \nabla \cdot (x\, r_t p_t) - r_t \nabla \cdot (x\, p_t) + \Delta(r_t p_t) - r_t \Delta p_t - 2\nabla \cdot (r_t p_t\, \nabla\mathcal{L}_t) \\ &= p_t\, x \cdot \nabla r_t + \Delta(r_t p_t) - r_t \Delta p_t - 2\nabla \cdot (r_t p_t\, \nabla\mathcal{L}_t), \end{aligned} \tag{44}$$

where the second line uses the identity $\nabla \cdot (x\, r_t p_t) - r_t \nabla \cdot (x\, p_t) = p_t\, x \cdot \nabla r_t$. Expanding $\Delta(r_t p_t) = p_t \Delta r_t + 2\nabla r_t \cdot \nabla p_t + r_t \Delta p_t$ and $\nabla \cdot (r_t p_t\, \nabla\mathcal{L}_t) = p_t \nabla r_t \cdot \nabla\mathcal{L}_t + r_t \nabla p_t \cdot \nabla\mathcal{L}_t + r_t p_t \Delta\mathcal{L}_t$, then dividing by p_t, yields

$$\partial_t r_t = x \cdot \nabla r_t + \Delta r_t + 2\nabla r_t \cdot \nabla \log p_t - 2\nabla r_t \cdot \nabla\mathcal{L}_t - 2 r_t\, \nabla\mathcal{L}_t \cdot \nabla \log p_t - 2 r_t \Delta\mathcal{L}_t. \tag{45}$$

Now substitute $r_t = e^{\mathcal{L}_t}/Z_t$ into Eq. (45). Since Z_t depends only on t,

$$\partial_t r_t = \frac{e^{\mathcal{L}_t}}{Z_t}\Big(\partial_t\mathcal{L}_t - \frac{\dot{Z}_t}{Z_t}\Big), \qquad \nabla r_t = \frac{e^{\mathcal{L}_t}}{Z_t}\nabla\mathcal{L}_t, \qquad \Delta r_t = \frac{e^{\mathcal{L}_t}}{Z_t}\big(\Delta\mathcal{L}_t + \|\nabla\mathcal{L}_t\|^2\big). \tag{46}$$

Substituting these into Eq. (45) and dividing by $e^{\mathcal{L}_t}/Z_t$:

$$\begin{aligned} \partial_t\mathcal{L}_t - \frac{\dot{Z}_t}{Z_t} = x \cdot \nabla\mathcal{L}_t + \big(\Delta\mathcal{L}_t + \|\nabla\mathcal{L}_t\|^2\big) + 2\nabla\mathcal{L}_t \cdot \nabla \log p_t \\ - 2\|\nabla\mathcal{L}_t\|^2 - 2\nabla\mathcal{L}_t \cdot \nabla \log p_t - 2\Delta\mathcal{L}_t. \end{aligned}$$

The $2\nabla\mathcal{L}_t \cdot \nabla \log p_t$ terms cancel exactly, and collecting the remaining terms gives

$$\partial_t \mathcal{L}_t - \frac{\dot{Z}_t}{Z_t} = x \cdot \nabla\mathcal{L}_t - \Delta\mathcal{L}_t - \|\nabla\mathcal{L}_t\|^2. \tag{47}$$

Setting $c(t) \coloneqq \dot{Z}_t/Z_t$, which depends only on t, and rearranging:

$$\partial_t \mathcal{L}_t + \Delta\mathcal{L}_t - x^\top \nabla\mathcal{L}_t + \|\nabla\mathcal{L}_t\|^2 = c(t). \tag{48}$$

Step 2: The Girsanov weight is endpoint-only. By Assumption 1(iii), the Novikov condition holds, so by Girsanov's theorem the tilted forward measure $\mathbb{Q}_x$ is absolutely continuous with respect to the OU measure $\mathbb{P}_x^{\mathrm{OU}}$, with Radon–Nikodým density

$$\mathcal{E}_{T^*} \coloneqq \frac{d\mathbb{Q}_x}{d\mathbb{P}_x^{\mathrm{OU}}} = \exp\left(\sqrt{2}\int_0^{T^*} \nabla\mathcal{L}_t(X_t)^\top dW_t - \int_0^{T^*} \|\nabla\mathcal{L}_t(X_t)\|^2\, dt\right), \tag{49}$$

where the factor $\sqrt{2}$ arises from the diffusion coefficient of the OU process, and the quadratic variation term is $\frac{1}{2}\cdot(\sqrt{2})^2\|\nabla\mathcal{L}_t\|^2 = \|\nabla\mathcal{L}_t\|^2$.

We apply Itô's formula to $\mathcal{L}_t(X_t)$ along the OU process $dX_t = -X_t\,dt + \sqrt{2}\,dW_t$:

$$d\mathcal{L}_t(X_t) = \sqrt{2}\,\nabla\mathcal{L}_t^\top dW_t + \big(\partial_t\mathcal{L}_t + \Delta\mathcal{L}_t - X_t^\top\nabla\mathcal{L}_t\big)dt, \tag{50}$$

where the $\Delta\mathcal{L}_t$ arises from the Itô correction with diffusion coefficient $\sqrt{2}$, i.e., $\frac{1}{2}\cdot 2\cdot\Delta\mathcal{L}_t = \Delta\mathcal{L}_t$. Integrating and rearranging:

$$\sqrt{2}\int_0^{T^*} \nabla\mathcal{L}_t^\top dW_t = \mathcal{L}_{T^*}(X_{T^*}) - \mathcal{L}_0(X_0) - \int_0^{T^*}\big(\partial_t\mathcal{L}_t + \Delta\mathcal{L}_t - X_t^\top\nabla\mathcal{L}_t\big)dt. \tag{51}$$

Substituting the PDE Eq. (48), which gives $\partial_t\mathcal{L}_t + \Delta\mathcal{L}_t - X_t^\top\nabla\mathcal{L}_t = c(t) - \|\nabla\mathcal{L}_t\|^2$:

$$\sqrt{2}\int_0^{T^*} \nabla\mathcal{L}_t^\top dW_t = \mathcal{L}_{T^*}(X_{T^*}) - \mathcal{L}_0(X_0) - \int_0^{T^*} c(t)\,dt + \int_0^{T^*}\|\nabla\mathcal{L}_t\|^2\,dt. \tag{52}$$

