統計数理研究所

Scaling Limits & Early Diagnostics for PDMP Samplers

Hirofumi Shiba & Kengo Kamatani

Institute of Statistical Mathematics, Tokyo

2602.17087

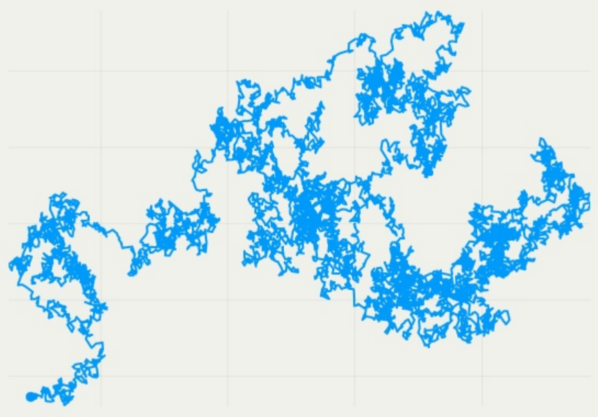
What is PDMP?

Faster convergence through ballistic movement

	Drift	Diffusion	Jump	$t\in[0,\infty)$
Diffusion	$dX_t = a(X_t)\,dt$	$+\,b(X_t)\,dB_t$		
PDMP	$dX_t = a(X_t)\,dt$		$+\int_{\mathbb{R}^d} c(X_{t-},u)\eta(dtdu)$	

Piecewise Deterministic ↑ Markov Process ↑

$$\|\mathcal{L}(X_t)-\mu\|_{\mathrm{TV}} = \sup_{\|f\|_\infty\le 1}\left|\mathrm{E}[f(X_t)]-\mu(f)\right| \le C_1 e^{-C_2 t}$$ for a broader class of targets μ



Langevin Diffusion

- Overdamped Dynamics
  → SDE discretization
- Equilibrium Steady State

VS.

Randomized Hamiltonian Monte Carlo

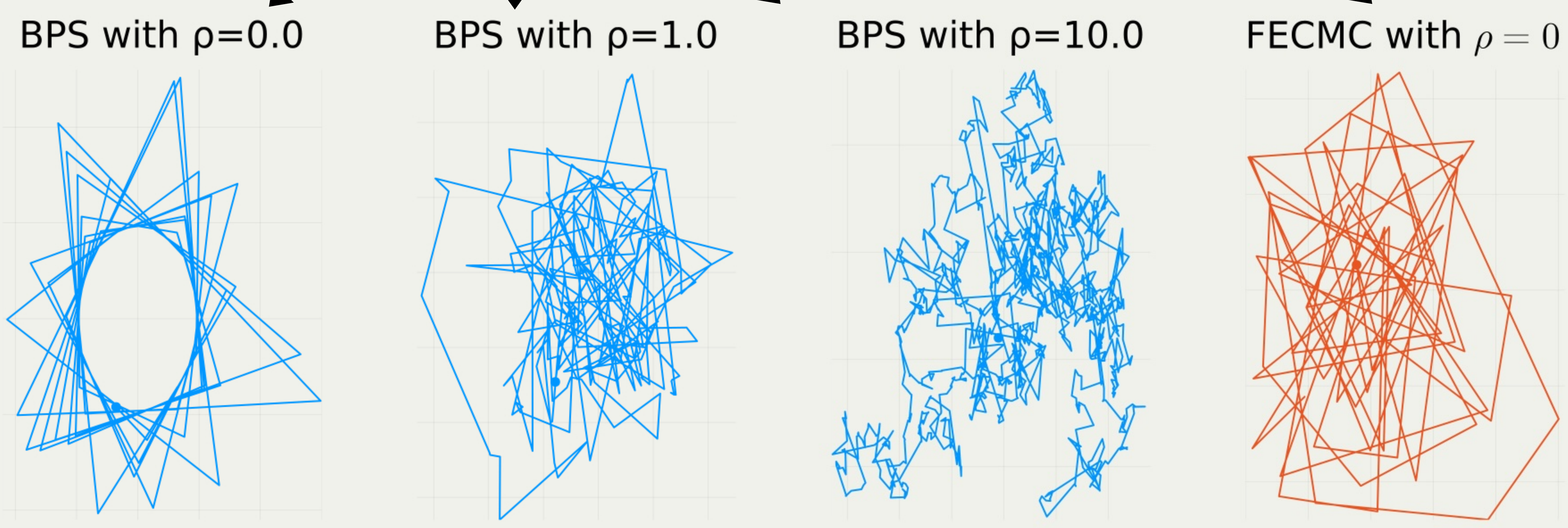
- Kinetic Dynamics
  → ODE discretization
- Non-equilibrium Steady State
  → Faster Relaxation

Which method is faster?

