Stochastic Localization via Iterative Posterior Sampling

Louis Grenioux*, Maxence Noble*, Marylou Gabrié, Alain Oliviero Durmus

CMAP, École Polytechnique, France

Sampling with Markov Chain Monte Carlo (MCMC)

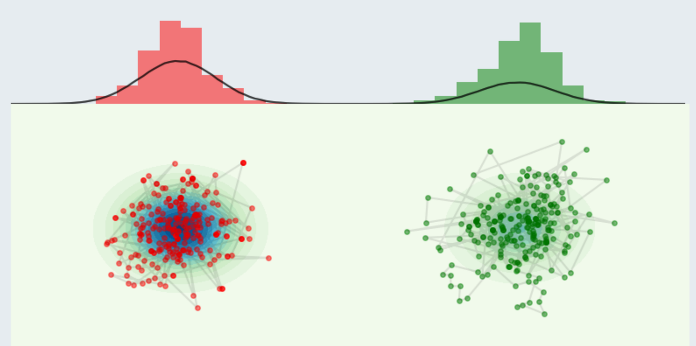
Setting. We want **samples** from a target probability distribution π supported on $\mathbb{R}^d$, while only having access to its log-density $\log \pi$.

MCMC approach. Relying on a π-invariant Markov kernel, these methods generate a sequence of samples $\{X_k\}_{k=1}^N$, which are approximately distributed according to π. A very popular example is the *Unadjusted Langevin Algorithm* (ULA), which uses the score $\nabla \log \pi$,

$$X_{k+1} = X_k + \gamma \nabla \log \pi(X_k) + \sqrt{2\gamma} Z_{k+1}, \ \gamma > 0, \ Z_{k+1} \sim \mathrm{N}(0, \mathrm{I}_d)$$

Suitable case	**Unfitting** case
• d small	• d large
• log-concave target ($\nabla^2 \log \pi \prec 0$)	• multi-modal target

Explanation: as the score $\nabla \log \pi$ is driving the ULA particles to the closest mode of π, escaping the attraction of the modes (to ensure *mixing*) takes a very long time.



Our goal

Sampling from **multi-modal** distributions in **high-dimension** with **MCMC** tools under a **low computational** cost.

Introducing our Stochastic Localization (SL) scheme

Observation process. Given $T > 0$, we consider the stochastic process $(Y_t)_{t \in [0,T]}$ defined by

$$Y_t = \alpha(t) X + \sigma W_t, \ X \sim \pi$$

where $(W_t)_{t \geq 0}$ is a Brownian motion and $\alpha : [0,T] \to \mathbb{R}_+$ is such that:

- $\alpha(0) = 0 \implies Y_0 \perp X$ (*full* noise at $t = 0$)
- $\alpha(t)/\sqrt{t} \xrightarrow{t \to T} \infty \implies$ the signal predominates over the noise
- α is strictly increasing $\implies$ the signal is increasingly informative

Stochastic Localization principle

It holds approximately that $Y_T/\alpha(T) \sim \pi$.

$\implies$ If we are able to simulate $(Y_t)_{t \in [0,T]}$, we can sample from π.

Sampling from the observation process with a Markovian scheme

Markovian projection. Under mild assumptions on π, $(Y_t)_{t \in [0,T]}$ has its marginals given by the *Stochastic Differential Equation* (SDE)

$$\mathrm{d}Y_t = \dot{\alpha}(t) u_t(Y_t) \mathrm{d}t + \sigma \mathrm{d}B_t \ , \ Y_0 = 0 \quad (1)$$

where $u_t(y) = \mathbb{E}[X | Y_t = y]$, the *denoiser*, is the expectation of the *posterior* of the SL model defined (up to a normalizing constant) by

$$\mathrm{d}q_t(x|y) \propto \mathrm{N}(y; \alpha(t)x, \sigma^2 t \, \mathrm{I}_d) \mathrm{d}\pi(x) \ .$$

$\implies$ We use the SDE (1) to simulate the observation process.

In practice. Given a time grid of $(0, T)$, we rather use a discretized version of the SDE (1) obtained with the *Euler–Maruyama* scheme.