Substituting into the Girsanov exponent Eq. (49):

$$\begin{aligned}
\log\mathcal{E}_{T^*} &= \sqrt{2}\int_0^{T^*}\nabla\mathcal{L}_t^\top dW_t - \int_0^{T^*}\|\nabla\mathcal{L}_t\|^2\,dt \\
&= \mathcal{L}_{T^*}(X_{T^*}) - \mathcal{L}_0(X_0) - \int_0^{T^*} c(t)\,dt + \int_0^{T^*}\|\nabla\mathcal{L}_t\|^2\,dt - \int_0^{T^*}\|\nabla\mathcal{L}_t\|^2\,dt \\
&= \mathcal{L}_{T^*}(X_{T^*}) - \mathcal{L}_0(X_0) - \int_0^{T^*} c(t)\,dt,
\end{aligned}$$

where the $\|\nabla\mathcal{L}_t\|^2$ terms cancel exactly. Since $\mathcal{L}_0(X_0) = \mathcal{L}_0(x)$ and $\int_0^{T^*} c(t)\,dt = \log(Z_{T^*}/Z_0)$ are both constants independent of the path, the Girsanov weight depends only on the endpoints:

$$\mathcal{E}_{T^*} = \frac{Z_0}{Z_{T^*}}\, e^{\mathcal{L}_{T^*}(X_{T^*}) - \mathcal{L}_0(x)}. \tag{53}$$

Step 3: Deriving F_{T^*}. The forward transition kernel is the marginal law of X_{T^*} under $\mathbb{Q}_x$. By Eq. (53):

$$\begin{aligned}
F_{T^*}(x,z) &= p_{T^*}^{\mathrm{OU}}(z \mid x)\cdot\mathbb{E}^{\mathbb{P}_x^{\mathrm{OU}}}\Big[\mathcal{E}_{T^*}\;\Big|\; X_{T^*} = z\Big] \\
&= p_{T^*}^{\mathrm{OU}}(z \mid x)\cdot\frac{Z_0}{Z_{T^*}}\, e^{\mathcal{L}_{T^*}(z) - \mathcal{L}_0(x)},
\end{aligned}$$

where the conditional expectation is trivial because $\mathcal{E}_{T^*}$ depends on the path only through its endpoint $X_{T^*} = z$. Absorbing the constants Z_0/Z_{T^*} and $e^{-\mathcal{L}_0(x)}$ into the normalizing factor

$$Z_{\mathrm{tilt}}(x) \coloneqq \frac{Z_0}{Z_{T^*}}\, e^{-\mathcal{L}_0(x)}\int_{\mathbb{R}^d} p_{T^*}^{\mathrm{OU}}(z \mid x)\, e^{\mathcal{L}_{T^*}(z)}\,dz, \tag{54}$$

which can be verified to equal $\int p_{T^*}^{\mathrm{OU}}(z \mid x)\, e^{\mathcal{L}_{T^*}(z)}\,dz$ up to a global constant independent of x, we obtain Eq. (34). Finiteness of $Z_{\mathrm{tilt}}(x)$ follows from Assumption 1(ii) and the Gaussian decay of $p_{T^*}^{\mathrm{OU}}(\cdot \mid x)$. □

C.6 Proof of Theorem 3

Before proving the theorem, we first state the key lemma, whose proof could be found in Appendix C.7.

Lemma 4 (χ^2-contractivity of K_{T^*})**.** *Under Assumption 1, there exists $\lambda_{T^*} = \lambda(q_0, T^*) \in (0,1)$ such that for any distribution p with $\chi^2(p \,\|\, q_0) < \infty$,*

$$\chi^2\Big(K_{T^*}^{(N)} p \,\Big\|\, q_0\Big) \le \lambda_{T^*}^{2N}\, \chi^2(p \,\|\, q_0), \qquad \forall\, N \ge 1. \tag{55}$$

Proof of Theorem 3. Combining the standard comparison inequality (Polyanskiy & Wu, 2025)

$$\mathrm{TV}(p, q) \le \sqrt{\chi^2(p \,\|\, q)} \tag{56}$$

with the χ^2-contractivity from Lemma 4:

$$\begin{aligned} \Big\| K_{T^*}^{(N)} p - q_0 \Big\|_{\mathrm{TV}} &\le \sqrt{\chi^2\Big(K_{T^*}^{(N)} p \,\Big\|\, q_0\Big)} \\ &\le \lambda_{T^*}^{N}\, \sqrt{\chi^2(p \,\|\, q_0)}. \end{aligned} \tag{57}$$

It remains to choose N so that $\lambda_{T^*}^N \le \varepsilon_{\mathrm{acc}}$. Taking logarithms, this requires

$$N \ge \frac{\log(1/\varepsilon_{\mathrm{acc}})}{\log(1/\lambda_{T^*})}. \tag{58}$$

Since $\log(1/\lambda_{T^*}) \ge 1 - \lambda_{T^*}$ for $\lambda_{T^*} \in (0,1)$, a sufficient condition is $N \asymp \frac{\log(1/\varepsilon_{\mathrm{acc}})}{1-\lambda_{T^*}}$, yielding Eq. (12). □

C.7 Proof of Lemma 4

Proof. **Step 1: Spectral structure of K_{T^*}.** Define the integral operator $\mathcal{K} : L^2(q_0) \to L^2(q_0)$ by

$$(\mathcal{K} f)(x) \coloneqq \int K_{T^*}(x, x')\, f(x')\, dx'. \tag{59}$$

By the detailed balance condition Eq. (32) from Lemma 2(i), $\mathcal{K}$ is self-adjoint on $L^2(q_0)$. Compactness of $\mathcal{K}$ follows from the square-integrability of $K_{T^*}(x, x')/q_0(x')$ with respect to $q_0 \otimes q_0$ in view of (Saloff-Coste, 2006).

Since $K_{T^*}(x, x') > 0$ for all $x, x' \in \mathbb{R}^d$ (Lemma 1), the Jentzsch–Perron theorem for strictly positive operators (Schaefer et al., 2012) implies:

(a) 1 is a simple eigenvalue of $\mathcal{K}$, with the constant function $\mathbf{1}$ as the unique (up to scaling) right eigenfunction.

(b) Every other eigenvalue μ satisfies $|\mu| \le \lambda_{T^*}$ for some $\lambda_{T^*} = \lambda(q_0, T^*) \in (0,1)$.

