Provable Posterior Sampling with Denoising Oracles via Tilted Transport

Joan Bruna [1] Jiequn Han [2]

[1]NYU [2]Flatiron Institute

Central Question of Interest

Given a prior distribution π on $\mathbb{R}^d$, we assume known its *Denoising Oracle*: $\mathsf{DO}_\pi(y,t) = \mathbb{E}[X|Y=y]$, where $X \sim \pi$ and $Y = X + tZ$, $Z \sim \gamma_d = \mathcal{N}(0, I_d)$.

By Tweedie's formula, DO_π is equivalent to score along Ornstein-Uhlenbeck (or Heat) semigroup.

$$\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \qquad X_0 \sim \pi \qquad \text{(forwardSDE)}$$
$$\mathrm{d}X_t^{\leftarrow} = (-X_t^{\leftarrow} - 2\nabla \log \pi_t(X_t^{\leftarrow}))\mathrm{d}t + \sqrt{2}\mathrm{d}\overline{W}_t, \qquad X_T^{\leftarrow} \sim \gamma_d\,. \qquad \text{(reverseSDE)}$$

Given linear observations

$$y = Ax + \sigma w, \quad \sigma > 0, x \sim \pi, w \sim \gamma_{d'},$$

we aim to solve the Bayesian inverse problem by provably sampling the (possibly multimodal) posterior distribution

$$\nu(x) := p(x|y) \propto \pi(x) p(y|x) \propto \pi(x) \exp\left\{-\frac{1}{2\sigma^2}\|Ax - y\|^2\right\}.$$

Notation: For $Q \succeq 0$ in $\mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$, the quadratic tilt of π is the measure $\mathsf{T}_{Q,b}\pi \ll \pi$ with $\frac{\mathrm{d}\mathsf{T}_{Q,b}\pi}{\mathrm{d}\pi}(x) \propto \exp\left\{-\frac{1}{2}x^\top Qx + x^\top b\right\}$. So we aim to sample

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

Background

- If $\lambda_{\min}(Q) \gg 1$, ν becomes strongly log-concave, allowing fast relaxation of Langevin dynamics (Logarithmic Sobolev Inequality and Bakry-Emery criterion).
- If $\lambda_{\max}(Q) \ll 1$, $\nu \approx \pi$, so samples from π can be efficiently perturbed into samples from ν via *importance sampling*.
- If A is unitary, $Q = \sigma^{-2}\mathrm{Id}$, reducing the problem to isotropic Gaussian denoising, seems compatible with the denoising oracle (see next block).

Two key problem parameters: $\mathrm{SNR} := \lambda_{\min}(Q) = \lambda_{\min}(A)^2/\sigma^2$ and $\kappa(A) := \lambda_{\max}(A)/\lambda_{\min}(A)$.

Denoising as a Motivating Example

When the task is denoising

$$y = x + \sigma w,$$

the observation has a similar structure to the forward process

$$X_s \overset{d}{=} e^{-s}X_0 + (1 - e^{-2s})w.$$

By defining

$$T^* = \frac{1}{2}\log(1+\sigma^2), \quad \tilde{y} = e^{-T^*}y,$$

we have

$$(x, \tilde{y}) \overset{d}{=} (X_0, X_{T^*}).$$

Therefore, sampling $p(x|y)$ is equivalent to $p(X_0|X_{T^*})$, which can be achieved by the following:

1. Initialize $X_{T^*}^{\leftarrow} = e^{-T^*}y$
2. Run the original reverse SDE from T^* to 0 to get the desired sample

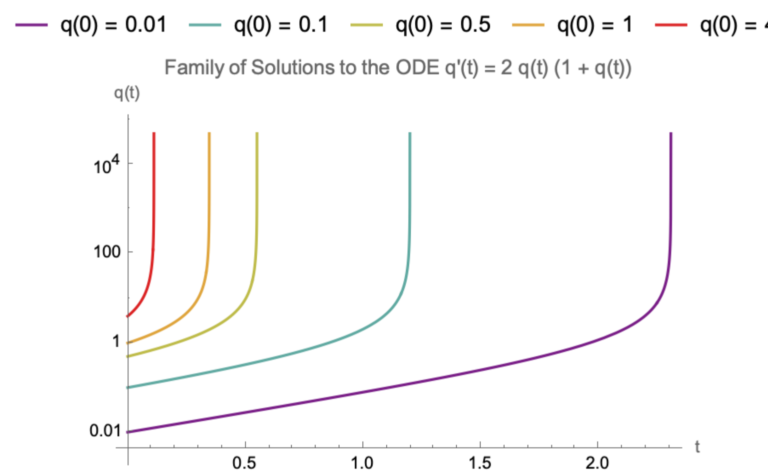
General Cases via Tilted Transport

We consider a one-parameter family of distributions ν_t of the form

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

with π_t denoting the density of X_t in the forward process and Q_t, b_t satisfying the first-order ODE:

