Diffusive Gibbs Sampling

Wenlin Chen*[,1,2], Mingtian Zhang*[,3], Brooks Paige[3], José Miguel Hernández-Lobato[1], David Barber[3]

[1]University of Cambridge, UK; [2]MPI for Intelligent Systems, Germany; [3]University College London, UK.

Paper QR Code

Challenge of Multi-modal Sampling

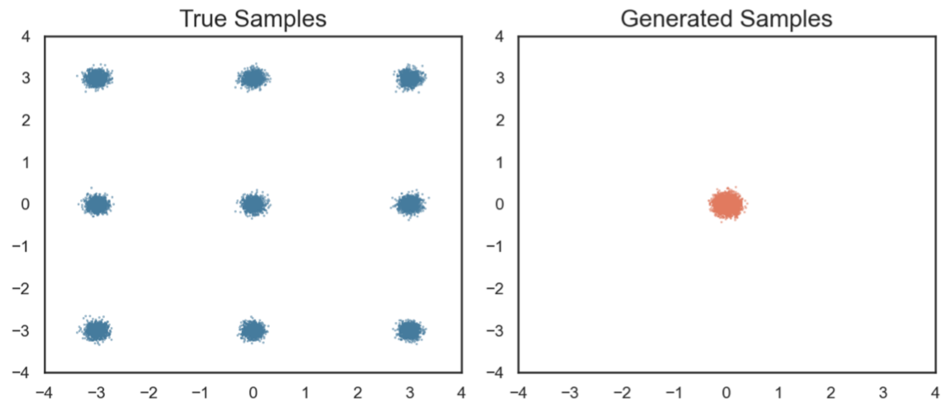
· **Goal:** Generate independent samples from an unnormalized distribution:

$$p(x) = \frac{\exp(-E(x))}{Z}$$

· **Challenge:** Sampling from distributions with disconnected/distant modes is extremely difficult for many MCMC samplers.

· **Example:** Score-based Langevin dynamics cannot capture distant modes.

$$x_{k+1} = x_k + \eta \nabla_x \log p(x_k) + \sqrt{2\eta}\epsilon_k$$



Convolution-based Sampling

· **Observation:** Diffusion models leverage Gaussian convolution kernels to bridge disconnected modes in multi-modal distributions.

$$p(\tilde{x}|x) = \mathcal{N}(\tilde{x}|\alpha x, \sigma^2 I) \qquad p(\tilde{x}) = \int p(\tilde{x}|x)p(x)\,\mathrm{d}x.$$

· **Property:** The modes in convolved noisy distributions $p(\tilde{x})$ are connected via non-negligible density paths.

· **Idea:** Can we use samples $\tilde{x} \sim p(\tilde{x})$ as initial points for sampling from $p(x)$ since they are more likely to cover more modes.

· **Issue:** Difficult to sample from $p(\tilde{x})$ since its score function is intractable.

$$\nabla_{\tilde{x}} \log p(\tilde{x}) = \nabla_{\tilde{x}} \log \int \exp\left(-E(x) - \frac{\|\tilde{x} - \alpha x\|^2}{2\sigma^2}\right) \mathrm{d}x$$

Accepted at ICML 2024. Paper link: https://arxiv.org/abs/2402.03008

Diffusive Gibbs Sampling (DiGS)

· **Idea:** Draw samples from the joint $p(x, \tilde{x}) = p(\tilde{x}|x)p(x)$ via Gibbs sampling by repeating the following two steps.

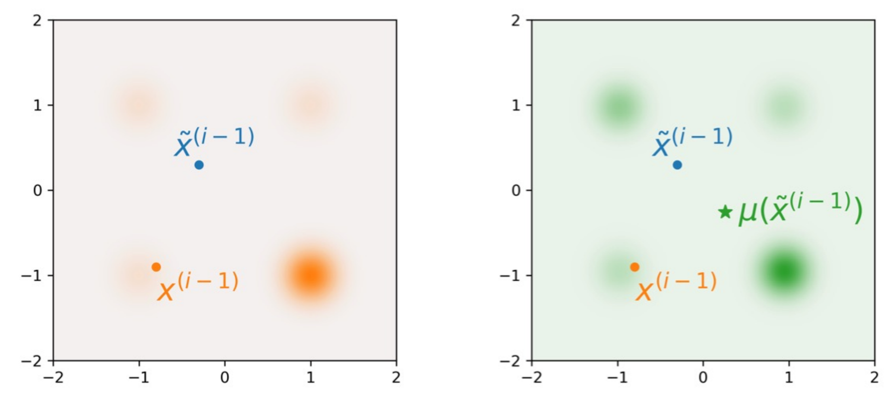
1. Adding noise $\tilde{x} \sim p(\tilde{x}|x)$.

