Posterior Sampling with Denoising Oracles via Tilted Transport

Joan Bruna[1,2,3] and Jiequn Han[3]

[1]Courant Institute of Mathematical Sciences, New York University
[2]Center for Data Science, New York University
[3]Flatiron Institute

July 2, 2024

Abstract

Score-based diffusion models have significantly advanced high-dimensional data generation across various domains, by learning a denoising oracle (or score) from datasets. From a Bayesian perspective, they offer a realistic modeling of data priors and facilitate solving inverse problems through posterior sampling. Although many heuristic methods have been developed recently for this purpose, they lack the quantitative guarantees needed in many scientific applications.

In this work, we introduce the *tilted transport* technique, which leverages the quadratic structure of the log-likelihood in linear inverse problems in combination with the prior denoising oracle to transform the original posterior sampling problem into a new 'boosted' posterior that is provably easier to sample from. We quantify the conditions under which this boosted posterior is strongly log-concave, highlighting the dependencies on the condition number of the measurement matrix and the signal-to-noise ratio. The resulting posterior sampling scheme is shown to reach the computational threshold predicted for sampling Ising models [Kun23] with a direct analysis, and is further validated on high-dimensional Gaussian mixture models and scalar field φ^4 models.

Contents

1 Introduction

Inverse problems consist in reconstructing a signal of interest from noisy measurements. As such, they are a central object of study across many scientific domains, including signal processing, imaging, astrophysics or computational biology. In the common settings where the measurement information is limited, a reliable solution for these problems usually depends on prior knowledge of the data. One popular approach is to choose a regularizer that utilizes data properties such as smoothness or sparseness, and then solve a regularized optimization problem to obtain *a point estimate* of the original data. However, this approach often struggles with selecting an appropriate regularizer and might be unstable in the presence of large measurement noise. A more robust approach takes a statistical formulation and seeks to sample the *posterior distribution* of data based on Bayes's theorem, which allows for uncertainty quantification in the reconstructed data by leveraging a model for the prior data distribution.

While accurate models for high-dimensional distributions are notoriously complex to estimate, the resurgence of deep neural networks has provided unprecedented capabilities for modeling complex data distributions in certain high-dimensional regimes. Specifically, score-based diffusion models [SDWMG15, HJA20, SSDK+20] have achieved remarkable empirical success in generating high-dimensional data across various domains, including images, video, text, and audio. These models implicitly parameterize data distributions through an iterative denoising process that builds up data from noise. Furthermore, there is a growing literature developing theoretical foundations of score-based diffusion models [CCL+23, BDBDD23, LLT23, CLL23, CKS24], giving a comprehensive error analysis including score estimation, initialization error and time-discretization error. By generating high-fidelity data, these models can also serve as data prior for posterior sampling in inverse problems in high dimensions. Following this idea, many studies (see, e.g.,

[KVE21, CKM+23]) have leveraged diffusion models for posterior sampling. However, as discussed below, various categories of approaches for posterior sampling introduce different uncontrollable errors, such as those arising from the approximation of the conditional score or the use of a limited variational family. This abundance of heuristics contrasts with the principled sampling used in prior data generation, and is often at odds with the statistical guarantees needed in many scientific applications.

In this work, we aim to bridge the gap between principled diffusion-based algorithms for both prior and posterior distributions. Focusing on the canonical setting of linear inverse problems, where measurements are of the form $y = Ax + w$, with $x \sim \pi$ the signal to be estimated and w an independent noise, we first illustrate a negative result, revealing that no method can efficiently sample from the posterior distribution in general cases, even with the prior denoising oracle. We next develop the *tilted transport* technique, which utilizes the quadratic structure of the log-likelihood in linear inverse problems in combination with the prior denoising oracle to exactly transform the original posterior sampling problem into a new one that is easier to sample. Figure 1 illustrates a schematic plot of the method using two-dimensional Gaussian mixture examples, showing that while the original target posterior problem remains multimodal, the boosted posterior resembles a unimodal distribution.

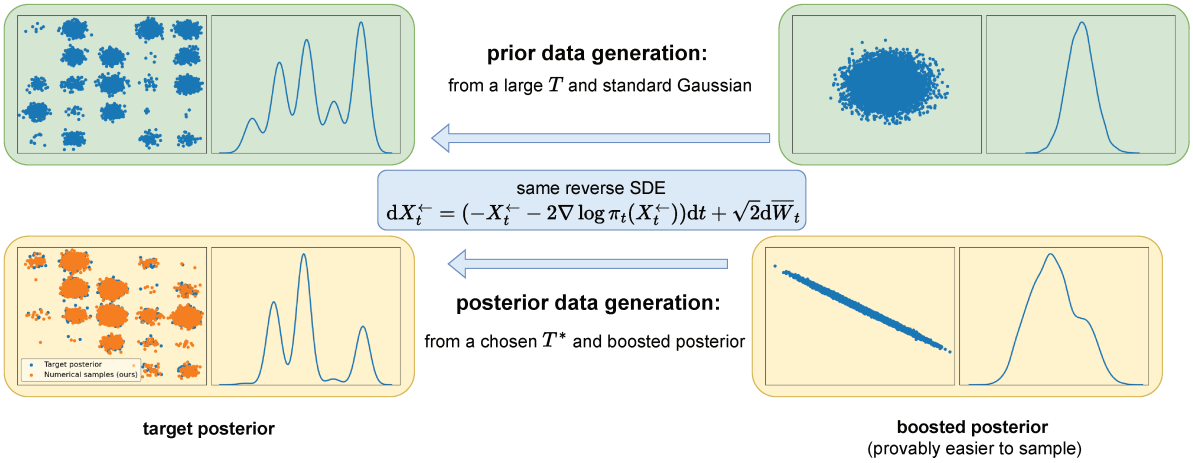

Figure 1: Schematic plot of tilted transport boosting posterior sampling with a 2D Gaussian mixture example. The density plot shows the first variable's density, and the scatter plot displays the samples.

We establish a sufficient condition where the density of the transformed posterior problem becomes strongly log-concave, making it suitable for efficient sampling via Langevin dynamics thanks to the Bakry-Emery criterion [BÉ85]. This condition showcases the interplay between a geometric property of the prior (the so-called *susceptibility* $\chi_t(\pi)$; see Section 5) and the conditioning and noise level of the measurements. Interestingly, the condition can be satisfied when the signal-to-noise ratio (SNR) is either moderately low *or* moderately high, in contrast with traditional sampling methods, which typically excel only within a specific regime.

As a first application, we show that tilted transport can sample from Ising models of the form $\nu(x) \propto e^{-\frac{1}{2}x^\top Q x}$, where $x \in \{\pm 1\}^d$ is supported in the hypercube, up to the critical threshold determined by the gap $\lambda_{\max}(Q) - \lambda_{\min}(Q) = 1$, thus matching the performance of Glauber dynamics [EAMS22, AJK+21] as well as the computational threshold predicted by the low-degree method [Kun23]. The sampling guarantees are directly based on the Bakry-Emery criterion applied to the tilted posterior, and avoid the need to perform multiscale extensions of this criterion based on localization schemes [CE22] or the Polchinksy renormalization group [BBD23].

More generally, even when the boosted posterior is not strongly log-concave, it is more amenable to sample

than the original one. Thus, tilted transport can be combined with any existing black-box posterior sampling methods to enhance their performance. This technique operates without any additional computational cost and functions in a plug-and-play fashion, allowing for straightforward integration into various frameworks. When working with high-dimensional Gaussian mixtures, where an analytical solution to the posterior is available, we numerically validate our theory and demonstrate enhanced posterior sampling performance. In the two-dimensional lattice scalar φ^4 field model, we also observe that the Langevin dynamics exhibit accelerated relaxation times with the boosted distribution compared to the original distribution.

1.1 Related Work

Numerous studies in recent years have explored score-based priors for posterior sampling. We note that several recent works [SSXE22, CKJ+21, GMJS22, SLK22, CSY22] introduce hyperparameters to balance the influence of the prior and measurements, resulting in sampling strategies that guide output to regions where the given observation is more likely. These strategies typically deviate from the principles of Bayesian posterior sampling and often lack a precise definition of the resulting distribution. In contrast, other approaches adhere more closely to Bayesian principles. One such approach is variational inference, which involves designing variational objectives and optimization methods based on the structure of score-based diffusion [KEES22, MSKV23, FSR+23, JDMO24]. However, even with an accurate prior score, the accuracy of posterior sampling heavily depends on the choice of variational family and optimization procedures, not to mention the additional optimization cost. Another popular strategy focuses on approximating the score conditional on the measurement using various heuristics [SSDK+20, KVE21, JAD+21, CKM+23, MK22, SVMK22, SZY+23]. In this approach, approximation errors typically remain largely uncontrollable due to the challenges associated with tracking the conditional distribution for intermediate states. Recently, some studies have adopted sequential Monte-Carlo methods to systematically approximate the conditional score [WTN+23, CJCM24, DS24], providing consistency as the number of particles used to approximate the conditional distribution of the intermediate states increases. However, this particle-based method still struggles with high-dimensional problems due to the curse of dimensionality [BLB08]. We note that [CJCM24, HDMOB24] also intuitively explores the possibility of reducing the original posterior to an equivalent one under restrictive conditions in the discrete time setting or purely denoising setting. In contrast, our tilted transport technique operates in a fairly generic setting and is supported by a clear theoretical foundation.

Notations:. $\mathcal{P}(\mathbb{R}^d)$ denotes the space of probability measures over $\mathbb{R}^d$, and $\mathcal{P}_2(\mathbb{R}^d)$ denotes those measures with finite second-order moments. γ_d denotes the d-dimensional standard Gaussian measure, and by slight abuse of notation, γ_δ or γ_Σ denote the centered Gaussian measure with covariance δI_d or Σ when the context is clear. For $Q \geq 0$ in $\mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$ in the span of Q, the *quadratic tilt* of π is the measure $\mathsf{T}_{Q,b}\pi \ll \pi$ with density proportional to $\frac{d\mathsf{T}_{Q,b}\pi}{d\pi}(x) \propto \exp\left\{-\frac{1}{2}x^\top Qx + x^\top b\right\}$. We also use the notation T_Q when $b = 0$. $\|Q\|$ denotes its operator norm. $\pi * \gamma$ denotes the convolution of two measures π and γ. For $\alpha \geq 0$ and $\pi \in \mathcal{P}(\mathbb{R}^d)$, we define $\mathsf{D}_\alpha\pi(x) := \alpha^d\pi(\alpha x)$ as the dilation of π. For $\beta \geq 0$ and $\pi \in \mathcal{P}(\mathbb{R}^d)$, we define $\mathsf{C}_\beta\pi(x) := \pi * (\mathsf{D}_{\beta^{-1/2}}\gamma_d)$ as the Gaussian convolution of π.

Acknowledgements:. We thank Stéphane Mallat and Maarten de Hoop for fruitful discussions during the earlier phase of this project. We thank Ahmed El Alaoui for pointing out the relevant work [Kun23] and for engaging and stimulating discussions. We thank Eric Vanden-Eijnden, Jonathan Niles-Weed, Andrea Montanari, Nick Boffi, Jianfeng Lu, Florentin Guth, Mark Goldstein, Brian Trippe, Benoit Dagallier and Roland Bauerschmidt for useful feedback during the completion of this work. JB was partially supported by the Alfred P. Sloan Foundation and awards NSF RI-1816753, NSF CAREER CIF 1845360, NSF CHS-1901091

and NSF DMS-MoDL 2134216.

2 Preliminaries

Problem Setup. Consider a high-dimensional object of interest $x \in \mathbb{R}^d$, drawn from a certain probability distribution $\pi \in \mathcal{P}(\mathbb{R}^d)$. We suppose that one has managed to learn a generative model for π via the DDPM objective [HJA20]; in other words, for any $y \in \mathbb{R}^d$ and $\sigma \geq 0$, we have access to the *denoising oracle* $\mathsf{DO}_\pi(y, \sigma) := \mathbb{E}[x|y]$, where $y = x + \sigma w$, with $x \sim \pi$ and $w \sim \gamma_d$ independent. It is by now well-established that such denoising oracle enables efficient sampling of π, well beyond the classic isoperimetric assumptions for fast relaxation of Langevin dynamics [CCL+23].

Suppose that we now measure $y = Ax + \sigma w$, where again $x \sim \pi$ and $w \sim \gamma_d$ are independent, but now $A \in \mathbb{R}^{d' \times d}$ is a *known* linear operator different from the identity. Given these linear measurements, we are now interested in the *posterior sampling* of x given y. This corresponds to the basic setup of linear inverse problems, encompassing many applications such as image inpainting, super-resolution, tomography, or source separation, to name a few. We are interested in the following natural question: can the power of denoising oracles be provably transferred to posterior sampling?

By Bayes' rule, the posterior distribution $\nu_{y,A}$ (denoted simply by ν when the context is clear) has density proportional to $\pi(x)p(y|x) \propto \exp\left\{-\frac{1}{2\sigma^2}\|Ax - y\|^2\right\}\pi(x)$, and thus we can write it as a quadratic tilt of π:

$$\nu = \mathsf{T}_{Q,b}\pi\,, \text{ with } Q = \sigma^{-2}A^\top A\,,\ b = -\sigma^{-2}A^\top y\,.$$

We readily identify certain regimes where sampling from ν might be easy:

- If $\lambda_{\min}(Q)$ is sufficiently large, $\lambda_{\min}(Q) \gg 1$, then one expects ν to be strongly log-concave, enabling fast relaxation of Langevin dynamics.
- If $\lambda_{\max}(Q)$ is sufficiently small, $\lambda_{\max}(Q) \ll 1$, then one expects $\nu \approx \pi$ in the appropriate sense, and therefore that samples from π (which can be produced efficiently thanks to DO_π) may be perturbed into samples from ν.
- If $A \in \mathcal{O}_d$ is a unitary transformation, then $Q = \mathrm{Id}$ and the inverse problem reduces to isotropic Gaussian denoising, and is thus at first glance 'compatible' with the structure of the denoising oracle (such observation will be formalized later).

At this stage, we can already identify two key parameters of the problem that are likely to drive the difficulty of posterior sampling: on one hand, a proxy for the signal-to-noise ratio, measured e.g., by $\mathrm{SNR} := \lambda_{\min}(Q) = \frac{\lambda_{\min}(A)^2}{\sigma^2}$. On the other hand, the conditioning of the measurement operator A, $\kappa(A) := \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$. As we shall see, these two characteristics of the linear measurement system will characterize necessary and sufficient conditions for probable posterior sampling. In the following, we assume the log of prior density π is smooth and its Hessian exists $\forall x \in \mathbb{R}^d$.

Denoising Oracles and Score-Based Diffusion. Let us first review the natural connection between denoising and score-based generative modeling. Score-based diffusion models consist of two processes: a forward process that gradually adds noise to input data and a reverse process that learns to generate data by iteratively removing this noise. For example, one widely used family for the forward process is the Ornstein–Uhlenbeck

(OU) process[1]:

$$\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \,, \tag{1}$$

where W_t is the standard Wiener process. We use π_t to denote the density of X_t, given by the action of the OU semigroup $\pi_t = \mathsf{O}_t^* \pi$, defined by $\mathsf{O}_t f(x) = \mathbb{E}[f(X_t)|X_0 = x]$, and explicitly given by dilated Gaussian convolutions, $\mathsf{O}_t^* := \mathsf{C}_{\beta_t}\mathsf{D}_{\alpha_t}$, with $\beta_t = 1 - e^{-2t}$ and $\alpha_t = e^t$. With a sufficiently large T, we know that π_T is close to the density of standard Gaussian γ_d, owing to the exponential contraction of the OU semigroup: $\mathrm{KL}(\pi_T||\gamma_d) \leq e^{-T}\mathrm{KL}(\pi||\gamma_d)$.