Step 2: Operator norm bound. Let $\mathcal{P}_{\mathbf{1}}$ denote the orthogonal projection onto $\mathbf{1}$ in $L^2(\mathbb{R}^d, q_0)$, defined by $\mathcal{P}_{\mathbf{1}} f(x) \coloneqq \int_{\mathbb{R}^d} f(x')\, q_0(x')\, dx'$. Since $\mathcal{K}\mathcal{P}_{\mathbf{1}} = \mathcal{P}_{\mathbf{1}}\mathcal{K} = \mathcal{P}_{\mathbf{1}}$, expanding the product yields $(\mathcal{K} - \mathcal{P}_{\mathbf{1}})^N = \mathcal{K}^N - \mathcal{P}_{\mathbf{1}}$. By the spectral bound from Step 1, the operator $\mathcal{K} - \mathcal{P}_{\mathbf{1}}$ has all eigenvalues bounded in modulus by λ_{T^*}, so

$$\big\| \mathcal{K}^N - \mathcal{P}_{\mathbf{1}} \big\|_{L^2(\mathbb{R}^d, q_0) \to L^2(\mathbb{R}^d, q_0)} \le \lambda_{T^*}^N. \tag{60}$$

Step 3: Bounding the inner product. For any $f \in L^2(q_0)$ with $\|f\|_{L^2(q_0)} \le 1$, we compute

$$\begin{aligned}
&\left| \int \left(K_{T^*}^{(N)} p(x) - q_0(x) \right) f(x)\, dx \right| \\
&\quad = \left| \int p(x)\, (\mathcal{K}^N f)(x)\, dx - \int q_0(x)\, f(x)\, dx \right| \\
&\quad = \left| \int \left(\frac{p(x)}{q_0(x)} - 1 \right) \left(\mathcal{K}^N - \mathcal{P}_{\mathbf{1}} \right) f(x)\, q_0(x)\, dx \right| \\
&\quad \le \left\| \frac{p}{q_0} - 1 \right\|_{L^2(q_0)} \left\| \left(\mathcal{K}^N - \mathcal{P}_{\mathbf{1}} \right) f \right\|_{L^2(q_0)} \\
&\quad \le \lambda_{T^*}^N \sqrt{\chi^2(p \,\|\, q_0)},
\end{aligned} \tag{61}$$

where the first inequality is Cauchy–Schwarz and the second uses Eq. (60) together with $\|p/q_0 - 1\|_{L^2(q_0)}^2 = \chi^2(p \,\|\, q_0)$.

Step 4: Optimal test function. Set

$$f(x) := \frac{K_{T^*}^{(N)} p(x) - q_0(x)}{q_0(x)}, \tag{62}$$

whose $L^2(q_0)$-norm satisfies $\|f\|_{L^2(q_0)} = \sqrt{\chi^2(K_{T^*}^{(N)} p \,\|\, q_0)}$. Then

$$\int \left(K_{T^*}^{(N)} p(x) - q_0(x) \right) f(x)\, dx = \chi^2 \left(K_{T^*}^{(N)} p \,\|\, q_0 \right). \tag{63}$$

Substituting into Eq. (61):

$$\chi^2 \left(K_{T^*}^{(N)} p \,\|\, q_0 \right) \le \lambda_{T^*}^N \sqrt{\chi^2(p \,\|\, q_0)} \cdot \sqrt{\chi^2 \left(K_{T^*}^{(N)} p \,\|\, q_0 \right)}. \tag{64}$$

Dividing both sides by $\sqrt{\chi^2(K_{T^*}^{(N)} p \,\|\, q_0)}$ (which is positive unless $K_{T^*}^{(N)} p = q_0$, in which case the bound holds trivially):

$$\chi^2 \left(K_{T^*}^{(N)} p \,\middle\|\, q_0 \right) \le \lambda_{T^*}^{2N}\, \chi^2(p \,\|\, q_0), \tag{65}$$

which corresponds to Eq. (55). □

C.8 Proof of Corollary 1 and Corollary 2

As the proofs of Theorem 2 and Theorem 3 do not depend on the value of T^*, both theorems hold when one substitutes T^* by any $t \in (0, T^*)$. Thus Corollary 1 and Corollary 2 follow naturally. One should note that $\lambda_t \neq \lambda_{T^*}$ in general, thus the convergence speed is different in Corollary 2.

C.9 Robustness to score approximation error

Throughout the main text we assumed exact score estimation. Here we relax that assumption and quantify how an imperfect score network degrades the guarantees of Theorem 3. Two structural features of FRITES make the analysis unusually clean.

First, the score enters only through the reverse kernel. In the linear setting the forward kernel F_t is the tilted Ornstein–Uhlenbeck process of Lemma 3, whose transition density is available in closed form through $\mathcal{L}_t$ and involves no learned quantity. Only R_t must be replaced by its empirical counterpart $\widehat{R}_t$, so the perturbed kernel is

$$\widehat{K}_t \;=\; F_t \circ \widehat{R}_t, \qquad \mathcal{E}_t \;:=\; K_t - \widehat{K}_t. \tag{66}$$

Second, the reverse process of FRITES is initialized at the boosted posterior and therefore has marginals q_t, whereas the score network is trained under the prior marginals p_t. The relevant error is thus measured under q_t rather than p_t. Because the tilted family has an *explicit* density ratio with respect to p_t, this change of measure costs only the constant identified in Lemma 5 below, with no additional moment conditions.

We emphasise at the outset that $\widehat{K}_t$ is no longer self-adjoint on $L^2(q_0)$: the detailed balance identity of Lemma 2 rests on the exact time-reversal of Theorem 1, which fails once $s_\theta \neq s$. Consequently q_0 is no longer invariant, and the correct statement is not convergence to q_0 but convergence to within a bias floor of it. We therefore treat $\widehat{K}_t$ as a perturbation of K_t and invoke the spectral gap of the *exact* kernel, rather than attempting to establish a gap for $\widehat{K}_t$ directly.

