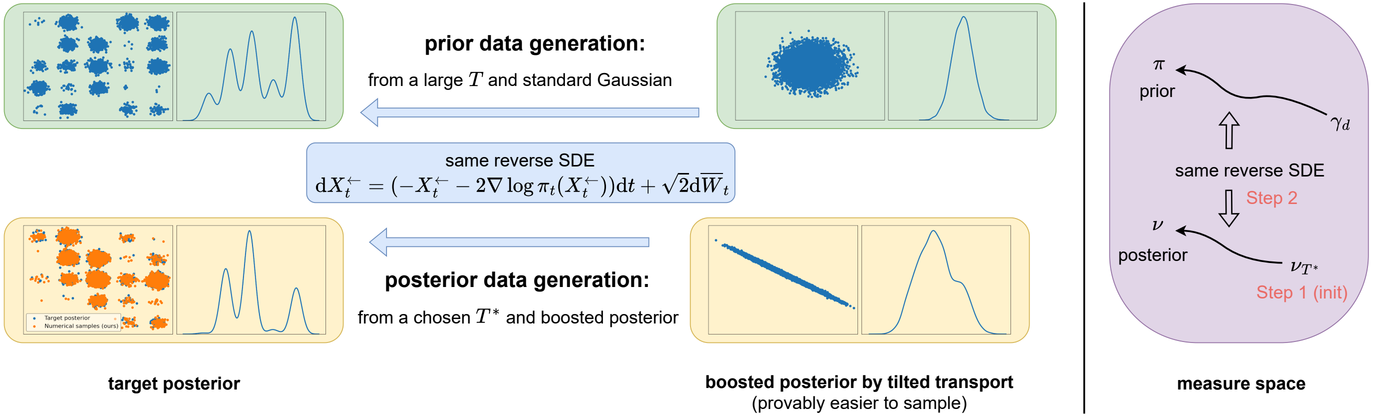
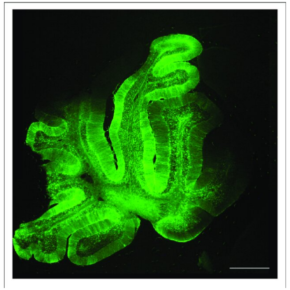
Provable Posterior Sampling With Denoising Oracles via Tilted Transport

Joan Bruna (NYU) and Jiequn Han (Flatiron Institute)

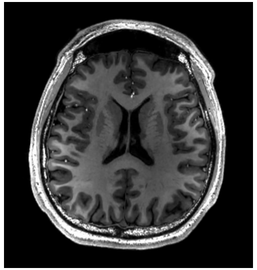
NeurIPS 2024

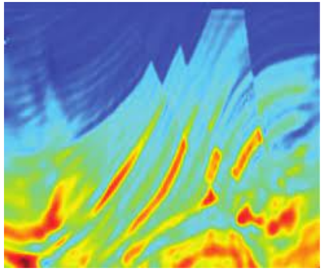


Inverse Problem



recover x from $y = \mathcal{F}(x) + \varepsilon$





Bayesian sampling: $x \sim p(x|y) \propto p(y|\mathcal{F}(x))p_{\text{prior}}(x)$

