

Sticky Manifold PDMP

低次元多様体上に台を持つ分布からのサンプリング

司馬 博文^{1,3}, Luke Hardcastle², & 鎌谷 研吾³

¹総合研究大学院大学

² MRC Biostatistics Unit, University of Cambridge

³統計数理研究所

PDMP とは？

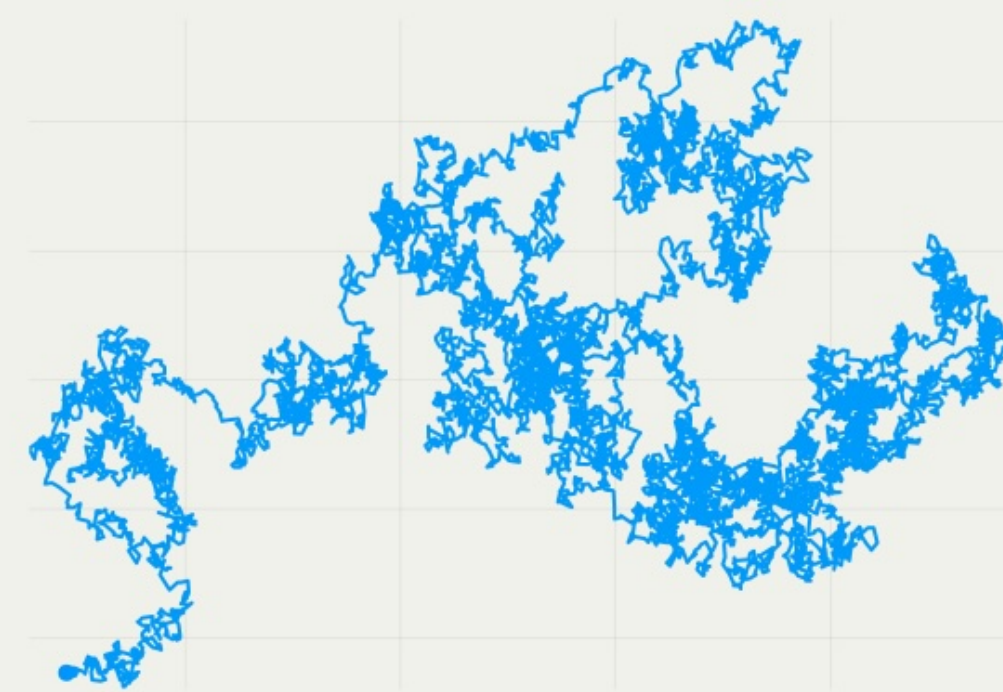
運動論的なダイナミクスで速い収束を狙う

Drift **Diffusion** **Jump** $t \in [0, \infty)$

Diffusion $dX_t = a(X_t) dt + b(X_t) dB_t$

PDMP $dX_t = a(X_t) dt + \int_{\mathbb{R}^d} c(X_{t-}, u) \eta(dt du)$

Piecewise Deterministic **Markov Process**



Langevin 拡散

- SDE で記述される過程
→ 数値離散化で誤差が入る
- 平衡状態のダイナミクス

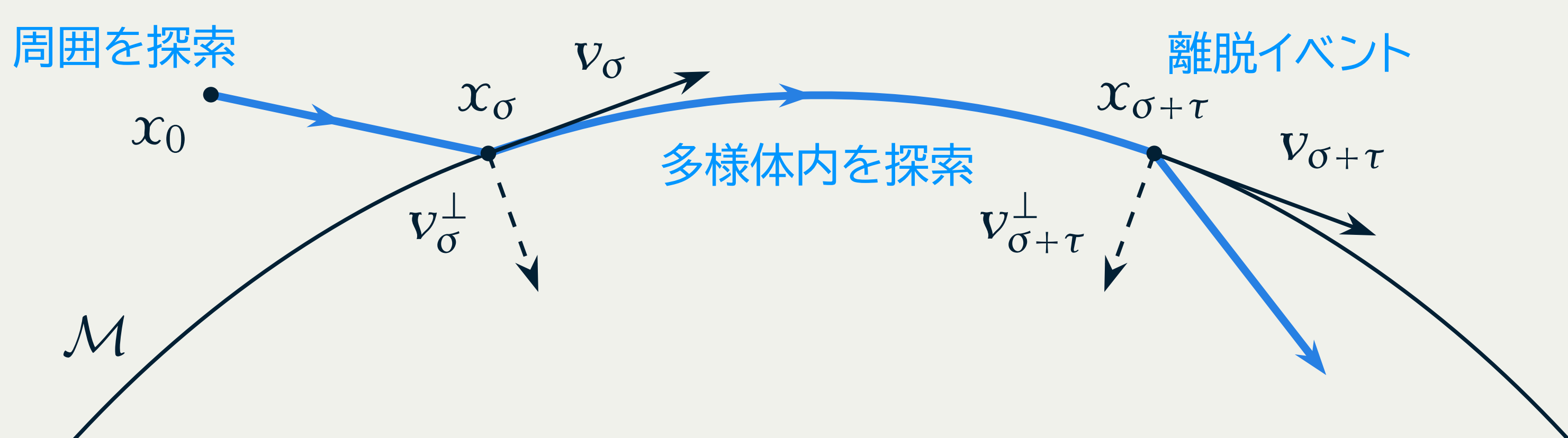
VS.