SL as a denoising method. Denote by p_t the *marginal* distribution of Y_t. The denoiser is linked to the score of p_t by *Tweedie's formula*

$$u_t(y) = \frac{y}{\alpha(t)} + \frac{\sigma^2 t}{\alpha(t)} \nabla \log p_t(y) \ . \quad (2)$$

$\implies$ The SL model is actually a **score-based diffusion model**.

Defining our sampling algorithm

Estimating the denoiser. In practice, u_t cannot be exactly computed. Hence, we need to approximate it. In our setting, we estimate u_t with MCMC (*non-parametric* approach).

Stochastic Localization via Iterative Posterior Sampling

Given a time grid $\{t_k\}_{k=0}^K$ of $[t_0, T)$, with $t_0 \in (0,T)$, $t_K = T$, and $Y_0 \sim p_{t_0}$, we define

$$Y_{k+1} = Y_k + (\alpha(t_{k+1}) - \alpha(t_k)) \hat{U}_k + \sigma \sqrt{t_{k+1} - t_k} Z_{k+1}$$

$$\hat{U}_k = \textbf{MCMC-Est}(u_{t_k}(Y_k)), Z_{k+1} \sim \mathrm{N}(0, \mathrm{I}_d)$$

Here, $\hat{U}_k$ is computed by sampling from $q_{t_k}(\cdot | Y_k)$ with ULA.

Target framework (for theoretical results)

Let $X \sim \pi$. We assume there exist $R > 0$ and $\tau > 0$ such that

$$X = U + N, \ \|U\| \leq R, \ N \sim \mathrm{N}(0, \tau^2 \mathrm{I}_d) \ .$$

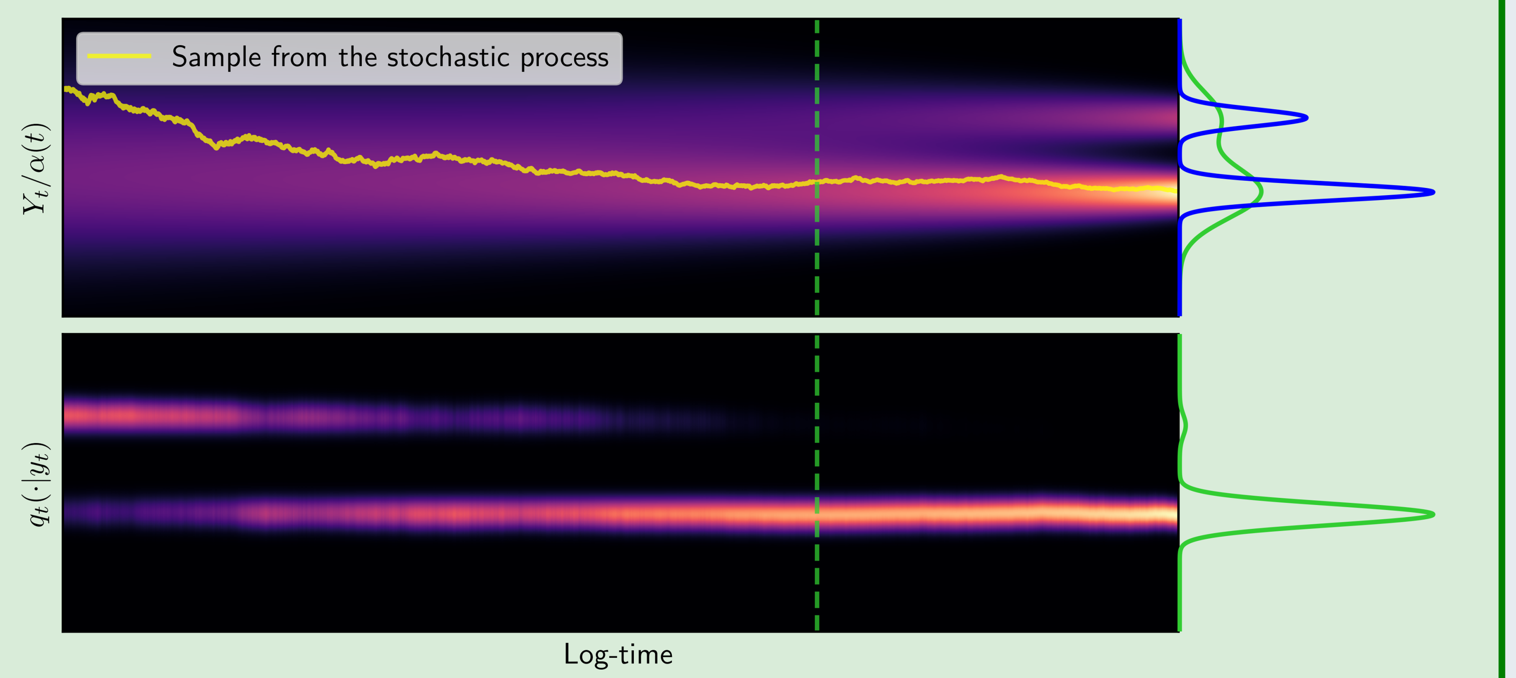
This class of distributions includes **Gaussian mixtures**: for $a > 0$, if $\pi = w\mathrm{N}(-a\mathbf{1}_d, \sigma^2 \mathrm{I}_d) + (1-w)\mathrm{N}(+a\mathbf{1}_d, \sigma^2 \mathrm{I}_d)$, $R = a\sqrt{d}$, $\tau = \sigma$.

