Improving the evaluation of samplers on multi-modal targets

Louis Grenioux*[,1,2], Maxence Noble*[,1], Marylou Gabrié[2]

[1] CMAP, École Polytechnique, France, [2] LPENS, École Normale Supérieure, France

- Louis Grenioux : louis.grenioux@polytechnique.edu
- Maxence Noble : maxence.noble-bourillot@polytechnique.edu

Sampling from multi-modal distributions

Setting. We want **samples** from a target probability distribution π supported on $\mathbb{R}^d$, while only having access to its log-density $\log \pi$.

Multi-modality. The densities of multi-modal distributions have multiple local minima (called *modes*). This give rise to two main challenges **(1)** finding each mode and **(2)** Ensuring that the samples are drawn with the right mode proportions.

Mode finding. Because finding all the local minima of an unknown function is NP hard, we focus here on the second challenge.

Mode weight recovery : the underlying challenging task

Even with known modes, estimating their proportions remains challenging. *Markov Chain Monte Carlo* (MCMC) uses local transition kernels to guide particles toward high-density areas. While effective in concave regions, it often traps the chain within modes, limiting exploration across the full distribution.

Mode weights estimation. Given a partition $\{S_k\}_{k=1}^K$ of the support of a multi-modal target π with K modes, mode weights are defined as

$$w_k = \pi(S_k) = \int \mathbb{1}_{S_k}(x)\pi(\mathrm{d}x).$$

Mode weights give simple, interpretable insight into the global structure of a multi-modal target; we recommend estimating them whenever possible.

A simple yet challenging experiment

Target. We introduce the following bi-modal Gaussian mixture

$$\pi = \frac{2}{3}\mathcal{N}(-a\mathbf{1}_d, \Sigma_1) + \frac{1}{3}\mathcal{N}(+a\mathbf{1}_d, \Sigma_2)$$

where $a > 0$ and Σ_1, Σ_2 are two diagonal matrices with the same challenging conditioning. For this target, we consider the mode partition

$$S_1 = \{x \in \mathbb{R}^d : \mathcal{N}(x; -a\mathbf{1}_d, \Sigma_1) > \mathcal{N}(x; +a\mathbf{1}_d, \Sigma_2)\}\,,\ S_2 = S_1^c$$

which verifies $w_1 = 2/3$ if a is taken large.

This distribution is hard to sample from both at the local scale (difficult mode conditioning) and at the global scale (the modes are not equally balanced).

Metric. We report the bias and variance of the Monte Carlo estimate of w_1, while varying two hyperparameters: the **distance** between the modes a as well as the **dimension** d.

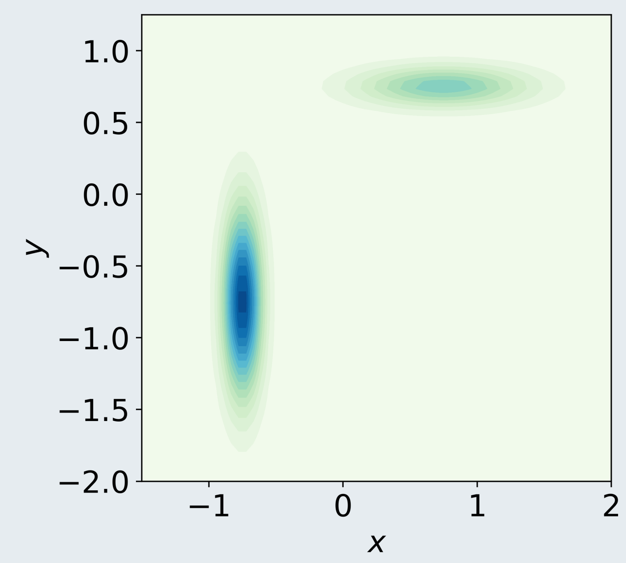
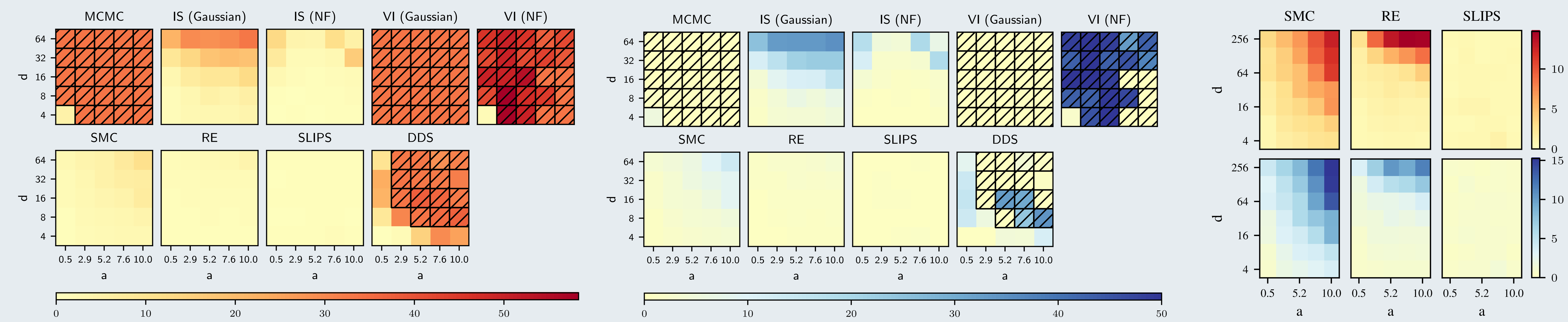