$$\begin{cases} \dot{Q}_t = 2(I + Q_t)Q_t\,, & Q_0 = Q\,, \\ \dot{b}_t = (I + 2Q_t)b_t\,, & b_0 = b\,. \end{cases} \qquad (1)$$



Theorem 1 (Tilted Transport) *Assume $t < T^* := \frac{1}{2}\log(1 + \lambda_{\max}(Q)^{-1})$ such that the ODE (1) is well-defined on $[0, t]$. By initializing $X_t^{\leftarrow} \sim \nu_t$ and running reverseSDE from t to 0, we have $X_s^{\leftarrow} \sim \nu_s$ for $s \in [0, t]$; specifically, $X_0^{\leftarrow}$ gives a sample from the desired posterior.*

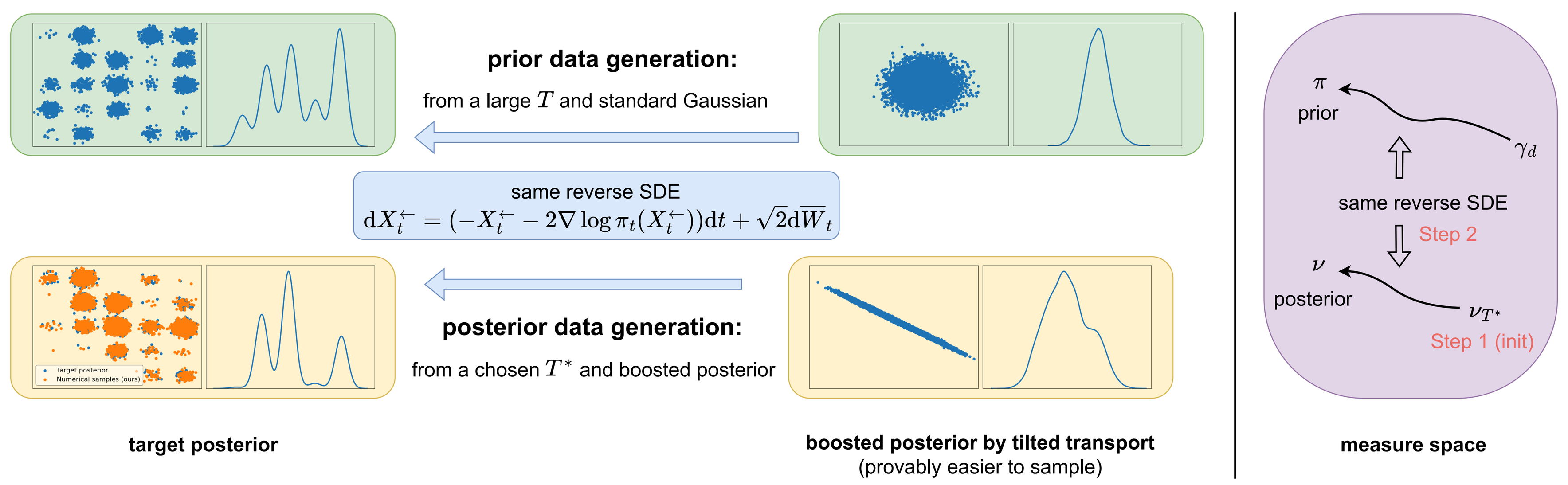
Key takeaway: The same reverse SDE allows us to move samples along ν_t backward, and ν_t becomes easier to sample from as t increases since

$$\nu_t(x) = \underbrace{\pi_t(x)}_{\text{easier prior}} \underbrace{\exp\left\{-\frac{1}{2}x^\top Q_t x + x^\top b_t\right\}}_{\text{easier likelihood}}$$

Resulting Numerical Algorithm

Given a baseline sampling algorithm **Alg** (e.g. Langevin Diffusion) and starting time $\tilde{T} = T^* - \epsilon$ (for stable ODE solutions), the tilted transport works in two steps:

1. Use the baseline sampling algorithm **Alg** to sample $X_{\tilde{T}}^{\leftarrow}$ from $\pi_{\tilde{T}}(x)\exp\left\{-\frac{1}{2}x^\top Q_{\tilde{T}}x + x^\top b_{\tilde{T}}\right\}$
2. Run the original reverse SDE from $\tilde{T}$ to 0 to get the desired sample



Remark: A similar two-step approach (marginal sampling + conditional sampling) was recently proposed for posterior sampling in sparse linear regression [Montanari and Wu, 2024].

