Assisting sampling with generative models

04/12/2025

Journée scientifique du groupe SMAI-SIGMA

Marylou Gabrié

École Normale Supérieure

Work done with

Grant Rotskoff
Stanford university

Eric Vanden-Eijnden
NYU

École Polytechnique

Louis Grenioux

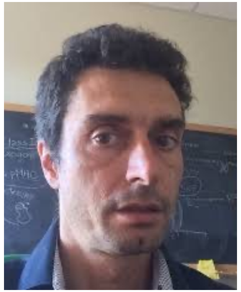
Maxence Noble

Alain Oliviero-Durmus

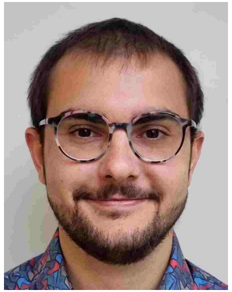
Christoph Schönle

Pilar Cossio
Flatiron Institute

Olga Lopez-Acevedo
U. de Antioquia



Alessandro Laio



Samuel Tamagnone

SISSA

Tony Lelièvre

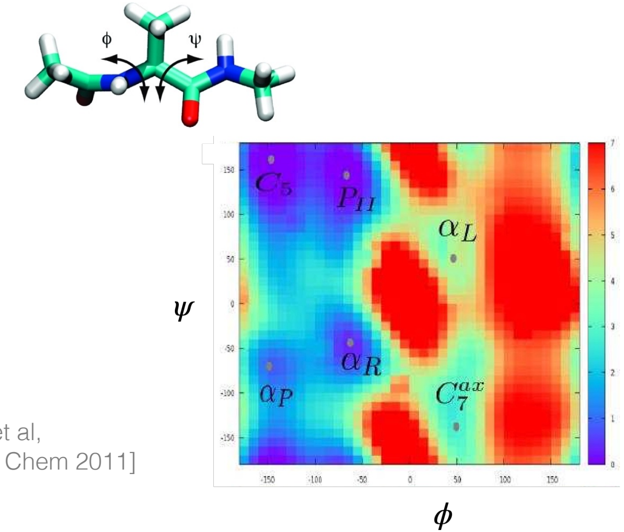
Gabriel Stoltz

ENPC CERMICS

Examples of scientific applications of sampling

▷ Statistical mechanics:
Gibbs-Boltzmann distribution

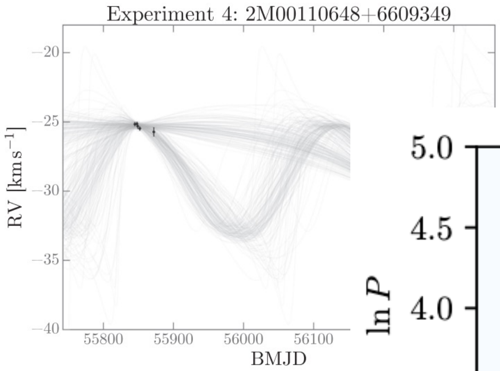
$$\rho_*(x) = \frac{1}{\mathcal{Z}_\beta} e^{-\beta U(x)}$$



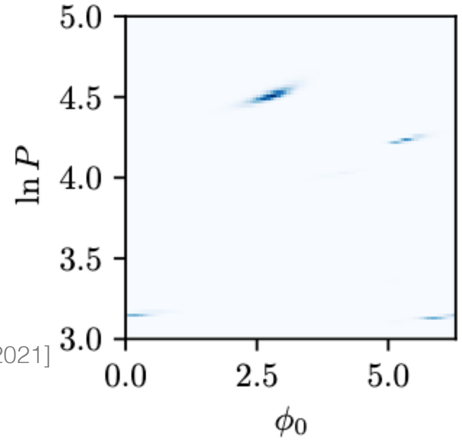
[Jaillet et al,
J Comp Chem 2011]

▷ Data analysis: Bayesian posteriors

$$\rho_*(\theta) \propto \ell(\mathcal{D}|\theta)\rho_0(\theta)$$



[Price-Whelan et al.
The Astrophysical Journal 2017]



[Gabrié et al, ICML workshop 2021]

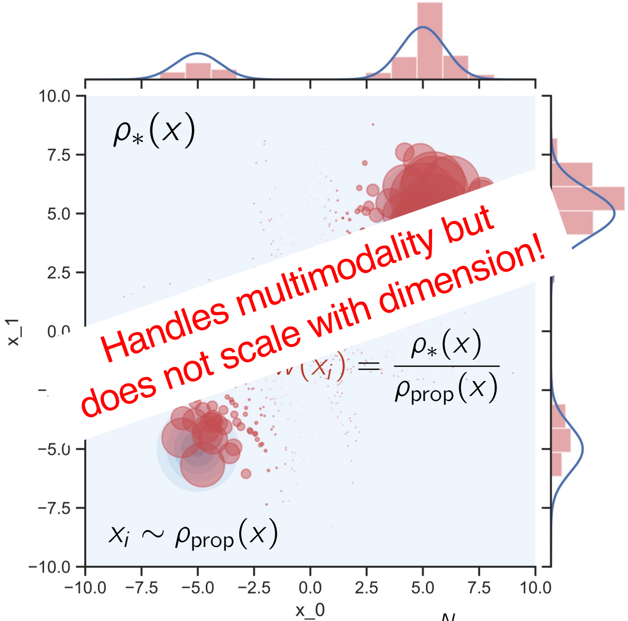
▷ Distribution is known up to a normalization constant.

▷ I am particularly interested in these multimodal distributions/metastable systems.

Why sampling multimodal distributions is hard?

Two fundamental approaches to sampling:

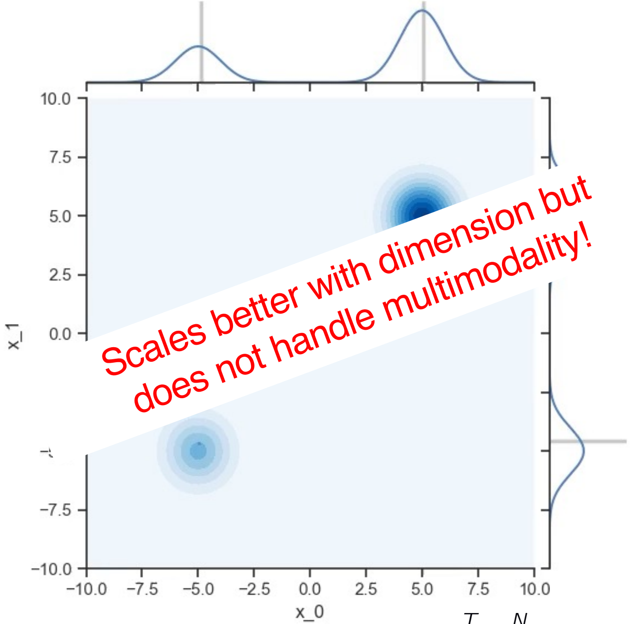
- Shoot and reject/reweight algorithms: *(e.g. Importance Sampling IS)*



$$\mathbb{E}_{\rho_*}[f(x)] = \int_\Omega f(x)\rho_*(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^{N} w(x_i) f(x_i)$$

High variance!

- Local exploration Markov chain Monte Carlo (MCMC) *(e.g. Metropolis Adjusted Langevin)*



$$\mathbb{E}_{\rho_*}[f(x)] = \int_\Omega f(x)\rho_*(x)\mathrm{d}x \approx \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} f(x_i^t)$$

High bias!

Enhanced samplers preferred by practitioners for high-dimensional multimodal distributions

- ▷ Decades of research that is impossible to summarize completely in a slide.
- ▷ **Annealing methods:** a path of distributions bridging an easy to sample distribution to the target

  e.g. Parallel tempering/replica exchange, Annealed Importance Sampling (AIS), Sequential Monte Carlo (SMC)

  [Marinari & Parisi (1992), Geyer & Thomson (1995), Neal (1998), Del Moral, Doucet & Ajay (2006) etc.]

