Computational Complexity Comparison via Scaling Limits for Emerging MCMC Samplers

Hirofumi Shiba

Sampling Problem

Obtain samples from $\pi(x) \propto e^{-U(x)}$ by only evaluating $U, \nabla U$

Piecewise Deterministic Monte Carlo

cf. Fearnhead+ (2018)

Running Example: $U^{(d)}(x) := \frac{1}{2}\sum_{i=1}^{d} x_i^2 \quad (x \in \mathbb{R}^d)$ Standard Gaussian

① Auxiliary Variable V

& its distribution $\mu(v) \propto e^{-K(v)}$ are introduced: $\widetilde{\pi}(x,v) := \pi(x)\mu(v)$ (augmented)

② Deterministic Flow ③ Random Times

ODE

$$\begin{cases} \dot{x}_t = f(x_t, v_t) \\ \dot{v}_t = g(x_t, v_t) \end{cases}$$

the solution

$$t \mapsto (x_t, v_t)$$

Rate function

$$\lambda(x,v) = (v|\nabla U(x))_+ + \rho$$

gives rise to random times

$$T_1, T_2, T_3, \cdots$$

③-A Reflections

$$V_{T_i} \sim Q(x_{T_i-}, v_{T_i-})$$

New velocity, given by the law Q

③-B Refreshments

$$V_{T_j} \sim \mu(v)\,dv$$

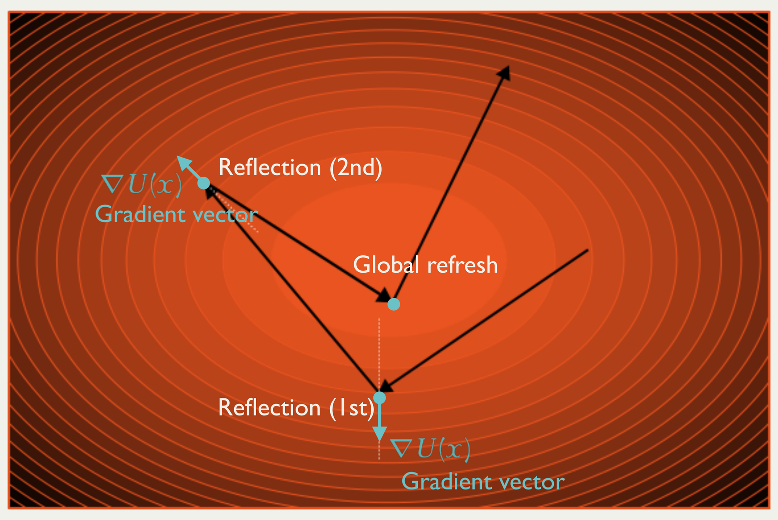
Regeneration

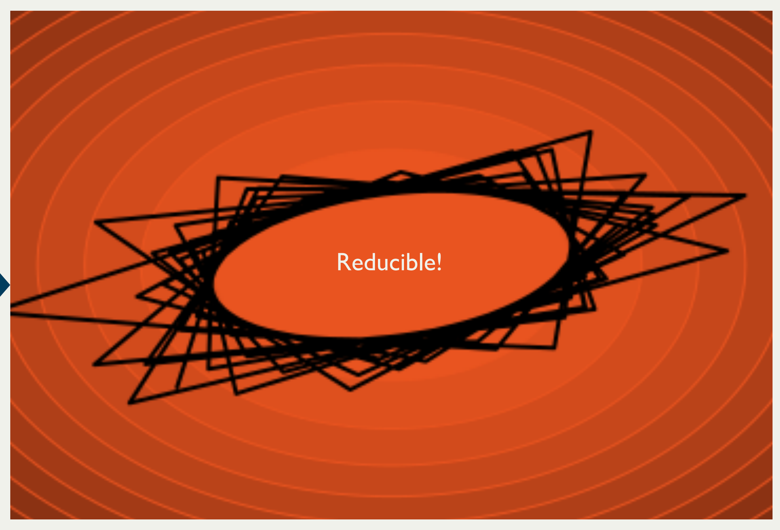
BPS

Bouncy Particle Sampler

cf. Bouchard-Côté+ (2018)

employs
③-A deterministically
③-B globally





③-A: Deterministic Reflection

$$v_{T_i} \leftarrow v_{T_i-} - 2\frac{(\nabla U(x_{T_i})|v_{T_i-})}{\|\nabla U(x_{T_i})\|^2}\nabla U(x_{T_i})$$