Finally, the measure π_t solves the Fokker-Plank equation

$$\partial_t \pi_t = \nabla \cdot (x\pi_t) + \Delta \pi_t \,, \ \pi_0 = \pi \,. \tag{2}$$

By writing (2) as a transport equation $\partial_t \pi_t = \nabla \cdot ((x + \nabla \log \pi_t)\pi_t)$, we can formally reverse the transport starting at a large time T and solving

$$\partial_t \tilde{\pi}_t = \nabla \cdot (-(x + \nabla \log \pi_{T-t})\tilde{\pi}_t) \,, \ \tilde{\pi}_0 = \pi_T \,. \tag{3}$$

Since $\tilde{\pi}_t = \pi_{T-t}$ for $0 \leq t \leq T$, introducing again the dissipative term leads to $\partial_t \tilde{\pi}_t = \nabla \cdot (-(x + 2\nabla \log \tilde{\pi}_t)\tilde{\pi}_t) + \Delta \tilde{\pi}_t \,, \ \tilde{\pi}_0 = \pi_T$, which admits the SDE representation

$$\mathrm{d}\tilde{X}_t = (\tilde{X}_t + 2\nabla \log \pi_{T-t}(\tilde{X}_t))\mathrm{d}t + \sqrt{2}\mathrm{d}\overline{W}_t, \quad \tilde{X}_0 \sim \pi_T \,. \tag{4}$$

In practice, one runs this reverse diffusion starting from $\tilde{X}_0 \sim \gamma_d$ rather than $\tilde{X}_0 \sim \pi_T$. However, by the data-processing inequality, we have that $\mathrm{KL}(\pi||\tilde{\pi}_T) \leq \mathrm{KL}(\pi_T||\gamma_d) = O(e^{-T})$, thus incurring in insignificant error. To facilitate later exposition, we write the above process reverse in time [And82, HP86]

$$\mathrm{d}X_t^{\leftarrow} = (-X_t^{\leftarrow} - 2\nabla \log \pi_t(X_t^{\leftarrow}))\mathrm{d}t + \sqrt{2}\mathrm{d}\overline{W}_t, \quad X_T^{\leftarrow} \sim \gamma_d, \tag{5}$$

and interpret the data generation process as running the reverse SDE from T back to 0.

By the well-known Tweedie's formula, and up to time reparametrisation, the denoising oracle is equivalent to the time-dependent score $\nabla \log \pi_t$:

Fact 1 (Tweedie's formula, [Her56])**.** *We have* $\nabla \log \pi_t(x) = -(1 - e^{-2t})^{-1}(x - e^{-2t}\mathsf{DO}_\pi(x, 1 - e^{2t}))$.

The OU semigroup corresponds to the so-called 'variance-preserving' diffusion scheme. Instead one could also consider the 'variance-exploding' scheme, given directly by the Heat semigroup $\mathsf{H}_t^* \pi = \pi * \gamma_{2t}$. One can immediately verify that denoising oracles are equally valid to reverse the associated Fokker-Plank transport equation, given by $\partial_t \pi_t = \Delta \pi_t$. Indeed, the reverse SDE now reads

$$\mathrm{d}X_t^{\leftarrow} = (-2\nabla \log \pi_t(X_t^{\leftarrow}))\mathrm{d}t + \sqrt{2}\mathrm{d}\overline{W}_t, \quad X_T^{\leftarrow} \sim \gamma_d \,. \tag{6}$$

Log-Sobolev Inequality and Fast Relaxation of Langevin Dynamics. Given a Gibbs distribution $\pi \in \mathcal{P}(\mathbb{R}^d)$ of the form $\pi \propto e^{-f}$, a powerful and versatile method to sample from π is to consider the Langevin dynamics

$$\mathrm{d}X_t = -\nabla f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t \,, \ X_0 \sim \mu_0 \,, \tag{7}$$

[1] In practice, it is also common to introduce a positive smooth function β: $\mathbb{R}_+ \to \mathbb{R}_+$ and consider the time-rescaled OU process $\mathrm{d}X_t = -\beta(t)X_t\mathrm{d}t + \sqrt{2\beta(t)}\mathrm{d}W_t$. Our results could be applied directly to these variants by rescaling time. For the ease of notation, we keep $\beta(t) \equiv 1$ in the main text.

where μ_0 is an arbitrary initial distribution. It is easy to verify that these dynamics define a Markov process that admits π as its unique invariant measure. Perhaps less obvious is the fact that the Fokker-Plank equation associated with eq. (7), given by $\partial_t \mu = \nabla \cdot (\nabla f \mu) + \Delta \mu$ (and where μ_t is the law of X_t) is in fact a Wasserstein gradient flow for the relative entropy functional $\mathrm{KL}(\mu||\pi)$ [JKO98]. Under this interpretation, one can quantify the convergence of Langevin dynamics to their invariant measure, i.e., its time to relaxation, by establishing a sharpness or *Polyak-Lowacjevitz* (PL)-type inequality. Indeed, by noticing that $\frac{d}{dt}\mathrm{KL}(\mu||\pi) = -\mathrm{I}(\mu||\pi)$, where $\mathrm{I}(\mu||\pi) = \mathbb{E}_\mu[\|\nabla \log \mu - \nabla \log \pi\|^2]$ is the Fisher divergence, the PL-type inequality in this setting is given by the *Logarithmic Sobolev Inequality* (LSI) [Gro75, Led01]: we say that a measure π satisfies $\mathrm{LSI}(\rho)$ if for any $\mu \in \mathcal{P}(\mathbb{R}^d)$ it holds

$$\mathrm{KL}(\mu||\pi) \leq \frac{1}{2\rho}\mathrm{I}(\mu||\pi) \ . \tag{8}$$

This functional inequality [2] directly implies $\mathrm{KL}(\mu_t||\pi) \leq e^{-2\rho t}\mathrm{KL}(\mu_0||\pi)$. While for general π it is typically hard to establish, there are two important sources of structure that lead to quantitative (i.e., $\rho = \Omega_d(1)$) bounds: when π is a product measure $\pi = \tilde{\pi}^{\otimes d}$ (in which case π satisfies LSI with the same constant as $\tilde{\pi}$), and when π is strongly log-concave[3], i.e., $-\nabla^2 \log \pi(x) \succeq \alpha I$ for all x, in which case the celebrated Bakry-Emery criterion [BÉ85] states that $\rho \geq \alpha$.

3 Evidence of Computational Hardness in the Generic Case

We start our analysis of posterior sampling by discussing negative results for the general case. Recently, [GJP+24] established computational lower bounds for this task using cryptographic hardness assumptions. In this section, we complement these results by illustrating a correspondence with sampling problems on Ising models, leading to an arguably simpler conclusion.

For this purpose, consider $\bar{\pi} = \mathrm{Unif}(\{\pm 1\}^d)$ the uniform measure of the hypercube. Quadratic tilts of $\bar{\pi}$ define generic Ising models, a rich and intricate class of high-dimensional distributions. Since $\bar{\pi}$ is a product measure, its associated denoising oracle becomes a separable function that can be computed in closed-form:

Fact 2 (Denoising Oracle for $\bar{\pi}$)**.** *Let* $\gamma(t;\mu,\sigma) = \exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\}$. *Then we have*

$$\mathsf{DO}_{\bar{\pi}}(y,\sigma) = (\phi(y_i;\sigma))_{i=1\ldots d} \ , \ \textit{with} \ \ \phi(t,\sigma) = \frac{\gamma(t,+1,\sigma) - \gamma(t,-1,\sigma)}{\gamma(t,+1,\sigma) + \gamma(t,-1,\sigma)} \ . \tag{9}$$

Given a symmetric matrix $Q \in \mathbb{R}^{d\times d}$, an Ising model is given by the tilt $\mathsf{T}_Q\bar{\pi} \in \mathcal{P}(\{\pm 1\}^d)$. In our setting, we can thus view such models as the posterior distribution of a linear inverse problem associated with the uniform prior $\bar{\pi}$. Efficiently sampling from Ising models is a fundamental question at the interface of statistical physics and high-dimensional probability, and several works provide evidence of computational hardness under a variety of settings.

Notably, by treating Q as the adjacency matrix of a regular graph, [GŠV16] establishes that sampling from ν is impossible for $\lambda_{\max}(Q) - \lambda_{\min}(Q) \geq 2 + \varepsilon$, for any $\varepsilon > 0$, unless $\mathsf{NP} = \mathsf{RP}$. In other words, for poorly conditioned tilt Q (in the sense that there is a large gap between the smallest and largest eigenvalue), there is no efficient posterior sampling algorithm, *even with the knowledge of the prior denoising oracle*. The threshold can even be reduced to $1 + \varepsilon$ by using a weaker notion of computational hardness [Kun23], given by the *low-degree polynomial method* [BHK+19, KWB19]. Remarkably, this threshold agrees with the current best-known algorithmic results for sampling generic Ising models with Glauber dynamics

[2] Note that we use the convention from [MV00] where the relevant direction is to lower bound ρ, as opposed to the other common direction $\mathrm{KL}(\mu||\pi) \leq \frac{\tilde{\rho}}{2}\mathrm{I}(\mu||\pi)$, e.g. [BGL14, Definition 5.1.7].

[3] or a suitable perturbation of it via the Hooley-Strook perturbation principle [HS87]

[EKZ22, AJK+21, BB19]. Finally, we also mention that when Q is a random Gaussian symmetric matrix, the associated model is the so-called Sherrington-Kirkpatrick (SK), which has been analyzed in [EAMS22, Cel24]. In this setting, these works establish that 'stable' sampling algorithms [4] fail to sample from the SK model as soon as $\lambda_{\max}(Q) - \lambda_{\min}(Q) > 4$, and that this threshold can be reached with dedicated sampling algorithms based on AMP.

In summary, we have:

Fact 3 (Computational Hardness of Sampling Ising Models, [Kun23, GŠV16])**.** *There exist no general-purpose, efficient posterior sampling algorithms, for Q sufficiently ill-conditioned, even under the knowledge of the prior denoising oracle.*

One could wonder whether this computational hardness comes from the discrete nature of the hypercube. It is not hard to observe that this is not the case: the following proposition, proved in Appendix A, shows a simple reduction from a model where the prior $\bar{\pi}$ is replaced by a smooth mixture of Gaussians π centered at the corners of the hypercube, with variance δ.

Proposition 4 (Hardness extends to smooth priors)**.** *Assume a posterior sampler exists for the smooth prior with TV error ϵ and $\delta = o(d^{-1/2})$. Then there exists a sampler for the associated Ising model with TV error 1.1ϵ.*

In conclusion, one cannot hope for a generic method that leverages the prior denoising oracle to perform efficient posterior sampling, as soon as A is mildly ill-conditioned. Thus, in order to perform provable posterior sampling, one needs to either (i) constraint the measurements, or (ii) exploit structural properties of the prior measure. In the following, we focus on (i), namely providing guarantees for well-conditioned A that leverage the OU semigroup for generic prior distributions.

4 Posterior Sampling via Tilted Transport

We now present a simple method that reduces the original posterior sampling problem to another posterior sampling problem with more benign geometry, by leveraging the shared quadratic structure of the posterior tilt and the OU semigroup. The power of the denoising oracle to perform sampling of the prior π comes from its ability to run the transport equation (3) in either direction, and leveraging the fact that sampling from π_T is easy. To transfer this power to posterior sampling, we can thus attempt to replicate this scheme: can we implement a transport between the posterior ν and a terminal measure ν_T that is easy to sample, that only relies on the pre-trained prior DO_π?

A Motivating Example. Consider first the denoising setting: $y = x + \sigma w$. According to the forward OU process, we have $p(X_s|X_0) \overset{d}{=} \mathcal{N}(e^{-s}X_0, (1-e^{-2s})I_d)$. Introduce $T > 0$ and define $\tilde{y} = e^{-T}y = e^{-T}x + e^{-T}\sigma w$ such that $p(\tilde{y}|x) \overset{d}{=} \mathcal{N}(e^{-T}x, e^{-2T}\sigma^2 I_d)$. We match the variance by letting $e^{-2T}\sigma^2 = 1 - e^{-2T}$, i.e., $T = \frac{1}{2}\log(1+\sigma^2)$, so $p(\tilde{y}|x) = p(X_s|X_0)$, and thus $(x, \tilde{y}) \overset{d}{=} (X_0, X_T)$. Therefore, to perform the posterior sampling $p(x|\tilde{y})$, we only need to do the sampling $p(X_0|X_T)$, which can be achieved through the reverse SDE. Specifically, let $X_T = e^{-T}y$ and run the reverse SDE (5) from T to 0, then X_0 will be the desired posterior.

Hamilton-Jacobi Equation and Quadratic Tilts. If π_t solves the Fokker-Plank eq. (2), then one can verify that the time-varying potentials $f_t := \log \pi_t$ solve the viscous Hamilton-Jacobi PDE (HJE)

$$\partial_t f_t = \Delta f_t + \|\nabla f_t\|^2 + x \cdot \nabla f_t \ , \ f_0 = f \ . \tag{10}$$

[4] Defined as sampling algorithms whose output law depends smoothly on Q in the Wasserstein metric

In the heat semigroup setting, one obtains a closely related HJE without the last linear term. Now, the posterior $\nu = \mathsf{T}_{Q,b}\pi$ creates an additional quadratic term in the potential $\log \nu = f - \frac{1}{2}x^\top Qx + x \cdot b$. One could naively hope that this additive quadratic term would still define a solution of the HJE with the tilted initial condition $\tilde{f}_0 = \log \nu$ — or equivalently that the measure $\mathsf{T}_{Q,b}\pi_t$ solves the transport equation (3). Unfortunately, due to the nonlinearity in (10) brought by the terms $\|\nabla f_t\|^2$, this is not the case. However, as we shall see now, this is not far from being true: one just needs to consider *time-varying* quadratic tilts in order to satisfy the HJE.

Tilt Transport Equation. We consider then a one-parameter family of distributions ν_t of the form

$$\nu_t := \mathsf{T}_{Q_t,b_t}\pi_t \ , \quad \text{with} \ \ Q_0 = Q \ , \ b_0 = b \ . \tag{11}$$

As it turns out, one can ensure that $\log \nu_t$ solves the HJE associated with the reverse OU process by asking that Q_t, b_t satisfy the first-order ODE:

$$\begin{cases} \dot{Q}_t = 2(I + Q_t)Q_t \ , & Q_0 = Q \\ \dot{b}_t = (I + 2Q_t)b_t \ , & b_0 = b \end{cases} \tag{12}$$

Theorem 5 (Tilted Transport under OU Semigroup)**.** *Assume $t < T$ such that the ODE (12) is well-defined on $[0, t]$. By initializing $X_t \sim \nu_t$ and run the reverse SDE (5) from t to 0, we have $X_s \sim \nu_s$ for $s \in [0, t]$, specifically, X_0 gives the desired posterior.*

Solution to eq. (12). Without loss of generality, we assume $d' \le d$, and the observation operator $A \in \mathbb{R}^{d' \times d}$ has a general singular value decomposition form $A = U\Sigma V^\top$ with non-zero singular values $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{d'} > 0$. By diagonalizing Q and solving the scalar ODE $\dot{q}_t = 2(1 + q_t)q_t$ for diagonal entries, we have $Q_t = V\mathrm{diag}\left(\frac{e^{2t}}{1+\sigma^2/\lambda_1^2 - e^{2t}}, \cdots, \frac{e^{2t}}{1+\sigma^2/\lambda_{d'}^2 - e^{2t}}, 0, \cdots, 0\right) V^\top$, where the solution is defined up to the blowup time $T := \frac{1}{2}\log(1 + \sigma^2/\lambda_1^2) = \frac{1}{2}\log(1 + \lambda_{\max}(Q)^{-1})$. b_t can be further solved from the solution Q_t; see Appendix B.2 for more details.

With the explicit solution of Q_t, b_t, we can interpret the term $\exp(-\frac{1}{2}x^\top Q_t x + x^\top b_t)$ as the likelihood of the inverse problem with respect to the new prior distribution π_t and the corresponding operator. Based on this observation and Theorem 5, we have the following corollary, transforming the original posterior sampling problem to a new posterior sampling problem exactly. We remark that when A is identity, the corollary recovers the analysis we have in the motivating example; see Appendix B.2 for the proof and more discussions.

Corollary 6 (Posterior Sampling via Tilted Transport)**.** *Fix $t \le T$. Sampling from the original posterior $\nu = \mathsf{T}_Q\pi$ is equivalent to a two-step process: first, sample from a new posterior $X_t \sim \nu_t$, and then run the reverse SDE (5) from time t to 0.*

Remark 7 (Tilted Transport under Heat Semigroup)**.** *If we consider the 'variance-exploding' setting, in which the prior evolves along the Heat semigroup $\pi_t = \pi * \gamma_{2t}$, the tilted transport equation has the same quadratic structure:*

$$\begin{cases} \dot{Q}_t = 2Q_t^2 \ , & Q_0 = Q \\ \dot{b}_t = 2Q_t b_t \ , & b_0 = b \end{cases} \tag{13}$$

which blows up in time $T = \|Q\|^{-1}/2$, and has the analytic form $Q_t = V\mathrm{diag}(\lambda_i(1 - 2t\lambda_i)^{-1})V^\top$. See Appendix B.3 for more details.