- ▷ **Enhanced samplers with "collective variables":** drive exploration along a low dimensional projection

  e.g. Metadynamics, Adaptive Biasing Force, Umbrella Sampling

  [Fu et al. "Enhanced Sampling Based on Collective Variables." 2023]

- ▷ **Samplers assisted by generative models:** Use a deep learning model approximating the target within MC

  mainly with Normalizing Flows, Autoregressive Models (discrete variables)

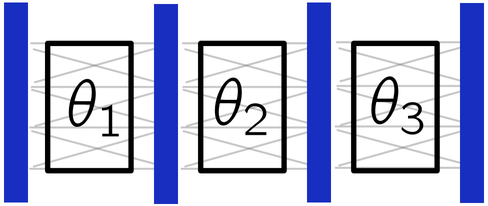
  [Rezende & Mohamed ICML 2015, Albergo et al PRD 2019, Wu et al. PRL 2019, Noé et al. Science 2019, Gabrié et al. PNAS 2019 etc.]

  & Diffusion Models, Flow Matchings [Vargas et al. arXiv:2302.13834, RDMC - Huang et al. 2307.02037, Berner et al. arXiv:2307.01198, Vargas et al. arXiv:arXiv:2307.01050, SLIPS – Grenioux et al. 2402.10758, Akhound-Sadegh et al. 2402.06121, etc…]

Transport based deep generative models

$T_\theta : \Omega \mapsto \Omega$ ← learnable diffeomorphism = transport map

$z \sim \rho_B(z)$



$x = T_\theta(z) \quad \sim \rho_\theta(x)$

new noise
→ new dog

[Song et al. ICLR 2021]

i.e. model of complex distributions with
1) simple sampling protocol
2) (more or less easy) access to likelihood

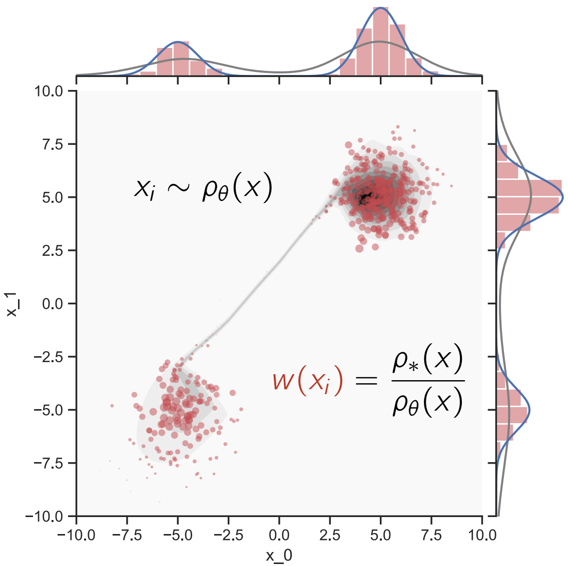
Change of variable formula: $\rho_\theta(x) = \rho_B(T_\theta^{-1}(x)) \det \left| \nabla_x T_\theta^{-1} \right|$

- The transport map can be parametrized in different ways:
  - Explicit parametrization of the map (deep learning: normalizing flows – or simpler: triangular, etc…)
  - Parametrization of a velocity or drift field to be included in an ODE/SDE (continuous normalizing flows, diffusion models, flow matchings/stochastic interpolants)

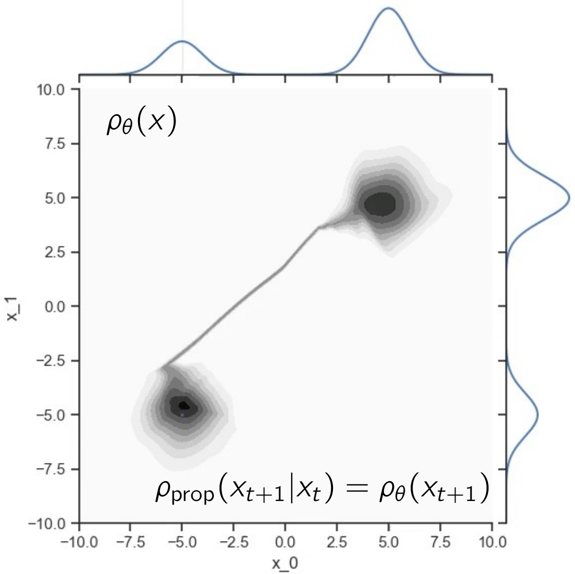
Approximate transport map for sampling

▷ Integrate approximate maps to build into Monte Carlo strategies to obtain exact samples

Importance sampling



Metropolis-Hastings



Pioneering works in statistics/machine learning: [Rezende & Mohamed ICML 2015, Parno & Marzouk 2018 SIAM Journal of Uncertainty,]
Pioneering works in physics: [Albergo et al. PRD 2019, Wu et al. PRL 2019, Noé et al. Science 2019]

Two caveats:

▷ Exact samples require exact evaluation of the likelihood $\rho_\theta(x) = \rho_B(T_\theta^{-1}(x)) \det |\nabla_x T_\theta^{-1}|$

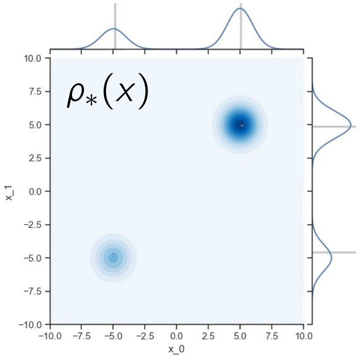
i.e. explicit parametrization to ensure access to Jacobian and no continuous-time model

▷ How do we train?

One training strategy: Adaptive learning of map while sampling

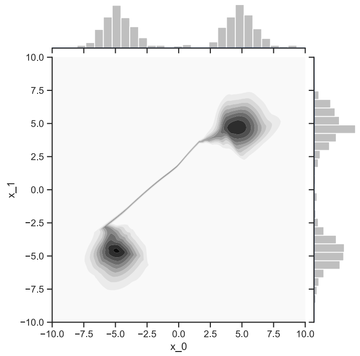
▷ Loop over 3 steps:

1. local sampler



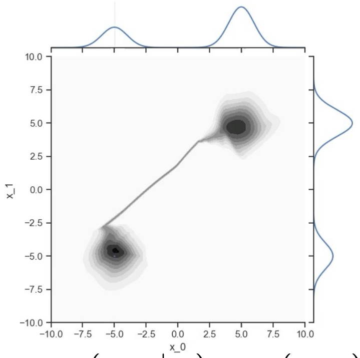
$$x^i_{t+1} \sim \pi_{\text{local}}(x^i_{t+1}|x^i_t)$$

2. maximum likelihood



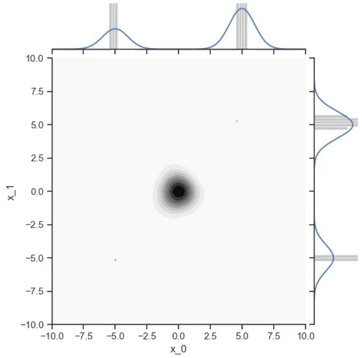
$$\theta^* = \arg\max_\theta \sum_{i,t} \log \rho_\theta(x^i_t)$$

3. non-local sampler



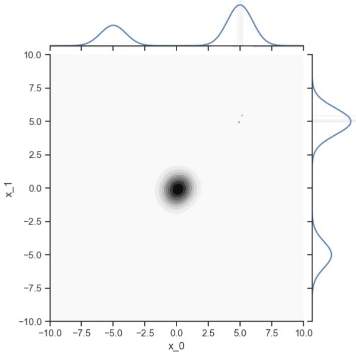
$$\rho_{\text{prop}}(x_{t+1}|x_t) = \rho_\theta(x_{t+1})$$

▷ Converging all steps in parallel:



▷ Requires the knowledge of the modes location!

a.k.a no free lunch!