Provable Sampling

Definition: (known as the susceptibility in the literature of stochastic localization and Polchinsky renormalisation group)

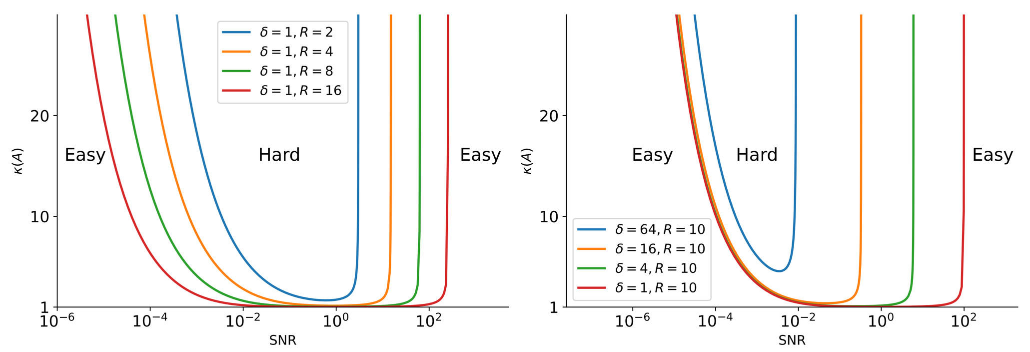
$$\chi_t(\pi) := \sup_{x\in\mathbb{R}^d} \|\mathrm{Cov}[\mathsf{T}_{tI,tx}\pi]\|,$$

Theorem 2 (Strong Log-Concavity of ν_{T^*}) *ν_{T^*} is strongly log-concave if*

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

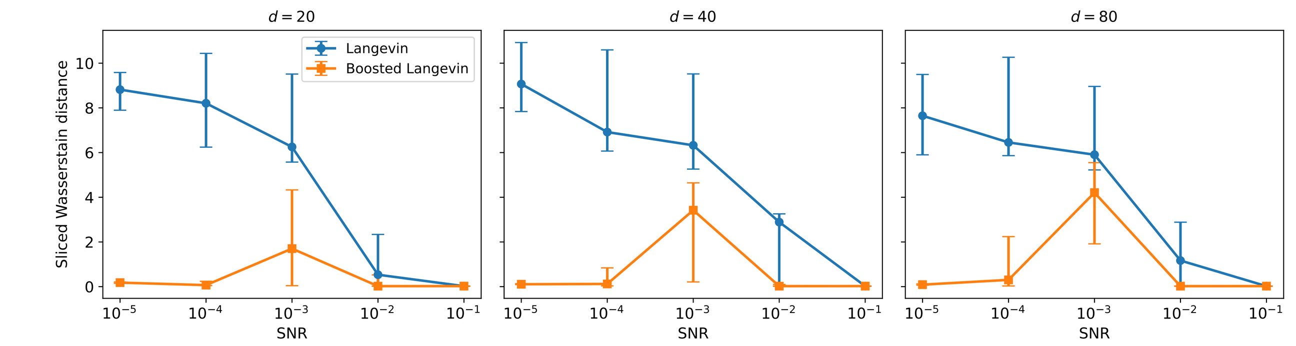
- **Gaussian Mixtures:** Let $\pi = \mu * \gamma_\delta$ and $\mathrm{diam}(\mathrm{supp}(\mu)) \le R$, then ν_{T^*} is strongly log-concave if

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



- **Ising Models:** Let π be the uniform measure on the hypercube $\{\pm 1\}^d$, and Q such that $\lambda_{\max}(Q) - \lambda_{\min}(Q) < 1$. Then ν_{T^*} is strongly log-concave, and therefore the Ising model $\nu = \mathsf{T}_Q\pi$ can be sampled efficiently (in continuous-time). This bound precisely matches the computational lower bound in [Kunisky, 2023].

Numerical Results



Gaussian mixture

Future work:

- Flow Matching/Stochastic Interpolant Oracles [Chen et al., Albergo et al.]
- Iterated Tilted Transport
- From Linear to Nonlinear inverse problems