Remark 8 (Covariance Decomposition of [BB19] and Polchinsky Flow)**.** *In [BB19], the authors develop a transformation of the tilt by Q via a decomposition of its associated covariance Q^{-1}. Specifically, they consider a decomposition of the form $Q^{-1} = \|Q\|^{-1}\mathrm{Id} + B^{-1}$, which expresses the Gaussian measure with covariance Q^{-1} as the convolution of two Gaussian measures, with covariance $\|Q\|^{-1}\mathrm{Id}$ and B^{-1} respectively. Our modified tilt at blowup $Q^\star := Q_T$ is precisely $Q^\star = B$ in the Heat semigroup setting (and $Q^\star = B(1 + \|Q\|^{-1})$ in the OU setting).*

In that case, in the language of the Polchinsky flow of [BBD23], given the original measure $\nu = \mathsf{T}_Q\pi$, the measure at blowup is $\nu_* = \mathsf{T}_{Q^\star}(\pi * \gamma_{\|Q\|^{-1}})$, which can be viewed as the renormalised measure *$\mathsf{T}_{(C_\infty - C_{\tilde{t}})^{-1}}(\pi * \gamma_{C_{\tilde{t}}})$ corresponding to $C_\infty = Q^{-1}$ and $C_{\tilde{t}} = \tilde{t}\mathrm{Id}$. In other words, we run the isotropic Polchinsky flow $\dot{C}_t = \mathrm{Id}$ for $t \le \tilde{t}$ as long as $C_\infty - C_t \ge 0$, which happens precisely at $\tilde{t} = \|Q\|^{-1}$.*

Collecting these remarks thus leads to the following.

Corollary 9 (Posterior Sampling via Tilted Transport, Heat Semigroup Setting)**.** *Sampling from the posterior $\nu = \mathsf{T}_Q\pi$ can be achieved by first sampling from $\nu_* := \mathsf{T}_{Q_*}(\pi * \gamma_{\|Q\|^{-1}})$, where $Q_*^{-1} = Q^{-1} - \|Q\|^{-1}\mathrm{Id}$, and then running the reverse SDE (6) from time $\|Q\|^{-1}/2$ to 0.*

5 Quantitative Conditions for Provable Sampling

The new posterior sampling problem described above may be easier to sample than the original posterior sampling problem due to two separate aspects. On the one hand, the (negative) eigenvalues of the quadratic tilt $-\frac{1}{2}x^\top Q_t x + x^\top b_t$ become more negative, essentially meaning that the SNR of the new observation model becomes larger. To be more specific, as $t \to T$, $\lambda_{\min}(Q_t) \to \frac{1+\lambda_{\max}(Q)^{-1}}{\lambda_{\min}(Q)^{-1}-\lambda_{\max}(Q)^{-1}} > \lambda_{\min}(Q)$. On the other hand, the new prior distribution π_T becomes closer to a single-mode Gaussian (recall that $\mathrm{KL}(\pi_t||\gamma_d) = O(e^{-t})$), which is also easier to sample. Let us now quantify the above intuition by leveraging the Bakry-Emery criterion.

5.1 Sufficient Conditions via Barky-Emery

We start by giving a simple sufficient condition that ensures that ν_T is strongly log-concave. As discussed earlier, by the Bakry-Emery criterion, this ensures fast relaxation of the Langevin dynamics, enabling efficient sampling from ν_T – and therefore of ν as per Corollary 6. For that purpose, given the prior $\pi \in \mathcal{P}(\mathbb{R}^d)$ and $t \ge 0$, we define

$$\chi_t(\pi) := \sup_{x \in \mathbb{R}^d} \|\mathrm{Cov}[\mathsf{T}_{tI,tx}\pi]\|, \tag{14}$$

where the covariance is given by $\mathrm{Cov}[\mu] = \mathbb{E}_{x\sim\mu}[xx^\top] - (\mathbb{E}_{x\sim\mu}[x])(\mathbb{E}_{x\sim\mu}[x])^\top$. $\chi_t(\pi)$ thus measures the largest 'spread' of any tilted measure of the form $\mathsf{T}_{t,x}\pi$, and is also known as the *susceptibility* in certain field models [BD22]. The covariance of isotropic tilts is a central object in the *stochastic localization* framework of Eldan [Eld20, CE22], as well as the Polchinsky renormalisation group approach of [BB19, BBD23]. Equipped with this definition, we have the following simple sufficient condition to ensure that ν_T is strongly log-concave:

Proposition 10 (Strong Log-Concavity of ν_T)**.** *Let $\kappa = \lambda_{\max}(Q)/\lambda_{\min}(Q)$ denote the condition number of Q. Then ν_T is strongly log-concave if*

$$\chi_{\|Q\|}(\pi) < \|Q\|^{-1}\frac{\kappa}{\kappa - 1}\,. \tag{15}$$

When Q is degenerate ($\lambda_{\min}(Q) = 0$, or equivalently $\kappa = \infty$), ν_T is strongly log-concave if

$$\chi_{\|Q\|}(\pi) < \|Q\|^{-1}. \tag{16}$$

The proof is in Appendix C.1. By the Barky-Emery criterion, Equation (15) is sufficient to ensure a fast relaxation of the Langevin dynamics with drift $\nabla \log \nu_T$, which can then be used to produce samples of ν thanks to Corollary 6. One can also verify that the same exact condition implies that ν_* (defined in Corollary 9) is strongly log-concave, thus the conclusion also applies to the Heat semigroup setting.

Proposition 10 relates two parameters of the measurement process, the condition number of A (equal to $\sqrt{\kappa}$ in the above definition) and the signal-to-noise ratio in terms of $\|Q\|$, with a geometric property of the prior, the susceptibility $\chi_t(\pi)$.

Controlling $\chi_t(\pi)$. The susceptibility is not generally explicit, and may not exist for general distributions in $\mathcal{P}_2(\mathbb{R}^d)$.[5] One can nevertheless consider a simple sufficient condition from [MCJ+19], that guarantees that $\chi_t(\pi) < \infty$ for all t, capturing several representative high-dimensional settings:

Proposition 11 (Sufficient Condition for Finite Susceptibility)**.** *If $\pi \in \mathcal{P}_2(\mathbb{R}^d)$ is such that $\pi = e^{-f}$, with $f \in C^1$ m-strongly convex outside a ball of radius R, and ∇f is L-Lipschitz, then $\chi_t(\pi) \leq (m/2+t)^{-1} e^{16LR^2}$ for all t.*

This sufficient condition, proved in Appendix C.2, thus imposes both concentration, obtained here through strong log-concavity (up to perturbation), and smoothness. These conditions are also (jointly) necessary, as illustrated by the following counter-examples, proved in Appendix C.3:

Proposition 12 (Susceptibility Blow-up)**.** *There exists $\pi \in \mathcal{P}_2(\mathbb{R})$ such that $\chi_t(\pi) = \infty$ for any $t > 0$. Moreover, there exists $\pi \in \mathcal{P}_2(\mathbb{R})$ with subgaussian tails, i.e., $\mathbb{P}_\pi(|Y| > z) \lesssim e^{-\lambda z^2}$, such that $\chi_t(\pi) = \infty$ for any $t > \lambda$.*

We can additionally establish simple, yet useful, properties of the susceptibility (proved in Appendix C.4):

Proposition 13 (Properties of $\chi_t(\pi)$)**.** *We have the following:*

(i) Tensorization: If $\mu = \mu_1 \otimes \mu_2 \cdots \otimes \mu_d$, then $\chi_t(\mu) = \max_i \chi_t(\mu_i)$.

(ii) Asymptotic behavior: If $\pi = e^f$ and ∇f is Lipschitz, then $\chi_t(\pi) = 1/t + o(1/t)$ as $t \to \infty$.

Finally, we conclude with explicit examples that will be used later:

Example 14 (Examples of $\chi_t(\pi)$)**.** *We have the following:*

(i) Gaussian measure: $\chi_t(\gamma_d) = \frac{1}{1+t}$.

(ii) Compactly Supported Gaussian Mixture: If μ is compactly-supported in a ball of radius R and $\delta \geq 0$, then $\chi_t(\mu * \gamma_\delta) \leq \left(\frac{R}{1+\delta t}\right)^2 + \frac{\delta}{1+\delta t}$.

(iii) Uniform measure on hypercube: If π is uniform on the hypercube $\mathcal{H}_d$, then $\chi_t(\pi) = 1$.

In light of these simple properties of the susceptibility, we can already extract useful information out of Proposition 10: in the regime where $\|Q\| \ll 1$, corresponding to the low SNR setting, and under compact support assumptions, the LHS converges to a finite value, while the RHS diverges, leading to log-concavity of ν_T. On the other hand, in the regime where $\|Q\| \gg 1$, corresponding to the high SNR setting, under minimal regularity assumptions we will have $\chi_{\|Q\|}(\pi) \simeq \|Q\|^{-1} < \|Q\|^{-1} \frac{\kappa}{\kappa-1}$, leading also to a sampling guarantee.

[5] In contrast to the *expected* covariance in the Stochastic Localization framework, which contracts thanks to the martingale property [Eld20, Equation 11].

5.2 Comparisons

With Langevin dynamics. As introduced above, Langevin dynamics and its discretized version, Langevin Monte Carlo (LMC) [RT96, MCF15] serve as natural baselines for efficient posterior sampling. As such, to assess whether sampling from ν_T is easier than sampling from ν, ideally one would like to compare the LSI constants $\rho(\nu)$ and $\rho(\nu_T)$ associated respectively with ν and ν_T.

While this direct comparison is not available in general, one can resort to comparing lower bounds. The starting point is to compare conditions for log-concavity, which imply lower bounds for ρ via Bakry-Emery. At low SNR regimes where $\|Q\| \ll 1$, Proposition 10 and Example 14 show that ν_T becomes log-concave under mild assumptions on π. On the other hand, ν will be generally non-log-concave in this regime, since the Hessian $\nabla^2 \log \nu = -Q + \nabla^2 \log \pi$ will converge to the prior.

Alternatively, by Remark 8, we can relate lower bounds for $\rho(\nu)$ and $\rho(\nu_T)$ via the multiscale Bakry-Emery criterion obtained with the Polchinksy flow using a specific covariance decomposition. Indeed, by setting $q = \|Q\|^{-1}$, $\dot{C}_t = \mathrm{Id}$ for $0 \le t \le q$, and $\dot{C}_t = (Q^{-1} - q\mathrm{Id})^{-1}\mathbf{1}(q < t \le q+1)$, from [BBD23, Theorem 3.6, Remark 3.7] (noting the difference in time parameterization by a factor of 2 between [BBD23] and our dynamics in the heat semigroup) we obtain the following lower bounds on $\rho(\nu)$ and $\rho(\nu_T)$:

$$\rho(\nu) \ge \frac{1}{2}\left(\int_0^{q+1} e^{-2\lambda_t}\mathrm{d}t\right)^{-1}, \quad \rho(\nu_T) \ge \frac{1}{2}\left(e^{2\lambda_q}\int_q^{q+1} e^{-2\lambda_t}\mathrm{d}t\right)^{-1}, \tag{17}$$

with $\lambda_t = \int_0^t \dot{\lambda}_s \mathrm{d}s$ and $(\dot{\lambda}_t)_t$ any sequence satisfying, for $t \in (0, q+1)$,

$$\dot{C}_t \nabla^2 \log(\pi * \gamma_{C_t}) \dot{C}_t \succeq \dot{\lambda}_t \dot{C}_t \,. \tag{18}$$

Now, observe that for $t \in (0, q)$, the optimal choice for $\dot{\lambda}_t$ is simply $\inf_x \lambda_{\min}\left[\nabla^2 \log(\pi * \gamma_t(x))\right]$. Thus, if the prior is such that

$$\inf_x \lambda_{\min}\left[\nabla^2 \log(\pi * \gamma_t(x))\right] \le 0 \quad \text{for} \quad t \in (0, q)\,, \tag{19}$$

we have $\lambda_q < 0$, and therefore

$$\left(e^{2\lambda_q}\int_q^{q+1} e^{-2\lambda_t}\mathrm{d}t\right)^{-1} \ge \left(\int_q^{q+1} e^{-2\lambda_t}\mathrm{d}t\right)^{-1} \tag{20}$$

$$\ge \left(\int_0^{q+1} e^{-2\lambda_t}\mathrm{d}t\right)^{-1}, \tag{21}$$

showing that the lower bound for $\rho(\nu_T)$ in (17) dominates that of $\rho(\nu)$. The condition (19) describes the generic setting where the heat semigroup, run for $t \in (0, \|Q\|^{-1})$, is not sufficient to make the prior measure log-concave, as the latter requires a much stronger condition:

$$\exists\, t \in (0, q) \;\text{ such that }\; \sup_x \lambda_{\max}\left[\nabla^2 \log(\pi * \gamma_t(x))\right] \le 0\,. \tag{22}$$

We emphasize however that the lower bound for $\rho(\nu)$ in (17) might not be tight for this particular choice of covariance decomposition, and that often one needs to adapt the decomposition to the specific model.

With Importance Sampling. In the low SNR regime with a well-conditioned A, the posterior measure can be viewed as a small perturbation of the prior. As such, a natural baseline for posterior sampling is Importance Sampling (IS), using the prior as a proposal — for which samples can be efficiently obtained thanks to the denoising oracle. In the low SNR regime, one thus expects the variance of the sampled weights to be small.

However, we now argue that, while IS is a nature baseline in this low SNR regime, it generally suffers from exponential complexity when the SNR is high. In order to estimate an integral of a function f with respect to the posterior measure ν

$$I(f) \coloneqq \int_{\mathbb{R}^d} f(x)\mathrm{d}\nu(x),$$

the idea of importance sampling is to independently sample $X_1, \ldots, X_n$ from the prior π and calculate

$$I_n(f) \coloneqq \frac{\sum_{i=1}^n f(X_i)\tau(X_i)}{\sum_{i=1}^n \tau(X_i)},$$

where $\tau(x)$ is the observation likelihood $\exp\left(-\frac{1}{2}x^\top Qx + x^\top r\right)$. With DO_π, we can sample from the prior efficiently. When σ is large such that the ratio τ is close to 1, $I_n(f)$ computed from prior samples can efficiently approximate $I(f)$. On the contrary, if σ is small, $\tau(x)$ can have very large variance and the importance sampling can be inefficient since many prior proposals have very small weights. The work [CD18, Theorem 1.2] proves that, in a fairly general setting, a sample of size approximately $\exp(\mathrm{KL}(\nu||\pi))$ is necessary and sufficient for accurate estimation by importance sampling, where $\mathrm{KL}(\nu||\pi)$ is the Kullback–Leibler divergence of π from ν:

$$\mathrm{KL}(\nu||\pi) = \int_{\mathbb{R}^d} \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\pi}\right)\mathrm{d}\nu = \int_{\mathbb{R}^d}\left(-\frac{1}{2}x^\top Qx + x^\top b\right)\mathrm{d}\nu(x) \ = -\frac{1}{2}\langle\Sigma, Q\rangle + c^\top b\ ,$$

where $\Sigma = \mathbb{E}_\nu[xx^\top]$ and $c = \mathbb{E}_\nu[x]$ are the first two moments of ν. This result confirms one part of the intuition above: if σ is sufficiently large, then the magnitude of Q and r will be sufficiently small, and so is $\mathrm{KL}(\nu||\pi)$ and the number of samples needed in the importance sampling. Next we show that for a fairly generic prior distribution π, when the SNR is large, $\mathrm{KL}(\nu||\pi)$ will be also large such that we need approximately $O(e^{d\cdot\mathrm{SNR}})$ examples to implement importance sampling, which is unachievable.

Without loss of generality, we assume the covariance of the prior π is Id.

Proposition 15 (Importance Sampling Sample Complexity Lower Bound)**.** *Assume $\nabla \log \pi(x)$ is L-Lipschitz:*

$$\|\nabla \log \pi(x) - \nabla \log \pi(z)\| \le L\|x - z\|\ . \tag{23}$$

Then, when $\mathrm{SNR} > L + 2$*, we have*

$$\mathrm{KL}(\nu||\pi) \ge O(\exp(d\cdot\mathrm{SNR}))\ , \tag{24}$$

and therefore the sample complexity of IS is exponential in dimension.

The proof of Proposition 15 is in Appendix C.6, and leverages Talagrand's transport inequality to relate the KL divergence term to the Wasserstein distance $W_2^2(\nu, \pi)$, which can be effectively bounded in the regime of high SNR.