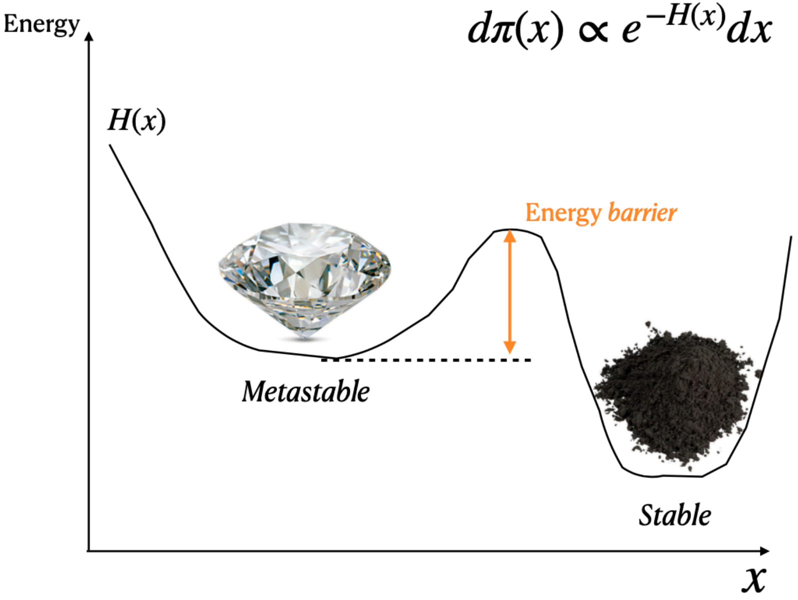
Langevin Dynamics and Metastability

Langevin Dynamics

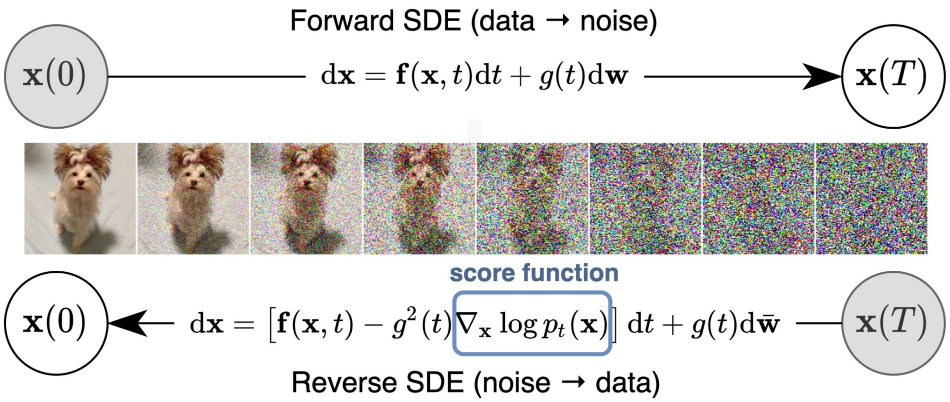
$$\text{sample } x \text{ from } \pi(x) \propto \exp(-H(x))$$

$$\mathrm{d}X_t = -\nabla H(x_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

Issues: time to relaxation 'cursed' by the presence of energy barriers and dimension

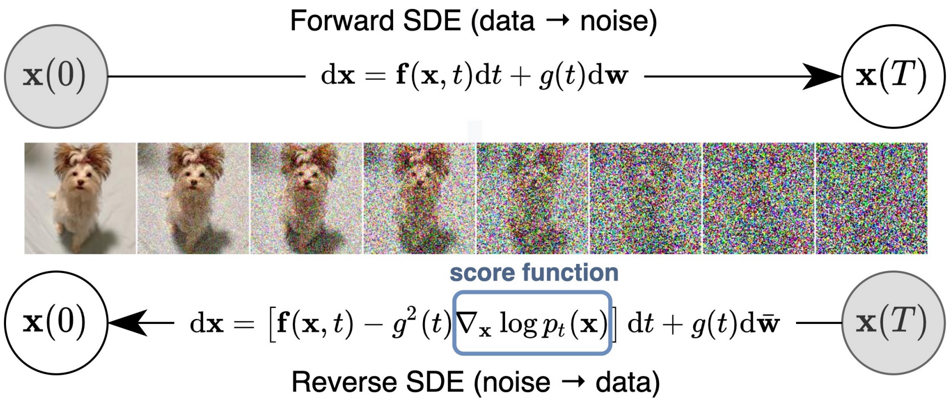


Score-Based Diffusion and Denoising Oracles



Score-Based Generative Modeling through Stochastic Differential Equations, Song et al. (2021)