Randomized Hamiltonian Monte Carlo

- 分子運動論的なダイナミクス
→ ODE としての離散化はいくらか簡単
- 非平衡状態のダイナミクス
→ より速い緩和



多様体に吸着するダイナミクス の設計



- ① 多様体 $\mathcal{M}$ に到達すると吸着
- ② 接方向に動き続け, 法線方向の速度は保持する
- ③ Poisson 到着に従って確率的に離脱, 周囲の探索を再開する

強度関数 $\Lambda(x) = \|v_\sigma^\perp\| \frac{\omega}{1-\omega} \frac{\pi_{\mathbb{R}^d}(x)}{\pi_{\mathcal{M}}(x)}, \quad x \in \mathcal{M}$

目標分布

複数次元にまたがる
mixed-dimensional target

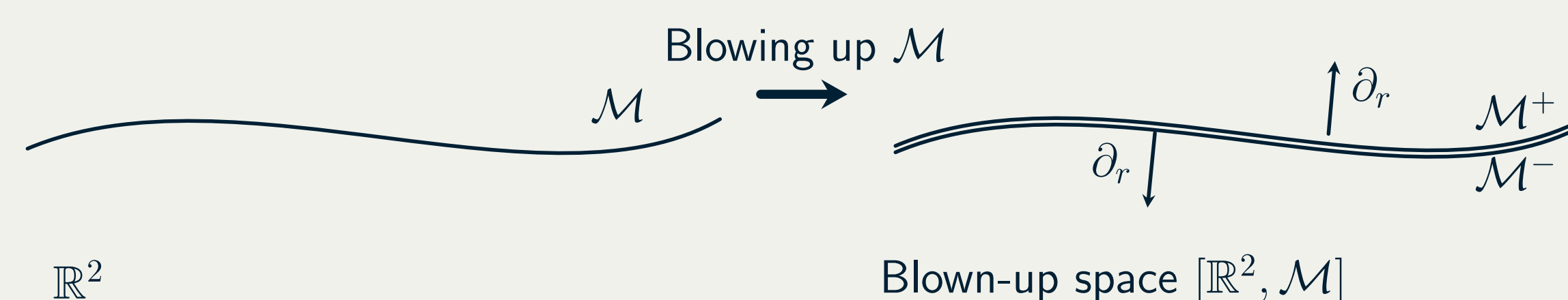
$$\pi(dx) \propto \exp(-U(x)) \left(\omega \pi_{\mathbb{R}^d}(x) dx + (1 - \omega) \pi_{\mathcal{M}}(x) \mathcal{H}_{\mathcal{M}}^{d-1}(dx) \right),$$

ベイズ統計での設定

$$L(x) \propto \exp(-U(x)) \quad \text{尤度} \quad + \quad \pi_{\mathbb{R}^d}, \pi_{\mathcal{M}} \quad \text{事前分布}$$

状態空間の定義

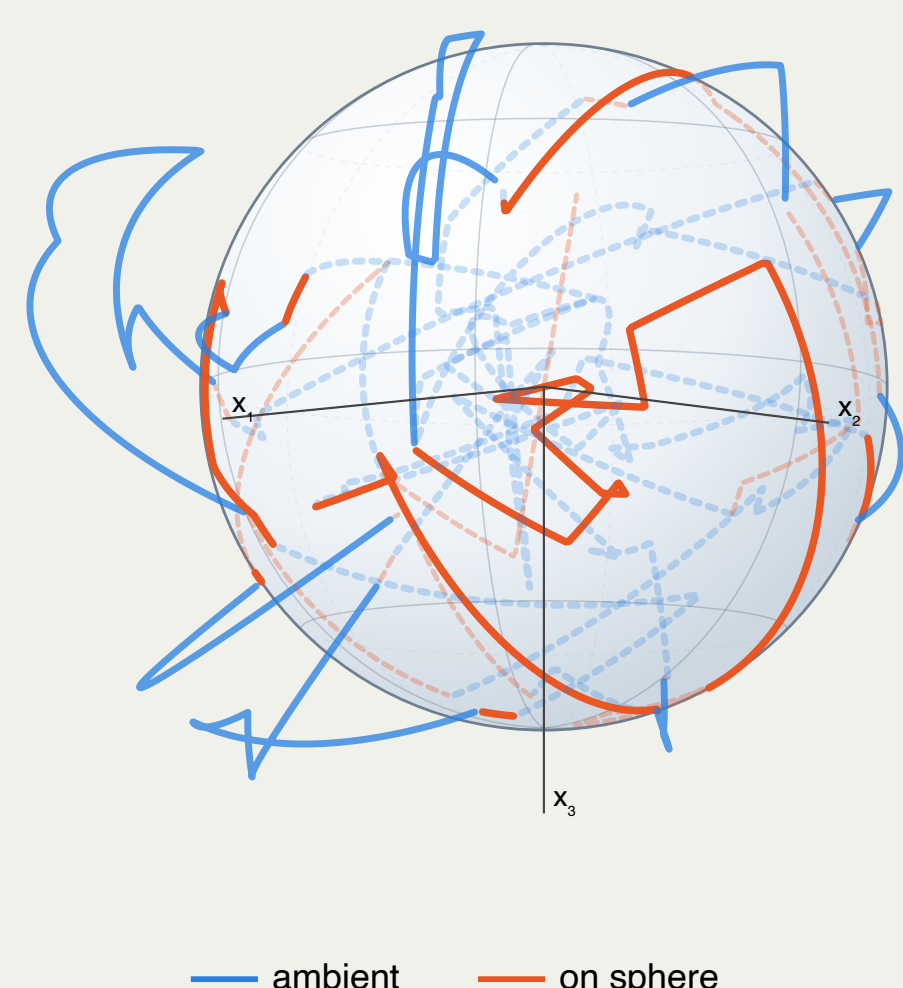
Markov 過程の状態空間としては, 侵入した側と法線方向を区別するために blown-up space を採用する