5.3 Stability Analysis

In the numerical implementation of boosted posterior, we typically encounter certain errors. Especially, we may have imperfect score subject to certain L^2 errors, and we may not be able to sample the boosted posterior ν_t exactly. Suppose that instead of starting from ν_t at t and run the exact reverse SDE (5), we start from an approximate distribution $q_t \approx \nu_t$ and run the reverse SDE (5) with approximating score $s_\theta(x, t) \approx \nabla \log \pi_t(x)$ where θ denote the parameters parametrizing the score. Denote the distribution of the final samples by q_0. We have the following error estimate

Proposition 16 (Stability of Tilted Transport)**.** *Suppose $\nu_t, q_t, \nabla \log \pi_t, s_\theta(x,t)$ has enough regularities such that the reverse SDEs exist, if the Novikov's condition*

$$\mathbb{E}\left[\exp(\int_0^t \|\nabla \log \pi_\tau(x) - s_\theta(x,\tau)\|^2 \mathrm{d}\tau)\right] < \infty$$

holds, then

$$\mathrm{KL}(\nu||q_0) \leq \int_0^t \mathbb{E}_{\nu_\tau}\|\nabla \log \pi_\tau(x) - s_\theta(x,\tau)\|^2 \mathrm{d}\tau + \mathrm{KL}(\nu_t||q_t). \tag{25}$$

The above proposition ensures that if both the initialization error and score L^2 error (over the posterior paths) are small, then the distribution of our final samples is close to the target posterior. Note that the score L^2 error in the RHS is *not* a Fisher divergence, since we are estimating the error using the posterior ν_τ rather than the prior π_τ. The stability guarantee in Proposition 16 thus requires a Denoising Oracle robust to Out-of-Distribution errors. That said, this OOD robustness appears to be unavoidable in our setting of posterior sampling.

The proof is provided in Appendix C.7. Note that we consider the reverse dynamics in continuous-time without time discretization error. There are various works [LLT22, LLT23, CCL+23, BDBDD24] analyzing the time discretization error and those techniques can be further incorporated into the above error estimate.

6 Case Studies

In this section we specialize the results of Section 5 to representative models. Posterior distributions of the form $\nu = \mathsf{T}_Q \pi$ where π is a product measure provide particularly explicit calculations.

6.1 Gaussian Mixtures

By applying Proposition 10 to Example 14 (ii), we directly obtain the following guarantee for generic comptactly supported Gaussian mixtures (proof in Appendix C.5):

Corollary 17 (Tilted Transport for Gaussian Mixtures)**.** *If $\pi = \mu * \gamma_\delta$ and $diam(supp(\mu)) \leq R$, then ν_T is strongly log-concave if*

$$R^2 < \frac{(1+\delta \mathrm{SNR}^2)(\delta \kappa(A)^2 + \mathrm{SNR}^{-2})}{\kappa(A)^2 - 1}. \tag{26}$$

It also holds when $\delta = 0$ and the prior π is any distribution with a bounded support radius R.

Figure 2 displays several contours of the condition in eq. (26) as a function of SNR and $\kappa(A)$. Each U-shaped contour is determined by a combination of δ and R, which uniquely characterizes the prior. For all points ((SNR), $\kappa(A)$) outside of a contour, representing a specific inverse problem, ν_T is strongly log-concave and thus easy to sample. Given an observation model where both SNR and $\kappa(A)$ are fixed, it is straightforward to see that the condition in eq. (26) is more readily satisfied as δ increases and R decreases. Figure 2 also confirms this result since as δ increases or R decreases, the U-shaped contour shrinks and the region of easy to sample expands. Now we discuss the implications in the reverse scenario where the prior is fixed and the observation model is adjusted. If we look at Figure 2 horizontally, we know that given a prior and $\kappa(A)$, the target posterior can be reliably sampled if the SNR is either sufficiently low or high, with the region of mid-SNR being challenging. The closer $\kappa(A)$ is to 1, the smaller this challenging region is. When the problem is denoising such that $\kappa(A) = 1$, the challenging region vanishes, and sampling the posterior is straightforward using the denoising oracle, as previously explained.

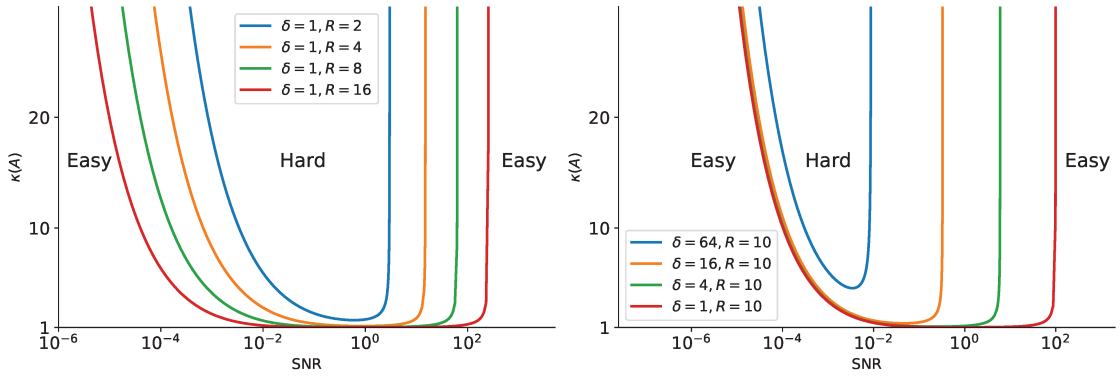

Figure 2: Phase diagram for the boosted posterior ν_T being strongly log-concave in Corollary 17.

6.2 Ising Models

As a direct consequence of Proposition 10 and Example 14 (iv), we establish a sampling guarantee for Ising models:

Corollary 18 (Tilted Transport for the Ising Model)**.** *Let π be the uniform measure on the hypercube, and Q such that $\lambda_{\max}(Q) - \lambda_{\min}(Q) < 1$. Then ν_T is strongly log-concave, and therefore $\nu = \mathsf{T}_Q \pi$ can be sampled efficiently (in continuous-time).*

This result thus establishes that Ising models may be sampled using Tilted Transport, provided their spectrum satisfies $\lambda_{\max}(Q) - \lambda_{\min}(Q) < 1$, thus precisely matching the computational lower bound of [Kun23]. We remark though that our procedure is not (yet) algorithmic; a careful analysis of the discrete-time complexity and the approximation rates to avoid the singularity of the score of π_t as $t \to 0$ (via Proposition 4) should be formalized.

In itself, this sampling guarantee should not come as a surprise, since, as discussed previously, Glauber dynamics are already known to mix efficiently in this regime [EKZ22, AJK+21], and a Log-Sobolev Inequality was established even earlier than these in [BB19]. That said, these positive guarantees both require a 'multiscale' decomposition of entropy that goes beyond the Bakry-Emery criterion, using either the framework of stochastic localization [Eld20, CE22] or the similar Polchinsky renormalisation group framework [BBD23]. In that sense, an arguably interesting feature of our result lies in its simplicity: we only exploit the Bakry-Emery criterion at the tilted measure ν_T.

If one specializes Corollary 18 to the Sherrington-Kirkpatrick (SK) model, the equivalent inverse temperature that guarantees sampling is $\beta = 1/4$, which remains below $\beta = 1$, the threshold of the hard phase. For this threshold, dedicated AMP-based sampling succeeds [EAMS22, Cel24]. As future work, it would be interesting to explore whether the Polchinsky-based multiscale Bakry-Emery criterion could be applied to ν_T to improve upon Proposition 10 in this model. Similarly as in the referred prior works [Eld20, CE22, BB19], the key ingredient that would remove the roadblock is a sharper bound on the susceptibility $\chi_t(\pi)$; and in particular extending the existing bounds on the covariance of the SK model [EAG24, BXY23], valid for $\beta < 1$, to arbitrary external fields.

6.3 Scalar Field φ^4 model

As a further illustration, we now consider the two-dimensional lattice scalar φ^4 field model, where ν is given by

$$\nu = \mathsf{T}_{\beta\Delta}\pi_\beta \in \mathcal{P}(\Omega)\,, \tag{27}$$

where Ω is a discrete two-dimensional lattice of total size d, β is the inverse temperature, Δ is the discrete Laplacian on Ω and $\pi_\beta = \mu_\beta^{\otimes d}$ is a product measure, whose marginal in each site is given by

$$d\mu_\beta(\varphi) \propto e^{-\varphi^4 + (1+2\beta)\varphi^2} d\varphi \ .$$

The scalar potential thus has the familiar double-well profile that enforces configurations close to ± 1.

This model is known to satisfy a Log-Sobolev Inequality with ρ independent of d as long as $\beta < \beta_c \approx 0.68$, by exploiting the multiscale Bakry-Emery criterion along a Polchinsky flow with carefully chosen covariance decomposition [BD22, BBD23]. We can alternatively apply Proposition 10 directly to ν_T. Using $\|\beta\Delta\| = 4\beta$, and $\lambda_{\min}(\Delta) = 0$, we obtain that ν can be efficiently sampled using tilted transport whenever

$$\eta < \frac{1}{4\beta} \ , \ \text{where} \ \ \eta = \sup_\alpha \mathrm{Cov}[\mathsf{T}_{0,\alpha}\mu_0] \ , \tag{28}$$

where $\mathsf{T}_{0,\alpha}$ is thus a linear tilt. Numerically estimating η gives $\eta \approx 0.52$ and therefore ν_T is log-concave whenever $\beta < 0.48$. This is slightly below the phase transition $\beta_c \approx 0.68$, indicating that in this example the direct Barky-Emery criterion, even if it is applied to the tilted measure, is not as sharp as the multiscale criterion.

A natural question is thus whether one could apply the techniques of [BD22] to ν_T. Let us briefly describe the main technical challenge that needs to be overcome for this purpose. We recall that the tilted posterior is $\nu_* = \mathsf{T}_{Q_*}(\pi * \gamma_{\|Q\|})$ (using the Heat semigroup w.l.o.g.), where $Q_* = (\beta^{-1}\Delta^{-1} - (4\beta)^{-1}\mathrm{Id})^{-1}$. The covariance decomposition used to establish the sharp LSI bound for ν exploits a key structural property of the operator $Q = \Delta$, namely that it is *ferromagnetic*, i.e., $Q_{ij} \leq 0$ for all $i \neq j$, enabling a sharp control of the associated susceptibility $\chi_t(\nu)$; see [BD22] and also [CE22, Section 5.1.2] for further details. The ferromagnetic property is unfortunately lost in Q_*, despite being a more 'convex' potential.

7 Iterated Tilted Transport

We have shown that posterior sampling of $\nu = \mathsf{T}_Q\pi$ can be reduced to sampling from ν_T by running the reverse SDE. While ν_T is easy to sample under the conditions presented in Section 5, these may not be verified in several situations of interest. In this context, a natural question is whether one could still leverage the tilted transport, at the expense of introducing sampling error. This is what we address in this section.

Let $\lambda_1, \dots \lambda_d$ be the eigenvalues of Q. Let us assume for simplicity that all eigenvalues have multiplicity 1, so $\lambda_i > \lambda_{i+1}$. We also adopt the heat semigroup to simplify the exposition without loss of generality. We define the events T_j for $j = 1 \dots d$ given by

$$T_j := \frac{1}{2}\lambda_j^{-1} \ . \tag{29}$$

Denote by

$$\bar{\lambda}_j(t) = \begin{cases} \infty & \text{if } t \geq T_j \ , \\ \lambda_j(t) & \text{otherwise,} \end{cases}$$

where $\lambda_j(t) = \frac{\lambda_j}{1-2t\lambda_j}$ is the solution to the ODE $\dot{q}_t = 2q_t^2$ for tilted transport in the heat semigroup setting. By abusing notation, we denote by $\bar{Q}_t$ the matrix that shares eigenvectors with Q, and with eigenvalues $(\bar{\lambda}_1(t), \dots, \bar{\lambda}_d(t))$. Denote by $V_k = [v_{d-k} \dots v_d] \in \mathbb{R}^{d\times k}$ the orthogonal projection onto the last k eigenvectors.

While previously we considered only the transport between ν and $\nu_1 := \nu_{T_1}$, now we can consider the sequence $\nu_k := \mathsf{T}_{\bar{Q}_{T_k}}\pi_{T_k}$ for $k = 1, \dots, d$, where we denote $\pi_t = \pi * \gamma_{2t}$ the action of the heat semigroup

on π. Observe that ν_k is a measure supported on a subspace Ω_k of dimension $d-k$; in other words, k directions are singular, corresponding to the eigenvectors associated with the ∞ eigenvalues of $\bar{Q}_{T_k}$, and thus $\Omega_k = \{x \in \mathbb{R}^d; V_k^\top x = \mathbf{y}_k\}$ for some $\mathbf{y}_k \in \mathbb{R}^k$.

Now, let us consider $k^\star = \min\{k; \nu_k \text{ is s.l.c.}\}$; that is, the first k such that ν_k is strongly log-concave, and therefore efficiently sampleable by Langevin dynamics. By defining $\kappa_k := \frac{\lambda_k}{\lambda_d}$ as the condition number of the truncated Q, we immediately obtain from Proposition 10 the bound

$$k^\star \le \min\left\{k; \chi_{\lambda_k}(\pi) \le \lambda_k^{-1}\frac{\kappa_k}{\kappa_k - 1}\right\} . \tag{30}$$

For $k < k^\star$, assume first that one had sampling access to ν_{k+1}. Running the reverse tilted transport for time $\eta_k := T_{k+1} - T_k$ produces samples from a tilted measure $\tilde{\nu}_k := \mathsf{T}_{\tilde{Q}_k}\pi_{T_k}$, where $\tilde{Q}_k$ has eigenvalues

$$(\tilde{\lambda}_1(T_k), \dots, \tilde{\lambda}_k(T_k), \tilde{\lambda}_{k+1}(T_k), \tilde{\lambda}_{k+2}(T_k), \dots, \tilde{\lambda}_d(T_k)) \tag{31}$$

$$:= \left(\underbrace{\frac{\lambda_{k+1}}{1-\lambda_k^{-1}\lambda_{k+1}}, \dots, \frac{\lambda_{k+1}}{1-\lambda_k^{-1}\lambda_{k+1}}}_{k \text{ times}}, \frac{\lambda_{k+1}}{1-\lambda_k^{-1}\lambda_{k+1}}, \frac{\lambda_{k+2}}{1-\lambda_k^{-1}\lambda_{k+2}}, \dots, \frac{\lambda_d}{1-\lambda_k^{-1}\lambda_d} \right) . \tag{32}$$

One can easily verify the above fact by noting that under the ODE $\dot{q} = 2q^2$ with initial condition $q(T_k) = \tilde{\lambda}_j(T_k)$, then the solution at time T_{k+1} will be exactly $q(T_{k+1}) = \bar{\lambda}_j(T_{k+1})$. Since all $\tilde{\lambda}_j(T_k) < \infty$, $\tilde{\nu}_k$ is thus a non-singular measure in $\mathbb{R}^d$, capturing the fact that the Brownian motion driving the reverse dynamics is isotropic, and thus oblivious to the existence of the singular support of ν_k.

We thus need a procedure to transform samples from $\tilde{\nu}_k$ to approximate samples of ν_k. The simplest way is to simply marginalize the coordinates $(x_1, \dots, x_k) = V_k^\top x \in \mathbb{R}^k$, i.e.,

$$\bar{\nu}_k(x_{k+1}, \dots, x_d) = \int_{\mathbb{R}^k} \tilde{\nu}_k(\mathrm{d}x_1, \dots, \mathrm{d}x_k, x_{k+1}, \dots, x_d) \in \mathcal{P}(\mathbb{R}^{d-k}) ,$$

and then 'lift' this measure in the subspace Ω_k, i.e.,

$$\hat{\nu}_k(\mathbf{x}_k; \mathbf{x}_{-k}) := \delta(\mathbf{x}_k - \mathbf{y}_k)\bar{\nu}_k(\mathbf{x}_{-k}) , \tag{33}$$

where we defined $\mathbf{x}_k = (x_1, \dots, x_k)$ and $\mathbf{x}_{-k} = (x_{k+1}, \dots, x_d)$. Under this definition, given a sample from ν_k, we just project its k components to $\boldsymbol{y}_k$ to get a sample from $\hat{\nu}_k$, as an approximation to the sample from ν_k.