Score-Based Diffusion and Denoising Oracles

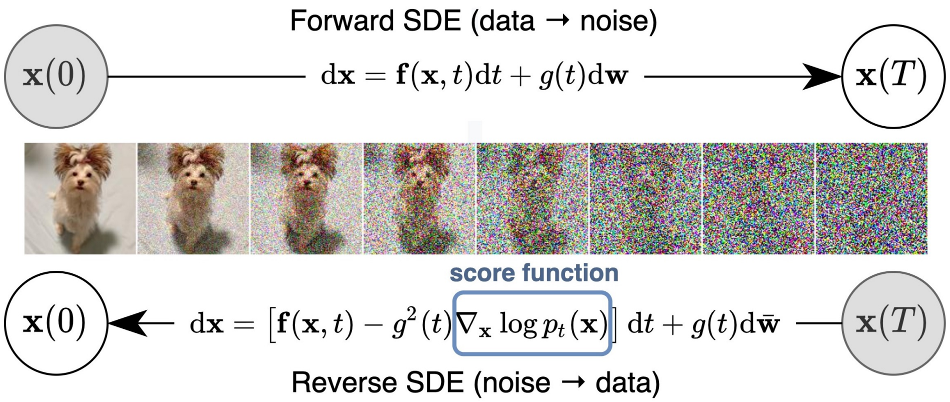


By Tweedie's formula, the time-dependent score along OU (or Heat) semigroup is equivalent to denoising oracle

$$\mathrm{DO}_\pi(y, t) = \mathbb{E}[X \,|\, y = X + tZ, \text{ where } X \sim \pi, Z \sim \mathcal{N}(0, I_d)$$

Score-Based Generative Modeling through Stochastic Differential Equations, Song et al. (2021)

Score-Based Diffusion and Denoising Oracles



By Tweedie's formula, the time-dependent score along OU (or Heat) semigroup is equivalent to denoising oracle

$$\mathrm{DO}_\pi(y, t) = \mathbb{E}[X \,|\, y = X + tZ, \text{ where } X \sim \pi, Z \sim \mathcal{N}(0, I_d)$$

Viewing the learned diffusion model or denoising oracle as a prior, how can we leverage it to sample from the posterior?

Score-Based Generative Modeling through Stochastic Differential Equations, Song et al. (2021)

Posterior Sampling

$$x \sim p(x|y) \propto p(y|\mathcal{F}(x))p_{\text{prior}}(x)$$

Existing works (see arXiv:2410.00083 for a recent survey):

- Approximating posterior conditional score (DPS, DMPS, ΠGDM, LGD, etc, inexact even for denoising problem or Gaussian mixture prior)
- Variational inference (RED-Diff, Score Prior etc, additional optimization needed)
- Combined with Sequential Monte Carlo/Particle Filtering/Plug and Play (asymptotically correct)

Posterior Sampling

$$x \sim p(x|y) \propto p(y|\mathcal{F}(x))p_{\text{prior}}(x)$$

Existing works (see arXiv:2410.00083 for a recent survey):

- Approximating posterior conditional score (DPS, DMPS, ΠGDM, LGD, etc, inexact even for denoising problem or Gaussian mixture prior)
- Variational inference (RED-Diff, Score Prior etc, additional optimization needed)
- Combined with Sequential Monte Carlo/Particle Filtering/Plug and Play (asymptotically correct)

How to transfer the power of diffusion model/denoising oracle prior to sample posterior, provably?