Figure 1: Target Gaussian mixture with $a = 1$, $d = 2$

Results on mode weight estimation for our proposed bi-modal Gaussian mixture



(Orange): Averaged absolute error of the estimation. **(Blue)**: Standard deviation of the estimation. **(Left)** Up to dimension 64. **(Right)** Up to dimension 256. *Hashed areas indicate settings with systematic mode collapse in the sampling process.*

Take-home messages

MCMC/VI methods struggle with multi-modal distributions (as they are limited to local exploration). **IS** suffers from increasing between-mode distance (leading to high variance). While **annealed MCMC** methods degrade with both dimensionality and mode separation, **diffusion-based samplers** offer a promising alternative, but they require for now careful tuning and their general applicability is still under exploration.

Evaluated sampling methods

- **Local MCMC** (MCMC): Standard and widely-used, but they are intrinsically limited to local exploration;
- **Importance Sampling** (IS): Estimates expectations based on reweighted samples taken from an easy-to-sample proposal. We test Gaussian and flow-based proposals that are fitted to true samples;
- **Variational Inference** (VI): Optimizes parametric distribution that approximates bests the target among a fixed variational family. We explore both Gaussian and flow-based variational families;
- *Annealed samplers*: Use a sequence of explicitly defined intermediate distributions to bridge an easy-to-sample distribution to the target.
  - **Sequential Monte Carlo** (SMC): Sequential MCMC with reweighting steps [Del Moral et al., 2006];
  - **Replica Exchange** (RE): Parallel MCMC chains with Metropolis-Hastings swaps [Swendsen and Wang, 1986].
- *Diffusion-based samplers*: Following diffusion models, these methods simulate a SDE bridging an easy-to-sample distribution to the target
  - **Stochastic Localization via Iterative Posterior Sampling** (SLIPS) : Estimates on-the-fly the intractable SDE drift using MCMC [Grenioux et al., 2024];
  - **Denoising Diffusion Sampler** (DDS) : Learns drift via a neural network solving a variational problem on path measures [Vargas et al., 2023].

References

[Del Moral et al., 2006] Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436.

[Grenioux et al., 2024] Grenioux, L., Noble, M., Gabrié, M., and Oliviero Durmus, A. (2024). Stochastic localization via iterative posterior sampling. In *The 41st International Conference on Machine Learning*.

[Swendsen and Wang, 1986] Swendsen, R. H. and Wang, J.-S. (1986). Replica monte carlo simulation of spin-glasses. *Phys. Rev. Lett.*, 57:2607–2609.

[Vargas et al., 2023] Vargas, F., Grathwohl, W. S., and Doucet, A. (2023). Denoising diffusion samplers. In *The Eleventh International Conference on Learning Representations*.