Assumption 2 (Score error)**.** *The score network s_θ satisfies the following.*

(i) (Prior-weighted accuracy.) *For every $t \in [0, T^*]$,*

$$\mathbb{E}_{p_t}\Big[\, \big\| s_\theta(\cdot, t) - \nabla \log p_t(\cdot) \big\|^2 \,\Big] \;\le\; \varepsilon_{\text{score}}^2. \tag{67}$$

(ii) (Bounded perturbation.) *The perturbation operator $\mathcal{E}_t$ of Eq.* (66)*, acting on signed measures with density in $L^2(q_0)$, is bounded: $\|\mathcal{E}_t\|_{L^2(q_0)\to L^2(q_0)} \le \delta_t < \infty$.*

Assumption 2 (i) is the standard denoising-score-matching guarantee and is the same quantity controlled in the analysis of Xu & Chi (2024).

Lemma 5 (Tilt-induced change of measure)**.** *Let $Z_t := \int_{\mathbb{R}^d} e^{\mathcal{L}_t(x)} p_t(x)\, dx$ and*

$$M \;:=\; \sup_{t\in[0,T^*]} \frac{\exp\big(\sup_{x\in\mathbb{R}^d} \mathcal{L}_t(x)\big)}{Z_t}. \tag{68}$$

Under Assumption 1 (ii) we have $M < \infty$, and for every measurable $g \ge 0$ and every $t \in [0, T^*]$,

$$\mathbb{E}_{q_t}[\, g \,] \;\le\; M\, \mathbb{E}_{p_t}[\, g \,]. \tag{69}$$

In particular, under Assumption 2 (i), $\mathbb{E}_{q_t}\big[\|\Delta s_t\|^2\big] \le M\, \varepsilon_{\text{score}}^2$, where $\Delta s_t := s_\theta(\cdot, t) - \nabla \log p_t(\cdot)$.

Proof. By construction $q_t = (e^{\mathcal{L}_t}/Z_t)\, p_t$, so the density ratio q_t/p_t is bounded above by $\exp(\sup_x \mathcal{L}_t)/Z_t \le M$, which is finite by Assumption 1 (ii) together with $Z_t > 0$ (a consequence of Assumption 1 (iv), which gives $p_t > 0$ everywhere). Eq. (69) is then immediate, and the final claim follows by taking $g = \|\Delta s_t\|^2$ and applying Eq. (67). □

In the linear setting M is computable in closed form from (Q_t, b_t) via Eq. (21), so it is an explicit function of the measurement signal-to-noise ratio rather than an abstract constant. It degenerates only as $t \to T^*$, where $\mathcal{L}_t$ blows up; this is a further reason to take t bounded away from T^* in the linear case.

Step 1: per-step reverse-kernel error.

Lemma 6 (Reverse kernel perturbation)**.** *Under Assumption 1 and Assumption 2 (i), for every $t \in (0, T^*]$,*

$$\mathrm{KL}\big(R_t \,\big\|\, \widehat{R}_t\big) \;\le\; \int_0^t \mathbb{E}_{q_s}\big[\|\Delta s_s\|^2\big]\, ds \;\le\; M\, t\, \varepsilon_{\text{score}}^2, \tag{70}$$

where the divergence is between the laws of the two reverse processes on $[0, t]$ started from the common law q_t. Consequently

$$\big\|R_t - \widehat{R}_t\big\|_{\text{TV}} \;\le\; \varepsilon_{\text{score}} \sqrt{\tfrac{1}{2} M t}, \qquad \textit{and} \qquad \big\|K_t - \widehat{K}_t\big\|_{\text{TV}} \;\le\; \varepsilon_{\text{score}} \sqrt{\tfrac{1}{2} M t}. \tag{71}$$

Proof. The exact and approximate reverse processes solve

$$d\overleftarrow{X}_u = \big(\overleftarrow{X}_u + 2\, s(\overleftarrow{X}_u, t-u)\big)du + \sqrt{2}\, d\overleftarrow{W}_u, \qquad d\overleftarrow{Y}_u = \big(\overleftarrow{Y}_u + 2\, s_\theta(\overleftarrow{Y}_u, t-u)\big)du + \sqrt{2}\, d\overleftarrow{W}_u,$$

which differ only in drift, by $2\Delta s_{t-u}$, and share the diffusion coefficient $\sqrt{2}\, I_d$. By Girsanov's theorem the Kullback–Leibler divergence between the two path measures is

$$\frac{1}{2\cdot 2}\int_0^t \mathbb{E}\big\|2\,\Delta s_{t-u}\big\|^2 du \;=\; \int_0^t \mathbb{E}\big\|\Delta s_u\big\|^2 du,$$

where the expectation is under the exact process, whose marginal at time s is q_s by Theorem 1. Lemma 5 bounds the integrand by $M\varepsilon^2_{\text{score}}$ uniformly in s, giving Eq. (70). The data-processing inequality transfers the bound from path measures to the terminal marginals, and Pinsker's inequality (Tsybakov & Zaiats, 2009) yields the first bound in Eq. (71). The second follows because $K_t = F_t \circ R_t$ and $\widehat{K}_t = F_t \circ \widehat{R}_t$ share the exact forward kernel F_t, which is a contraction in total variation. □

Step 2: propagation through the iteration.

Theorem 5 (Finite-step bound under score error)**.** *Let $\lambda_t \in (0,1)$ be the contraction factor of Corollary 2 and let δ_t be as in Assumption 2 (ii). Then for any initial distribution p with $\chi^2(p\,\|\,q_0) < \infty$ and any $N \geq 1$,*

$$\Big\|\widehat{K}_t^{(N)}p - q_0\Big\|_{\text{TV}} \;\leq\; \underbrace{\lambda_t^N \sqrt{\chi^2(p\,\|\,q_0)}}_{\textit{mixing}} \;+\; \underbrace{\frac{\delta_t}{1-\lambda_t}}_{\textit{bias floor}}. \tag{72}$$

Proof. Write $\mathcal{E}_t = K_t - \widehat{K}_t$ and telescope:

$$\widehat{K}_t^{(N)}p - K_t^{(N)}p \;=\; -\sum_{k=0}^{N-1} K_t^{(N-1-k)}\, \mathcal{E}_t\, \widehat{K}_t^{(k)}p. \tag{73}$$