Problem Setup and Warmup

Given time-dependent score for OU $\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \text{ (prior)}$

$$y = Ax + \sigma\varepsilon, \quad x \sim \pi, \; \varepsilon \sim \gamma_d, \; \sigma > 0$$

Target posterior:

$$\nu \propto \pi(x) \exp\{-\frac{1}{2}x^\top Q x + x^\top b\} := \mathsf{T}_{Q,b}\pi \;, \text{ with } Q = \frac{1}{\sigma^2}A^\top A \,, \, b = -\frac{1}{\sigma^2}A^\top y$$

Problem Setup and Warmup

Given time-dependent score for OU $\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \text{ (prior)}$

$$y = Ax + \sigma\varepsilon, \quad x \sim \pi, \; \varepsilon \sim \gamma_d, \; \sigma > 0$$

Target posterior:

$$\nu \propto \pi(x) \exp\{-\frac{1}{2}x^\top Q x + x^\top b\} := \mathsf{T}_{Q,b}\pi \;, \; \text{with } Q = \frac{1}{\sigma^2}A^\top A \,, \, b = -\frac{1}{\sigma^2}A^\top y$$

Warmup: when $Q \propto \mathrm{Id}$ the task seems 'compatible' with the denoising oracle.

$$T^* = \frac{1}{2}\log(1+\sigma^2), \; \tilde{y} = e^{-T^*}y \quad \Longrightarrow \quad p(x|\tilde{y}) \overset{d}{=} p(X_0 | X_{T^*} = \tilde{y})$$

Problem Setup and Warmup

Given time-dependent score for OU $\quad \mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0 \sim \pi \text{ (prior)}$

$$y = Ax + \sigma\varepsilon, \quad x \sim \pi, \ \varepsilon \sim \gamma_d, \ \sigma > 0$$

Target posterior:

$$\nu \propto \pi(x) \exp\{-\frac{1}{2} x^\top Q x + x^\top b\} := \mathsf{T}_{Q,b}\pi \ , \ \text{with } Q = \frac{1}{\sigma^2} A^\top A \, , \, b = -\frac{1}{\sigma^2} A^\top y$$

Warmup: when $Q \propto \mathrm{Id}$ the task seems 'compatible' with the denoising oracle.

$$T^* = \frac{1}{2}\log(1+\sigma^2), \ \tilde{y} = e^{-T^*} y \quad \Longrightarrow \quad p(x|\tilde{y}) \overset{d}{=} p(X_0 | X_{T^*} = \tilde{y})$$

We can (1) first initialize $X_{T*} = e^{-T^*} y$ and (2) run the original reverse SDE from T^* to 0 to get the exact posterior

Tilted Transport for Posterior Sampling

Consider a time-varying quadratic tilt

$$\nu_t \propto \pi_t(x) \exp\{-\frac{1}{2}x^\top Q_t x + x^\top b_t\}$$

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

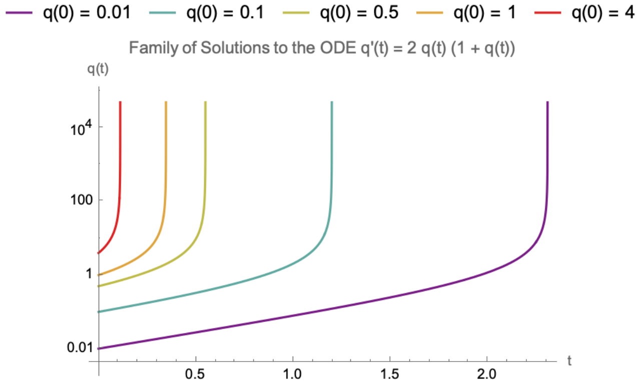
Theorem (titled transport) Assume $t < T^*$ such that the ODE is well-defined on $[0, t]$. By initializing $X_t \sim \nu_t$ and run the reverse SDE from t to 0, we have $X_s \sim \nu_s$ for $s \in [0, t]$, specifically, X_0 gives the desired posterior.

