Provable Convergence and Limitations of Geometric Tempering for Langevin Dynamics

O. Chehab, A. Korba, A. Stromme, A. Vacher
{adrien.vacher,emir.chehab}@ensae.fr

Background

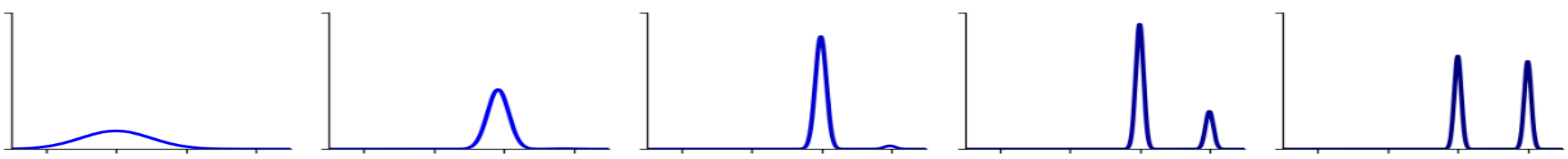
Sampling process: Annealed Langevin Dynamics

$$dx_t = \nabla \log \mu_t(x_t)dt + \sqrt{2}dw_t$$

Prescribed path: Interpolate target and Gaussian, using the geometric mean.

$$\mu_t(x) = c_t \pi(x)^{\lambda_t} \times \mathcal{N}(x; 0, I)^{1-\lambda_t}$$, increasing schedule $\lambda_t \in [0,1]$

Multimodality appears in small (inverse) log-Sobolev constant $\alpha_t \geq 0$.



This path is **computationally convenient**: the scores $\nabla \log \mu_t(x)$ are known when π is known up to a normalizing constant. But does it **accelerate convergence?**

Our results

Convergence is guaranteed: upper bound

$$\mathrm{KL}(p_t, \pi) \leq \underbrace{\exp(-2\int_0^t \alpha_s ds) \cdot \mathrm{KL}(p_0, \mu_0)}_{\text{Initial condition}} + \underbrace{(1-\lambda_t)A}_{\text{Terminal condition}} + \underbrace{A\int_0^t \dot{\lambda}_s \exp(-2\int_s^t \alpha_v dv)ds}_{\text{Speed of traversal vs. Multimodality}}$$

Convergence can be accelerated for a unimodal and peaky target: optimal schedule

$$\lambda(t) = \begin{cases} 1 & \text{for a peakier target, } \alpha_\pi \geq \alpha_\nu/2 \\ \dfrac{\alpha_\nu}{\alpha_\nu - \alpha_\pi}\dfrac{1+\alpha_\nu t}{2+\alpha_\nu t} & \text{for a flatter target, } \alpha_\pi < \alpha_\nu/2 \end{cases}$$

Convergence is provably slow for a multimodal target**:** lower bound

$$\mathrm{TV}(p_t, \pi) \geq \frac{1}{20} - 16 \cdot t \cdot e^{-m^2/64}$$

Convergence time is exponential in the distance m between target modes.