Guarantees via the duality of log-concavity

Duality of log-concavity

There exist time-steps t_q and t_p in $(0,T)$ such that

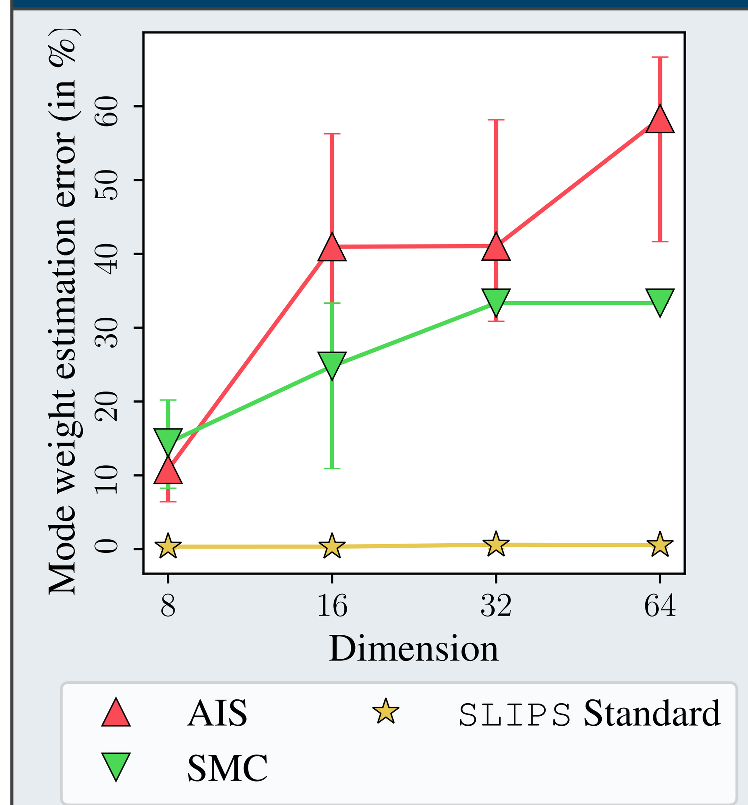
q_t is log-concave for $t > t_q$	p_t is log-concave for $t < t_p$
• If $t > t_q$, sample with ULA.	• If $t < t_p$, sample with ULA.
• u_t can be well estimated.	• But $\nabla \log p_t$ is not tractable.



Key idea: set $t_0 \in (t_q, t_p)$.

- Y_0 can be simulated by running ULA with the estimation of $\nabla \log p_{t_0}$ given by samples from q_{t_0}, see (2).
- $\hat{U}_k$ is well approximated with ULA as q_{t_k} will remain log-concave.

Numerical results



SLIPS is able to recover the *relative weight* of a bi-modal Gaussian mixture in high-dimension, where gold-standard MCMC methods fail.

- The **only limitation** of SLIPS is to find a suitable t_0, which needs to be tuned in practice.
- More experiments on Bayesian tasks and physics field systems are available **in the paper**.

How to contact us:

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