Each summand is the image under $K_t^{(N-1-k)}$ of the signed measure $\mathcal{E}_t\widehat{K}_t^{(k)}p$, which has zero total mass since both K_t and $\widehat{K}_t$ are Markov. Its density with respect to q_0 therefore lies in $L^2_0(q_0)$, on which K_t contracts with factor λ_t by Lemma 4. Using Assumption 2 (ii) to bound each $\|\mathcal{E}_t\widehat{K}_t^{(k)}p\|_{L^2(q_0)} \leq \delta_t$,

$$\Big\|\widehat{K}_t^{(N)}p - K_t^{(N)}p\Big\| \;\leq\; \sum_{k=0}^{N-1} \lambda_t^{N-1-k}\,\delta_t \;\leq\; \frac{\delta_t}{1-\lambda_t}.$$

Combining with Lemma 4 and the comparison inequality Eq. (56), which together give $K_t^{(N)}p - q_{0\,\text{TV}} \leq \lambda_t^N\sqrt{(\chi^2(p\|q_0))}$ for every $N \geq 1$ (valid for any $t \in (0, T^*]$ by Appendix C.8), and applying the triangle inequality gives Eq. (72). □

Corollary 3 (Bias of the perturbed invariant measure)**.** *Under the hypotheses of Theorem 5, $\widehat{K}_t$ admits an invariant distribution $\widehat{q}_0$ satisfying*

$$\big\|\widehat{q}_0 - q_0\big\|_{\text{TV}} \;\leq\; \frac{\delta_t}{1-\lambda_t}. \tag{74}$$

Proof. Let $N \to \infty$ in Eq. (72) with $p = \widehat{q}_0$, for which $\widehat{K}_t^{(N)}p = \widehat{q}_0$ for all N; the mixing term vanishes and the bias floor remains. □

C.10 Proof of Theorem 4

Proof. The proof uses the variational characterization of the spectral gap together with a conditional Jensen inequality along the forward Markov chain, and concludes with a strictness argument based on the non-degeneracy of the forward transition.

Step 1: Variational characterization. Since $\mathcal{K}_t$ is self-adjoint on $L^2(q_0)$ with 1 as a simple eigenvalue (Lemma 4), the second largest eigenvalue admits the variational representation

$$\lambda_t = \sup_{\substack{f \in L_0^2(q_0) \\ \|f\|_{L^2(q_0)}=1}} \langle f, \mathcal{K}_t f \rangle_{L^2(q_0)}, \tag{75}$$

where $L_0^2(q_0) \coloneqq \{f \in L^2(q_0) : \int f(x)\, q_0(x)\, dx = 0\}$ is the subspace orthogonal to the constant eigenfunction.

Step 2: Conditional expectation representation. Fix $f \in L_0^2(q_0)$ and consider the forward-reverse chain with $X_0 \sim q_0$: sample $X_t \sim F_t(X_0, \cdot)$ via the tilted forward SDE, then $X_0' \sim R_t(X_t, \cdot)$ via the reverse SDE. Since X_0' depends on X_0 only through X_t, the pair (X_0, X_0') is conditionally independent given X_t.

Define $f_t \coloneqq \mathbb{E}[f(X_0) \mid X_t]$. By the time-reversal identity of Theorem 1, when $X_0 \sim q_0$ the conditional distribution of X_0 given X_t under the forward kernel coincides with $R_t(X_t, \cdot)$. Therefore $\mathbb{E}[f(X_0') \mid X_t] = f_t$ as well, and by conditional independence:

$$\begin{aligned} \langle f, \mathcal{K}_t f \rangle &= \mathbb{E}\big[f(X_0)\, f(X_0')\big] \\ &= \mathbb{E}\big[\mathbb{E}[f(X_0) \mid X_t] \cdot \mathbb{E}[f(X_0') \mid X_t]\big] \\ &= \mathbb{E}\big[f_t^2\big]. \end{aligned} \tag{76}$$

Step 3: Jensen's inequality along the forward chain. For $0 < t \le t' \le T^*$, the forward process defines a Markov chain $X_0 \to X_t \to X_{t'}$. By the Markov property, $X_0 \perp X_{t'} \mid X_t$, so

$$f_{t'} = \mathbb{E}[f(X_0) \mid X_{t'}] = \mathbb{E}\big[\mathbb{E}[f(X_0) \mid X_t, X_{t'}] \,\big|\, X_{t'}\big] = \mathbb{E}[f_t \mid X_{t'}]. \tag{77}$$

By Jensen's inequality applied to the convex function $u \mapsto u^2$:

$$f_{t'}^2 = \big(\mathbb{E}[f_t \mid X_{t'}]\big)^2 \le \mathbb{E}[f_t^2 \mid X_{t'}]. \tag{78}$$

Taking expectations of both sides:

$$\mathbb{E}[f_{t'}^2] \le \mathbb{E}[f_t^2]. \tag{79}$$

Step 4: Non-degenerate inequality. Combining Eq. (76) applied at times t and t' with Eq. (79):

$$\langle f, \mathcal{K}_{t'} f \rangle = \mathbb{E}[f_{t'}^2] \le \mathbb{E}[f_t^2] = \langle f, \mathcal{K}_t f \rangle. \tag{80}$$

Since this holds for all $f \in L_0^2(q_0)$, the variational characterization Eq. (75) yields the weak bound

$$\lambda_{t'} \le \lambda_t. \tag{81}$$

Step 5: Strictness. It remains to show the inequality is strict when $t < t'$. Equality in Jensen's inequality Eq. (78) holds if and only if f_t is almost surely constant conditionally on $X_{t'}$; that is, $\mathrm{Var}(f_t \mid X_{t'}) = 0$ almost surely.

Because the tilted forward SDE is driven by a uniformly elliptic diffusion term $\sqrt{2}\, I_d$, standard parabolic regularity guarantees that the transition kernel $p(x_{t'} \mid x_t)$ is strictly positive everywhere. Consequently, the backward conditional distribution of X_t given $X_{t'}$ is fully supported on $\mathbb{R}^d$. Therefore, any non-constant function of X_t must have strictly positive conditional variance given $X_{t'}$.

We verify that for any non-zero eigenfunction $f \in L^2_0(q_0)$, the corresponding f_t is non-constant. If f_t were a constant c, then $c = \mathbb{E}[f_t] = \mathbb{E}[f(X_0)] = 0$, implying $f_t = 0$ almost surely. This would mean $\langle f, \mathcal{K}_t f\rangle = \mathbb{E}[f_t^2] = 0$, which contradicts the assumption that $\lambda_t > 0$ for the true eigenfunction.

Thus, $\mathrm{Var}(f_t \mid X_{t'}) > 0$, making the inequality Eq. (79) strictly less than:

$$\mathbb{E}[f_{t'}^2] < \mathbb{E}[f_t^2]. \tag{82}$$

Evaluating this at the specific eigenfunction f^* that achieves the supremum for $\lambda_{t'}$ in Eq. (75):

$$\lambda_{t'} = \langle f^*,\, \mathcal{K}_{t'} f^*\rangle_{L^2(q_0)} = \mathbb{E}\big[(f^*_{t'})^2\big] < \mathbb{E}\big[(f^*_t)^2\big] = \langle f^*,\, \mathcal{K}_t f^*\rangle_{L^2(q_0)} \le \lambda_t. \tag{83}$$

This establishes $\lambda_{t'} < \lambda_t$, completing the proof. □

D Metropolis-Hastings Correction

Proposition 1 (Metropolis-Hastings Correction)**.** *Let the exact and tilted target densities be $q_0(x) \propto p_0(x)\exp(\mathcal{L}_0(x))$ and $q_t(x) \propto p_t(x)\exp(\mathcal{L}_0(x))$. Define the forward and reverse transition proposals as*

$$q_{0t}(x, x') = \mathcal{N}(x';\, (1-t)x + 2t\nabla\mathcal{L}_0(x),\, 2t\mathbb{I}_d)\,, \tag{84}$$

$$q_{t0}(x', x) = \mathcal{N}(x;\, (1-t)x' - 2ts(x', t),\, 2t\mathbb{I}_d)\,. \tag{85}$$

Then the exact Metropolis-Hastings acceptance probabilities for the forward and reverse steps are, respectively,

$$\alpha_{0t}(x, x') = \min\left\{1,\, \frac{q_{t0}(x', x)}{q_{0t}(x, x')}\frac{q_t(x')}{q_0(x)}\right\}, \quad \alpha_{t0}(x', x) = \min\left\{1,\, \frac{q_{0t}(x, x')}{q_{t0}(x', x)}\frac{q_0(x)}{q_t(x')}\right\}. \tag{86}$$

While the log-proposal ratio can be computed exactly from the two Gaussian densities, the exact marginal ratio $q_t(x')/q_0(x)$ is analytically intractable. We therefore compute an empirical acceptance probability $\hat{\alpha}$ by approximating the log-density ratio via a first-order Taylor expansion in space and a Fokker-Planck based expansion in time (evaluated at small t):

$$\log q_t(x') - \log q_0(x) \approx s(x, 0)^\top (x' - x) + t\,\nabla_{x'} \cdot (x' + s(x', 0)) + \mathcal{L}_0(x) - \mathcal{L}_0(x'). \tag{87}$$

To avoid the prohibitive cost of computing the full Jacobian matrix $\mathbf{J}_f$ in high dimensions, the divergence of the vector field $f(x') \coloneqq x' + s(x', 0)$ is estimated stochastically via the Hutchinson trace estimator (Hutchinson, 1989):

$$\nabla_{x'} \cdot f(x') \approx \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\mathbb{I}_d)}\left[\varepsilon^\top \mathbf{J}_f(x')\,\varepsilon\right]. \tag{88}$$

Proof. The proof proceeds in three steps: we first compute the log proposal ratio, then approximate the log density ratio using small-t expansions, and finally combine the two to obtain the acceptance probability.

Step 1: Log proposal ratio. Since $q_{0t}(x, x')$ and $q_{t0}(x', x)$ are both Gaussian with the same covariance $2t\, I_d$, their log-ratio involves only the difference of squared means:

$$\begin{aligned}\log\frac{q_{t0}(x', x)}{q_{0t}(x, x')} = &-\frac{1}{4t}\big\|x - (1-t)x' + 2t\, s(x', t)\big\|^2 \\ &+ \frac{1}{4t}\big\|x' - (1-t)x - 2t\nabla\mathcal{L}_0(x)\big\|^2.\end{aligned} \tag{89}$$

This quantity is always tractable given the score network s_θ and the likelihood gradient $\nabla\mathcal{L}_0$.

Step 2: Approximating the log density ratio. The log density ratio $\log q_t(x') - \log q_0(x)$ is not directly available. We decompose it using $q_t \propto p_t\, e^{-\mathcal{L}}$:

$$\log q_t(x') - \log q_0(x) = \big[\log p_t(x') - \log p_0(x)\big] + \big[\mathcal{L}_0(x) - \mathcal{L}_0(x')\big], \tag{90}$$

where the second bracket uses the approximation $\mathcal{L}_t(x') \approx \mathcal{L}_0(x')$ for small t. It remains to approximate $\log p_t(x') - \log p_0(x)$, which we expand to first order in both space and time.

Spatial expansion. A first-order Taylor expansion of $\log p_0$ around x gives

$$\log p_0(x') \approx \log p_0(x) + s(x,0)^\top (x' - x), \tag{91}$$

where $s(x,0) = \nabla \log p_0(x)$ is the score at time 0.

Temporal expansion. Since the Metropolis–Hastings correction restores exact detailed balance regardless of the quality of the proposal, we use the zeroth-order approximation

$$\log p_t(x') \approx \log p_0(x'), \tag{92}$$

which holds up to an $O(t)$ error for small t.

Combining. Substituting Eq. (91) and Eq. (92) into Eq. (90):

$$\log q_t(x') - \log q_0(x) \approx s(x,0)^\top (x' - x) + t\,\nabla \cdot \big(x' + s(x',0)\big) + \mathcal{L}_0(x) - \mathcal{L}_0(x'). \tag{93}$$

Step 3: Acceptance probability. The Metropolis–Hastings acceptance probability for the forward step is defined as

$$\alpha_{0t}(x,x') = \min\left\{1,\ \frac{q_{t0}(x',x)}{q_{0t}(x,x')} \cdot \frac{q_t(x')}{q_0(x)}\right\}, \tag{94}$$

whose log-argument is the sum of the log proposal ratio Eq. (89) and the log density ratio Eq. (93). Both quantities are tractable: the proposal ratio is exact given s and $\nabla\mathcal{L}_0$, while the density ratio is approximated via Eq. (93). The acceptance probability for the reverse step follows by symmetry:

$$\alpha_{t0}(x',x) = \min\left\{1,\ \frac{q_{0t}(x,x')}{q_{t0}(x',x)} \cdot \frac{q_0(x)}{q_t(x')}\right\}. \tag{95}$$

Divergence estimation. The divergence $\nabla \cdot (x' + s(x',0)) = d + \nabla \cdot s(x',0)$ involves a trace of the Jacobian of s, which is expensive to compute exactly in high dimensions. Following (Hutchinson, 1989), we estimate it via the Hutchinson trace estimator:

$$\nabla \cdot f(x) = \mathrm{tr}(\nabla f(x)) = \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, I_d)}\left[\varepsilon^\top \nabla f(x)\, \varepsilon\right], \tag{96}$$

which requires only a single Jacobian-vector product $\nabla f(x)\,\varepsilon$, computable efficiently via automatic differentiation. □

E Algorithm

Here we detail the FRITES in Algorithm 1.