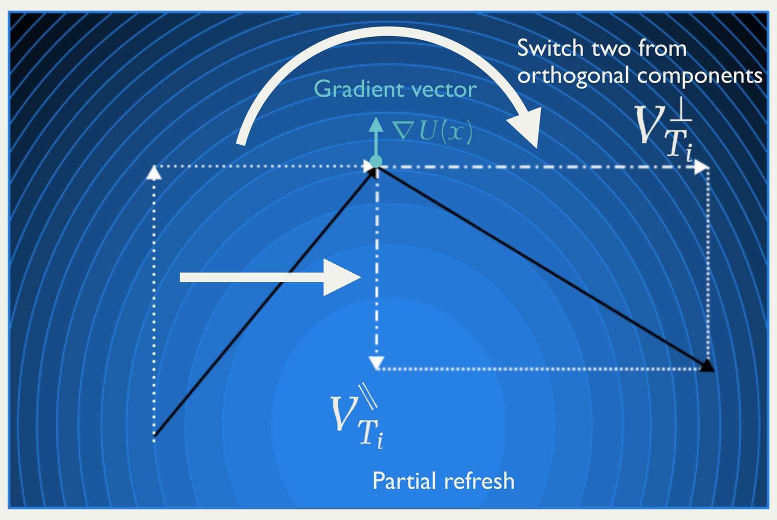
③-B: Global Refreshments (required!) $\rho > 0$

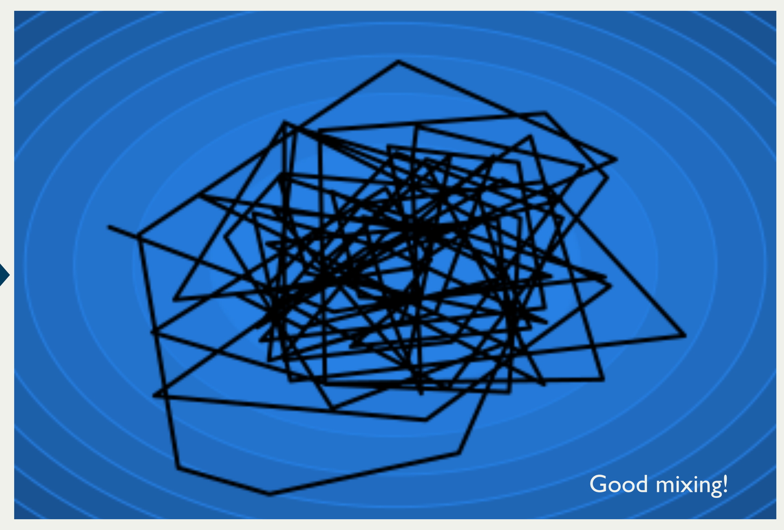
FECMC

Forward Event-Chain Monte Carlo

cf. Michel+ (2020)

combines
③-A stochastically
③-B partially





③-C: Stochastic Refreshment

Parallel Refresh
Orthogonal Switch

$$\begin{cases} V^{\parallel}_{T_i} \sim Q^{\parallel}(x_{T_i-}) \\ V^{\perp}_{T_i} \leftarrow A v^{\perp}_{T_i-} \end{cases} \Rightarrow \text{Combine} \quad V_{T_i} \leftarrow V^{\parallel}_{T_i} + V^{\perp}_{T_i}$$

(③-B: NO GLOBAL REFRESHMENT REQUIRED) $\rho = 0$

Scaling Analysis: Compare Limiting Dynamics as $d \to \infty$

Potential (= negative log-likelihood) $Y_t^{(d)} := \frac{U^{(d)}(X^d_{dt}) - d}{\sqrt{d}}$ converges to an Ornstein-Uhlenbeck diffusion $dY_t = -\frac{\sigma^2}{4}Y_t\,dt + \sigma\,dB_t$