Algorithm 1 Sampling Using Iterated Tilted Transport

1: Start by sampling $X_{k^\star} \sim \nu_{k^\star}$.
2: **for** $k = k^\star - 1$ **to** 0 **do**
3: Run tilted transport starting at X_{k+1} for time η_k, resulting in $\tilde{X}$.
4: Set $X_k = (\mathbf{y}_k; \tilde{\mathbf{X}}_{-k})$.
5: **end for**
6: **return** X_0

We can then iteratively run the tilted transport backwards, from $k = k^\star - 1$ to $k = 0$, as illustrated in Algorithm 1 and Figure 3. By the data-processing inequality, the TV error will accumulate linearly at each step. Denoting $\hat{\nu}$ the law of X_0, we have

$$\mathrm{TV}(\hat{\nu}, \nu) \le \sum_{0<k<k^\star} \mathrm{TV}(\hat{\nu}_k, \nu_k) . \tag{34}$$

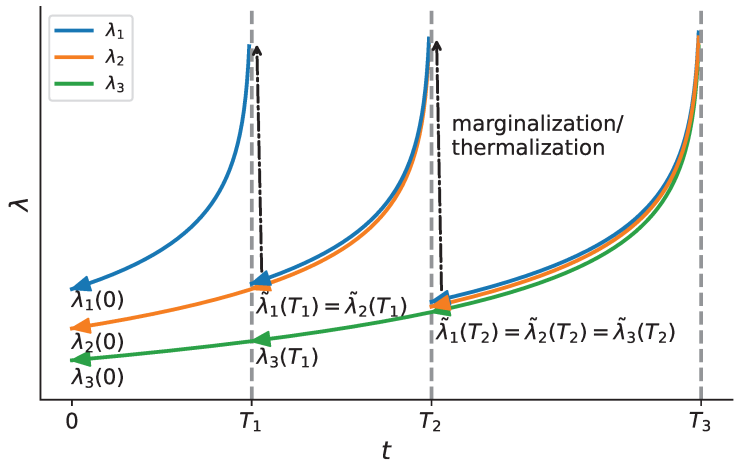

Figure 3: Scheme plot of iterated titled transport in an example with $d = 3, k^\star = 3$. The tilted transport from T_{k+1} to T_k transforms samples of ν_{k+1} (corresponding to eigenvalues $\bar{\lambda}_j(T_{k+1})$ to $\tilde{\nu}_k$ (corresponding to eigenvalues $\tilde{\lambda}_j(T_k)$), and marginalization/thermalization (denoted by the dashed lines) further transforms samples of $\tilde{\nu}_k$ toward ν_k.

This bound can be interpreted as the accumulation of errors arising from conditioning a measure by marginalizing over its first k components. To the extent that $\frac{\lambda_{k+1}}{\lambda_k} \approx 1$ such that $\eta_k \ll 1$, these variables are nearly deterministic, so one would expect that marginalization is a good approximation of conditioning. The outstanding question is to understand conditions when this error guarantee can be quantified.

Inspired by [CCL+24], a natural extension of this simple iterative procedure based on marginalization is to apply 'thermalization' towards the stationary measure ν_k after line 4 of Algorithm 1 above, by running Langevin dynamics in Ω_k with score $\nabla \log \nu_k$:

$$\mathrm{d}X_t = \nabla \log \nu_k(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t \ , \ X_0 \sim \hat{\nu}_k. \tag{35}$$

The drift of this diffusion is available, since both $\bar{Q}_{T_k}$ and $\nabla \log \pi_{T_k}$ are known, so is $\nabla \log \nu_k$.

Denote by $\check{\nu}_k$ the law of X_t after time $t = B_k$. While the time to relaxation of such Langevin dynamics is generally not quantitative (otherwise $k^\star \leq k$), even a short amount of thermalization is able to improve upon the previous method. Indeed, by the reverse transport inequality [BGL01, Lemma 4.2], a weaker Wasserstein guarantee $W_2(\hat{\nu}_k, \nu_k)$ can be 'upgraded' to a TV guarantee of the form $\mathrm{TV}(\check{\nu}_k, \nu_k) = O(\sqrt{L_k} W_2(\hat{\nu}_k, \nu_k))$ by running Langevin dynamics for time $B_k = \Theta(1/L_k)$, where $L_k = \sup_x \lambda_{\max}(\nabla^2 \log \nu_k(x)) > 0$ is the largest eigenvalue of $\nabla^2 \log \nu_k$, which is positive by definition of $k^\star$ and $k < k^\star$. Notice also that from Equation (72) we have the upper bound

$$L_k \leq (1 + \lambda_k)\left(\lambda_k \chi_{\lambda_k}(\pi) - \frac{\kappa_k}{\kappa_k - 1}\right) .$$

In summary, the 'thermalized' iterated tilted transport satisfies an error bound of the form

$$\mathrm{TV}(\hat{\nu}, \nu) \lesssim \sum_{0<k<k^\star} \sqrt{L_k} W_2(\hat{\nu}_k, \nu_k) \ . \tag{36}$$

We can interpret the error guarantee of Equation (36) as being able to leverage a warm-start in Wasserstein distance when running Langevin dynamics at each step k. In that sense, it would be interesting to understand whether the further structure of the warmstart could be further exploited.

Interpretation as a Homotopy 'Frequency-Marching' Method. If one considers an inverse problem where the data x is a signal over a physical domain, e.g. an image or a time-series, and the measurement operator A is a translation-invariant low-pass filter, then $Q = A^\top A$ diagonalises in the Fourier basis, and the eigenvalues of Q are sorted from low to high-frequencies. In this prototypical setting, the iterative tilted transport Algorithm 1 can be viewed as an homotopy method for sampling, whereby the low-frequency projection of the measurements (encoded in the vector $\mathbf{y}_k$) is used to condition the reverse diffusion process; this is reminscient of 'Frequency-Marching' methods, a powerful heuristic used in several imaging inverse problems [Che97, BGPS17].

8 Numerical Experiments

Gaussian Mixture Model. We verify our theoretical results using the Gaussian mixture model in high dimensions. Same to the models considered in [CJCM24], the prior distribution is a mixture of 25 components with known means and variances (see Figure 1 for a 2D visualization and Appendix D for detailed settings). We examine three cases where $d = 20, 40$, and 80. In each scenario, we set $d' = d$, fix $\kappa = 20$, and vary the SNR from 10^{-5} to 10^{-1}. We use the Sliced Wasserstein distance as a principled error metric, computed from samples obtained by our algorithms and samples directly from the analytically computed posterior Gaussian mixture. Figure 4 illustrates the comparison between LMC and LMC boosted by tilted transport. As analyzed earlier, LMC is effective when the SNR is high enough to render the target posterior strongly log-concave, but its error quickly increases as the SNR decreases. In contrast, the tilted transport enhances LMC to perform well in both low and high SNR regimes with small sampling errors. Its performance is weaker in the mid-SNR regime compared to the extremes, as predicted by Corollary 17. However, the tilted transport still improves upon LMC in this challenging regime by boosting effective SNR and simplifying the prior.

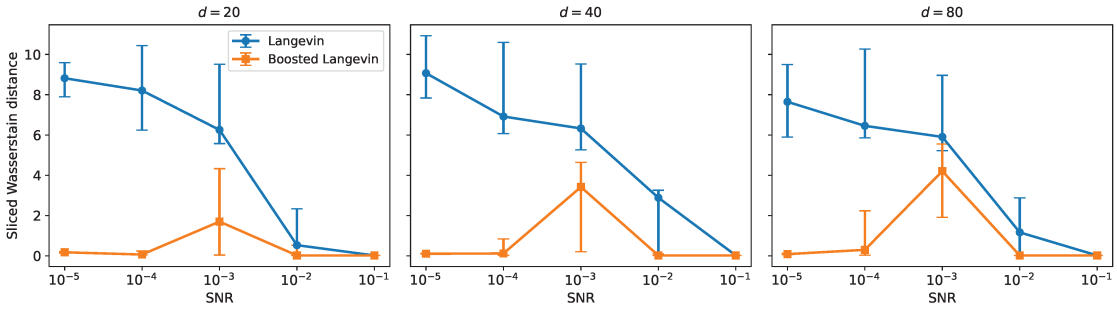

Figure 4: Comparison of Langevin and boosted Langevin for Gaussian mixture prior. We generate the prior, measurement and sample the posterior under 20 different instances in each setting. The sliced Wasserstein distances are plotted with the median in the middle, and the 25th and 75th percentiles indicated by the error bars.

We further test tilted transport when $d' < d$, in which $\lambda_{\min}(Q_t)$ remains zero but the signal corresponding to the non-zero eigenvalues still gets enhanced. Therefore, although it becomes more difficult for ν_T to be strongly log-concave, the tilted transport can still make the new posterior easier to sample even if it is not strongly log-concave yet. Detailed results are reported in Appendix D.1.

Scalar Field φ^4 model. We test the effect of tilted transport on the scalar field φ^4 model. We run the Langevin dynamics for the target measure ν in Equation (27) and its tilted version near blowup time, using a lattice size of 128×128. To compare the relaxation time speed, we compute the autocorrelation of each single site over 100,000 steps after the burn-in stage. Figure 5 shows the autocorrelation averaged over all

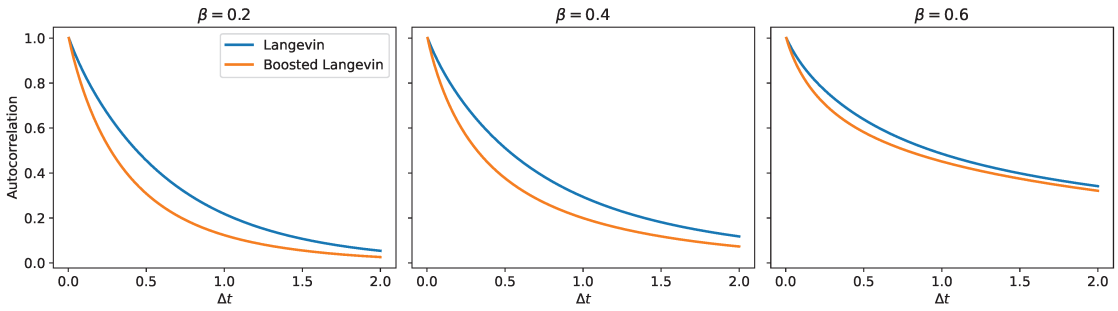

Figure 5: Comparison of autocorrelation in Langevin dynamics between the original posterior and boosted posterior distributions. The autocorrelation is computed for each site individually and then averaged.

sites for $\beta = 0.2, 0.4$, and 0.6. We clearly observe that the relaxation becomes slower as β increases, but in all three cases, the tilted transport accelerates the Langevin dynamics.

9 Discussion and Future Work

In this paper, we theoretically investigate posterior sampling using powerful priors provided by denoising oracles. We demonstrate that efficient posterior sampling can be challenging even with a perfect denoising oracle for the prior. To achieve provable posterior sampling, one must either constrain the measurements or leverage the structural properties of the prior. In this paper, we focused on the former, showing that well-conditioned measurements enable the proposed tilted transport technique to simplify the task significantly, providing a clear, verifiable condition for efficient sampling, as demonstrated on the Ising model.

That said, several questions remain open: Can this approach provably handle poorly-conditioned measurements, such as inpainting? Can it be extended from linear to nonlinear inverse problems? We showed in Section 7 how to extend the tilted transport beyond the condition of Proposition 10 via 'iterated tilts', at the expense of introducing approximation errors. A natural goal is to quantify these errors in practical situations. On the theory side, the key object underlying the success of the tilted transport is the susceptibility $\chi_t(\pi)$; in particular, understanding when one can remove dimension dependence is an interesting question. We also aim to systematically evaluate the empirical performance of tilted transport in imaging and scientific computing. Appendix D.2 provides a proof-of-concept for various imaging tasks. We suspect that tilted transport could even improve existing posterior point estimate methods by boosting SNR and enabling proper uncertainty quantification through the reverse process.

References

[AJK+21] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence I: Modified log-Sobolev inequalities for fractionally log-concave distributions and high-temperature ising models. *arXiv preprint arXiv:2106.04105*, 2021.

[And82] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

[BB19] Roland Bauerschmidt and Thierry Bodineau. A very simple proof of the LSI for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019.

[BBD23] Roland Bauerschmidt, Thierry Bodineau, and Benoit Dagallier. Stochastic dynamics and the Polchinski equation: an introduction. *arXiv preprint arXiv:2307.07619*, 2023.

[BD22] Roland Bauerschmidt and Benoit Dagallier. Log-Sobolev inequality for the φ_2^4 and φ_3^4 measures. *arXiv preprint arXiv:2202.02295*, 2022.

[BDBDD23] Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Linear convergence bounds for diffusion models via stochastic localization. *arXiv preprint arXiv:2308.03686*, 2023.

[BDBDD24] Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *The Twelfth International Conference on Learning Representations*, 2024.

[BÉ85] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In Jacques Azéma and Marc Yor, editors, *Séminaire de Probabilités XIX 1983/84*, pages 177–206. Springer Berlin Heidelberg, Berlin, Heidelberg, 1985.

[BGL01] Sergey G Bobkov, Ivan Gentil, and Michel Ledoux. Hypercontractivity of Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 80(7):669–696, 2001.

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

[BGPS17] Alex Barnett, Leslie Greengard, Andras Pataki, and Marina Spivak. Rapid solution of the cryo-EM reconstruction problem by frequency marching. *SIAM Journal on Imaging Sciences*, 10(3):1170–1195, 2017.

[BHK+19] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

[BL76] Herm Jan Brascamp and Elliott H Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of functional analysis*, 22(4):366–389, 1976.

[BLB08] Peter Bickel, Bo Li, and Thomas Bengtsson. Sharp failure rates for the bootstrap particle filter in high dimensions. In *Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K. Ghosh*, volume 3, pages 318–330. Institute of Mathematical Statistics, 2008.

[BXY23] Christian Brennecke, Changji Xu, and Horng-Tzer Yau. Operator norm bounds on the correlation matrix of the SK model at high temperature. *arXiv preprint arXiv:2307.12535*, 2023.

[CCL+23] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, 2023.

[CCL+24] Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast. *Advances in Neural Information Processing Systems*, 36, 2024.

[CCLL24] Alberto Cabezas, Adrien Corenflos, Junpeng Lao, and Rémi Louf. Blackjax: Composable Bayesian inference in JAX, 2024.

[CD18] Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122. IEEE, 2022.

[Cel24] Michael Celentano. Sudakov–Fernique post-AMP, and a new proof of the local convexity of the TAP free energy. *The Annals of Probability*, 52(3):923–954, 2024.

[Che97] Yu Chen. Inverse scattering via Heisenberg's uncertainty principle. *Inverse problems*, 13(2):253, 1997.

[CJCM24] Gabriel Cardoso, Yazid Janati El idrissi, Sylvain Le Corff, and Eric Moulines. Monte Carlo guided denoising diffusion models for Baysian linear inverse problems. In *The Twelfth International Conference on Learning Representations*, 2024.

[CKJ+21] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.

[CKM+23] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

[CKS24] Sitan Chen, Vasilis Kontonis, and Kulin Shah. Learning general gaussian mixtures with efficient score matching. *arXiv preprint arXiv:2404.18893*, 2024.

[CLL23] Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *International Conference on Machine Learning*, pages 4735–4763. PMLR, 2023.

[CSY22] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12413–12422, 2022.

[DS24] 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.

[EAG24] Ahmed El Alaoui and Jason Gaitonde. Bounds on the covariance matrix of the Sherrington–Kirkpatrick model. *Electronic Communications in Probability*, 29:1–13, 2024.

[EAMS22] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334. IEEE, 2022.

[EKZ22] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A spectral condition for spectral gap: fast mixing in high-temperature ising models. *Probability theory and related fields*, 182(3):1035–1051, 2022.

[Eld20] Ronen Eldan. Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation. *Probability Theory and Related Fields*, 176(3):737–755, 2020.

[FSR+23] Berthy T Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L Bouman, and William T Freeman. Score-based diffusion models as principled priors for inverse imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10520–10531, 2023.

[Gel90] Matthias Gelbrich. On a formula for the L2 wasserstein metric between measures on euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990.

[GJP+24] Shivam Gupta, Ajil Jalal, Aditya Parulekar, Eric Price, and Zhiyang Xun. Diffusion posterior sampling is computationally intractable. *arXiv preprint arXiv:2402.12727*, 2024.

[GMJS22] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. *Advances in Neural Information Processing Systems*, 35:14715–14728, 2022.

[Gro75] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[GŠV16] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic ising and hard-core models. *Combinatorics, Probability and Computing*, 25(4):500–559, 2016.

[HDMOB24] David Heurtel-Depeiges, Charles C Margossian, Ruben Ohana, and Bruno Régaldo-Saint Blancard. Listening to the noise: Blind denoising with gibbs diffusion. *arXiv preprint arXiv:2402.19455*, 2024.

[Her56] Robbins Herbert. An empirical bayes approach to statistics. In *Proceedings of the third berkeley symposium on mathematical statistics and probability*, volume 1, pages 157–163, 1956.

[HG14] Matthew D Hoffman and Andrew Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[HJA20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[HP86] Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pages 1188–1205, 1986.

[HS87] Richard Holley and Daniel Stroock. Logarithmic Sobolev inequalities and stochastic ising models. *Journal of Statistical Physics*, 46(5–6):1159–1194, March 1987.

[JAD+21] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. *Advances in Neural Information Processing Systems*, 34:14938–14954, 2021.

[JDMO24] Yazid Janati, Alain Durmus, Eric Moulines, and Jimmy Olsson. Divide-and-conquer posterior sampling for denoising diffusion priors. *arXiv preprint arXiv:2403.11407*, 2024.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[KEES22] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022.

[KLA19] 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*, pages 4401–4410, 2019.