[Parno & Marzouk 2018 SIAM Journal of Uncertainty; Naesseth et al. Neurips 2020, MG, Rotskoff, Vanden-Eijnden PNAS 2022]

Enhancing sampling with flows: state of affairs

▷ Proofs of concepts of using generative models for sampling physical systems in many different fields
(an incomplete selection)

biomolecules

Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning

Frank Noé*†, Simon Olsson*, Jonas Köhler*, Hao Wu

DeepBAR: A Fast and Exact Method for Binding Free Energy Computation

Xinqiang Ding and Bin Zhang*

(supercooled) liquids

Learning mappings between equilibrium states of liquid systems using normalizing flows

Special Collection: Molecular Dynamics, Methods and Applications 60 Years after Rahman

Alessandro Coretti ; Sebastian Falkner ; Phillip L. Geissler ; Christoph Dellago

Normalizing flows as an enhanced sampling method for atomistic supercooled liquids

Gerhard Jung*, Giulio Biroli and Ludovic Berthier

crystals

Targeted free energy estimation via learned mappings

Cite as: J. Chem. Phys. 153, 144112 (2020); doi: 10.1063/5.0018903
Submitted: 18 June 2020 • Accepted: 25 September 2020 •
Published Online: 13 October 2020

Peter Wirnsberger,[a] Andrew J. Ballard,[a] George Papamakarios, Stuart Abercrombie, Sébastien Racanière, Alexander Pritzel, Danilo Jimenez Rezende, and Charles Blundell

quantum chromodynamcics

Perspective | Published: 04 August 2023

Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics

Kyle Cranmer, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende & Phiala E. Shanahan

▷ However, limits arise when it becomes difficult to obtain $\rho_\theta(x) \approx \rho_*(x)$ through training, either because of the **complexity of the landscape** or because of the **dimension**

[Del Debbio et al PRD 2021; Ciarella et al MLST 2023; Grenioux, Durmus, Moulines & MG ICML 2023]

Two directions of current research

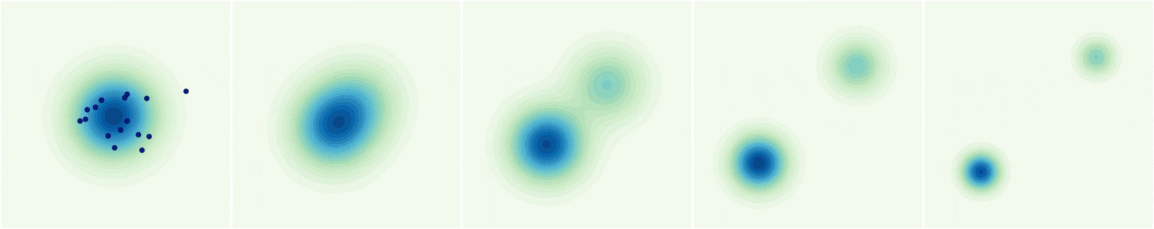
▷ Coupling generative models with traditional enhanced sampling algorithms

Enhancing tempering algorithms [Arbel et al. 2102.07501, Invernizzi et al. 2210.14104, Zhanger, Zahm et al. 2402.17943]

Non-equilibrium transport sampler [Vargas et al. 2307.01050, Albergo et al. 2410.02711]

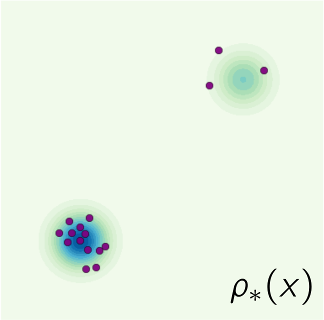
Collective-variable based samplers [Tamagnone at al. 2406.01524, Schönle et al. 2405.18160]

▷ Leveraging the more powerful diffusion models (continuous in time models) ← discussed next



[Vargas et al. arXiv:2302.13834, RDMC - Huang et al. 2307.02037, Berner et al. arXiv:2307.01198,
Vargas et al. arXiv:arXiv:2307.01050, SLIPS – Grenioux et al. 2402.10758, Akhound-Sadegh et al. 2402.06121]

Sampling by denoising

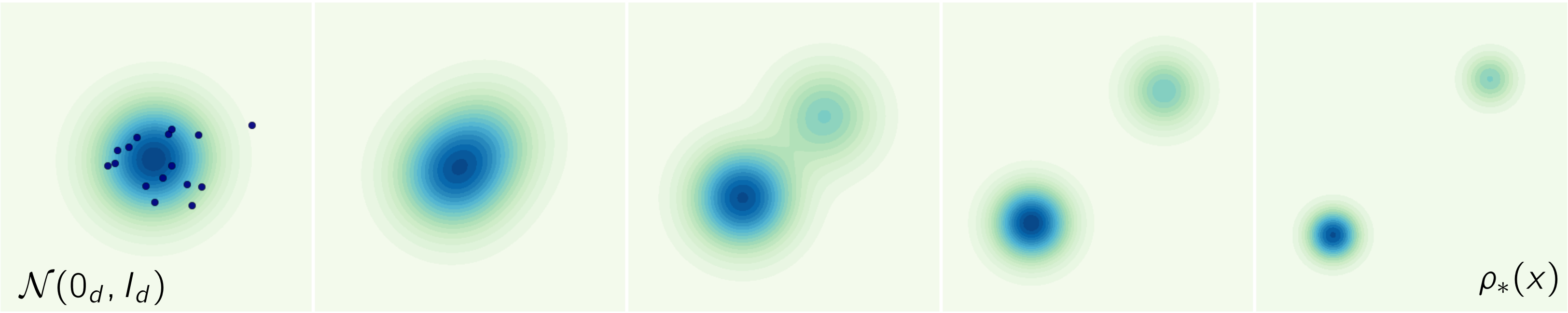


Animations by Louis Grenioux!

Sampling by denoising

Time marginals: $p_t(x) = \int \mathcal{N}\left(x; e^{-t}x_0, (1 - e^{-2t})\, I_d\right) \rho_*(x_0)\, \mathrm{d}x_0$

Forward process: $\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t$

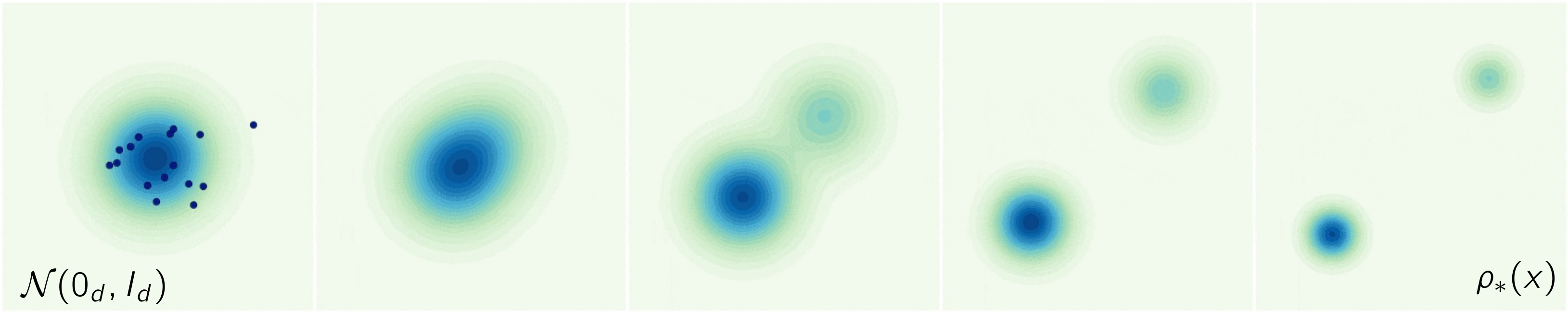


Animations by Louis Grenioux!

Sampling by denoising

Time marginals: $p_t(x) = \int \mathcal{N}(x; e^{-t}x_0, (1-e^{-2t})\, I_d)\, \rho_*(x_0)\, dx_0$

Forward process: $dX_t = -X_t dt + \sqrt{2} dW_t$



Backward « denoising process » $dY_t = (Y_t + \nabla \log p_{T-t}(Y_t))dt + \sqrt{2}dB_t$

non-tractable score function

▷ In diffusion models, score learned from an extensive set of samples from $\rho_*(x)$. [Ho et al. 2006.11239 NeurIPS 2020, Song et al. 2011.13456 ICLR 2021]

▷ Can this idea be adapted to sampling where no data is available a priori?

- In stat mech community, theorems for score estimation with message passing algorithms on canonical models [El Alaoui et al.. FOCS 2022 & arXiv:2310.08912, Montanari arxiv/2305.10690, Ghio et al. PNAS 2024]

What about an arbitrary target?

Investigated strategies to estimate the score

▷ **Non parametric:** Monte Carlo estimation of the score in Monte Carlo diffusion samplers

Saremi et al. "Chain of Log-Concave Markov Chains," ICLR 2024
Huang et al. "Reverse Diffusion Monte Carlo," ICLR 2024
Grenioux*, Noble*, MG, and Oliviero Durmus. "Stochastic Localization via Iterative Posterior Sampling." ICML 2024

$$\nabla \log p_t(y) = -\frac{y - e^{-t}\,\mathbb{E}[X_0 | X_t = y]}{(1 - e^{-2t})} \quad \text{where} \quad X_t \overset{\mathcal{L}}{=} X_0 e^{-t} + \sqrt{1 - e^{-2t}}\, Z$$

Computing the score = computing an expectation with respect to the posterior $q_t(x_0|y) \propto \underbrace{p_t(y|x_0)}_{\mathcal{N}\left(y;\, e^{-t}x_0,\, (1-e^{-2t})\, I_d\right)} \rho_*(x_0)$

▷ **Parametric:** Variational inference meets diffusion models

Zhang, et al. "Path Integral Sampler: A Stochastic Control Approach For Sampling." ICLR 2022
Vargas, et al. "Denoising Diffusion Samplers." ICLR 2023
Richter, et al. "Improved sampling via learned diffusions." ICLR 2024
Noble*, Grenioux* et MG and Oliviero Drumus. "Learned Reference-based Diffusion Sampler for multi-modal distributions" , ICLR 2025

path measures

$\mathbb{P}_* \longrightarrow$ fwd: $X_0 \sim \rho_*$ $\quad dX_t = -X_t dt + \sqrt{2} dW_t$ *(e.g. Ornstein-Uhlenbeck)*

bwd: $Y_0 \sim \mathcal{N}(0, I)$ $\quad dY_t = (Y_t + \nabla \log p_{T-t}(Y_t))dt + \sqrt{2} dB_t$

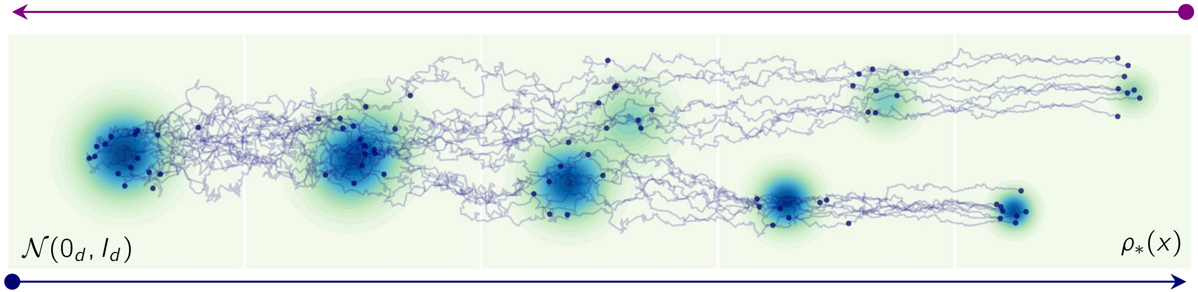
$\mathbb{P}_\theta \longrightarrow$ bwd: $Y_0 \sim \mathcal{N}(0, I)$ $\quad dY_t = (Y_t + s_t^\theta(Y_t))dt + \sqrt{2} dB_t$

Path space VI problem

$$\min_\theta \mathrm{KL}(\mathbb{P}_\theta || \mathbb{P}_*)$$

Monte Carlo diffusion samplers

Forward process: $\mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t$



Backward « denoising process » $\mathrm{d}Y_t = (Y_t + \nabla \log p_{T-t}(Y_t))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

▷ Main idea of Monte Carlo diffusion samplers: build statistical estimators of score to sample denoising process

▷ Tweedie's formula links the score to the denoising function

$$\nabla \log p_t(y) = -\frac{y - e^{-t}\,\mathbb{E}[X_0 | X_t = y]}{(1 - e^{-2t})}$$

where $X_t \overset{\mathcal{L}}{=} X_0 e^{-t} + \sqrt{1 - e^{-2t}}\, Z$

▷ Boils down to the MC estimation of the mean of the posterior $q_t(x_0|y) \propto \underbrace{p_t(y|x_0)}_{\mathcal{N}(y;\, e^{-t}x_0,\, (1-e^{-2t})\, I_d)}\, \rho_*(x_0)$

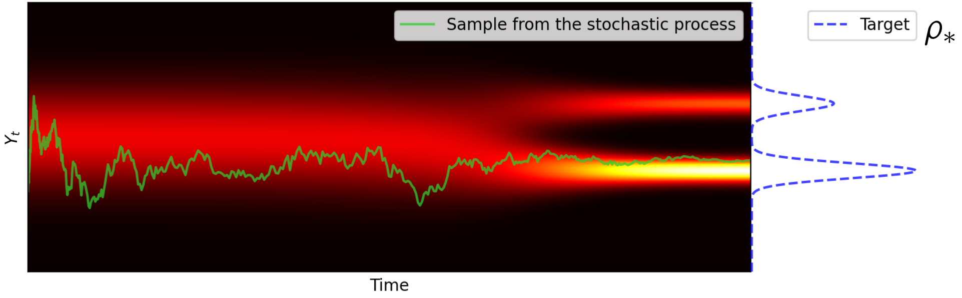
Saremi et al. "Chain of Log-Concave Markov Chains," ICLR 2024
Huang et al. "Reverse Diffusion Monte Carlo," ICLR 2024

Stochastic Localization via Iterative Posterior Sampling (SLIPS)[15]

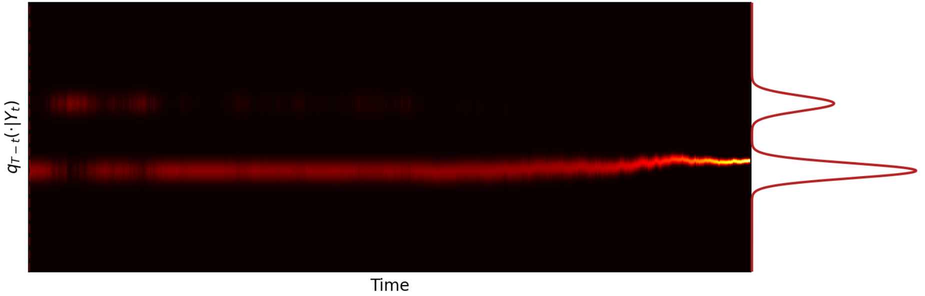
- Among Monte Carlo diffusion samplers, SLIPS estimates the posterior mean via **Markov Chain Monte Carlo**

$$\mathrm{d}Y_t = (Y_t + \nabla \log p_{T-t}(Y_t))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$$

$$p_t(y) = \int \mathcal{N}\big(y; e^{-t}x_0, (1-e^{-2t})\, I_d\big)\, \rho_*(x_0)\, \mathrm{d}x_0$$



$$q_t(x_0|y) \propto \mathcal{N}\big(y; e^{-t}x_0, (1-e^{-2t})\, I_d\big)\, \rho_*(x_0)$$

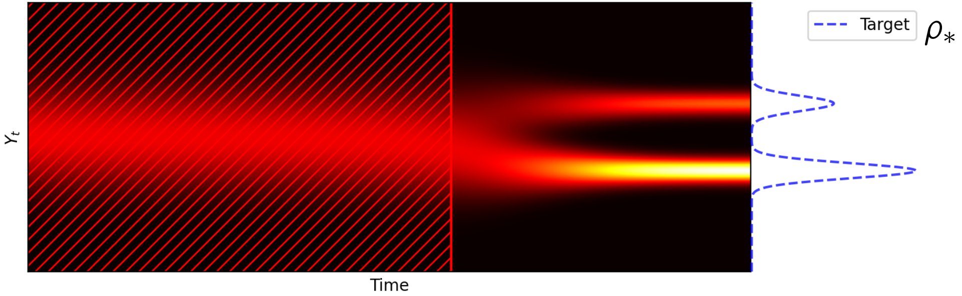


Stochastic Localization via Iterative Posterior Sampling (SLIPS)[16]

- ▷ Among Monte Carlo diffusion samplers, SLIPS estimates the posterior mean via **Markov Chain Monte Carlo**

$$\mathrm{d}Y_t = (Y_t + \nabla \log p_{T-t}(Y_t))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$$

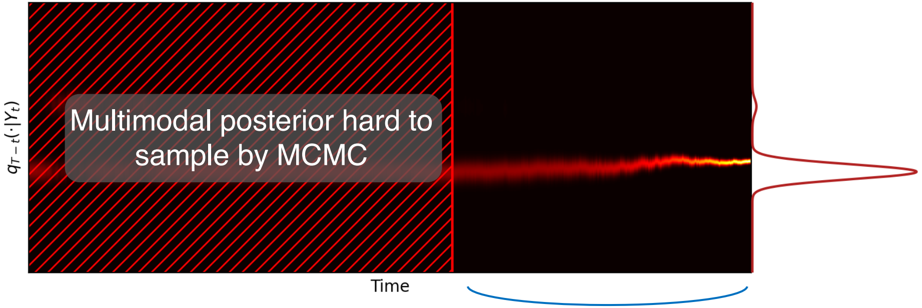
$$p_t(y) = \int \mathcal{N}\left(y; e^{-t}x_0, (1-e^{-2t})\, I_d\right) \rho_*(x_0)\, \mathrm{d}x_0$$



$Y_{t_0} \sim p_{t_0}$

t_0

$$q_t(x_0|y) \propto \mathcal{N}\left(y; e^{-t}x_0, (1-e^{-2t})\, I_d\right) \rho_*(x_0)$$



posterior mean (→ score) easy to compute by MCMC

- ▷ Commitment time ~ speciation time by Marc, Giulio & co-authors
- ▷ How to initialize the backward process integration at $t_0 > 0$?

[Biroli & Mézard J. Stat. Mech 2023 , Biroli et al Nat. Comm. 2024]

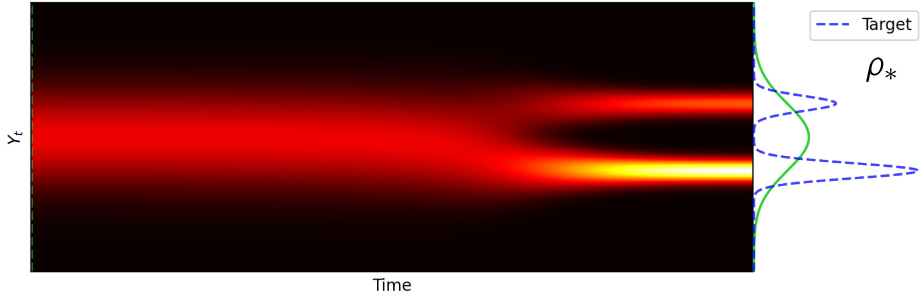
Duality of log-concavity & SLIPS method

▷ Score can be MCMC estimated for t **large enough**

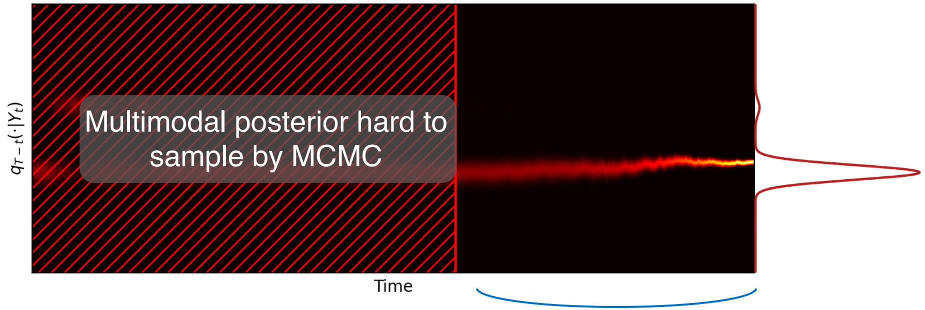
$$\nabla \log p_t(y) = -\frac{y - e^{-t}\,\mathbb{E}[X_0|X_t = y]}{(1 - e^{-2t})}$$

$$\downarrow$$

$$\nabla \widehat{\log p_t}(y) = -\frac{y - e^{-t}\,\widehat{X_0(y)}}{(1 - e^{-2t})}$$



t_0



posterior mean (→ score) easy to compute by MCMC

Grenioux, Noble, MG, and Oliviero Durmus. "Stochastic Localization via Iterative Posterior Sampling." ICML 2024

Duality of log-concavity & SLIPS method

- Score can be MCMC estimated for t **large enough**

$$\nabla \log p_t(y) = -\frac{y - e^{-t}\,\mathbb{E}[X_0|X_t = y]}{(1 - e^{-2t})}$$

$$\widehat{\nabla \log p_t(y)} = -\frac{y - e^{-t}\,\widehat{X_0(y)}}{(1 - e^{-2t})}$$

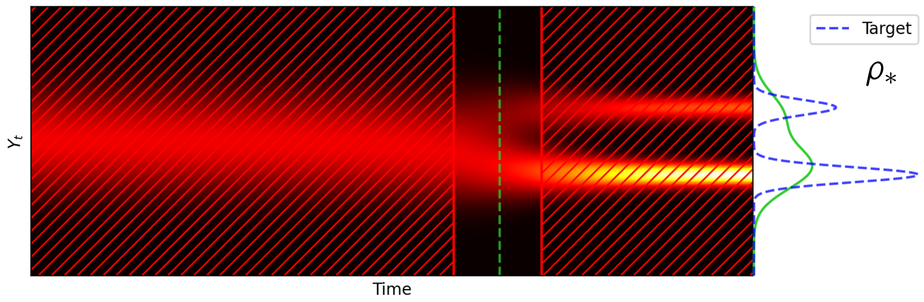
- Step 1: The estimated score is used in Langevin to initialize the backward SDE integration at $t_0 > 0$ **small enough!**

$$Y_{t_0}^{k+1} = Y_{t_0}^{k} + \gamma\widehat{\nabla \log p_{t_0}}(Y_{t_0}^{k}) + \sqrt{2\gamma}\, Z^{k+1}$$

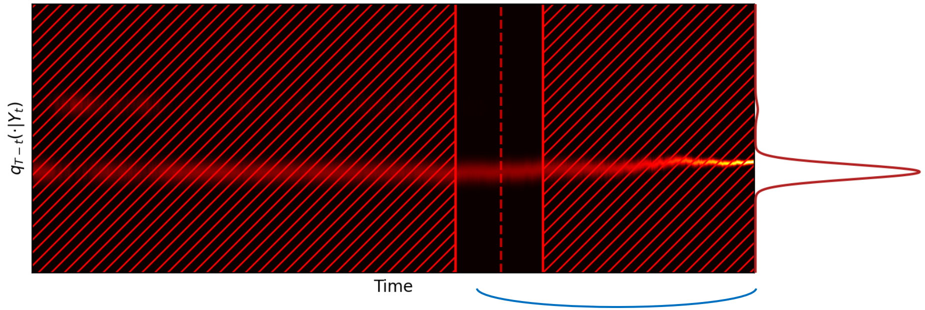
- Step 2: A discretized version of the SDE is integrated using the estimated scores

$$Y_{t_{n+1}} = Y_{t_n} + \Delta t\big(Y_{t_n} + \widehat{\nabla \log p_{T-t_n}}(Y_{t_n})\big) + \sqrt{2\Delta t}\, Z^{n+1}$$

marginal easily MCMC sampled



t_0

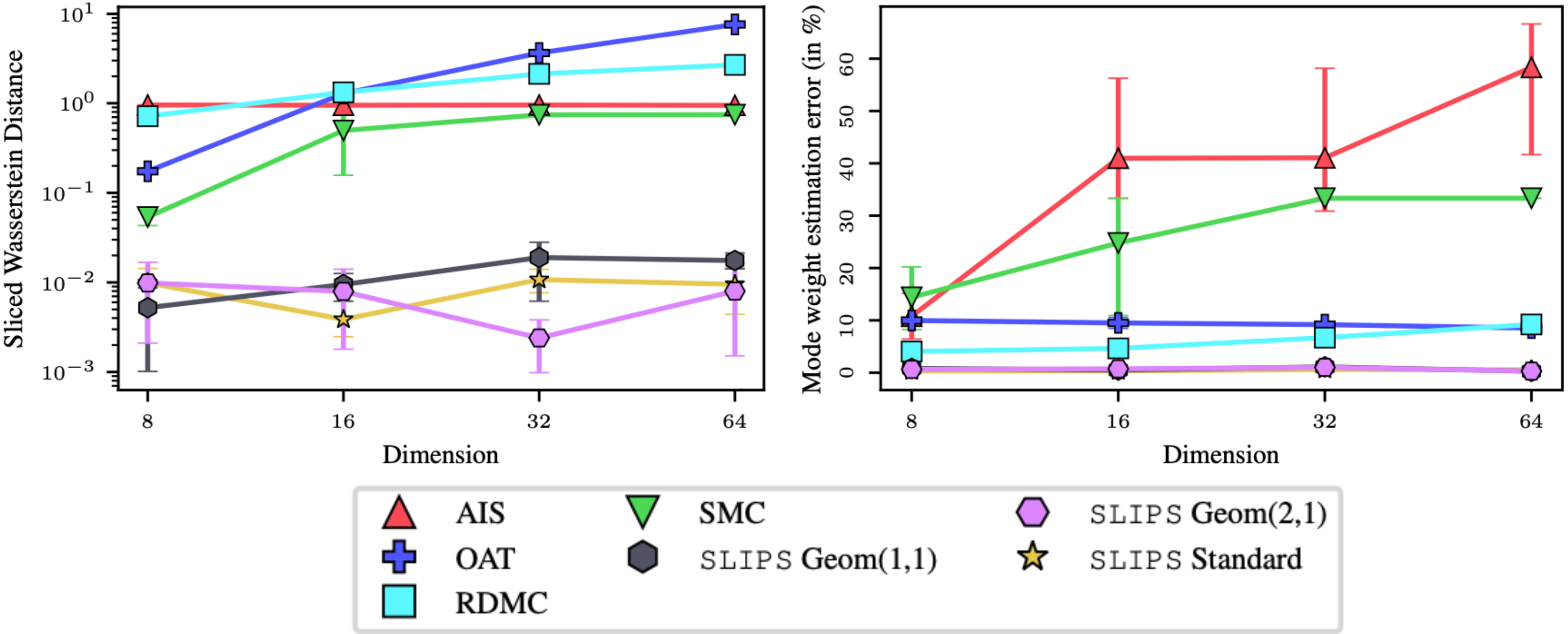


posterior mean (→ score) easy to compute by MCMC

→ Y_T is an approximate sample from ρ_*

Grenioux, Noble, MG, and Oliviero Durmus. "Stochastic Localization via Iterative Posterior Sampling." ICML 2024

Experiments on Gaussian mixtures



* SMC without adaptive resampling

Grenioux, Noble, MG, and Oliviero Durmus. "Stochastic Localization via Iterative Posterior Sampling." ICML 2024

Variational inference meets diffusion models

▷ To be tractable the VI problem on path measures $\min_\theta \mathrm{KL}(\mathbb{P}_\theta || \mathbb{P}_*)$ must be rewritten using a reference process

different init

Target $\mathbb{P}_*$ ➡ fwd: $X_0 \sim \rho_*$ $\quad \mathrm{d}X_t = -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t$ *(e.g. Ornstein-Uhlenbeck)*

bwd: $Y_0 \sim \mathcal{N}(0, I)$ $\quad \mathrm{d}Y_t = (Y_t + \nabla \log p_{T-t}(Y_t))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

same noising

Ref $\mathbb{P}_{\mathrm{ref}}$ ➡ fwd: $X_0^{\mathrm{ref}} \sim \rho_{\mathrm{ref}} \neq \rho_*$ $\quad \mathrm{d}X_t^{\mathrm{ref}} = -X_t^{\mathrm{ref}} \mathrm{d}t + \sqrt{2}\mathrm{d}W_t$ *(e.g. Ornstein-Uhlenbeck)*

bwd: $Y_0^{\mathrm{ref}} \sim \mathcal{N}(0, I)$ $\quad \mathrm{d}Y_t^{\mathrm{ref}} = (Y_t^{\mathrm{ref}} + \nabla \log p_{T-t}^{\mathrm{ref}}(Y_t^{\mathrm{ref}}))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

Variational $\mathbb{P}_\theta$ ➡ bwd: $Y_0^\theta \sim \mathcal{N}(0, I)$ $\quad \mathrm{d}Y_t^\theta = (Y_t^\theta + \nabla \log p_{T-t}^{\mathrm{ref}}(Y_t^\theta) + \phi_t^\theta(Y_t^\theta))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

learnable control

▷ Using the chain rule and Girsanov's theorem one obtains:

$$\mathrm{KL}(\mathbb{P}_\theta || \mathbb{P}_*) = \mathrm{KL}(\mathbb{P}_\theta || \mathbb{P}_{\mathrm{ref}}) + \mathbb{E}_{\mathbb{P}_\theta^T}\left[\log \frac{\rho_{\mathrm{ref}}(Y_T)}{\rho_*(Y_T)}\right] = \mathbb{E}_{\mathbb{P}_\theta}\left[\int_0^T \sqrt{2}\|\phi_t^\theta(Y_t)\|^2 \mathrm{d}t + \log \frac{\rho_{\mathrm{ref}}(Y_T)}{\rho_*(Y_T)}\right]$$

tractable loss!

Zhang, et al. Path Integral Sampler: A Stochastic Control Approach For Sampling. ICLR 2022
Vargas, et al. Denoising Diffusion Samplers. ICLR 2023
Richter, et al. Improved sampling via learned diffusions. ICLR 2024
Noble*, Grenioux* et al. Learned Reference-based Diffusion Sampler for multi-modal distributions. ICLR 2025

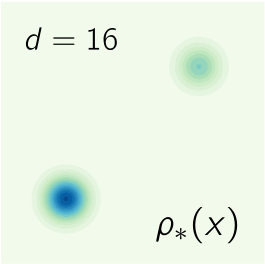
Choosing a simple initial reference keeps scores tractable

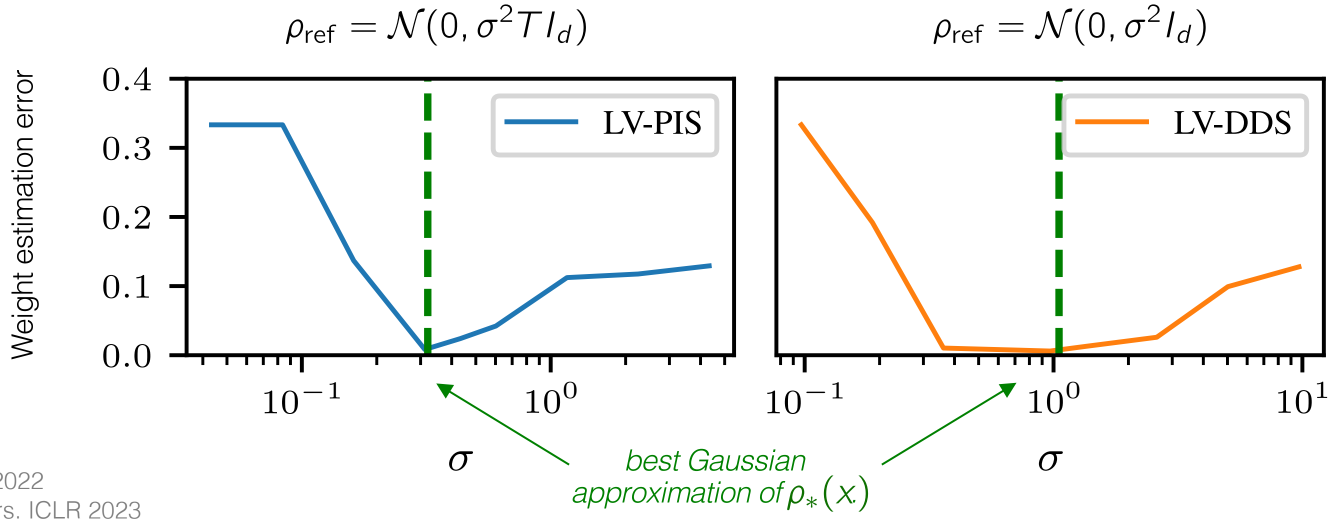
$= \mathcal{N}(0, \sigma^2 I)$ *Gaussian init → ref marginals Gaussian*

Ref $\mathbb{P}_{\text{ref}}$ → $X_0^{\text{ref}} \sim \rho_{\text{ref}}$ $\quad \mathrm{d}X_t^{\text{ref}} = -X_t^{\text{ref}}\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$ *(e.g. Ornstein-Uhlenbeck)*

bwd: $Y_0^{\text{ref}} \sim \mathcal{N}(0, I)$ $\quad \mathrm{d}Y_t^{\text{ref}} = (Y_t^{\text{ref}} + \nabla \log p_{T-t}^{\text{ref}}(Y_t^{\text{ref}}))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$

Variational $\mathbb{P}_\theta$ → bwd: $Y_0^\theta \sim \mathcal{N}(0, I)$ $\quad \mathrm{d}Y_t^\theta = (Y_t^\theta + \nabla \log p_{T-t}^{\text{ref}}(Y_t^\theta) + \phi_t^\theta(Y_t^\theta))\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$



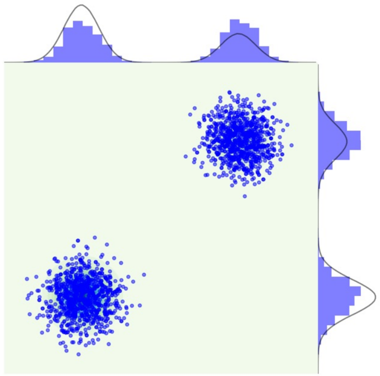


Zhang, et al. Path Integral Sampler ICLR 2022
Vargas, et al. Denoising Diffusion Samplers. ICLR 2023
Noble*, Grenioux* et al. Learned Reference-based Diffusion Sampler for multi-modal distributions, ICLR 2025

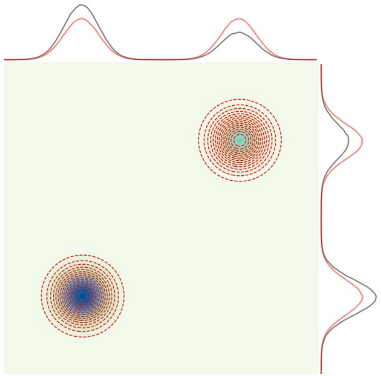
Learned Reference-based Diffusion Sampling (LRDS)

- ▷ Idea: learn the reference process from approximate samples to increase robustness
- ▷ In practice: $\rho_{\mathrm{ref}}^{\phi}$ as a Gaussian Mixture $\implies$ $\nabla \log \rho_{\mathrm{ref}}^{\phi}$ analytically tractable & well suited for multimodal ρ_*
  - Step 1: Run a local MCMC and fit ϕ parameters of $\rho_{\mathrm{ref}}^{\phi}$
  - Step 2: Fit parameters θ of the control term by minimizing VI objective

$$\mathrm{KL}(\mathbb{P}_\theta || \mathbb{P}_*) = \mathbb{E}_{\mathbb{P}_\theta}\left[\int_0^T \sqrt{2}\|\phi_t^\theta(Y_t)\|^2 \mathrm{d}t + \log \frac{\rho_{\mathrm{ref}}(Y_T)}{\rho_*(Y_T)}\right]$$



local-MCMC samples



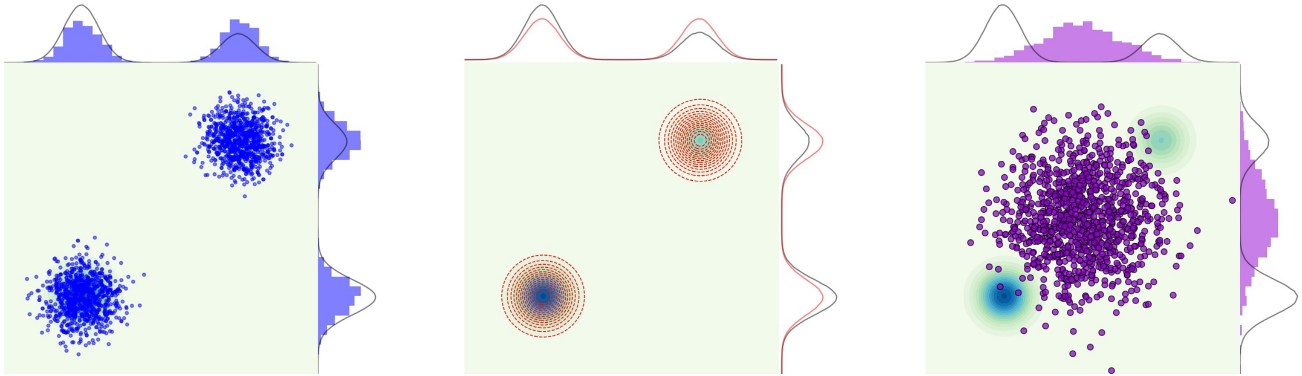
fitted GM $\rho_{\mathrm{ref}}^{\phi}$

Noble*, Grenioux* et al. Learned Reference-based Diffusion Sampler for multi-modal distributions, ICLR 2025

Learned Reference-based Diffusion Sampling (LRDS)

- ▷ Idea: learn the reference process from approximate samples to increase robustness
- ▷ In practice: ρ_{ref}^{ϕ} as a Gaussian Mixture $\implies \nabla \log \rho_{\text{ref}}^{\phi}$ analytically tractable & well suited for multimodal ρ_*
  - Step 1: Run a local MCMC and fit ϕ parameters of ρ_{ref}^{ϕ}
  - Step 2: Fit parameters θ of the control term by minimizing VI objective

$$\mathrm{KL}(\mathbb{P}_\theta || \mathbb{P}_*) = \mathbb{E}_{\mathbb{P}_\theta}\left[\int_0^T \sqrt{2}\|\phi_t^\theta(Y_t)\|^2 \mathrm{d}t + \log \frac{\rho_{\text{ref}}(Y_T)}{\rho_*(Y_T)}\right]$$



local-MCMC samples (mode weight blind) — *fitted GM ρ_{ref}^{ϕ}* — *backward ref process*

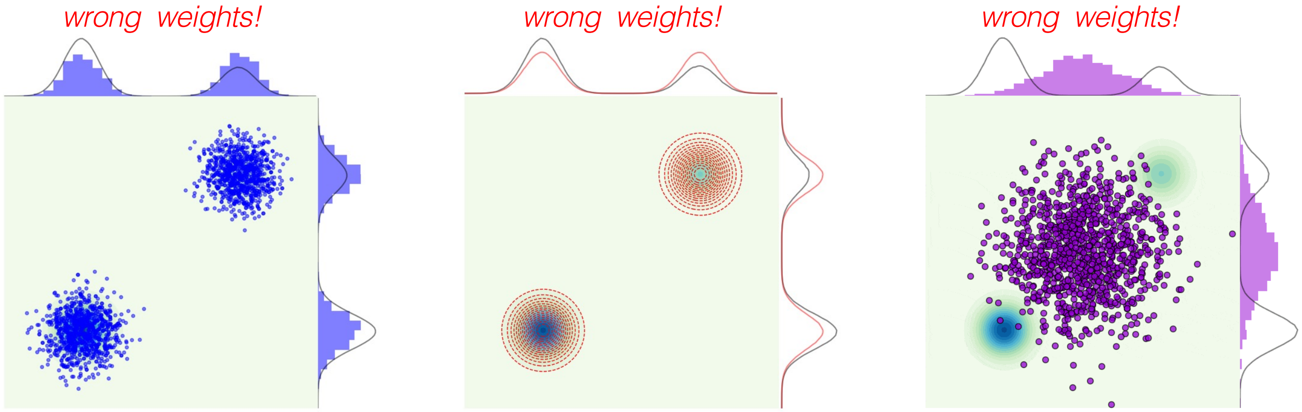
Learned Reference-based Diffusion Sampling (LRDS)

▷ Idea: learn the reference process from approximate samples to increase robustness

▷ In practice: $\rho_{\mathrm{ref}}^{\phi}$ as a Gaussian Mixture $\implies$ $\nabla \log \rho_{\mathrm{ref}}^{\phi}$ analytically tractable & well suited for multimodal ρ_*

- Step 1: Run a local MCMC and fit ϕ parameters of $\rho_{\mathrm{ref}}^{\phi}$
- Step 2: Fit parameters θ of the control term by minimizing VI objective

$$\mathrm{KL}(\mathbb{P}_\theta || \mathbb{P}_*) = \mathbb{E}_{\mathbb{P}_\theta}\left[\int_0^T \sqrt{2}\|\phi_t^\theta(Y_t)\|^2 \mathrm{d}t + \log \frac{\rho_{\mathrm{ref}}(Y_T)}{\rho_*(Y_T)}\right]$$



local-MCMC samples (mode weight blind) | *fitted GM $\rho_{\mathrm{ref}}^{\phi}$* | *backward ref process*

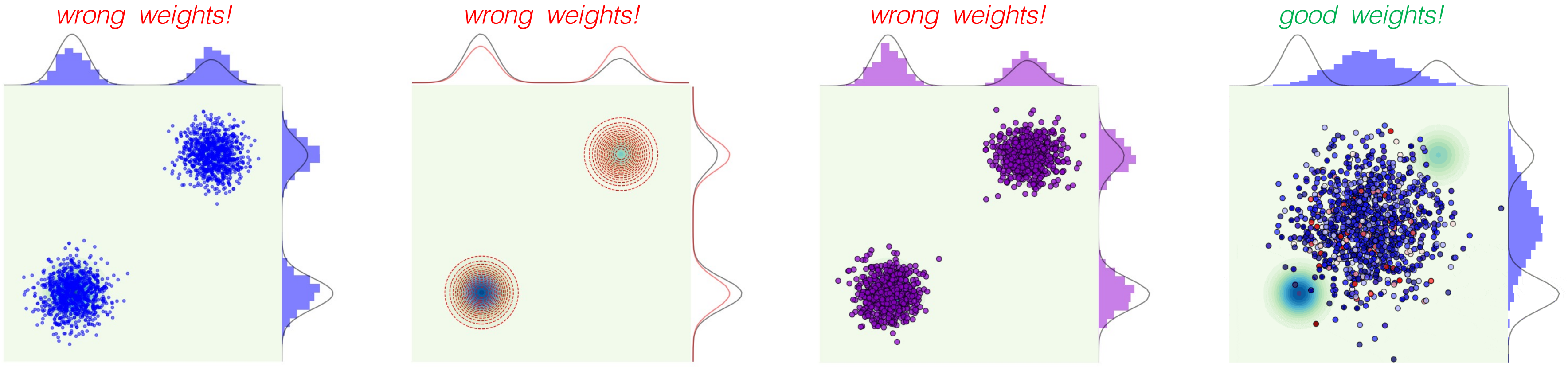
Noble*, Grenioux* et al. Learned Reference-based Diffusion Sampler for multi-modal distributions, ICLR 2025

Animations by Louis Grenioux!

Learned Reference-based Diffusion Sampling (LRDS)

- ▷ Idea: learn the reference process from approximate samples to increase robustness
- ▷ In practice: ρ_{ref}^{ϕ} as a Gaussian Mixture $\implies$ $\nabla \log \rho_{\text{ref}}^{\phi}$ analytically tractable & well suited for multimodal ρ_*
  - Step 1: Run a local MCMC and fit ϕ parameters of ρ_{ref}^{ϕ}
  - Step 2: Fit parameters θ of the control term by minimizing VI objective

$$\mathrm{KL}(\mathbb{P}_\theta||\mathbb{P}_*) = \mathbb{E}_{\mathbb{P}_\theta}\left[\int_0^T \sqrt{2}\|\phi_t^\theta(Y_t)\|^2 \mathrm{d}t + \log \frac{\rho_{\text{ref}}(Y_T)}{\rho_*(Y_T)}\right]$$



local-MCMC samples (mode weight blind) | fitted GM ρ_{ref}^{ϕ} | backward ref process | controlled $\mathbb{P}_\theta$ backward process

Promising numerical results

▷ Gaussian mixtures in increasing dimension

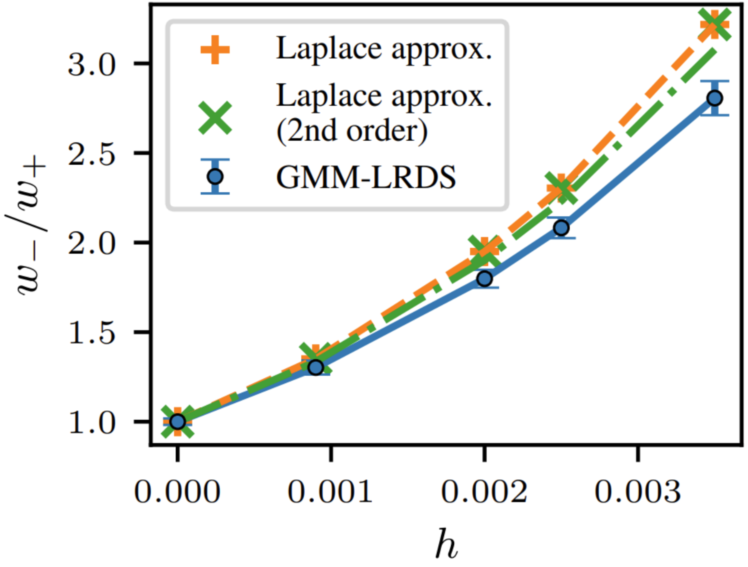
Absolute mode weight estimation error

Algorithm	$d = 16 \downarrow$	$d = 32 \downarrow$	$d = 64 \downarrow$
SMC	11.4%±9.1%	15.8%±8.5%	15.2%±7.5%
RE	16.5%±1.3%	15.9%±1.4%	17.0%±1.4%
LV-PIS	6.0%±3.4%	33.2%±0.1%	33.0%±0.1%
LV-DDS	11.8%±9.3%	31.5%±2.9%	33.1%±0.1%
LV-DIS	14.6%±1.0%	16.9%±1.1%	16.7%±0.7%
LV-CMCD	36.8%±18.9%	42.3%±24.4%	27.7%±22.6%
iDEM	33.3%±0.0%	66.7%±0.0%	11.7%±0.4%
PDDS	**0.8%±0.6%**	66.7%±0.0%	N/A
GMM-LRDS	1.7%±0.6%	**2.7%±0.8%**	**4.1%±0.6%**

orange – uninformative estimation
red – mode collapse

▷ 1d Φ^4 field theory on grid of size L=32

Mode weight ratio



▷ Reference initial distribution can be extended to an **Energy Based Model** for further flexibility but at the price of learning the score of the reference process as well

Noble*, Grenioux* et al. Learned Reference-based Diffusion Sampler for multi-modal distributions, ICLR 2025

Thanks for your attention!

- ▷ Progress in generative modelling can be used to facilitate the sampling of metastable systems.
- ▷ Limits arise to the scalability of a vanilla implementation going directly after the target distribution.
- ▷ Current research efforts seek to combine traditional enhance sampling with generative modelling and leverage the powerful diffusion models

Thanks to collaborators on these topics:

Christoph Schönle, Louis Grenioux, Maxence Noble & Alain Oliviero Durmus (École Polytechnique)

Grant Rotskoff (Stanford)

Eric Vanden-Eijnden (Courant Institute, NYU)

Gabriel Stoltz & Tony Lelièvre (École Nationale des Ponts & Chaussés)

Ana Molina Taborda, Olga Lopez Acevedo (Universidad de Antioquia)

Pilar Cossio (Flatiron, CCM)

Samuel Tamagnone & Alessandro Laio (SISSA)

Davide Carbone (ENS)