Tilted Transport for Posterior Sampling

Consider a time-varying quadratic tilt

$$\nu_t \propto \pi_t(x)\exp\{-\frac{1}{2}x^\top Q_t x + x^\top b_t\}$$

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

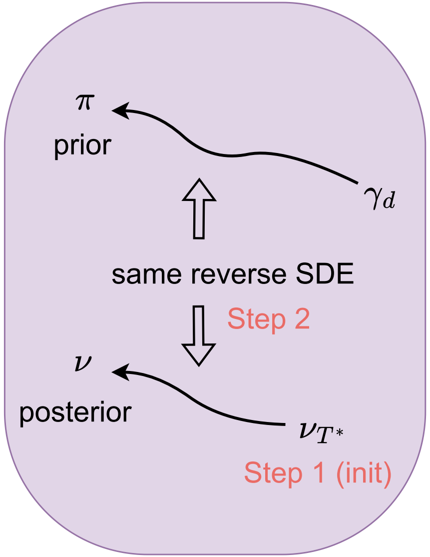


Theorem (titled transport) Assume $t < T^*$ such that the ODE is well-defined on $[0, t]$. By initializing $X_t \sim \nu_t$ and run the reverse SDE from t to 0, we have $X_s \sim \nu_s$ for $s \in [0, t]$, specifically, X_0 gives the desired posterior.

Tilted Transport for Posterior Sampling

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

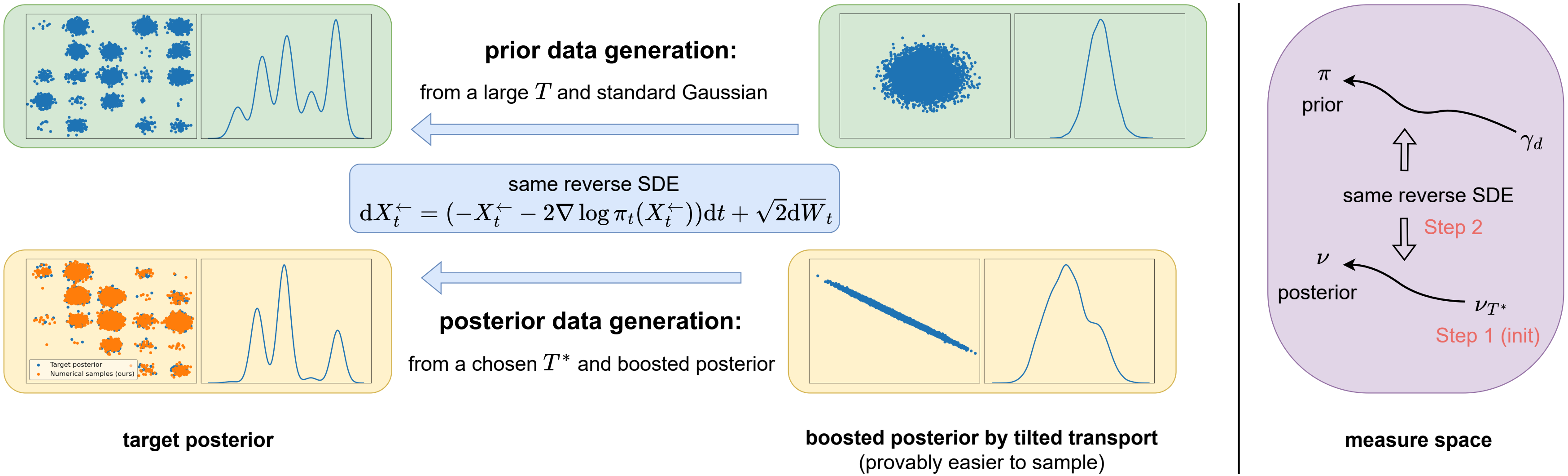


measure space

Intuition for Easier Sampling

Equivalent posterior sampling: $\nu_t \propto \boxed{\pi_t(x)} \boxed{\exp\{-\frac{1}{2}x^\top Q_t x + x^\top b_t\}}$

easier prior   easier likelihood



Provable Sampling

Theorem (Strong Log-Concavity of ν_T) For $t \geq 0$, let $\chi_t(\pi) := \sup_{x \in \mathbb{R}^d} \|\mathrm{Cov}[\mathsf{T}_{tI_d, tx}\pi]\|_{\mathrm{op}}$ denote the *susceptibility* of π, and let $\kappa = \lambda_{\max}(Q)/\lambda_{\min}(Q)$ denote the condition number of Q. Then ν_{T^*} is strongly log-concave if