[Kun23] Dmitriy Kunisky. Optimality of Glauber dynamics for general-purpose Ising model sampling and free energy approximation. *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5013–5028, 2023.

[KVE21] Bahjat Kawar, Gregory Vaksman, and Michael Elad. SNIPS: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.

[KWB19] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.

[Led01] M. Ledoux. *The concentration of measure phenomenon*. Mathematical surveys and monographs. American Mathematical Society, 2001.

[LLT22] Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence for score-based generative modeling with polynomial complexity. *Advances in Neural Information Processing Systems*, 35:22870–22882, 2022.

[LLT23] Holden Lee, Jianfeng Lu, and Yixin Tan. Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pages 946–985. PMLR, 2023.

[MCF15] Yi-An Ma, Tianqi Chen, and Emily Fox. A complete recipe for stochastic gradient MCMC. *Advances in neural information processing systems*, 28, 2015.

[MCJ+19] Yi-An Ma, Yuansi Chen, Chi Jin, Nicolas Flammarion, and Michael I Jordan. Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences*, 116(42):20881–20885, 2019.

[MK22] Xiangming Meng and Yoshiyuki Kabashima. Diffusion model based posterior sampling for noisy linear inverse problems. *arXiv preprint arXiv:2211.12343*, 2022.

[MSKV23] Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. *arXiv preprint arXiv:2305.04391*, 2023.

[MV00] Peter A Markowich and Cédric Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp*, 19:1–29, 2000.

[Oks13] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[OV00] Felix Otto and Cédric Villani. Generalization of an inequality by talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000.

[RT96] Gareth O Roberts and Richard L Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

[SCC+22] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022.

[SDME21] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in neural information processing systems*, 34:1415–1428, 2021.

[SDWMG15] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.

[SLK22] Shirin Shoushtari, Jiaming Liu, and Ulugbek S Kamilov. Dolph: Diffusion models for phase retrieval. *arXiv preprint arXiv:2211.00529*, 2022.

[SSDK+20] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[SSXE22] 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.

[SVMK22] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2022.

[SZY+23] Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *International Conference on Machine Learning*, pages 32483–32498. PMLR, 2023.

[WTN+23] 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, pages 31372–31403, 2023.

A Proof of Proposition 4

Let $\delta > 0$ and $\bar{\pi}$ be the uniform measure in the d-dimensional hypercube. Consider a Gaussian mixture prior π defined as $\pi = \bar{\pi} * \gamma_\delta$.

Since both $\bar{\pi}$ and γ_δ are product measures, it follows that π is also a product measure, and therefore its denoising oracle DO_π is explicitly given by $\mathrm{DO}_\pi(y,t)_i = \psi(y_i, t)$, with

$$\psi(v,t) = \int_{\mathbb{R}} u q_{v,t}(u) du \,, \tag{37}$$

$$q_{v,t}(u) = Z^{-1}\left(e^{-\frac{1}{2}(\delta^{-2}(u-1)^2 + t^{-2}(v-u)^2)} + e^{-\frac{1}{2}(\delta^{-2}(u+1)^2 + t^{-2}(v-u)^2)}\right) . \tag{38}$$

Observe that $q_{v,t}$ is the density of a Gaussian mixture in $\mathbb{R}$ of the form $\alpha\mathcal{N}(b_-, \sigma) + (1-\alpha)\mathcal{N}(b_+, \sigma)$, with parameters

$$\sigma^{-2} = \delta^{-2} + t^{-2} \tag{39}$$

$$b_\pm = \frac{\pm\delta^{-2} + t^{-2}v}{\sigma^{-2}} \tag{40}$$

$$\alpha = \frac{e^{\frac{(\delta^{-2}+t^{-2}v)^2}{2\sigma^2}}}{e^{\frac{(\delta^{-2}+t^{-2}v)^2}{2\sigma^2}} + e^{\frac{(-\delta^{-2}+t^{-2}v)^2}{2\sigma^2}}} , \tag{41}$$

and thus $\psi(v,t) = \alpha b_- + (1-\alpha)b_+$.

Let us now denote by μ_Q the target Ising model, supported in the d-dimensional hypercube, and define the approximation $\mu_Q^\sigma := \mathsf{T}_Q \pi$. Suppose that there is an algorithm $\mathcal{A}$ that leverages the denoising oracle of π that can efficiently sample from μ_Q^σ: its law $\hat{\mu}$ satisfies $\mathrm{TV}(\mu_Q^\sigma, \hat{\mu}) \le \epsilon$ with runtime polynomial in d and $\log(\epsilon^{-1})$.

Let now $R(x) = \mathrm{sign}(x)$, and consider the sampler $R \circ \mathcal{A}$, which is now supported in the hypercube. By the triangle and data-processing inequalities, we directly have

$$\mathrm{TV}(R_\# \hat{\mu}, \mu_Q) \le \mathrm{TV}(R_\# \hat{\mu}, R_\# \mu_Q^\delta) + \mathrm{TV}(R_\# \mu_Q^\delta, \mu_Q) \tag{42}$$

$$\le \mathrm{TV}(\hat{\mu}, \mu_Q^\delta) + \mathrm{TV}(R_\# \mu_Q^\delta, \mu_Q) \tag{43}$$

$$\le \epsilon + \mathrm{TV}(R_\# \mu_Q^\delta, \mu_Q). \tag{44}$$

It remains to bound the second term in the RHS. We have to compare two measures in the hypercube. For $\sigma \in \mathcal{H}_d := \{\pm 1\}^d$, they are given respectively by

$$\mu_Q(\sigma) = \frac{1}{Z} e^{-\frac{1}{2}\sigma^\top Q \sigma} \,, \tag{45}$$

$$R_\# \mu_Q^\delta(\sigma) = \frac{1}{\tilde{Z}} \sum_{z \in \mathcal{H}_d} \int_{R(x)=\sigma} e^{-\frac{1}{2}(x^\top Q x + \delta^{-2}\|x-z\|^2)} dx \,. \tag{46}$$

Applying the Laplace approximation into each integral we obtain, as $\delta \to 0$,

$$\int_{R(x)=\sigma} e^{-\frac{1}{2}(x^\top Q x + \delta^{-2}\|x-z\|^2)} dx = \begin{cases} \sim C_{d,\delta} e^{-\frac{1}{2}\sigma^\top Q \sigma} & \text{if } z = \sigma \,, \\ \sim C_{d,\delta} e^{-\frac{1}{2}(\sigma\oplus z)^\top Q (\sigma \oplus z)} e^{-\frac{|\sigma - z|}{2\delta^2}} & \text{otherwise} \,, \end{cases} \tag{47}$$

where $\sigma \oplus z$ is the XOR, and $|\sigma - z|$ is the Hamming distance. We thus have, for any $\sigma \in \mathcal{H}_d$,

$$\left|\tilde{C} R_\# \mu_Q^\delta(\sigma) - e^{-\frac{1}{2}\sigma^\top Q \sigma}\right| \le 2^d e^{d\lambda_{\min}(Q)/2} e^{-\frac{1}{2}\delta^{-2}} \tag{48}$$

$$\le e^{-\frac{1}{2}\sigma^\top Q \sigma} 2^d e^{d(\lambda_{\min}(Q) + \lambda_{\max}(Q))/2} e^{-\frac{1}{2}\delta^{-2}} . \tag{49}$$

It follows that we can write $R_{\#}\mu_Q^{\delta}(\sigma)$ as

$$R_{\#}\mu_Q^{\delta}(\sigma) = C(e^{-\frac{1}{2}\sigma^\top Q\sigma} + \eta_\sigma)\ ,$$

with a relative error

$$\frac{|\eta_\sigma|}{e^{-\frac{1}{2}\sigma^\top Q\sigma}} \le e^{d(1+\frac{1}{2}(\lambda_{\min}(Q)+\lambda_{\max}(Q)))-\delta^{-2}/2} := \theta\ . \tag{50}$$

It follows that

$$\mathrm{TV}(R_{\#}\mu_Q^{\delta}, \mu_Q) = O(\theta)\ , \tag{51}$$

and thus if $\delta \ll \frac{1}{\sqrt{d}}$, we have a negligible TV approximation.

B Proofs of Section 4

B.1 Proof of Theorem 5

Proof. We denote the time-dependent score function $\nabla \log \pi_t(x)$ by $s_t(x)$. As derived in Section 2, if we initialize X_τ according to density ρ_τ and run the reverse SDE eq. (5), the density of X_t for $t \le \tau$, denoted by ρ_t, satisfies the backward PDE:

$$\partial_t \rho_t = \nabla \cdot ((x + 2s_t)\rho_t) - \Delta\rho_t. \tag{52}$$

We need to verify that ν_t satisfies the above PDE.Note that a general positive function ρ_t satisfies this PDE is equivalent to that $h_t = \log \rho_t$ satisfies the following Hamilton-Jacobi PDE

$$\partial_t h_t = d + 2\nabla \cdot s_t + \nabla h_t \cdot (x + 2s_t) - (\Delta h_t + \|\nabla h_t\|^2). \tag{53}$$

By definition, we know $h_t = \log \pi_t$ satisfies the above PDE, and we need to prove $h_t = \log \nu_t = \log \pi_t - \frac{1}{2}x^\top Q_t x + x^\top b_t + F(t)$ satisfies this PDE as well. Here $F(t)$ denotes the normalizing constant. Taking the difference between two equations, we need

$$-\frac{1}{2}x^\top \dot{Q}_t x + x^\top \dot{b}_t + \dot{F} = (-Q_t x + b_t) \cdot (x + 2s_t) + \mathrm{trace}(Q_t) + \|s_t\|^2 - \|s_t - Q_t x + b_t\|^2 \tag{54}$$

$$\Leftrightarrow \quad -\frac{1}{2}x^\top \dot{Q}_t x + x^\top \dot{b}_t + \dot{F} = x^\top(-Q_t - Q_t^\top Q_t)x + x^\top(b_t + 2Q_t^\top b_t) + \|b_t\|^2 + \mathrm{trace}(Q_t) \tag{55}$$

which can be satisfied by the ODE dynamics (12). ∎

B.2 Derivation of Solution to Equation (12)

Sanity Check for the Motivating Example. In the denoising setting, we have $Q_0 = \frac{1}{\sigma^2} I_d, b_0 = \frac{1}{\sigma^2} y$. The ODE (12) has the explicit solution

$$Q_t = \frac{e^{2t}}{1 + \sigma^2 - e^{2t}} I_d.$$

Note that this solution has a finite blowup time when $1 + \sigma^2 - e^{2t} \to 0^+$, which is exactly at $T = \frac{1}{2}\log(1 + \sigma^2)$, as derived in the main text by matching the SNR. As $t \to T$, $Q_t \to \infty I_d$,

$$\nu_t = \exp\left(f_t(x) - \frac{1}{2}x^\top Q_t x + x^\top b_t + F(t)\right) \to \mathcal{N}(Q_t^{-1} b_t, Q_t^{-1}).$$

To see the limit of $Q_t^{-1}b_t$, we only need to consider the ODE for each component since Q is diagonal. So we view the above ODE as scalar ODEs. Considering the dynamics of

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{r}{Q} = \frac{\dot{r}Q - r\dot{Q}}{Q^2} = -\frac{r}{Q}, \tag{56}$$

gives $Q_t^{-1}b_t = e^{-t}Q_0^{-1}b_0$. Therefore

$$\lim_{t\to(T)^-} Q_t^{-1}b_t = e^{-T}Q_0^{-1}b_0 = e^{-T}y,$$

which matches the initial condition derived in the main text for the denoising case. Furthermore, we can explicitly verify that the intermediate distribution of X_t by running the reverse SDE from ν_T is ν_t:

$$\begin{aligned} p(X_t|X_T = e^{-T}y) &\propto p(X_t)p(X_T = e^{-T}y|X_t) \\ &\propto \pi_t(X_t)\exp\left(-\frac{1}{2}\frac{\|e^{-T}y - e^{-(T-t)}x\|^2}{1-e^{-2(T-t)}}\right) \\ &= \pi_t(X_t)\exp\left(-\frac{1}{2}\frac{\|e^{-t}y - x\|^2}{e^{2(T-t)}-1}\right) \end{aligned} \tag{57}$$

To match the form of ν_t, we have $Q_t = \frac{1}{e^{2(T-t)}-1} = \frac{e^{2t}}{1+\sigma^2-e^{2t}}$, $Q_t^{-1}b_t = e^{-t}y$, which are the solutions of the ODE (12).

Solution to eq. (12). We recall that the observation operator $A \in \mathbb{R}^{d'\times d}$ has a general singular value decomposition form $A = U\Sigma V^\top$ with non-zero singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{d'} > 0$. By definition, we have $Q_0 = V\mathrm{diag}(\lambda_1^2/\sigma^2, \cdots, \lambda_{d'}^2/\sigma^2, 0, \cdots, 0)V^\top$. By left multiplying $V^\top$ and right multiplying V to the first ODE in (12), we can diagonalize it to scalar equations $\dot{q}_t = 2(1+q_t)q_t$ for each diagonal entry. Solving this ODE gives

$$Q_t = V\mathrm{diag}\left(\frac{e^{2t}}{1+\sigma^2/\lambda_1^2 - e^{2t}}, \cdots, \frac{e^{2t}}{1+\sigma^2/\lambda_{d'}^2 - e^{2t}}, 0, \cdots, 0\right)V^\top. \tag{58}$$

Here we explain how to solve b_t from eq. (12). We denote $V = [v_1, \cdots, v_d]$, in which v_i are eigenvectors of Q (and Q_t as well), and denote the eigenvalues of Q_t ($0 \leq t < T$) by

$$\tilde{\lambda}_i(t) = \begin{cases} \dfrac{e^{2t}}{1+\sigma^2/\lambda_i^2 - e^{2t}}, & 1 \leq i \leq d' \\ 0, & d'+1 \leq i \leq d \end{cases} \tag{59}$$

By definition, we know $\tilde{\lambda}_i$ satisfies the ODE

$$\dot{\tilde{\lambda}} = 2(1+\tilde{\lambda})\tilde{\lambda}.$$

We rewrite the solution $b_t = \sum_i^d \xi_i(t)v_i$ and aim to solve $\xi_i(t)$. From $b_0 = V(\frac{1}{\sigma^2}\Sigma^\top U^\top y)$, we have the initial condition

$$\xi_i(0) = \begin{cases} \dfrac{\lambda_i}{\sigma^2}(U^\top y)_i, & 1 \leq i \leq d' \\ 0, & d'+1 \leq i \leq d \end{cases} \tag{60}$$

Taking the inner product between v_i and both sides of the ODE $\dot{r}_t = (I + 2Q_t)b_t$, we have

$$\dot{\xi}_i(t) = (1 + 2\tilde{\lambda}_i(t))\xi_i(t).$$

Therefore, for $d' + 1 \le i \le d$, $\xi_i(t) = 0$. For $1 \le i \le d'$, same to the derivation in eq. (56), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\xi_i}{\tilde{\lambda}_i} = -\frac{\xi_i}{\tilde{\lambda}_i},$$

which gives

$$\frac{\xi_i(t)}{\tilde{\lambda}_i(t)} = e^{-t}\frac{\xi_i(0)}{\tilde{\lambda}_i(0)}, \tag{61}$$

$$\Rightarrow \left(\frac{e^{2t}}{1+\sigma^2/\lambda_i^2 - e^{2t}}\right)^{-1}\xi_i(t) = e^{-t}\frac{\sigma^2}{\lambda_i^2}\frac{\lambda_i}{\sigma^2}(U^\top y)_i, \tag{62}$$

$$\Rightarrow \xi_i(t) = \frac{e^t}{\lambda_i(1+\sigma^2/\lambda_i^2 - e^{2t})}(U^\top y)_i. \tag{63}$$

Proof of Corollary 6. Given Theorem 5, we only need to prove that sampling from

$$\nu_t = \mathsf{T}_{Q_t,b_t}\pi_t \;\propto\; \pi_t(x)\exp\left(-\frac{1}{2}x^\top Q_t x + x^\top b_t\right)$$

is equivalent to sampling from a new posterior. Taking π_t as the corresponding prior, we only need to show that the factor $\exp\left(-\frac{1}{2}x^\top Q_t x + x^\top b_t\right)$ is the likelihood of certain observation model in the form of $\tilde{y} = A_t x + w$ with $w \sim \gamma_d$. To end, we need to ensure

$$\exp\left(-\frac{1}{2}x^\top Q_t x + x^\top b_t\right) \;\propto\; \exp(-\frac{1}{2}\|A_t x - \tilde{y}\|^2)$$

Choosing A_t in the standard SVD form $A_t = \Sigma_t V^\top$ where the singular values of A_t (the diagonal entries of Σ_t) are $\frac{e^t}{(1+\sigma^2/\lambda_i^2 - e^{2t})^{1/2}}$ for $1 \le i \le d'$, the quadratic term is matched. Matching the first order term requires $b_t = A^\top\tilde{y} = V\Sigma_t^\top\tilde{y}$. Further matching the coefficients in the basis of V requires that

$$(\Sigma_t^\top\tilde{y})_i = \xi_i(t) = \frac{e^t}{\lambda_i(1+\sigma^2/\lambda_i^2 - e^{2t})}(U^\top y)_i, \quad 1 \le i \le d'.$$

It is easy to verify that $\tilde{y} = \Sigma_t' U^\top y$ with $\Sigma_t' = \mathrm{diag}\left(\frac{1}{(\sigma^2+\lambda_1^2(1-e^{2t}))^{1/2}}, \ldots, \frac{1}{(\sigma^2+\lambda_{d'}^2(1-e^{2t}))^{1/2}}\right)$ satisfies the above requirement.