*Among piecewise linear methods

Bouchard-Côté, Vollmer & Doucet (2018) The Bouncy Particle Sampler

Michel, Durmus & Sénécal (2020) Forward Event-Chain Monte Carlo



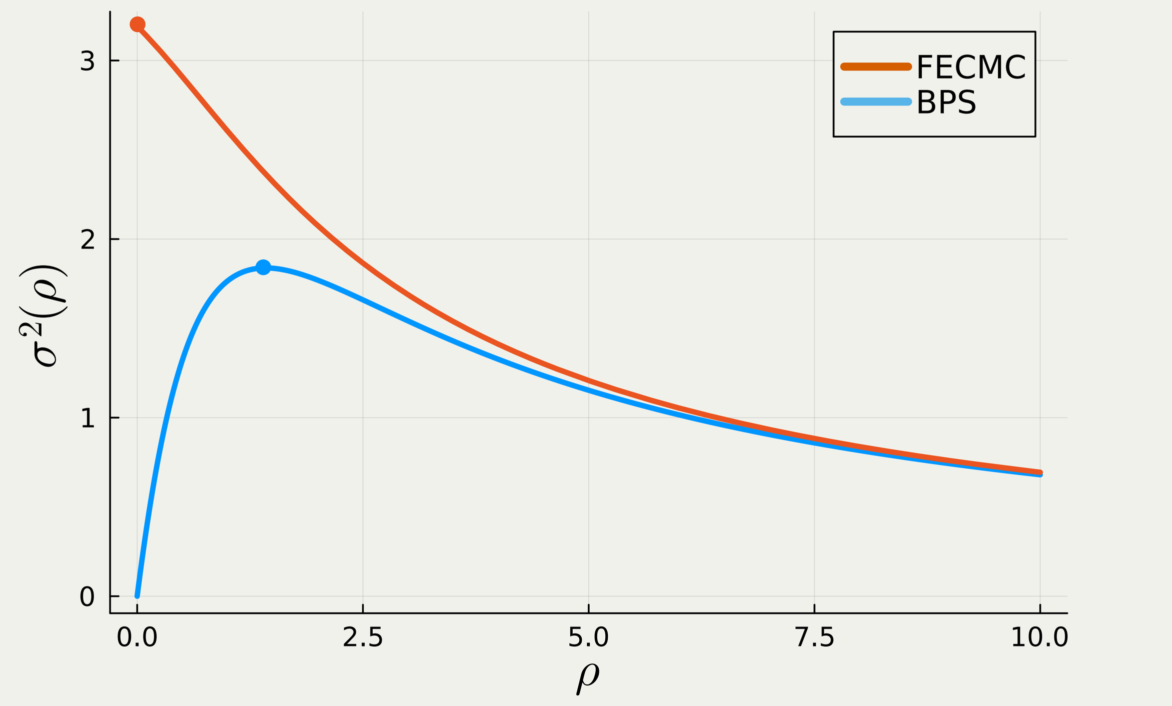
How to set the hyperparameter ρ?

Is this really more efficient?

For a standard Gaussian

$$\pi(x)\propto e^{-U^{(d)}(x)}\qquad U^{(d)}(x)=\frac{\|x\|_2^2}{2}$$

with $d\to\infty$



$\sigma^2 \approx$ the limiting speed of the process $U^{(d)}(X_t)$

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

Theorem 3.7 & 3.9 The potential of both methods converges to an Ornstein-Uhlenbeck diffusion

$$Y_t^{(d)} := \frac{2U^{(d)}(X^d_{dt})-d}{\sqrt{d}} \Longrightarrow dY_t = -\frac{\sigma^2(\rho)}{4}Y_t\,dt+\sigma(\rho)\,dB_t$$