F Additional experiment details

F.1 Additional results

F.1.1 Numerical illustration of invariant measure under K_{T^*}

Figure 3 shows empirically that the true posterior q_0 is the invariant measure for K_{T^*}.

F.1.2 Gaussian prior with non-linear measurement

Since the analytical posterior of non-linear inverse problems are intractable, we illustrate that FRITES could achieve posterior sampling using a toy model, and leave more results on non-linear measurement to Section 4.2. The prior we use for the illustrative experiment is a $2d$ standard Gaussian, and the measurement is generated by $y = A(x * x) + n$, where $A \in \mathbb{R}^{2\times 2}$ and $*$ denotes entry-wise product. We choose the measurement noise level to be 0.1 and 1.0 respectively, and compute the ground truth using the grid method (Johnson et al., 2022). Results are shown in Figure 4.

Algorithm 1: FRITES: Forward-Reverse Iterations of Tilted SDEs

Input: Initial distribution p, time horizon $t \in (0, T^*]$, number of iterations N, score function $s_\theta(\cdot, \cdot)$, measurement log-likelihood $\mathcal{L}_0(\cdot)$

Output: Approximate posterior sample $x^{(N)} \sim q_0$

Sample $x^{(0)} \sim p$

for $n = 0, 1, \ldots, N-1$ **do**

  `// Forward step: diffuse current state to time` t

  if *linear measurement* **then**

    Sample $x_t^{(n)} \sim F_t(x^{(n)}, \cdot)$ by solving the forward SDE Eq. (9) on $[0, t]$

  else

    Sample $x_t^{(n)}$ via the Euler discretization:

$$x_t^{(n)} = (1-t)\, x^{(n)} + 2t\, \nabla \mathcal{L}_0(x^{(n)}) + \sqrt{2t}\, w, \quad w \sim \mathcal{N}(0, I_d)$$

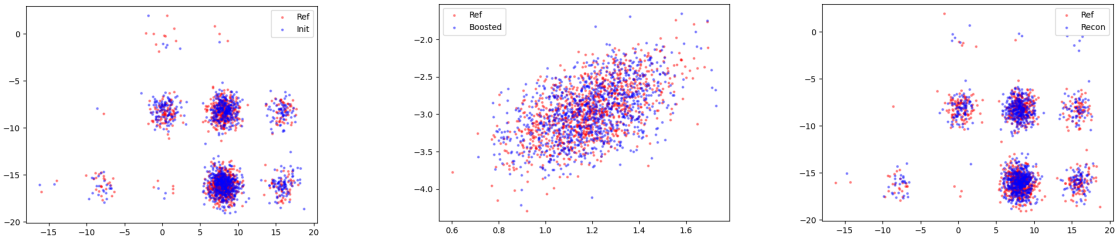
  end

  `// Reverse step: run reverse SDE from time` t `back to 0`

  Sample $\tilde{x}^{(n+1)} \sim R_t(x_t^{(n)}, \cdot)$ by solving the reverse SDE Eq. (3) on $[t, 0]$ using score $s_\theta(\cdot, \cdot)$

end

return $x^{(N)}$



Initialization Forward step Reverse step

Figure 3: Qualitative results illustrating K_{T^*} has invariant measure q_0. We plot the analytical reference in red. In the left figure we initialize with the analytical ground truth q_0 in blue. The blue dots in the middle figure is the boosted posterior after applying the forward kernel F_{T^*} to the initialization. The blue dots in the right figure is obtain after applying the reverse kernel R_{T^*} to the boosted posterior.

F.2 Implementation of FRITES in imaging

For imaging problems, the posterior distribution is intractable. Instead we use a number of metrics to evaluate the sampling quality. Hence we implement FRITES in a slightly-adjusted version.

Rather than applying FRITES (Algorithm 1) verbatim to imaging problems, we modify it to obtain sharper reconstructions. Consider the one-step update in Eq. (15), which decomposes into two parts: a gradient update $X_0' = X_0 - 2t\nabla\mathcal{L}_0(X_0)$ and a diffusion update $X_t = (1-t)X_0' + \sqrt{2t}\, w$. To enforce measurement consistency more strongly, we replace the single gradient update with M gradient-descent steps of step size ξ, in the spirit of Yan et al. (2025); M and ξ are treated as task-dependent hyperparameters. The practical implementation is summarized in Algorithm 2.

F.3 Implementation details

Hyperparameters for GMM experiments (Section 4.1). For the $d' = 0.9d$ setting, we use the following hyperparameters:

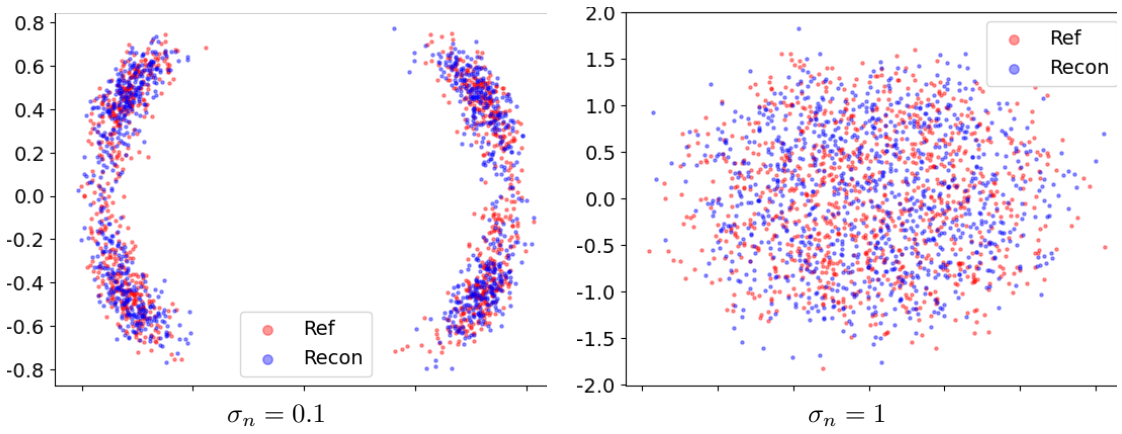


Figure 4: Qualitative results on the non-linear experiment with Gaussian prior. Reference (Ref) is computed by the grid method, and reconstruction (Recon) by FRITES.