統計学への応用

量子状態の推定

純粋状態か？混合状態か？

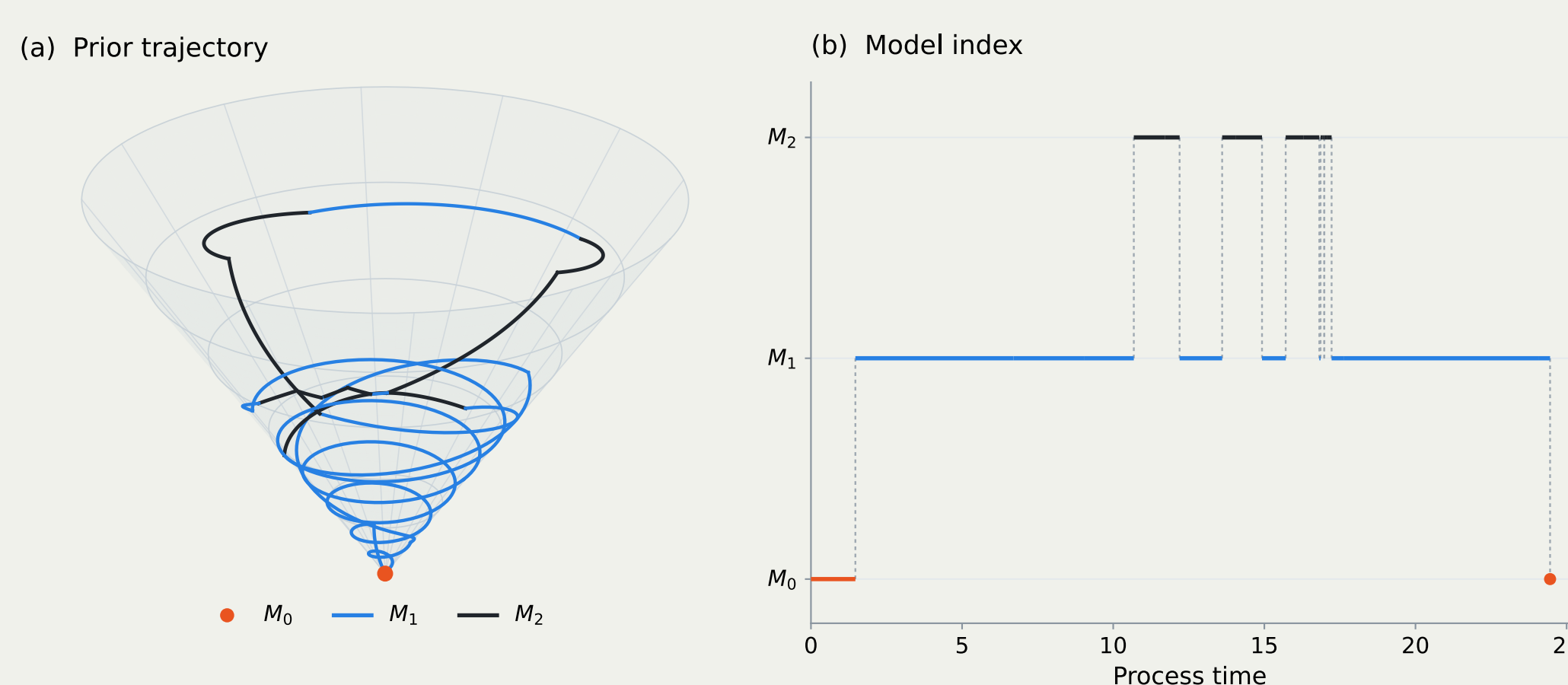


Bloch球の内部は混合状態, 球面は純粋状態
事前確率を置き, 両者を同時に推論する

$$L(x) = \prod_{i=1}^n \frac{1 + y_i a_i^\top x}{2}$$

主成分分析

潜在次元はいくつか？

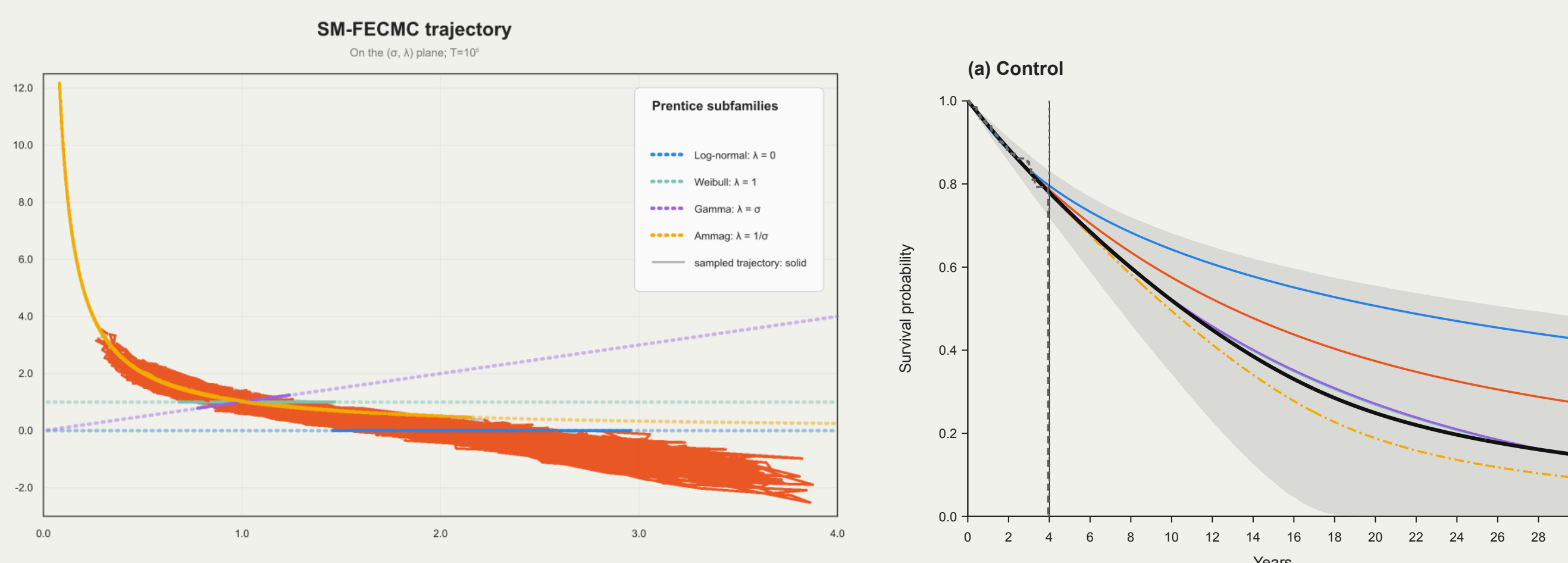


共分散行列のランクが異なるモデルを, 多様体の階層として扱う
ランクの不確実性とパラメータの不確実性を同時に評価する

$$z_i \sim N_d(0, G), \quad y_i | z_i \sim N_d(z_i, \sigma^2 I_d)$$

生存時間解析

未観測期間の生存関数はどのような形をしているか？



一般化ガンマ分布に標準的な分布族を多様体として埋め込む
モデル平均で分布族の不確実性を生存曲線に反映する

$$L(x) = \prod_{i=1}^n f_i(t_i; x)^{\delta_i} S_i(t_i; x)^{1-\delta_i}$$

尤度に依らない統一的な原理で, ベイズ事後分布からサンプリングできるダイナミクスを構成可能

Sticky Manifold PDMP

低次元多様体上に台を持つ分布からのサンプリング

Hirofumi Shiba & Kengo Kamatani

Institute of Statistical Mathematics, Tokyo

What is PDMP?