Theorem 4.1

$$\sigma^2_{\mathrm{FECMC}}(\rho)=\sqrt{\frac{32}{\pi}}\left(1-\frac{(\rho^2-\rho\sqrt{\frac{\pi}{2}}+\Omega(\rho))^2}{\rho^4\Omega(\rho)(2-\Omega(\rho))}\right)$$

$$\sigma^2_{\mathrm{BPS}}(\rho)=\frac{8}{\rho^4}\left(\rho^3-\rho^2\sqrt{\frac{8}{\pi}}+\rho-\sqrt{\frac{8}{\pi}}\frac{((1+\rho^2)\Omega(\rho)-\rho^2)^2}{\Omega(2\rho)}\right)$$

$\Omega(\rho):=\rho e^{\frac{\rho^2}{2}}\int_\rho^\infty e^{-\frac{t^2}{2}}dt$: exponentially scaled complementary error function

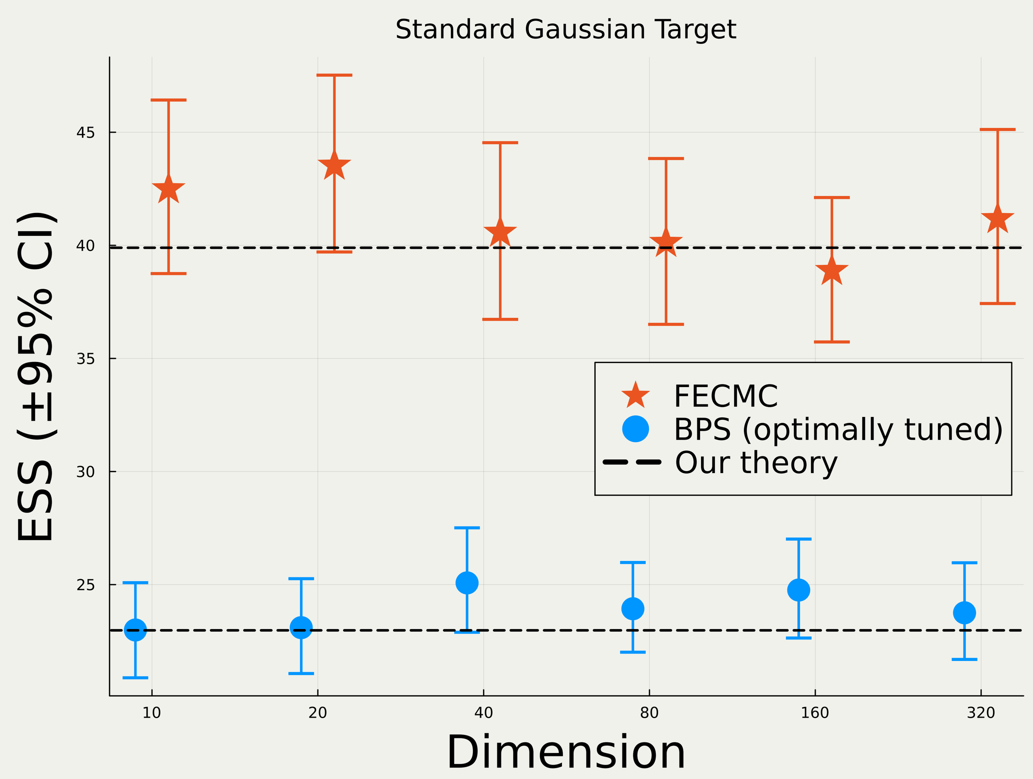
Corollary 4.3

$$\sigma^2_{\mathrm{FECMC}}(0)=\sqrt{\frac{32}{\pi}}$$

ESS: A benchmark For $U^{(d)}$, the effective sample size (ESS) is

$$\mathrm{ESS}(U)\approx T\frac{\sigma^2(\rho)}{8}$$

ESS / T roughly corresponds to mean squared error (MSE)



Complexity Scaling

	$\nabla U(x)$ evals / ESS	$\nabla U(x)$	Total Complexity
Potential			
BPS / FECMC	$O(d)$	$O(d^{0\sim 1})$	$O(d^{1\sim 2})$
Hamiltonian Monte Carlo	Unkown (Open problem)		
Marginal coordinate ≈			
Random Walk Metropolis	$O(d)$	$O(d)$	$O(d^2)$
Metropolis-adjusted Langevin	$O(d^{1/3})$	$O(d)$	$O(d^{4/3})$
Hamiltonian Monte Carlo	$O(d^{1/4})$	$O(d)$	$O(d^{5/4})$

An Early Diagnostic

In general, σ^2 depends on the target and needs to be estimated on the fly in practice

Monitor the time-derivative process!

