Scalable Discrete Diffusion Samplers: Combinatorial Optimization and Statistical Physics

TL;DR: Scalable Diffusion Sampler with the usage of Reinforcement Learning & unbiased sampling on Ising Models.

Sebastian Sanokowski[1], **Wilhelm Berghammer**[1], **Martin Ennemoser**[1], **Haoyu Peter Wang**[2], **Sepp Hochreiter**[1,3], **Sebastian Lehner**[1]

[1] ELLIS Unit Linz and LIT AI Lab, Institute for Machine Learning, Johannes Kepler University Linz
[2] Department of Electrical & Computer Engineering, Georgia Institute of Technology [3] NXAI

1. CONTRIBUTIONS

1. Principled framework based on the **Policy Gradient theorem** and **Neural Importance Sampling** yield memory efficient diffusion samplers
2. Framework to leverage diffusion samplers in **Self-Normalized Importance Sampling** and **Neural MCMC**
3. State-of-the-art results on **Combinatorial Optimization** problems and in unbiased sampling on **Ising models** at criticality

2. PROBLEM DESCRIPTION

Learn to generate samples $X \in \{0,1\}^N$ according to the target distribution at a given inverse temperature β:

$$p_B(X, \beta) = \frac{\exp(-\beta H(X))}{\mathcal{Z}}$$

$$\text{where} \quad \mathcal{Z} = \sum_X \exp(-\beta H(X))$$

only the energy function $H(X)$ is available and there are no samples form $p_B(X, \beta)$!

Alpha Divergences:

$$D_\alpha(p_B(X)||q_\theta(X)) = -\frac{\int p_B(X)^\alpha q_\theta(X)^{1-\alpha} dX}{\alpha(1-\alpha)}$$

$\alpha \geq 1$ mass-covering, $\alpha \to 1 : D_{KL}(p_B \,||\, q_\theta)$ (fKL)
$\alpha \leq 0$ mode seeking, $\alpha \to 0 : D_{KL}(q_\theta \,||\, p_B)$ (rKL)

Optimization with Diffusion Models:
Data Processing Inequality for α-Divergences.

$$D_\alpha(q_\theta(X_0) \,||\, p(X_0)) \leq D_\alpha(q_\theta(X_{0:T}) \,||\, p(X_{0:T}))$$

This is a tractable upper-bound that serves as the loss.

3. METHODS

rKL-based objective in previous discrete diffusion samplers:

$$D_{KL}(q_\theta(X_{0:T}) \,||\, p(X_{0:T})) = -\sum_{t=1}^{T} \mathbb{E}_{X_{T:t} \sim q_\theta(X_{T:t})}[S(q_\theta(X_{t-1}|X_t))]$$
$$- \sum_{t=1}^{T} \mathbb{E}_{X_{T:t-1} \sim q_\theta(X_{T:t-1})}[\log p(X_t|X_{t-1})]$$
$$+ \beta\, \mathbb{E}_{X_{T:0} \sim q_\theta(X_{T:0})}[H_Q(X_0)] + C$$

Problem:
Memory scales $\mathcal{O}(T)$ during training
→ method does not scale to many diffusion steps

Scalable rKL-based Objective:
rKL can alternatively be minimized using Reinforcement Learning methods:

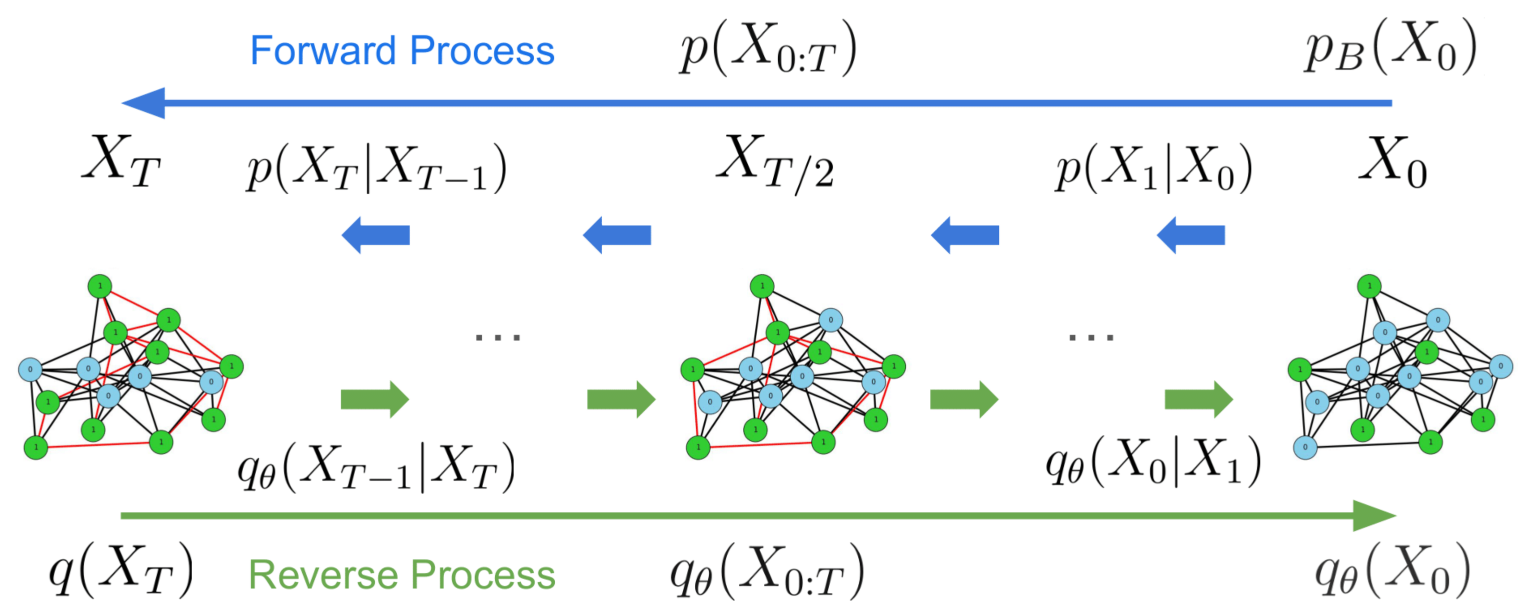
$$L(\theta) = D_{\mathrm{KL}}(q_\theta(X_{0:T}) \,||\, p(X_{0:T}))$$