Faster convergence through ballistic movement

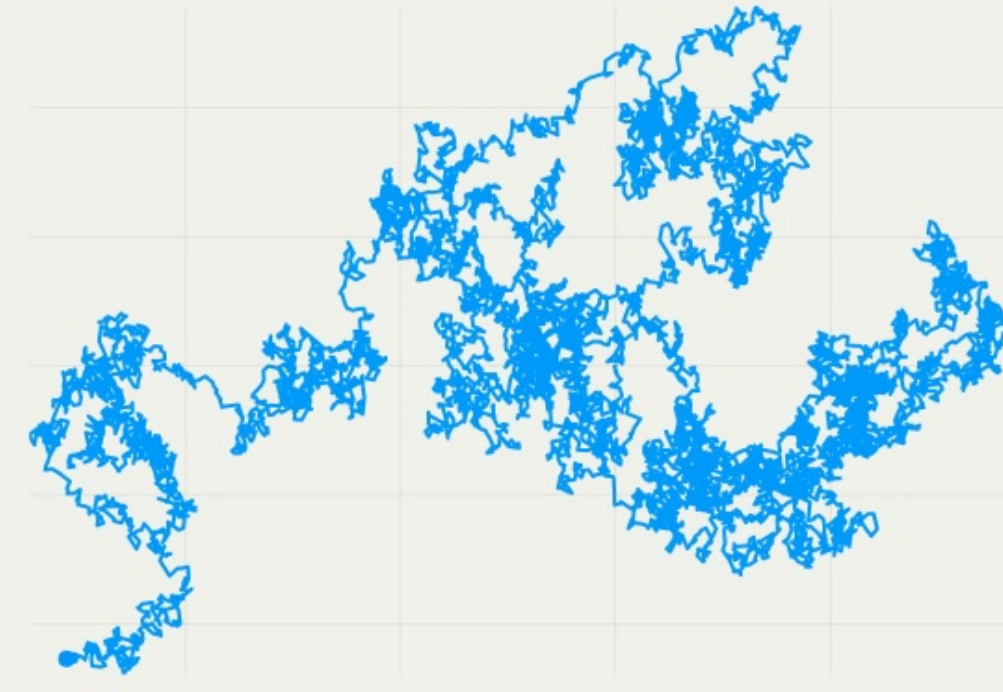
Drift **Diffusion** **Jump** $t \in [0, \infty)$