Th'm (Bierkens+2022)

$$\sigma_{\mathrm{BPS}}(\rho)^2 = 8\int_0^\infty e^{-\rho s}K(s,0)\,ds$$

where K is the transition kernel of the Gauss-Markov process

$$Gf(x) = f'(x) + x_+\big(f(-x) - f(x)\big)$$

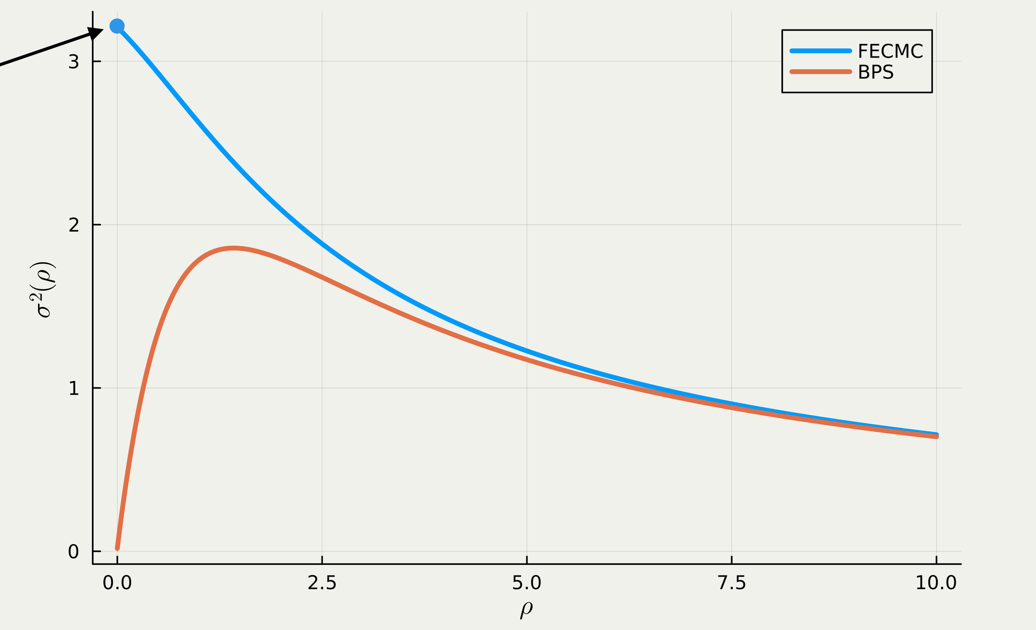
Theorem

$$\sigma^2_{\mathrm{FECMC}} = \sqrt{\frac{32}{\pi}}$$

Corollary

For all $\rho > 0$

$$\sigma^2_{\mathrm{BPS}}(\rho) < \sigma^2_{\mathrm{FECMC}}$$



Complexity

Number of gradient $\nabla U(x)$ evals scales as $O(d)$ (for the potential)

cf. Existing Computational Complexity Results

Random Walk Metropolis-Hastings	$O(d^2)$
Metropolis-adjusted Langevin	$O(d^{4/3})$
Hamiltonian Monte Carlo	$O(d^{5/4})$

MSE / Monte Carlo Variance

For spherically symmetric $f : \mathbb{R}^d \to \mathbb{R}$

$$\mathrm{Var}\left[\frac{1}{N}\sum_{n=1}^N f(X^{\mathrm{BPS}}_{\delta n})\right] \ge \mathrm{Var}\left[\frac{1}{N}\sum_{n=1}^N f(X^{\mathrm{FECMC}}_{\delta n})\right] = O\left(\frac{d}{N}\right)$$

in high dimensions $d \gg 1$ critical slow down!

Monitor the time-derivative process!

$$\frac{d}{dt}f(X_t) = (\nabla f(X_t)|V_t) =: R_t$$

An Early Diagnostic

→ MSE reduction by an $O(d^{-1})$ scale