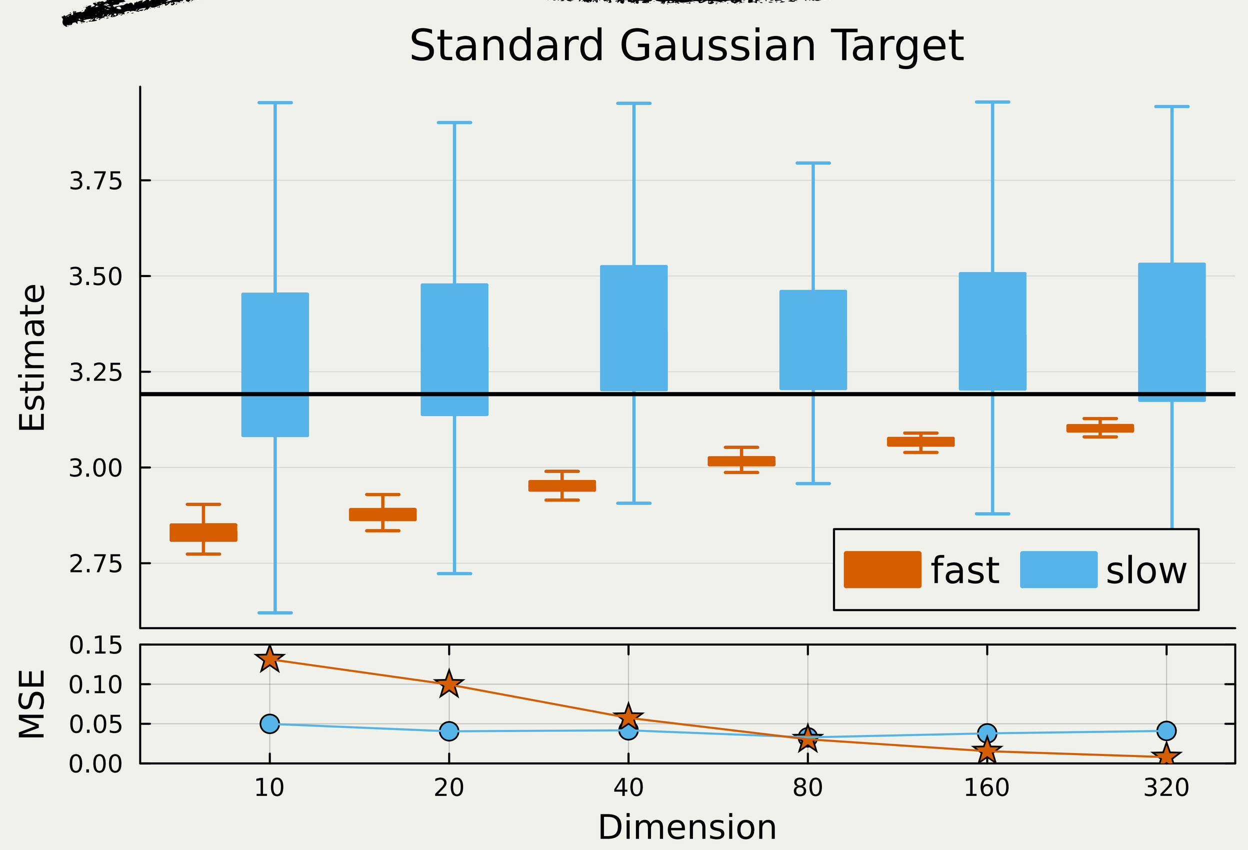
$$\frac{d}{dt}h(X_t)=(\nabla h(X_t)|V_t)=:R_t$$

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

Proposition 4.5

$$\frac{1}{\sqrt{T}}\int_0^{dT}h(X_t)\,dt \overset{d,T\to\infty}{\Rightarrow} N\left(0,\frac{8}{\sigma^2}\right)$$

$$\frac{1}{\sqrt{T}}\int_0^{T}R_t\,dt \overset{d,T\to\infty}{\Rightarrow} N\left(0,\frac{\sigma^2}{4}\right)$$



Comparison of two batch means estimators