Remark 19. *Also, one can directly use the backward transport equation (3) to generate samples with the probability flow ODE [SSDK+20] backward in time*

$$\mathrm{d}X_t^{\leftarrow} = (-X_t^{\leftarrow} - \nabla\log\pi_t(X_t^{\leftarrow}))\mathrm{d}t \tag{64}$$

from $X_T^{\leftarrow} \sim \gamma_d$. Combining the fact that ν_t satisfies the PDE (5) and $\Delta\nu_t = \nabla\cdot(\nabla\nu_t) = \nabla\cdot(s_t - Q_t x + b_t)$, we have that ν_t also satisfies the transport equation $\partial_t p_t = \nabla\cdot((s_t + (I+Q_t)x - b_t)p_t)$. Therefore, for any $t < T$, by initializing $X_t \sim \nu_t$ and run the reverse ODE $\mathrm{d}X_t^{\leftarrow} = (-(I+Q_t)X_t^{\leftarrow} - \nabla\log\pi_t(X_t^{\leftarrow}) + b_t)\mathrm{d}t$ then $X_0^{\leftarrow}$ also gives the desired posterior. However, we note that unlike the reverse SDE case, the corresponding transport PDE and the vector field in the reverse ODE case are different from those used in the prior data generation. As discussed below, both Q_t and b_t are singular near T. Therefore running the reverse SDE might be preferrable for better numerical stability.

B.3 Proof for the Heat Semigroup Setting

We first follow the proof of Theorem 5 to derive the ODEs for tilted transport in the heat semigroup setting. We again denote the time-dependent score function $\nabla \log \pi_t(x)$ by $s_t(x)$. As derived in Section 2, if we initialize X_τ according to density ρ_τ and run the reverse SDE eq. (6), the density of X_t for $t \le \tau$, denoted by ρ_t, satisfies the backward PDE:

$$\partial_t \rho_t = \nabla \cdot (2 s_t \rho_t) - \Delta \rho_t. \tag{65}$$

We again wish to find Q_t, b_t such that

$$\nu_t := \mathsf{T}_{Q_t, b_t} \pi_t$$

satisfies the above PDE. Note that ρ_t satisfies this PDE is equivalent to that $h_t = \log \rho_t$ satisfies the following Hamilton-Jacobi PDE

$$\partial_t h_t = 2\nabla \cdot s_t + 2\nabla h_t \cdot s_t - (\Delta h_t + \|\nabla h_t\|^2). \tag{66}$$

By definition, we know $h_t = \log \pi_t$ satisfies the above PDE, and we need to find $h_t = \log \nu_t = \log \pi_t - \frac{1}{2} x^\top Q_t x + x^\top b_t + F(t)$ satisfying this PDE as well. Here $F(t)$ denotes the normalizing constant. Taking the difference between two equations, we need

$$-\frac{1}{2} x^\top \dot{Q}_t x + x^\top \dot{b}_t + \dot{F} = 2(-Q_t x + b_t) \cdot s_t + \operatorname{trace}(Q_t) + \|s_t\|^2 - \|s_t - Q_t x + b_t\|^2 \tag{67}$$

$$\Leftrightarrow \quad -\frac{1}{2} x^\top \dot{Q}_t x + x^\top \dot{b}_t + \dot{F} = -x^\top (Q_t^\top Q_t) x + 2 x^\top (Q_t^\top b_t) + \|b_t\|^2 + \operatorname{trace}(Q_t) \tag{68}$$

which leads us to the ODE (13). Therefore, the eigenvalues of Q_t evolve according to the ODE $\dot{q} = 2q^2$, whose solution has the form

$$\lambda_i(t) = \frac{\lambda_i}{1 - 2t\lambda_i}. \tag{69}$$

So the ODE will blowup at time $T = \|Q\|^{-1}/2$. The full solution of Q_t and b_t can be solved similarly as in Appendix B.2.

C Proofs of Section 5

C.1 Proof of Proposition 10

Proof of Proposition 10. By definition, we need to show

$$-\lambda_{\min}(Q_T) + \sup_x \lambda_{\max}(\nabla^2 \log \pi_T(x)) < 0. \tag{70}$$

From the argument in the main text, we know $T = \frac{1}{2} \log(1 + \lambda_{\max}(Q)^{-1})$, and thus

$$\lambda_{\min}(Q_T) = \frac{1 + \lambda_{\max}(Q)^{-1}}{\lambda_{\min}(Q)^{-1} - \lambda_{\max}(Q)^{-1}}.$$

From Corollary 21, we have

$$\sup_x \lambda_{\max}(\nabla^2 \log \pi_T(x)) \le (1 + \|Q\|) \left(\|Q\| \chi_{\|Q\|}(\pi) - 1 \right). \tag{71}$$

Let $m = \lambda_{\min}(Q)$. Therefore, we can guarantee that ν_T is strongly log-concave if

$$\frac{1 + \|Q\|^{-1}}{m^{-1} - \|Q\|^{-1}} > (1 + \|Q\|) \left(\|Q\| \chi_{\|Q\|}(\pi) - 1 \right), \tag{72}$$

or equivalently

$$\chi_{\|Q\|}(\pi) < \|Q\|^{-1} \frac{\kappa^2(A)}{\kappa^2(A) - 1} \,. \tag{73}$$

■

Lemma 20 (Hessian of Gaussian Mixture Potential)**.** *Let $\pi = \mu * \gamma_\Sigma$ be a Gaussian mixture. Then* $\nabla^2 \log \pi(x) = \Sigma^{-1} \left(\mathrm{Cov} \left[\mathsf{T}_{\Sigma^{-1}, \Sigma^{-1} x} \mu \right] \Sigma^{-1} - I \right)$.

Proof. Let us first compute the score $\nabla \log \pi$. By definition we have

$$\nabla \log \pi(x) = -\Sigma^{-1} \left(x - \frac{\int y \mu(y) e^{-\frac{1}{2}(x-y)^\top \Sigma^{-1}(x-y)} dy}{\int \mu(y) e^{-\frac{1}{2}(x-y)^\top \Sigma^{-1}(x-y)} dy} \right) \tag{74}$$

$$= -\Sigma^{-1}(x - \mathbb{E}\left[\mathsf{T}_{\Sigma^{-1}, \Sigma^{-1} x} \mu \right]) \,, \tag{75}$$

and thus

$$\nabla^2 \log \pi(x) = \Sigma^{-1} \left(\mathrm{Cov} \left[\mathsf{T}_{\Sigma^{-1}, \Sigma^{-1} x} \mu \right] \Sigma^{-1} - I \right) \,, \tag{76}$$

where we defined $\mathrm{Cov}[\mu] = \mathbb{E}_{x \sim \mu}[x x^\top] - (\mathbb{E}_{x \sim \mu} x)(\mathbb{E}_{x \sim \mu} x)^\top$. ■

Corollary 21. *In particular, we have*

$$\sup_x \lambda_{\max}(\nabla^2 \log \pi_T(x)) \leq (1 + \|Q\|) \left(\|Q\| \chi_{\|Q\|}(\pi) - 1 \right) \,. \tag{77}$$

Proof. From Lemma 20 and $\pi_t = \mathsf{C}_{1-e^{-2t}}(\mathsf{D}_{e^t} \pi) = \mathsf{D}_{e^t} \pi * \gamma_{1-e^{-2t}}$, we directly have

$$\nabla^2 \log \pi_t(x) = (1 - e^{-2t})^{-1} \left((1 - e^{-2t})^{-1} \mathrm{Cov} \left[\mathsf{T}_{(1-e^{-2t})^{-1}, (1-e^{-2t})^{-1} x} (\mathsf{D}_{e^t} \pi) \right] - I \right) \,. \tag{78}$$

Now, using the commutation property between the isotropic tilt and the dilation $\mathsf{D}_\alpha \mathsf{T}_{\eta, \theta} = \mathsf{T}_{\alpha^2 \eta, \alpha \theta} \mathsf{D}_\alpha$, we have

$$\mathrm{Cov} \left[\mathsf{T}_{(1-e^{-2t})^{-1}, (1-e^{-2t})^{-1} x} (\mathsf{D}_{e^t} \pi) \right] = \mathrm{Cov} \left[\mathsf{D}_{e^t} \mathsf{T}_{(e^{2t}-1)^{-1}, e^{-t}(1-e^{-2t})^{-1} x} \pi \right] \tag{79}$$

$$= e^{-2t} \mathrm{Cov} \left[\mathsf{T}_{(e^{2t}-1)^{-1}, e^{-t}(1-e^{-2t})^{-1} x} \pi \right] \,, \tag{80}$$

and therefore

$$\sup_x \left\| \mathrm{Cov} \left[\mathsf{T}_{(1-e^{-2t})^{-1}, (1-e^{-2t})^{-1} x} (\mathsf{D}_{e^t} \pi) \right] \right\| \leq e^{-2t} \chi_{(e^{2t}-1)^{-1}}(\pi) \,. \tag{81}$$

Using that $e^{2T} - 1 = \|Q\|^{-1}$, we thus obtain

$$\sup_x \lambda_{\max}(\nabla^2 \log \pi_T(x)) \leq (1 + \|Q\|) \left(\|Q\| \chi_{\|Q\|}(\pi) - 1 \right) \,. \tag{82}$$

■

Lemma 22 (Isotropic Tilt of a Gaussian Mixture)**.** *If $\pi = \mu * \gamma_\delta$, then*

$$\mathsf{T}_{tI, z} \pi = \tilde{\mu} * \gamma_{\sigma^2} \,, \tag{83}$$

where $\sigma^{-2} = \delta^{-1} + t$ and

$$\tilde{\mu}(\tilde{y}) \propto \mu((\sigma^{-2} \tilde{y} - z)\delta) e^{\frac{1}{2}(\sigma^{-2} \|\tilde{y}\|^2 - \delta \|\sigma^{-2} \tilde{y} - z\|^2)} \,. \tag{84}$$

Proof. By definition, we have

$$\mathsf{T}_{tI,z}\pi \propto \int d\mu(y) e^{-\frac{1}{2}t\|x\|^2 + x\cdot z - \frac{1}{2}\delta^{-1}\|x-y\|^2} .$$

By expressing

$$-\frac{1}{2}t\|x\|^2 + x\cdot z - \frac{1}{2}\delta^{-1}\|x-y\|^2 = -\frac{1}{2}\sigma^{-2}\|x-\tilde{y}\|^2 + C$$

we have

$$\sigma^{-2} = \delta^{-1} + t \,, \tag{85}$$

$$\tilde{y} = \frac{\delta^{-1}y + z}{\delta^{-1} + t} \,, \tag{86}$$

$$C = \frac{1}{2}\left[\sigma^{-2}\|\tilde{y}\|^2 - \delta^{-1}\|y\|^2\right] \,, \tag{87}$$

which gives the desired result after performing the affine change of variables from y to $\tilde{y}$. ∎

C.2 Proof of Proposition 11

Proof. Applying [MCJ+19, Lemma 1], under our assumptions, we can approximate the potential f (recall that $\pi = e^{-f}$) with a $m/2$-strongly-convex potential $\tilde{f}$, satisfying $\mathrm{osc}[f - \tilde{f}] \le 16LR^2$, where the oscillation of a function is defined as $\mathrm{osc}(g) := \sup_x g(x) - \inf_x g(x)$.

Now, for any x, we approximate the potential of the quadratic tilt

$$d\mathsf{T}_{t,tx}\pi(y) = e^{-\frac{t}{2}\|y-x\|^2 - f(y) - \log Z_x}\mathrm{d}y$$

by

$$\tilde{f}_x(y) := \tilde{f}(y) + \frac{t}{2}\|y-x\|^2 + \log Z_x .$$

We verify that $\tilde{f}_x$ is $(m/2+t)$-strongly convex, and that

$$\mathrm{osc}\left[\tilde{f}_x - \log \mathsf{T}_{t,tx}\pi\right] = \mathrm{osc}\left[\tilde{f} - f\right] \le 16LR^2 .$$

By the Holley-Strook perturbation principle and the Barky-Emery criterion, we therefore have that, under our assumptions, the measures $\mathsf{T}_{t,tx}\pi$ satisfy a Poincaré inequality *uniformly* over x, with constant $C := (m/2+t)e^{-16LR^2}$.

Finally, observe

$$\|\mathrm{Cov}[\mathsf{T}_{t,tx}\pi]\| = \sup_{\|\theta\|=1} \mathrm{Var}_{\mathsf{T}_{t,tx}\pi}[\theta^\top Y] \tag{88}$$

$$\le C^{-1}\mathbb{E}_{\mathsf{T}_{t,tx}\pi}[\|\theta\|^2] = C^{-1} \,, \tag{89}$$

showing that $\chi_t(\pi) \le (m/2+t)^{-1}e^{16LR^2}$, as claimed. ∎

C.3 Proof of Proposition 12

Proof. Let $\pi \in \mathcal{P}(\mathbb{R})$ be of the form

$$\pi = \sum_{j\ge 0}\alpha_j\delta_{y_j} \,, \text{ with } y_j = \sum_{i\le j} b_i \,,$$

where $0 \le b_1 \le b_2 \le \dots$ is an non-decreasing sequence with $\lim_{j\to\infty} b_j = \infty$, and $\alpha_{2k} = \alpha_{2k+1}$, $b_{2k} = b_{2k+1}$ for all $k \in \mathbb{N}$, satisfying

$$\sum_j \alpha_j = 1 \,, \quad \sum_j \alpha_j j^2 b_j^2 < \infty \,. \tag{90}$$

Note that $\mathbb{E}_\pi[y^2] \le \sum_j \alpha_j j^2 b_j^2$ and thus $\pi \in \mathcal{P}_2(\mathbb{R})$.

We set $x = \frac{1}{2}(y_{2k} + y_{2k+1})$ and consider the tilted measure $\mathsf{T}_{t,tx}\pi$: it is also atomic and supported in $\{y_j\}_j$, with weights

$$\tilde{\alpha}_j = \frac{\alpha_j e^{-\frac{1}{2}t(y_j - x)^2}}{\sum_i \alpha_i e^{-\frac{1}{2}t(y_i - x)^2}} \,.$$

Let us now lower bound the weights $\tilde{\alpha}_{2k}$ and $\tilde{\alpha}_{2k+1}$. We have

$$\sum_i \alpha_i e^{-\frac{1}{2}t(y_i - x)^2} = (\alpha_{2k} + \alpha_{2k+1}) e^{-\frac{t}{8}b_{2k+1}^2} + \sum_{i \neq 2k, 2k+1} \alpha_i e^{-\frac{1}{2}t(y_i - x)^2} \tag{91}$$

$$\le e^{-\frac{t}{8}b_{2k+1}^2} 2\alpha_{2k} + e^{-\frac{t}{2}(b_{2k+1}/2 + b_{2k})^2} \tag{92}$$

$$= e^{-\frac{t}{8}b_{2k}^2} \left(2\alpha_{2k} + e^{-tb_{2k}^2}\right) \,, \tag{93}$$

thus

$$\tilde{\alpha}_{2k} \ge \frac{\alpha_{2k} e^{-\frac{t}{8}b_{2k}^2}}{e^{-\frac{t}{8}b_{2k}^2}\left(2\alpha_{2k} + e^{-tb_{2k}^2}\right)} \tag{94}$$

$$= \frac{1}{2 + \alpha_{2k}^{-1} e^{-tb_{2k}^2}} \,, \tag{95}$$

and clearly $\tilde{\alpha}_{2k+1} = \tilde{\alpha}_{2k}$.

Now, observe that

$$\mathrm{Var}(\mathsf{T}_{t,tx}\pi) \ge \frac{\tilde{\alpha}_{2k} + \tilde{\alpha}_{2k+1}}{4} b_{2k}^2 \,. \tag{96}$$

Therefore, the susceptibility $\chi_t(\pi)$ satisfies

$$\chi_t(\pi) \ge \sup_k \frac{b_{2k}^2}{2(2 + \alpha_{2k}^{-1} e^{-tb_{2k}^2})} \,. \tag{97}$$

We identify two regimes of blowup for $\chi_t(\pi)$: for any $t > 0$ when π has heavy tails, and for $t \ge \lambda > 0$ for subgaussian tails.