Diffusion $dX_t = a(X_t) dt + b(X_t) dB_t$

PDMP $dX_t = a(X_t) dt + \int_{\mathbb{R}^d} c(X_{t-}, u) \eta(dt du)$

Piecewise Deterministic **Markov Process**

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



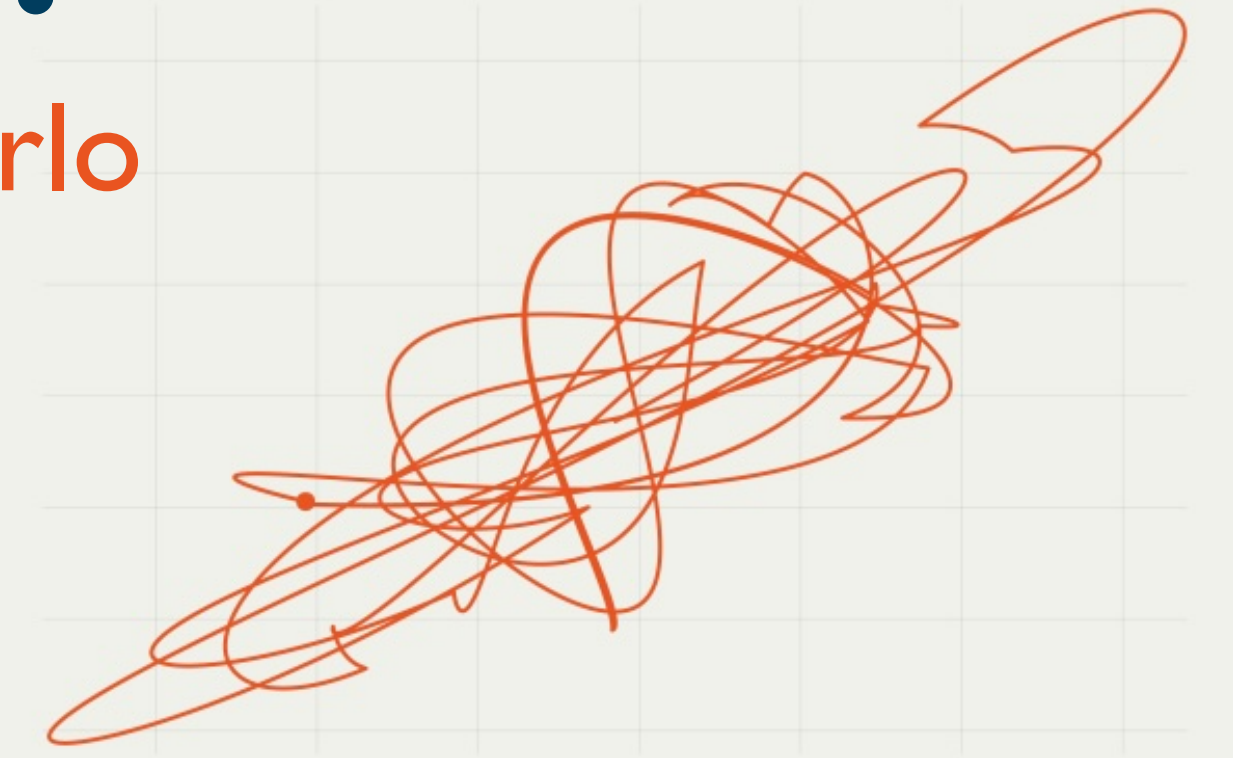