Algorithm 2: FRITES for imaging

Input : Measurement y, forward operator A, data-fidelity loss $\mathcal{L}_0(\cdot) \coloneqq \mathcal{L}(A(\cdot), y)$
Parameters : Outer iterations N, gradient steps M and step size ξ, noise level t, DDIM steps J
Output : Reconstructed image $X^{(N)}$

Initialize $X^{(0)}$ with a 500-step DPS reconstruction Chung et al. (2023);
for $n = 0$ **to** $N - 1$ **do**
  $X_0 \leftarrow X^{(n)}$;
  `// Measurement consistency:` M `gradient-descent steps`
  for $m = 1$ **to** M **do**
    $X_0 \leftarrow X_0 - \xi \nabla \mathcal{L}_0(X_0)$;
  end
  `// Forward diffusion to noise level` t `(Eq. 15)`
  Sample noise $w \sim \mathcal{N}(0, I)$;
  $X_t \leftarrow (1 - t)\, X_0 + \sqrt{2t}\, w$;
  `// Reverse process:` J `DDIM steps (Eq. 16)`
  $X^{(n+1)} \leftarrow \mathrm{DDIM}(X_t;\, J)$;
end
return $X^{(N)}$;

- **Unadjusted Langevin (ULA):** 500 Langevin steps.
- **Tilted transport:** 500 Langevin steps.
- **DiffPnP:** $\eta_{\max} = 1.0$, $\eta_{\min} = 0.01$, $k = 35$, $k_{\text{start}} = 5$, MALA scale $= 0.5$, and 1000 Langevin steps.
- **FRITES:** $N = 300$.

For the $d' = 1$ setting, the hyperparameters are configured as follows:

- **NUTS:** 100 NUTS steps.
- **Tilted transport:** 100 NUTS steps.
- **Annealed Langevin (ALD):** $\sigma_{\min} = 0.01$, 200 noise levels with 50 steps per level.
- **DiffPnP:** $\eta_{\max} = 1.0$, $\eta_{\min} = 0.01$, $k = 35$, $k_{\text{start}} = 5$, MALA scale of 0.5, and 1000 Langevin steps.
- **FRITES:** $N = 100$.

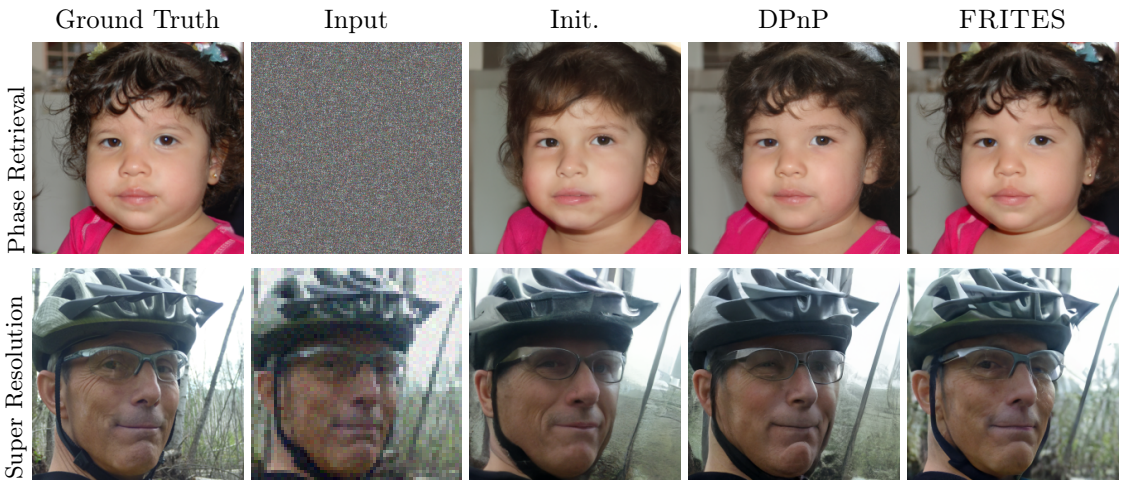


Figure 5: Qualitative imaging results on two inverse problems. The top row shows *Phase Retrieval* and the bottom row *Super Resolution*. From left to right: ground-truth label, measurement input, initialization, DPnP reconstruction, and our proposed FRITES method.

Hyperparameters for imaging experiments (Section 4.2). For the phase retrieval task, we utilize the following configurations for each respective model:

- **DPS:** guidance scale = 0.8, run for 500 steps to produce the shared initialization.
- **DPnP:** $\eta_{\max} = 0.4$, $\eta_{\min} = 0.016$, $k = 12$, $k_{\text{start}} = 4$, MALA scale = 0.8, and 1000 Langevin steps.
- **FRITES:** $M = 5$ gradient steps, learning rate $\xi = 0.05$, $t = 0.01$, $N = 100$ outer iterations, and $J = 20$ DDIM steps.

For the super-resolution task, the hyperparameters for each baseline and our proposed method are detailed below:

- **DPS:** guidance scale = 0.8, run for 500 steps to produce the shared initialization.
- **DPnP:** $\eta_{\max} = 0.4$, $\eta_{\min} = 0.016$, $k = 12$, $k_{\text{start}} = 4$, MALA scale = 0.6, and 1000 Langevin steps.
- **FRITES:** $M = 10$ gradient steps, learning rate $\xi = 0.05$, $t = 0.01$, $N = 100$ outer iterations, and $J = 20$ DDIM steps.

F.4 Numerical Results on Imaging

Figure 5 shows the recovered images for phase retrieval and super resolution. Quantitative comparison is available in Table 3.