Indeed, for the first setting, let $\alpha_{2k} \simeq k^{-s}$, $b_{2k} \simeq k^r$, with $s > 0$, $r > 0$ and $s - 2r > 3$. Then (90) can be satisfied, and we verify that for any $t > 0$ the RHS of (97) is unbounded. Finally, by choosing $\lambda > 0$, setting $b_{2k}^2 = k$ and $\alpha_{2k} = e^{-\lambda k}$ for $k > k^*$ with a large enough $k^* \in \mathbb{N}$ and appropriate α_{2k} for $k \le k^*$, we can have the condition (90) satisfied. We verify that for this choice of parameters, π has subgaussian tails; Indeed, for sufficiently large z,

$$\mathbb{P}_\pi(Y \ge z) \lesssim e^{-\lambda z^2} \,. \tag{98}$$

Meanwhile, we have a blowup for the susceptibility as soon as $t > \lambda$:

$$\chi_t(\pi) \ge \sup_{k > k^*} \frac{k}{2(2 + e^{(\lambda - t)k})} = \infty \,. \tag{99}$$

■

C.4 Proof of Proposition 13

Proof of Proposition 13. If μ is a product measure, we observe that the isotropic tilt $\mathsf{T}_t\mu$ is also a product measure, and therefore its covariance is diagonal.

If $\pi = e^f$ with ∇f L-Lipschitz, we apply the Brascamp-Lieb inequality to $\mathsf{T}_{t,tx}\pi$:

Theorem 23 (Brascamp-Lieb Inequality, [BL76])**.** *If π is a strongly-log-concave measure on $\mathbb{R}^d$, ie of the form $\pi = e^{-f}$ with $\nabla^2 f(x) \geq \alpha \mathrm{Id}$ for all $x \in \mathbb{R}^d$, then $\|\mathrm{Cov}(\pi)\| \leq \alpha^{-1}$.*

Observe that $\mathsf{T}_{t,tx}\pi(y) \propto e^{-t/2\|y\|^2 + tx.y + f}$, and thus

$$\nabla^2 \left(-\log \mathsf{T}_{t,tx}\pi\right) \geq (t - L)\mathrm{Id}\,,$$

and hence $\chi_t(\pi) \leq \frac{1}{t-L}$ provided $t > L$.

■

Proof of Example 14. The first example is immediate, after observing that $\mathsf{T}_t\gamma$ is a Gaussian of variance $(1+t)^{-1}$. For the Gaussian mixture example, we observe from Lemma 22 that $\mathsf{T}_t(\mu * \gamma_\delta)$ is a Gaussian mixture of variance $(t + \delta^{-1})^{-1}$, where the mixture distribution is supported in a ball of radius $R\frac{\delta^{-1}}{\delta^{-1}+t}$. Moreover, the covariance of a homogeneous mixture of the form $\mu * \gamma_\Sigma$ is $\Sigma + \mathrm{Cov}(\mu)$.

Finally, by the Proposition 13, if π is the uniform measure on the hypercube, then $\chi_t(\pi) = \chi_t(\frac{1}{2}(\delta_{-1} + \delta_{+1})) = 1$.

■

C.5 Proof of Corollary 17

Proof of Corollary 17. We plug the susceptibility $\chi_t(\pi) = \chi_t(\mu * \gamma_\delta) \leq \left(\frac{R}{1+\delta t}\right)^2 + \frac{\delta}{1+\delta t}$ from Example 14 (ii) into eq. (26) to get (we use κ to denote $\kappa(A)$ for simplicity)

$$\|Q\|^{-1}\frac{\kappa^2}{(\kappa^2-1)} > \left(\frac{R}{1+\delta\|Q\|}\right)^2 + \frac{\delta}{1+\delta\|Q\|} \tag{100}$$

$$\Leftrightarrow (1+\delta\|Q\|)\left(\frac{((1+\delta\|Q\|))\kappa^2}{\|Q\|(\kappa^2-1)} - \delta\right) > R^2 \tag{101}$$

$$\Leftrightarrow \frac{((1+\delta\|Q\|)(\kappa^2+\delta\|Q\|)}{\|Q\|(\kappa^2-1)} > R^2 \tag{102}$$

$$\Leftrightarrow \frac{(1+\delta\mathrm{SNR}^2)(\delta\kappa^2+\mathrm{SNR}^{-2})}{\kappa^2-1} > R^2\,. \tag{103}$$

■

C.6 Importance Sampling

Proof of Proposition 15. We wish to lower bound $\mathrm{KL}(\nu||\pi)$ with tools of functional inequalities for the concentration of measure. By the celebrated work [OV00, BGL01] that the log-Sobolev inequality implies the Talagrand transport-entropy inequality, we have

$$\mathrm{KL}(\nu||\pi) \geq \frac{\mathrm{LSI}(\nu)W(\nu,\pi)^2}{2}. \tag{104}$$

Here $W(\nu,\pi)$ denotes the Wassertain distance between ν and π:

$$W(\nu,\pi) = \sqrt{\inf_{\gamma\in\Gamma(\nu,\pi)} \|x-y\|^2 \mathrm{d}\gamma(x,y)}\,,$$

where $\Gamma(\nu, \pi)$ denotes the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals ν and π.

First by Equation (23), we have

$$\nabla^2 (-\log \nu) \geq (\mathrm{SNR} - L)\mathrm{Id}\,, \tag{105}$$

which gives

$$\mathrm{LSI}(\nu) \geq (\mathrm{SNR} - L)\,, \tag{106}$$

by Bakry-Emery criterion [BÉ85]. Furthermore, we have the lower bound for the Wasserstain distance [Gel90]

$$\begin{aligned} W^2(\nu, \pi) &\geq \|\mathrm{mean}(\nu) - \mathrm{mean}(\pi)\|^2 + \mathrm{trace}(\mathrm{Cov}(\nu) + \mathrm{Cov}(\pi) - 2(\mathrm{Cov}(\pi)^{\frac{1}{2}}\mathrm{Cov}(\nu)\mathrm{Cov}(\pi)^{\frac{1}{2}})^{\frac{1}{2}}) && (107)\\ &\geq \mathrm{trace}(\mathrm{Cov}(\nu) + I_d - 2\mathrm{Cov}(\nu)^{\frac{1}{2}}) && (108)\\ &= \mathrm{trace}(\mathrm{Diag}(\mathrm{Cov}(\nu)) + I_d - 2\mathrm{Diag}(\mathrm{Cov}(\nu))^{\frac{1}{2}}) && (109)\\ &= \sum_{i=1}^{d}((1 - \mathrm{Std}(\nu)_i)^2)\,, && (110) \end{aligned}$$

where the second-to-last equality comes from the fact that the trace remains unchanged under orthogonal transformation. By (105) and Brascamp-Lieb Inequality (Theorem 23), we have $\mathrm{Std}(\nu)_i \leq \sqrt{\frac{1}{\mathrm{SNR}-L}}$. With the condition $\mathrm{SNR} - L > 2$,

$$W^2(\nu, \pi) \geq \left(1 - \sqrt{\frac{1}{2}}\right)^2 d\,. \tag{111}$$

Collecting eqs. (104), (106) and (111), we obtain our final estimate

$$\mathrm{KL}(\nu||\pi) \geq O(d \cdot \mathrm{SNR})\,. \tag{112}$$

■

C.7 Stability Analysis

The proof is similar to that in [SDME21]. Here we provide the proof in our posterior sampling context for completeness.

Proof of Proposition 16. Consider the following two reverse dynamics needed for the error estimate: one is based on the exact score and starts from the exact boosted posterior

$$\mathrm{d}X_\tau = (-X_\tau - 2\nabla \log \pi_\tau(X_\tau))\mathrm{d}\tau + \sqrt{2}\mathrm{d}W_\tau, \quad X_t \sim \nu_t, \tag{113}$$

and another one is based on the approximate score and starts from the approximation to the boosted posterior

$$\mathrm{d}\tilde{X}_\tau = (-\tilde{X}_\tau - 2s_\theta(\tilde{X}_\tau, \tau))\mathrm{d}\tau + \sqrt{2}\mathrm{d}W_\tau, \quad \tilde{X}_t \sim q_t, \tag{114}$$

Note that these two dynamics are defined backwardly for $\tau \in [0, t]$ and we drop the superscript $^{\leftarrow}$ for notation simplicity. We denote the path measure of $\{X_\tau\}_{\tau\in[0,t]}$ and $\{\tilde{X}_\tau\}_{\tau\in[0,t]}$ by $\boldsymbol{\nu}$ and $\boldsymbol{q}$, respectively. Then ν and q_0 are the marginal distributions of the two path measures at $t = 0$. By data processing inequality and chain rule of KL divergence, we have

$$\begin{aligned} \mathrm{KL}(\nu||q_0) &\leq \mathrm{KL}(\boldsymbol{\nu}||\boldsymbol{q}) && (115)\\ &\leq \mathrm{KL}(\nu_\tau||q_\tau) + \mathbb{E}_{z\sim\nu_\tau}\mathrm{KL}(\boldsymbol{\nu}(\cdot|X_t = z)||\boldsymbol{q}(\cdot|\tilde{X}_t = z)) && (116) \end{aligned}$$

Given the Novikov's condition, we can apply the Girsanov theorem [Oks13] to eq. (113)114 to compute the second term above

$$\mathbb{E}_{z\sim\nu_\tau}\mathrm{KL}(\boldsymbol{\nu}(\cdot|X_t=z)||\boldsymbol{q}(\cdot|\tilde{X}_t=z)) \tag{117}$$

$$\leq -\mathbb{E}_{\boldsymbol{\nu}}\left[\log\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}\boldsymbol{\nu}}\right] \tag{118}$$

$$=\mathbb{E}_{\boldsymbol{\nu}}\left[2\int_0^t(\nabla\log\pi_\tau(x)-s_\theta(x,\tau))\mathrm{d}W_\tau+\int_0^t\|\nabla\log\pi_\tau(x)-s_\theta(x,\tau)\|^2\mathrm{d}\tau\right] \tag{119}$$

$$=\mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^t\|\nabla\log\pi_\tau(x)-s_\theta(x,\tau)\|^2\mathrm{d}\tau\right] \tag{120}$$

$$=\int_0^t\mathbb{E}_{\nu_\tau}\|\nabla\log\pi_\tau(x)-s_\theta(x,\tau)\|^2\mathrm{d}\tau\,. \tag{121}$$

■

D Experimental Details

D.1 Gaussian Mixture Models

For a given dimension d with $d \bmod 2 = 0$, we consider prior data a mixture of 25 Gaussian distributions, the same as considered in [CJCM24]. The Gaussian distribution has mean $(8i, 8j, \cdots, 8i, 8j) \in \mathbb{R}^d$ for $(i, j) \in \{-2, -1, 0, 1, 2\}^2$ and unit variance. Each (unnormalized) mixture weight is independently drawn according to a χ^2 distribution.

For the measurement model considered in Figure 4, we generate A in the following way. We first sample a $d \times d$ matrix with each entry sampled from the standard normal and compute its SVD to get U and V for A. The singular value is given by $[1, \cdots, 1/20]$ where each component in between is independently sampled from $\mathrm{Unif}([1/20, 1])$ such that the condition number of A is 20. The observation noise is then determined by SNR. For the measurement model considered in Table 1, the matrix U and V for the SVD form of A is the same to the above. Each singular value in S is independently sampled from $\mathrm{Unif}([0, 1])$, and σ is sampled from $\mathrm{Unif}([0.2 \max S, \max S])$.

For all the experiments we run the boosted posterior from $T - 0.01$ such that the ODE solution Q, b is well-defined. We use BlackJAX [CCLL24] to implement the No-U-turn sampler.

Besides results reported in Figure 4, we further test tilted transport when $d' < d$. In this setting, $\lambda_{\min}(Q_t)$ remains zero but the signal corresponding to the non-zero eigenvalues still gets enhanced. Therefore, although it becomes more difficult for ν_T to be strongly log-concave, the tilted transport can still make the new posterior easier to sample even if it is not strongly log-concave yet. As shown in Table 1, when $d' = 0.9d$, 10% percent eigenvalues of Q_t are zero, our tilted transport technique still reduces the statistical distance of the posterior samples significantly. We also consider an even more challenging case where $d' = 1$ such that the target posterior is still heavlily multimodal (as visualized in the 2D example in Figure 2). In this case, LMC suffers from the local maxima of the potential and thus cannot explore the multimodal distribution efficiently. We use the No-U-turn sampler[HG14], a Hamilton Monte Carlo (HMC) method, as the baseline method, which can move among different modes more efficiently than Langevin. We find that the tilted transport technique can still boost the performance of HMC in this challenging setting. We also verify Theorem 5 by sampling from the boosted posterior directly from its analytical formula and running the reverse SDE, and the obtained samples approximate the target posterior well, as reported in Table 1.

Table 1: Sliced Wasserstein distance for Gaussian mixture prior for degenerate case

	$d' = 0.9d$			$d' = 1$		
d	Langevin	Boosted Langevin	Analytic Boost	HMC	Boosted HMC	Analytic Boost
20	4.21 ± 1.87	$\mathbf{2.32 \pm 2.42}$	*0.02 ± 0.00*	1.33 ± 1.02	$\mathbf{1.11 \pm 0.83}$	*0.12 ± 0.07*
40	4.09 ± 2.02	$\mathbf{2.45 \pm 1.79}$	*0.02 ± 0.00*	2.04 ± 1.26	$\mathbf{1.81 \pm 1.03}$	*0.13 ± 0.07*
80	4.40 ± 2.31	$\mathbf{2.75 \pm 2.10}$	*0.02 ± 0.00*	2.98 ± 2.15	$\mathbf{2.77 \pm 2.32}$	*0.11 ± 0.06*

D.2 Imaging Problems

We perform four inverse tasks on the Flickr-Faces-HQ Dataset (FFHQ) [KLA19] to demonstrate the application of the tilted transport technique on imaging data as a proof of concept. To apply the proposed tilted transport technique to these problems, we still need to select a baseline method for sampling from the boosted posterior ν_T. In the case of ill-conditioned problems, sampling ν_T may still be challenging for principled algorithms like LMC, and we still need to rely on heuristic methods for imaging tasks. However, as noted in the introduction, most existing heuristic methods primarily facilitate conditional generation based on the measurement, lacking principled guarantees for posterior sampling. Consequently, we lack a principled interpretation for enhancing these methods with tilted transport. Nevertheless, we can still experiment with such methods as a proof of concept. We chose Diffusion Model Based Posterior Sampling (DMPS) [MK22] as the baseline method for the following reasons: The main assumption of DMPS in approximating the time-dependent conditional score is that the prior π is uninformative (flat) with respect to X_t, such that $p(X_0|X_t) \propto p(X_t|X_0)$. This assumption only holds approximately in early phases of the forward diffusion, and hopefully a higher SNR provided by tilted transport makes the effect of this approximation error smaller.

We conducted four tasks: (a) denoising; (b) inpainting with random masks from [SCC+22]; (c) 4× super-resolution; and (d) deblurring using a Gaussian kernel. Our algorithm was implemented using the NVIDIA codebase [MSKV23] with 1000 diffusion steps for posterior sampling, and utilized the score function from a pretrained diffusion model [CKJ+21]. Similar to our Gaussian mixture model experiments where we adjusted the timing for the boosted posterior to avoid the singularity of Q_t, we shifted 6 - 10 timesteps for setting the boosted posterior. Experiments show that the final performance is robust with respect to the number of shifted steps. Figure 6 showcases examples from the inpainting task, demonstrating how tilted transport enhances the baseline DMPS method. Additionally, we report various sample statistics including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Learned Perceptual Patch Similarity (LPIPS). However, it is important to note that while these statistics assess the quality of prior data generation, they may not accurately reflect the quality of posterior samples.

Table 2: Performance for tasks on FFHQ Dataset.

Task	**Denoising**		**Inpainting**		**Super-resolution**		**Deblur**	
Metrics	DMPS	Boost	DMPS	Boost	DMPS	Boost	DMPS	Boost
PSNR(dB) ↑	32.153	**32.350**	22.458	**23.312**	26.761	**26.899**	29.088	**29.155**
SSIM ↑	0.886	0.886	0.786	**0.800**	0.760	**0.754**	0.815	0.815
LPIPS ↓	0.060	**0.039**	0.131	**0.098**	0.129	**0.109**	0.098	**0.094**

Figure 6: Examples for inpainting with random masks over FFHQ dataset