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

Sufficient condition relates

1. prior susceptibility
2. signal-to-noise ratio
3. condition of measurement

Provable Sampling

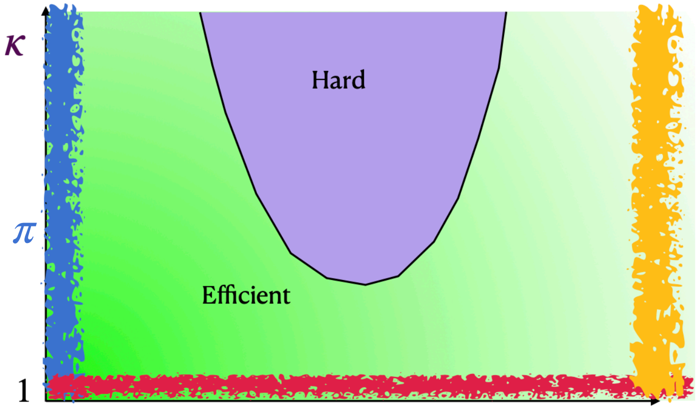
Theorem (Strong Log-Concavity of ν_T) For $t \geq 0$, let $\chi_t(\pi) := \sup_{x \in \mathbb{R}^d} \|\mathrm{Cov}[\mathsf{T}_{tI_d, tx}\pi]\|_{\mathrm{op}}$ denote the *susceptibility* of π, and let $\kappa = \lambda_{\max}(Q)/\lambda_{\min}(Q)$ denote the condition number of Q. Then ν_{T^*} is strongly log-concave if

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

Sufficient condition relates

1. prior susceptibility
2. signal-to-noise ratio
3. condition of measurement

low-SNR regime: $\nu \approx \pi$



high-SNR regime: ν is log-concave

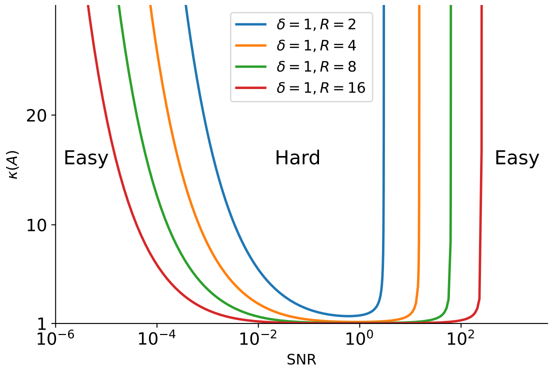
Denoising-type: efficient at any SNR level

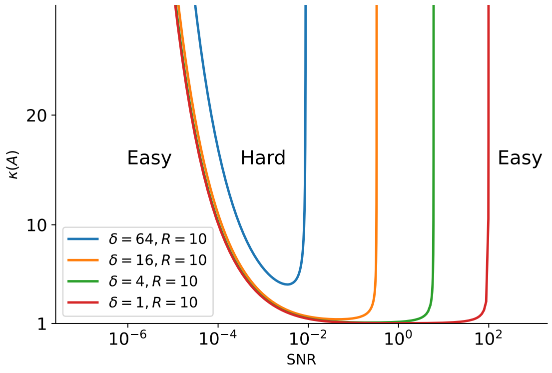
$\|Q\|$

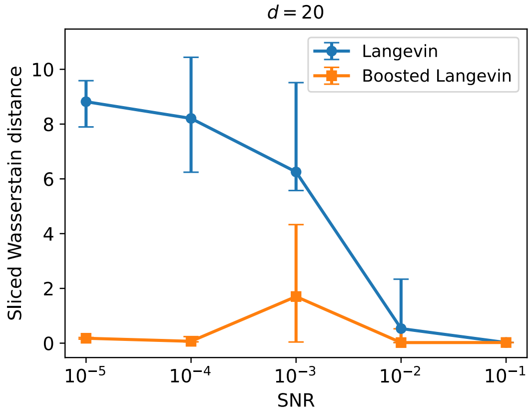
Provable Sampling for Gaussian Mixtures

Corollary (tilted transport for Gaussian mixtures) Let $\pi = \mu \star \gamma_\delta$ and $\mathrm{diam}(\mathrm{supp}(\mu)) \leq R$, then ν_{T^*} is strongly log-concave if $(\mathrm{SNR} := \lambda_{\min}(Q) = \lambda_{\min}(A)^2/\sigma^2)$

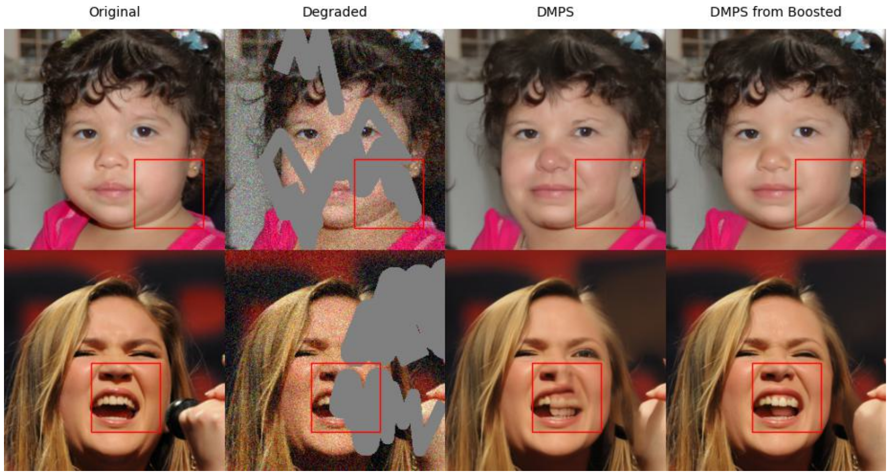
$$\frac{(1 + \delta \mathrm{SNR}^2)(\delta \kappa(A)^2 + \mathrm{SNR}^{-2})}{\kappa(A)^2 - 1} > R^2 \; .$$



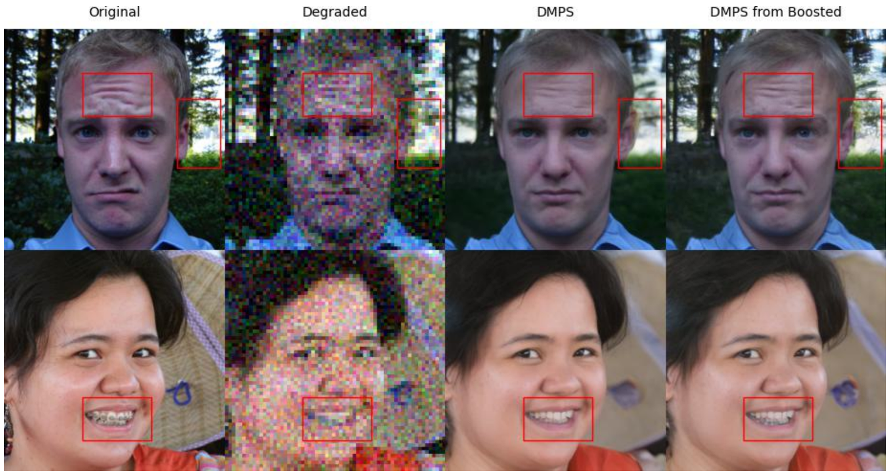




Imaging Problems



inpainting



deblur