Policy Gradient Theorem

⇓

$$\nabla_\theta L(\theta) = -\mathop{\mathbb{E}}_{X_t \sim d^\theta(\mathcal{X}, t), X_{t-1} \sim q_\theta}\left[Q^\theta(X_{t-1}, X_t) \nabla_\theta \log q_\theta(X_{t-1}|X_t)\right],$$

$$R(X_t, X_{t-1}) := \begin{cases} \mathcal{T}[\log \frac{p(X_t|X_{t-1})}{q_\theta(X_{t-1}|X_t)}] & \text{if } t > 1 \\ \mathcal{T}[\log \frac{p(X_t|X_{t-1})}{q_\theta(X_{t-1}|X_t)}] - H(X_{t-1}) & \text{else} \end{cases}$$



How to remove the bias between q_θ and p_B?

Use Neural Importance Sampling (NIS) for unbiased estimation of observables $\langle O(X) \rangle$

Exact Likelihood Model:

$$\langle O(X_0) \rangle_{p_B(X_0)} \approx \sum_{i=1}^{M} [w(X_0^i)\, O(X_0^i)]$$

where $w(X_0^i) = \frac{\widehat{w}(X_0^i)}{\sum_{j=1}^M \widehat{w}(X_0^j)}$ and $X_0^i \sim q_\theta(X_0^i)$ with $\widehat{w}(X_0^i) = \frac{\widehat{p}(X_0^i)}{q_\theta(X_0^i)}$.

Approximate Likelihood Model:

$$\langle O(X_0) \rangle_{p_B(X_0)} \approx \sum_{i=1}^{M} [w(X_{0:T}^i)\, O(X_0^i)]$$

where $w(X_{0:T}^i) = \frac{\widehat{w}(X_{0:T}^i)}{\sum_{j=1}^M \widehat{w}(X_{0:T}^j)}$ and $X_{0:T}^i \sim q_\theta(X_{0:T}^i)$ with $\widehat{w}(X_{0:T}^i) = \frac{\widehat{p}(X_{0:T}^i)}{q_\theta(X_{0:T}^i)}$.

Scalable fKL-based Objective:

$$L(\theta) = D_{\mathrm{KL}}(p(X_{0:T}) \,||\, q_\theta(X_{0:T}))$$

NIS and MC over diffusion time

⇓

$$\nabla_\theta L(\theta) = -T \sum_{i=1}^{M} \mathbb{E}_{t \sim U\{1,\dots,T\}}\left[w(X_{0:T}^i) \nabla_\theta \log q_\theta(X_{t-1}^i | X_t^i)\right],$$

where $w(X_{0:T}^i) = \frac{\widehat{w}(X_{0:T}^i)}{\sum_{j=1}^M \widehat{w}(X_{0:T}^j)}$ with $\widehat{w}(X_{0:T}^i) = \frac{\widehat{p}(X_{0:T}^i)}{q_\theta(X_{0:T}^i)}$, $X_{0:T}^i \sim q_\theta(X_{0:T})$, and $U\{1,\dots,T\}$ is the uniform distribution over the set $\{1,\dots,T\}$.

Experiments:

1. Experiments on popular benchmarks in **Unsupervised Combinatorial Optimization**
2. Experiments in unbiased sampling on the **Ising Model** and on **Spin Lattices**

Results:

1. SDDS outperforms previous diffusion based samplers on UCO benchmarks and in unbiased sampling
2. fKL-based objective works better than rKL-based objective in unbiased sampling, rKL works better in UCO

Link to Paper: