On Sampling with Approximate Transport Maps

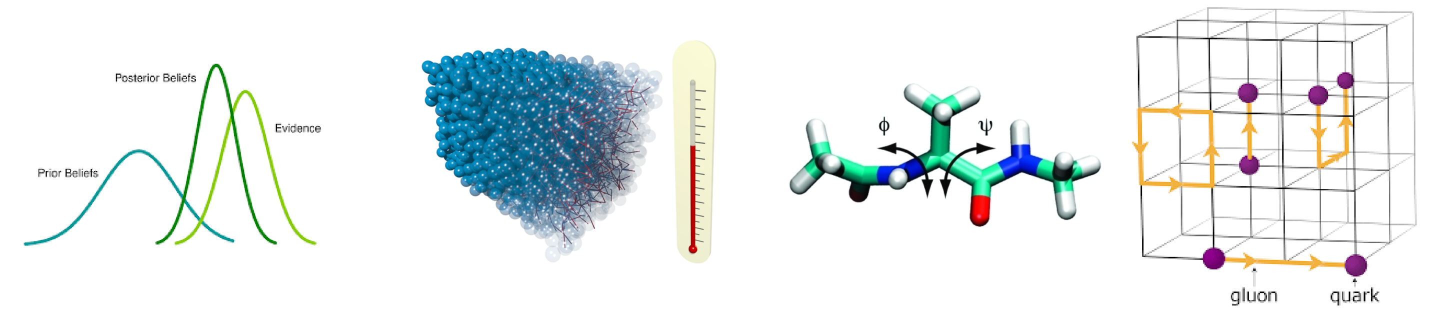
Louis Grenioux, Alain Oliviero Durmus, Éric Moulines, Marylou Gabrié

CMAP, École polytechnique, France

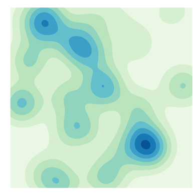
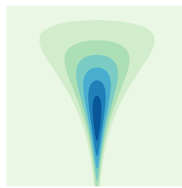
The sampling problem

Sampling is a **key** task in many domains

Bayesian Inference | Statistical Mechanics | Chemistry | Lattice QCD



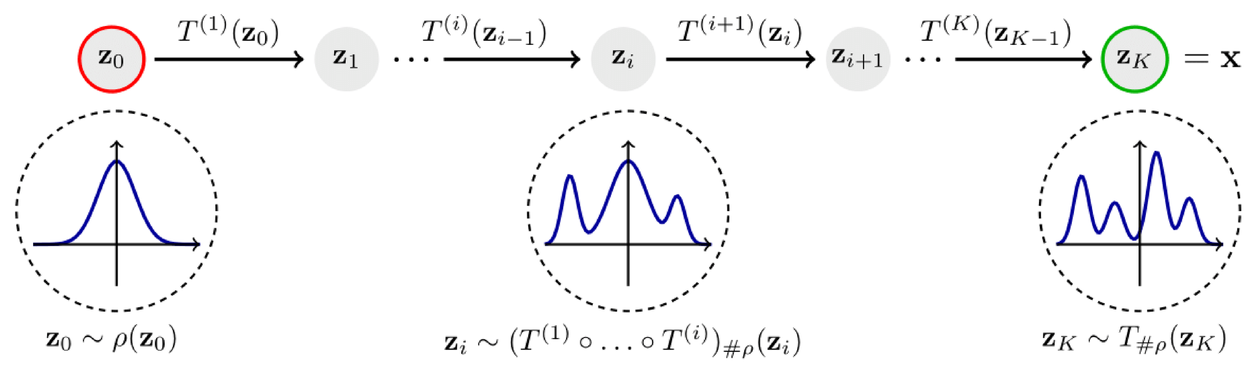
Given a target distribution π sampling can be difficult

multi-modal  or **ill-conditionned** 

and in very **high-dimension**.

Sampling with transport maps

To **ease sampling**, it was recently proposed to approximate the tagret distribution π as the push-forward of a simple distribution ρ through a transport map T : $T_{\#\rho} \simeq \pi$.

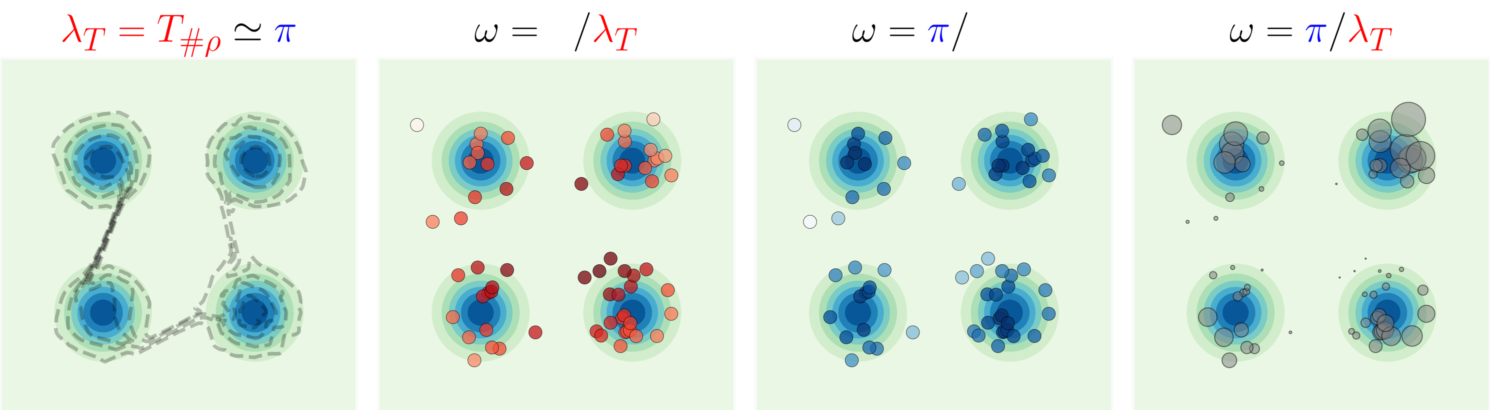


⚠ Learning T is difficult and leads to **approximation errors**.

✅ This can be mitigated by using T in **Monte-Carlo schemes**.

neural-IS [Müller et al., 2019] : flow as a proposal in IS

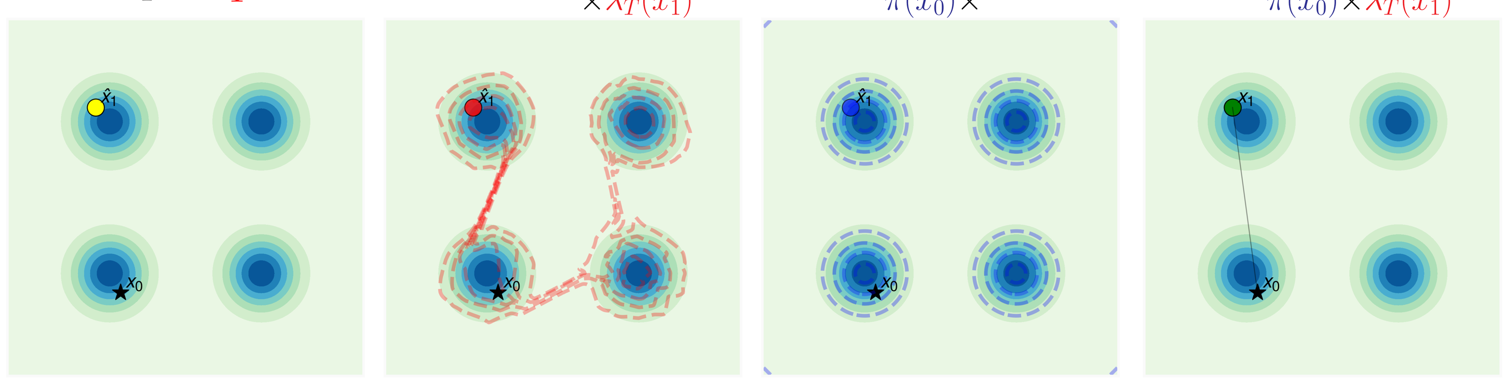
$\lambda_T = T_{\#\rho} \simeq \pi$ | $\omega = \quad / \lambda_T$ | $\omega = \pi /$ | $\omega = \pi / \lambda_T$



Expectations can be computed using $\mathbb{E}_{X \sim \pi}[f(X)] \simeq \sum_{i=1}^{N} \omega(X^{(i)}) f(X^{(i)})$

flow-MCMC [Gabrié et al., 2022] : flow as a proposal in MH

$\hat{x}_1 \sim \lambda_T$ | $\alpha = \frac{\times \lambda_T(x_0)}{\times \lambda_T(\hat{x}_1)}$ | $\alpha = \frac{\pi(\hat{x}_1) \times}{\pi(x_0) \times}$ | $\alpha = \frac{\pi(\hat{x}_1) \times \lambda_T(x_0)}{\pi(x_0) \times \lambda_T(\hat{x}_1)}$

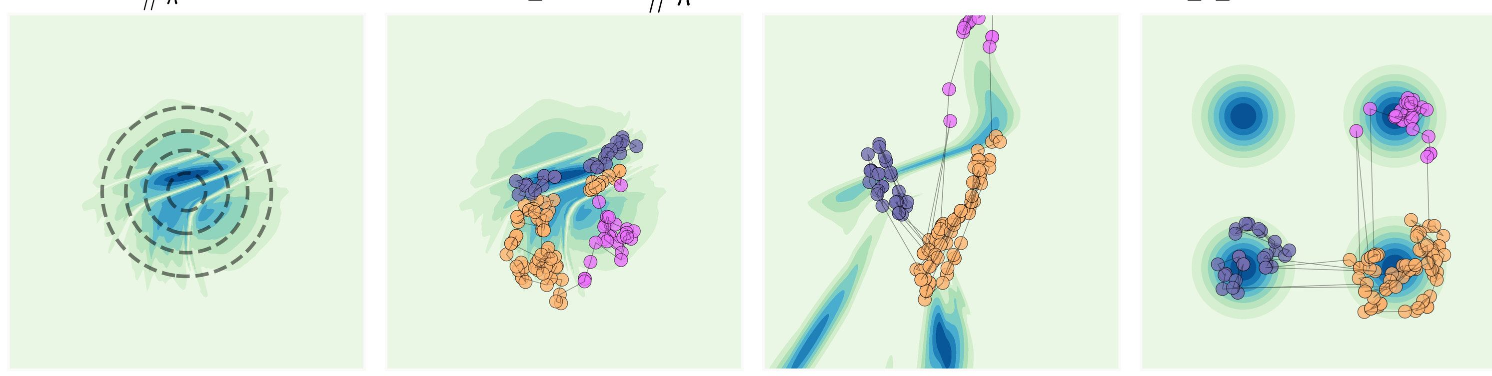


(1) Suggest $\hat{x}_k$ (2) Evaluate λ_T (3) Evaluate π (4) Set $x_k = \hat{x}_k$ with prob. α else $x_k = x_{k-1}$

Iterating (1) → (4) for N steps builds a Markov chain with π as invariant distribution

neutra-MCMC [Hoffman et al., 2019] : flow for reparametrization

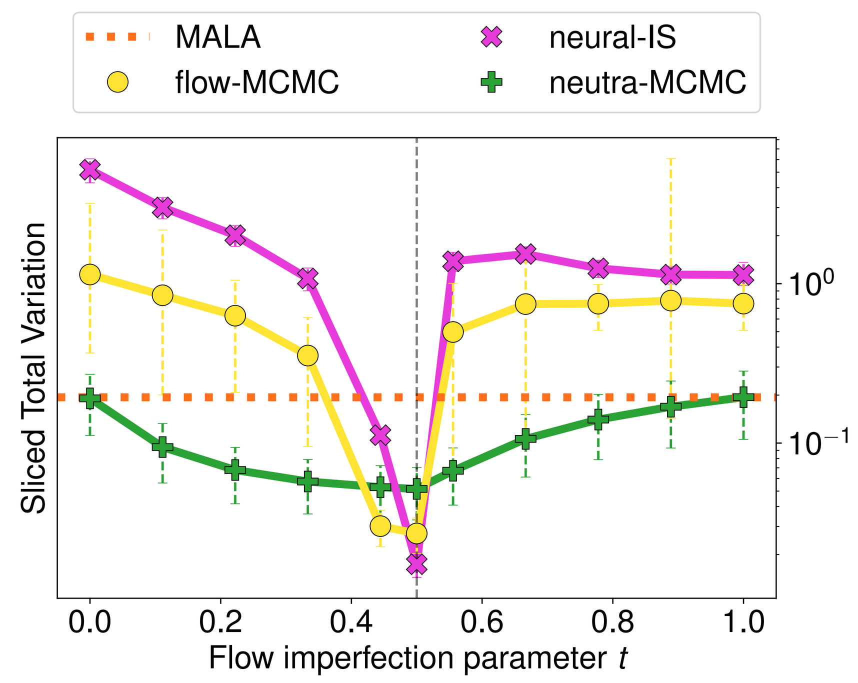
$T_{\#\pi}^{-1} \simeq \rho$ | Sample $T_{\#\pi}^{-1}$ | $\xrightarrow{T}$ | Approximate π



Contributions

- **Comparison** of the different algorithms depending on flow quality, multi-modality, poor conditioning and dimensionality
- **New theoretical result** on the mixing time of flow-MCMC/IMH
- Validation on **real-world experiments**

neutra-MCMC is more robust to imperfections

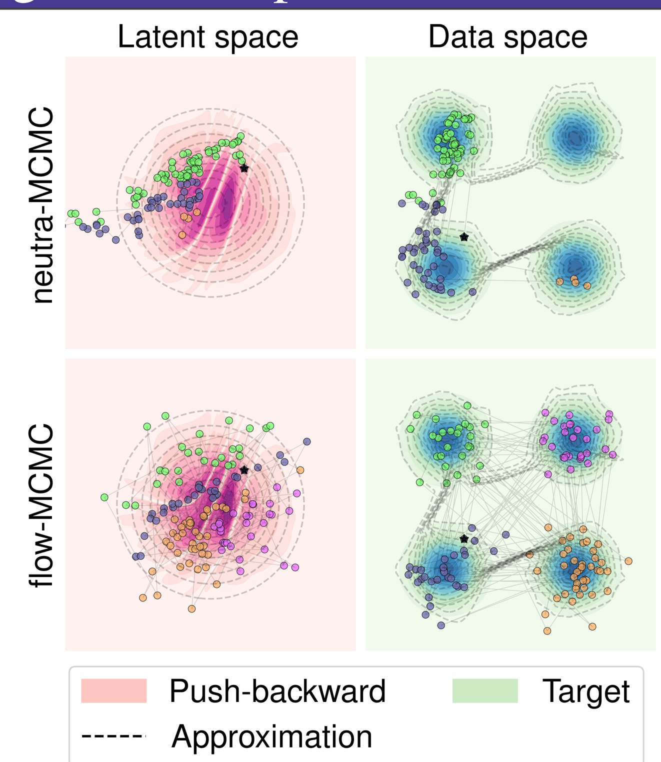


- π is a poorly conditioned Gaussian in dimension 128
- T_t is an analytical flow with quality parameter t
- **neutra-MCMC is less sensitive to t than flow-MCMC or neural-IS**

Real experiment : Confirmed on sparse logistic regression

neutra-MCMC doesn't ease sampling in latent space

- π is a mixture
- **T cannot transform something multimodal into something unimodal**
- This limitation is due to the constrains of the flow
- flow-MCMC/neural-IS **can mix between modes**



Real experiment : Confirmed on a molecular system and a field system

New flow-MCMC's mixing time bound

Assuming that π is **log-concave** and that the importance weights $\omega = \pi / \lambda$ verify

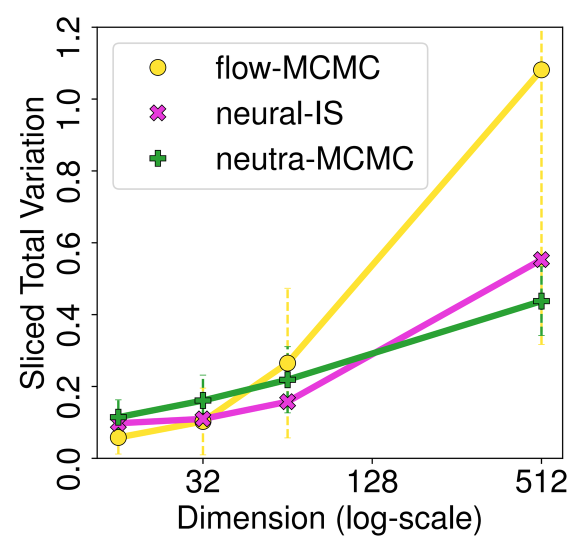
$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{B}(\mathbf{0}, \mathbf{R}), |\log \omega(\mathbf{x}) - \log \omega(\mathbf{y})| \leq \mathbf{C_R} \|\mathbf{x} - \mathbf{y}\|,$$

then the mixing time of **flow-MCMC** with proposal λ is bounded as

$$\tau_{mix}(\mu, \epsilon) \leq \beta C_R^2.$$

where β is a constant. As better λ lead to smaller C_R, **this results provides a quantitative bound depending on the quality of** λ.

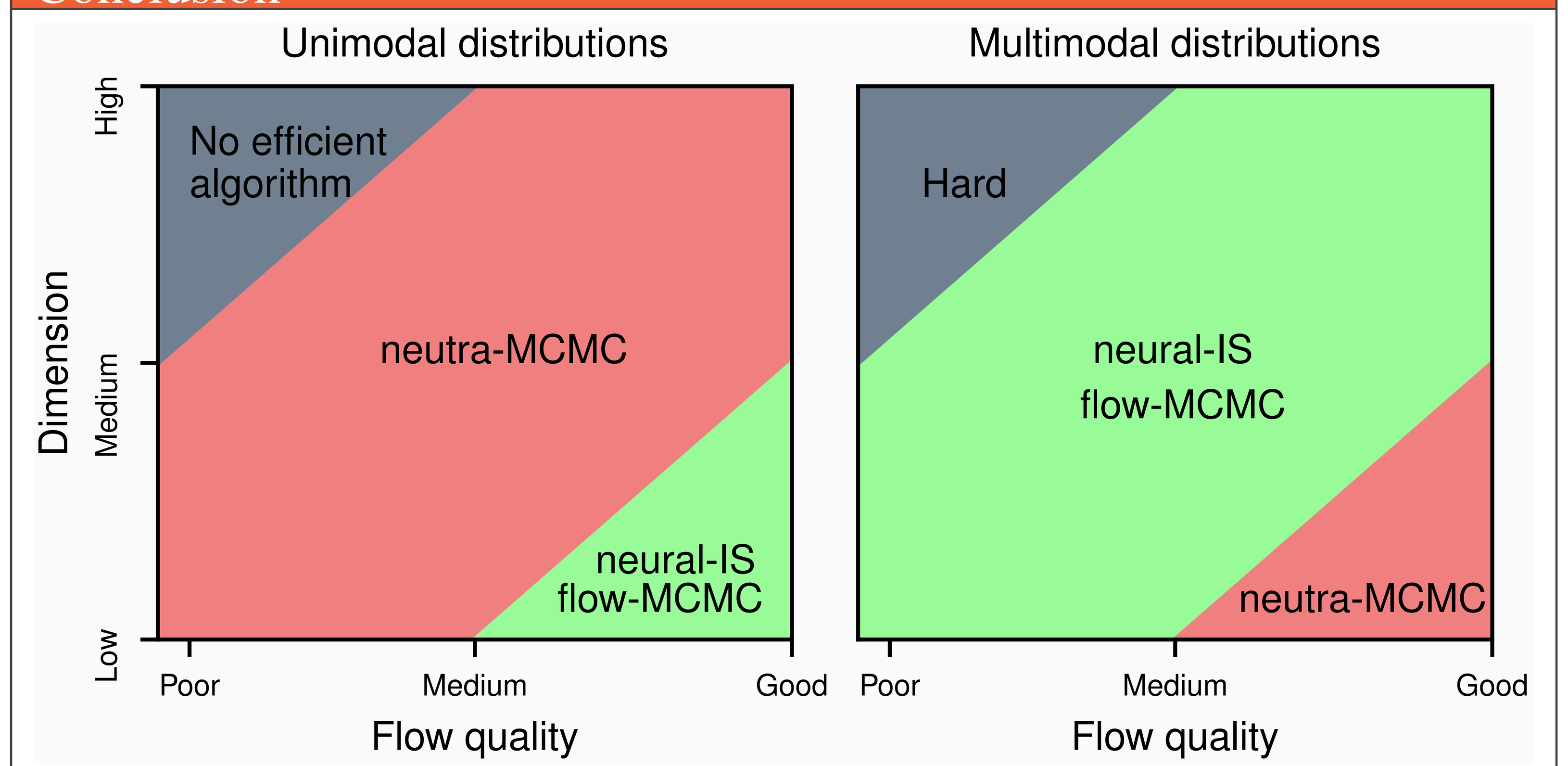
flow-MCMC doesn't scale in high-dimension



- π is a ill-conditioned distribution
- **flow-MCMC/neural-IS deteriorates faster than neutra-MCMC with d**
- The previous theorem applied with unit Gaussian as π and $\lambda = \mathcal{N}(0, (1+\epsilon)^2 I_d)$ gives that *the error* ϵ *should scale as* $\mathcal{O}(1/d)$ *at a fixed computational budget*

Real experiment : Confirmed on a field system and image data

Conclusion



References

[Gabrié et al., 2022] Gabrié, M., Rotskoff, G. M., and Vanden-Eijnden, E. (2022). Adaptive monte carlo augmented with normalizing flows. *Proceedings of the National Academy of Sciences*, 119(10):e2109420119.

[Grenioux et al., 2023a] Grenioux, L., Durmus, A., Éric Moulines, and Gabrié, M. (2023a). On sampling with approximate transport maps.

[Grenioux et al., 2023b] Grenioux, L., Éric Moulines, and Gabrié, M. (2023b). Balanced training of energy-based models with adaptive flow sampling.

[Hoffman et al., 2019] Hoffman, M. D., Sountsov, P., Dillon, J. V., Langmore, I., Tran, D., and Vasudevan, S. (2019). NeuTra-lizing Bad Geometry in Hamiltonian Monte Carlo Using Neural Transport. In *1st Symposium on Advances in Approximate Bayesian Inference, 2018 1–5*.

[Müller et al., 2019] Müller, T., Mcwilliams, B., Rousselle, F., Gross, M., and Novák, J. (2019). Neural importance sampling. *ACM Trans. Graph.*, 38(5).