Langevin Diffusion

- 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

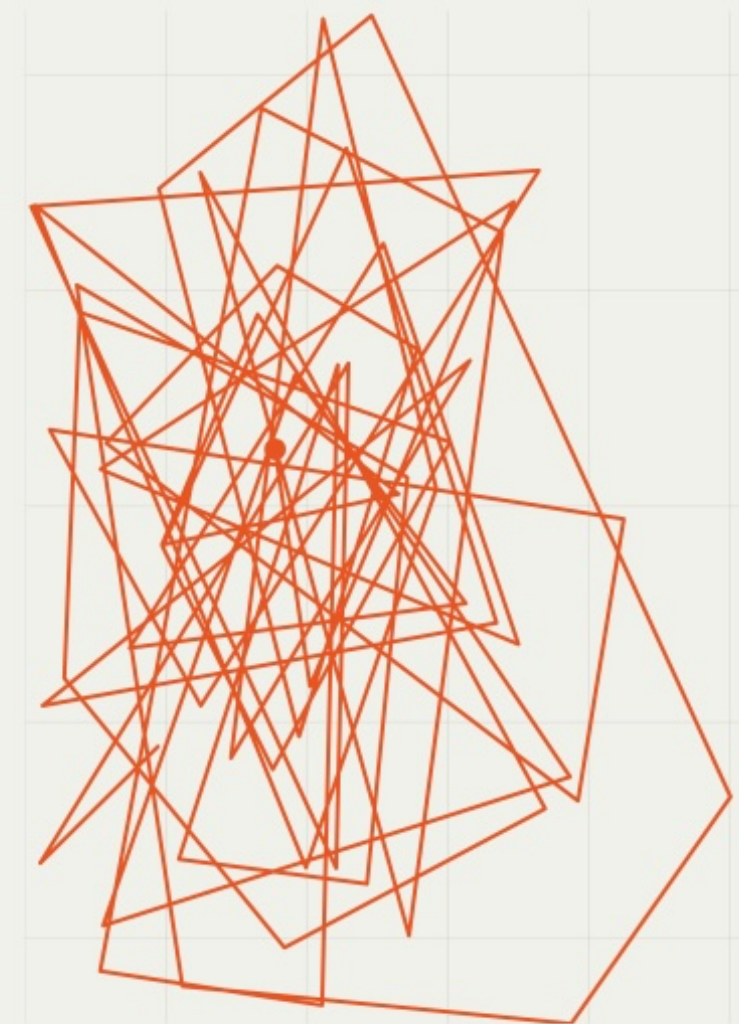
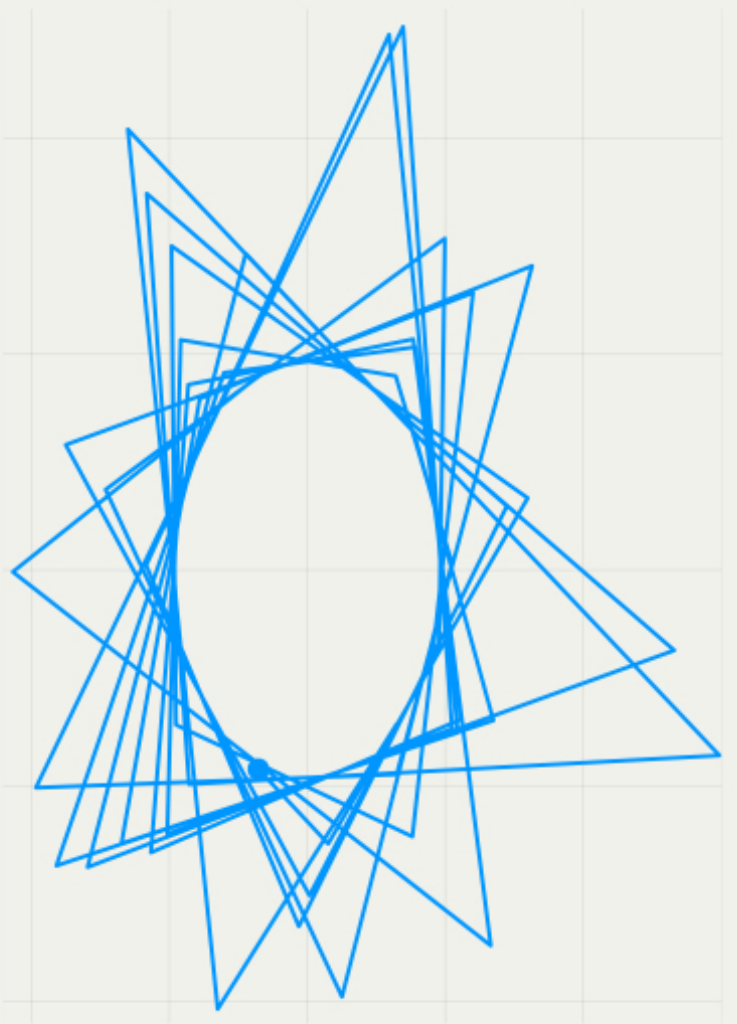
Michel, Durmus & Sènećal (2020) Forward Event-Chain Monte Carlo

BPS with $\rho=0.0$

BPS with $\rho=1.0$

BPS with $\rho=10.0$

FECMC with $\rho=0$



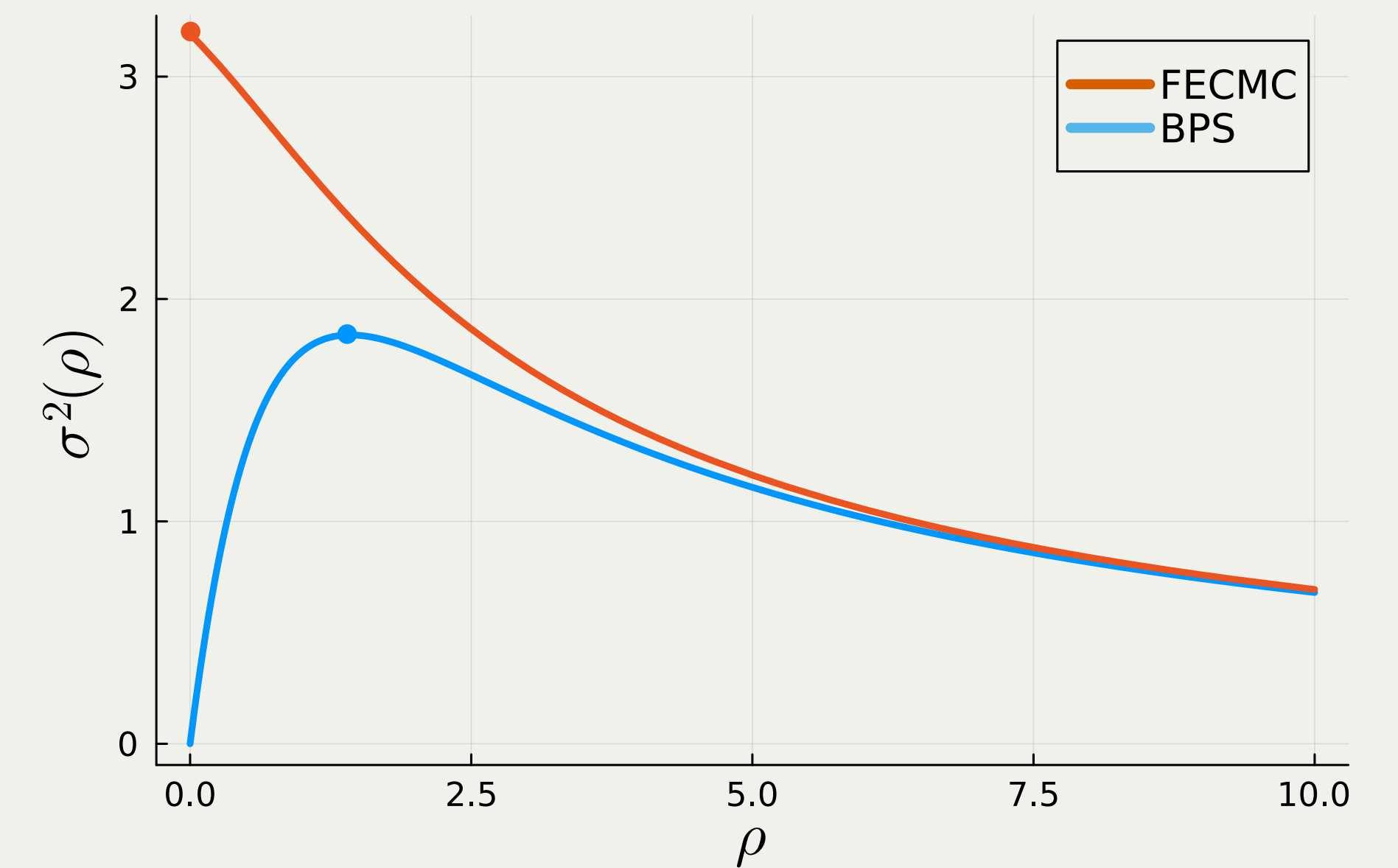
How to set the hyperparameter ρ ?

Is this really more efficient?

