BNEM: A Boltzmann Sampler Based On Bootstrapped Noised Energy Matching

RuiKang OuYang*, [1] Bo Qiang*, [1, 2] José Miguel Hernández-Lobato[1]

*Equal contribution [1]University of Cambridge [2]University of Washington

Check Our Paper!

NEURAL INFORMATION PROCESSING SYSTEMS

UNIVERSITY OF CAMBRIDGE

TL;DR

We propose ***NEM***, a ***neural sampler*** targets the ***noised energy*** and enables ***diffusion sampling***; We extend NEM to ***BNEM***, which self-enhances the learning target via ***bootstrapping***.

Motivation

- Sampling from an ***unnormalized density***, i.e. ***Boltzmann density***, is foundational in science.
- ***Diffusion*** sampling enjoys fast mode-mixing
- Standard diffusion model requires large amount of ***data*** to learn the marginal scores
- How to train learn the marginal scores ***without known data*** but ***access to the energy***?

Related Works

- iDEM [1] regresses an MC score estimator of the with only access to energy $\mathcal{E}(x)$

Methods

NEM

- Marginal density perturbed by VE noising process

$$p_t(x_t) \propto \int \exp(-\mathcal{E}(x))\mathcal{N}(x_t; x, \sigma_t^2 I)dx$$

- We define the ***noised energy*** as

$$\mathcal{E}_t(x_t) \coloneqq -\log\int \exp(-\mathcal{E}(x))\mathcal{N}(x_t; x, \sigma_t^2 I)dx$$

- We approximate it by an ***MC estimator***

$$E_k(x_t, t) = -\log\frac{1}{K}\sum_{i=1}^{K}\exp\left(-\mathcal{E}\left(x_{0|t}^{(i)}\right)\right)$$

with $x_{0|t}^{(i)} \sim \mathcal{N}(x; x_t, \sigma_t^2 I)$.

- ***Learning Objective***

$$\mathcal{L}_{NEM}(x_t, t) = ||E_\theta(x_t, t) - E_k(x_t, t)||^2$$

BNEM

- As t increase, $E_k(x_t, t)$ becomes inaccurate with high variance
- To remedy, we ***leverage the learned noised energy*** at lower time level s, i.e. $\mathcal{E}_s$, to estimate energy at t
- We propose the ***Bootstrapped energy estimator***

$$E_k(x_t, t, s; \theta) = -\log\frac{1}{K}\sum_{i=1}^{K}\exp\left(-E_\theta\left(x_{s|t}^{(i)}, s\right)\right)$$

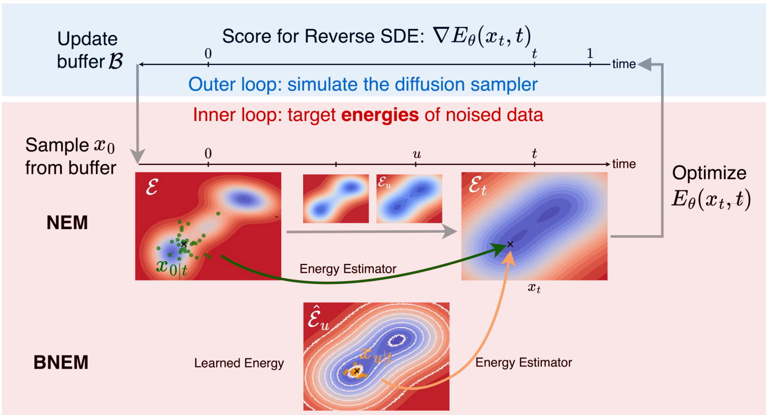
with $x_{s|t}^{(i)} \sim \mathcal{N}(x; x_t, (\sigma_t^2 - \sigma_s^2)I)$.

- ***Learning Objective***

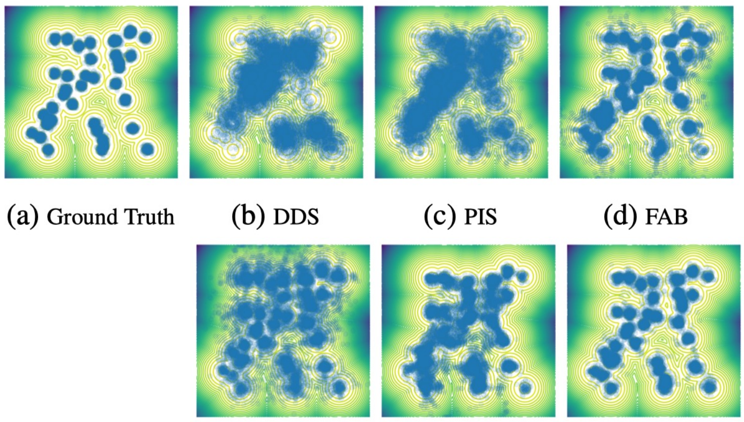
$$\mathcal{L}_{BNEM}(x_t, t, s) = ||E_\theta(x_t, t) - E_k(x_t, t, s; sg(\theta))||^2$$

where $sg(\theta)$ refers stopping gradient of θ.

Overview



Results



(a) Ground Truth (b) DDS (c) PIS (d) FAB

(e) iDEM (f) NEM (ours) (g) BNEM (ours)

Table 1: Neural sampler performance comparison for GMM-40 and DW-4 energy function. we measured the performance using data Wasserstein-2 distance (x-$\mathcal{W}_2$), Energy Wasserstein-2 distance ($\mathcal{E}$-$\mathcal{W}_2$), and Total Variation (TV). * indicates divergent training. **Bold** indicates the best values and underline indicates the second best ones.

Energy →	GMM-40 ($d=2$)			DW-4 ($d=8$)		
Sampler ↓	x-$\mathcal{W}_2$↓	$\mathcal{E}$-$\mathcal{W}_2$↓	TV↓	x-$\mathcal{W}_2$↓	$\mathcal{E}$-$\mathcal{W}_2$↓	TV↓
DDS	11.69	86.69	0.944	0.701	109.8	0.429
PIS	5.806	76.35	0.940	*	*	*
FAB	<u>3.828</u>	<u>64.23</u>	<u>0.824</u>	0.614	211.5	0.359
iDEM	8.512	562.7	0.909	0.532	2.109	0.161
NEM (ours)	5.192	85.05	0.906	<u>0.489</u>	<u>0.999</u>	<u>0.145</u>
BNEM (ours)	**3.652**	**2.973**	**0.830**	**0.467**	**0.458**	**0.134**

References

[1] Tara Akhound-Sadegh, Jarrid Rector-Brooks, Avishek Joey Bose, Sarthak Mittal, Pablo Lemos, Cheng-Hao Liu, Marcin Sendera, Siamak Ravanbakhsh, Gauthier Gidel, Yoshua Bengio, Nikolay Malkin, and Alexander Tong. Iterated denoising energy matching for sampling from boltzmann densities, 2024.