$$\tilde{x} = \alpha x + \sigma\epsilon, \quad \epsilon \sim N(0, I),$$

2. Denoising sampling $x \sim p(x|\tilde{x})$ with a score-based sampler.

$$\nabla_x \log p(x|\tilde{x}) = -\nabla_x E(x) - \frac{\alpha(\alpha x - \tilde{x})}{\sigma^2}.$$

· **Blindness Issue:** The score-based sampler cannot capture the correct density weighting of different modes in $p(x|\tilde{x})$.



(a) $p(x)$. (b) $p(x|\tilde{x}^{(i-1)})$.

· **Solution:** We propose a Metropolis-Hasting (MH) scheme to facilitate mixing across modes by initialize the denoising sampling step with the following proposal distribution.

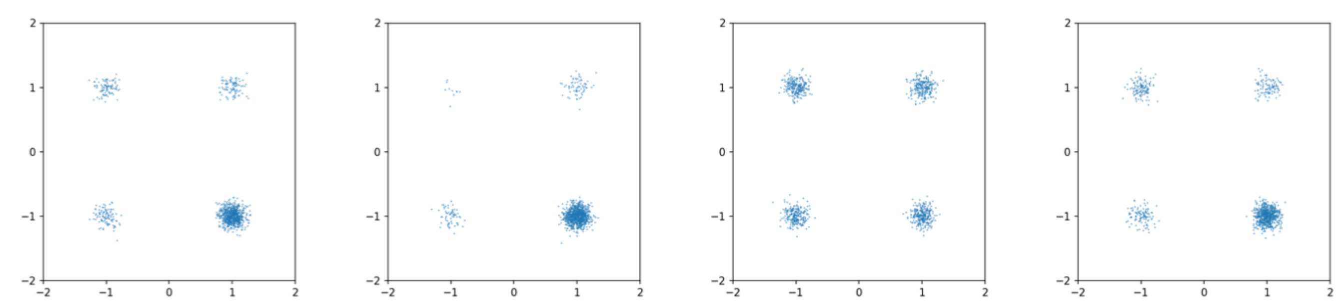
$$q(x|\tilde{x}^{(i-1)}) = \mathcal{N}(x|\tilde{x}^{(i-1)}/\alpha, (\sigma/\alpha)^2 I)$$

The MH proposal $x'_{init} \sim q(x|\tilde{x}^{(i-1)})$ is accepted with probability

$$a_{init} = \min\left(1, \frac{p(x'_{init}|\tilde{x}^{(i-1)})q(x^{(i-1)}|\tilde{x}^{(i-1)})}{p(x^{(i-1)}|\tilde{x}^{(i-1)})q(x'_{init}|\tilde{x}^{(i-1)})}\right)$$

Otherwise, the initial position is set to the previous clean sample $x^{(i-1)}$.

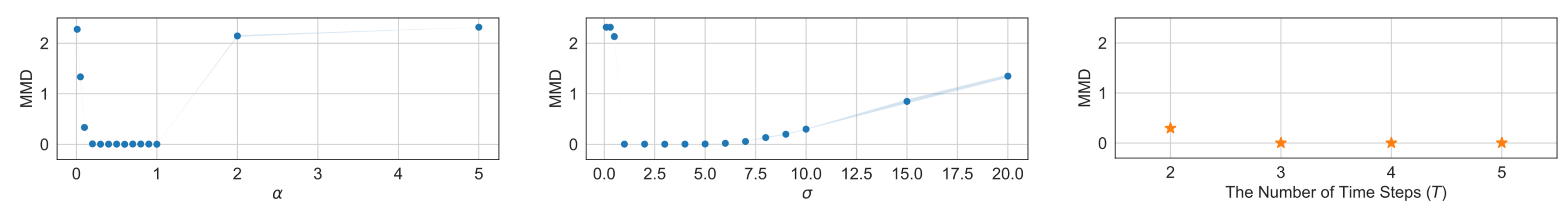
· **Comparison of Initializations for Denoising Sampling:**



(a) True. (b) $x^{(i-1)}$. (c) $\tilde{x}^{(i-1)}/\alpha$. (d) *MH*.

Multi-level Noise Scheduling

· **Observation:** The choice of the convolution hyperparameters α and σ significantly influence the performance of DiGS.



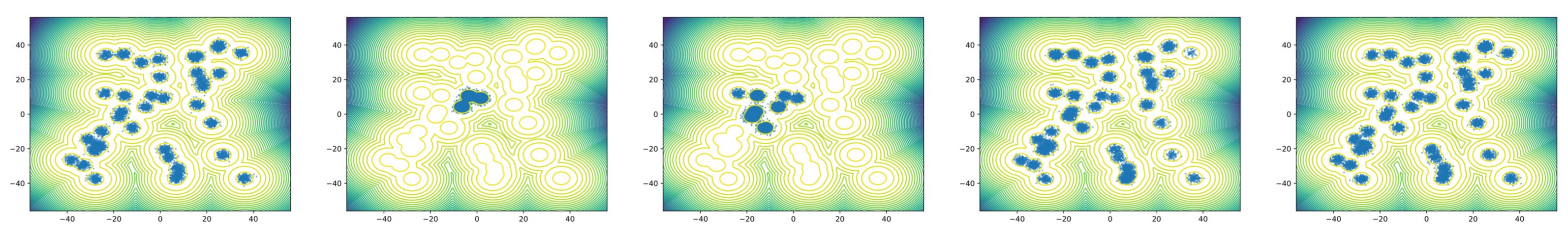
(a) Effects of α (fix $\sigma = 1$). (b) Effects of σ (fix $\alpha = 1$). (c) Effects of T in the VP schedule.

· **Idea:** We employ the multi-level noise scheduling from DDPM to eliminate the tuning of the convolution hyperparameters.

$$\alpha_t = \alpha_T + (\alpha_1 - \alpha_T)\frac{T - t}{T - 1}, \quad \sigma_t = \sqrt{1 - \alpha_t^2}.$$

Experiments

· **Toy demonstration: Mixture of 40 Gaussians**



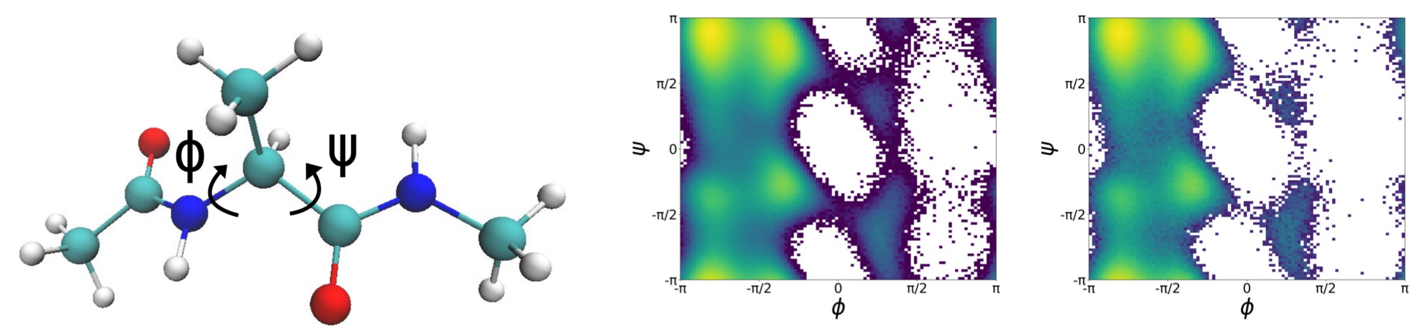
(a) True Samples (b) MALA (c) HMC (d) PT (e) *DiGS*

· **Bayesian Inference: Bayesian Neural Network**

Table 3. Average test predictive NLL for the BNN estimated by 10^3 samples generated by each sampler.

Sampler	NLL	#energy evaluations
MAP	0.548 ± 0.066	5.0×10^6
MALA	0.399 ± 0.014	5.0×10^6
HMC	0.315 ± 0.012	5.0×10^6
PT	0.241 ± 0.005	5.0×10^6
DiGS	$\mathbf{0.189} \pm \mathbf{0.002}$	5.0×10^6

· **AI for Science: Molecular Configuration Sampling**



(a) Ground-truth (b) *DiGS* (10^6 samples)