For a standard Gaussian

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

with $d \rightarrow \infty$



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

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

Theorem 3.7 & 3.9

The potential of both methods converges to an Ornstein-Uhlenbeck diffusion

$Y_t^{(d)} := \frac{2U^{(d)}(X_{dt}^d) - d}{\sqrt{d}} \Rightarrow dY_t = -\frac{\sigma^2(\rho)}{4} Y_t dt + \sigma(\rho) dB_t$

Theorem 4.1

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

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

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

Corollary 4.3

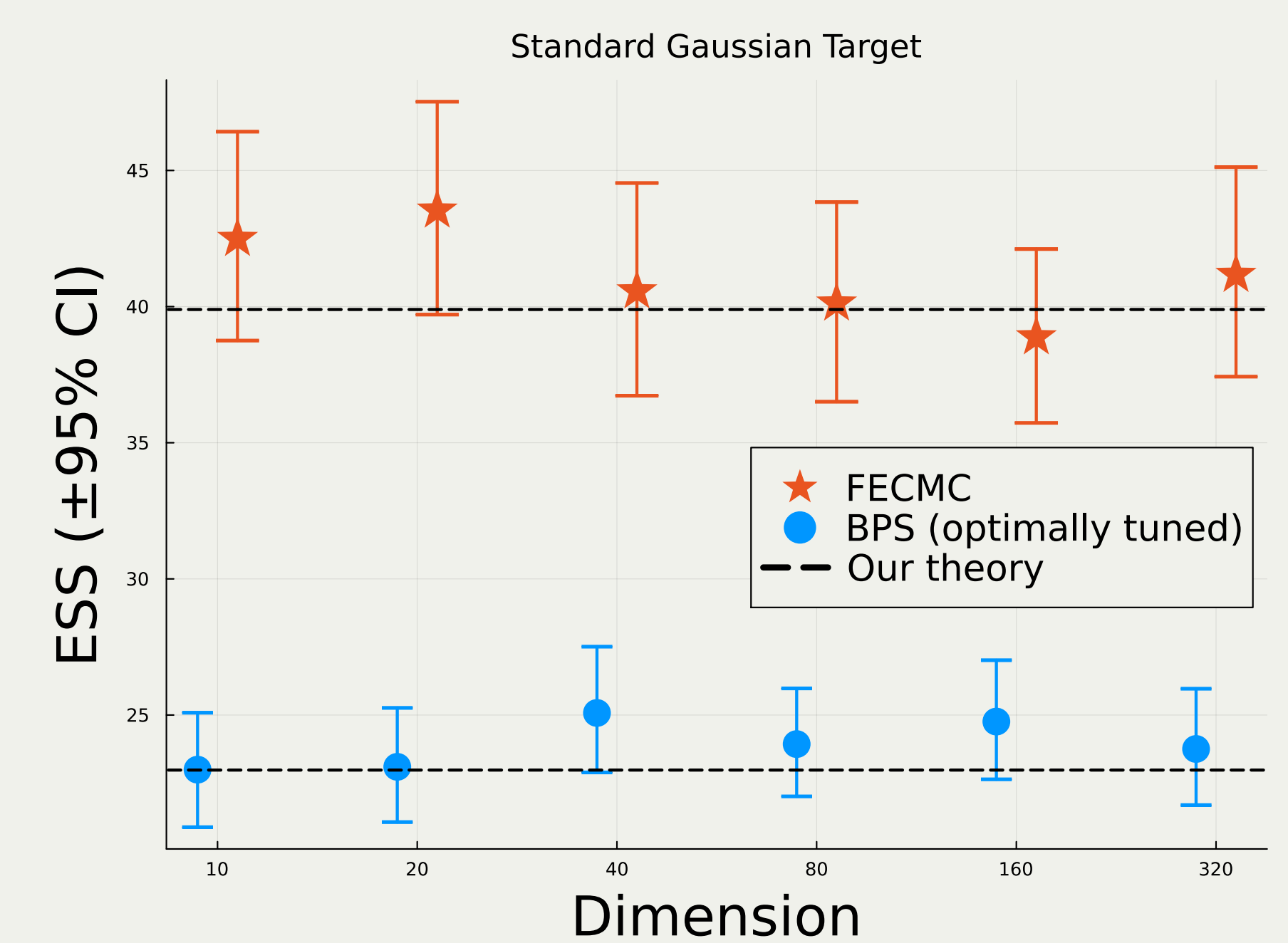
$\sigma_{\text{FECMC}}^2(0) = \sqrt{\frac{32}{\pi}}$

ESS: A benchmark

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

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

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



Complexity Scaling

Potential

BPS / FECMC

Hamiltonian Monte Carlo

Marginal coordinate

Random Walk Metropolis

Metropolis-adjusted Langevin

Hamiltonian Monte Carlo

$\nabla U(x)$ evals / ESS	$\nabla U(x)$	Total Complexity
$O(d)$	$O(d^{0 \sim 1})$	$O(d^{1 \sim 2})$
Unkown (Open problem)		

$\approx$

$O(d)$	$O(d)$	$O(d^2)$
$O(d^{1/3})$	$O(d)$	$O(d^{4/3})$
$O(d^{1/4})$	$O(d)$	$O(d^{5/4})$

An Early Diagnostic

Monitor the **time-derivative** process!

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

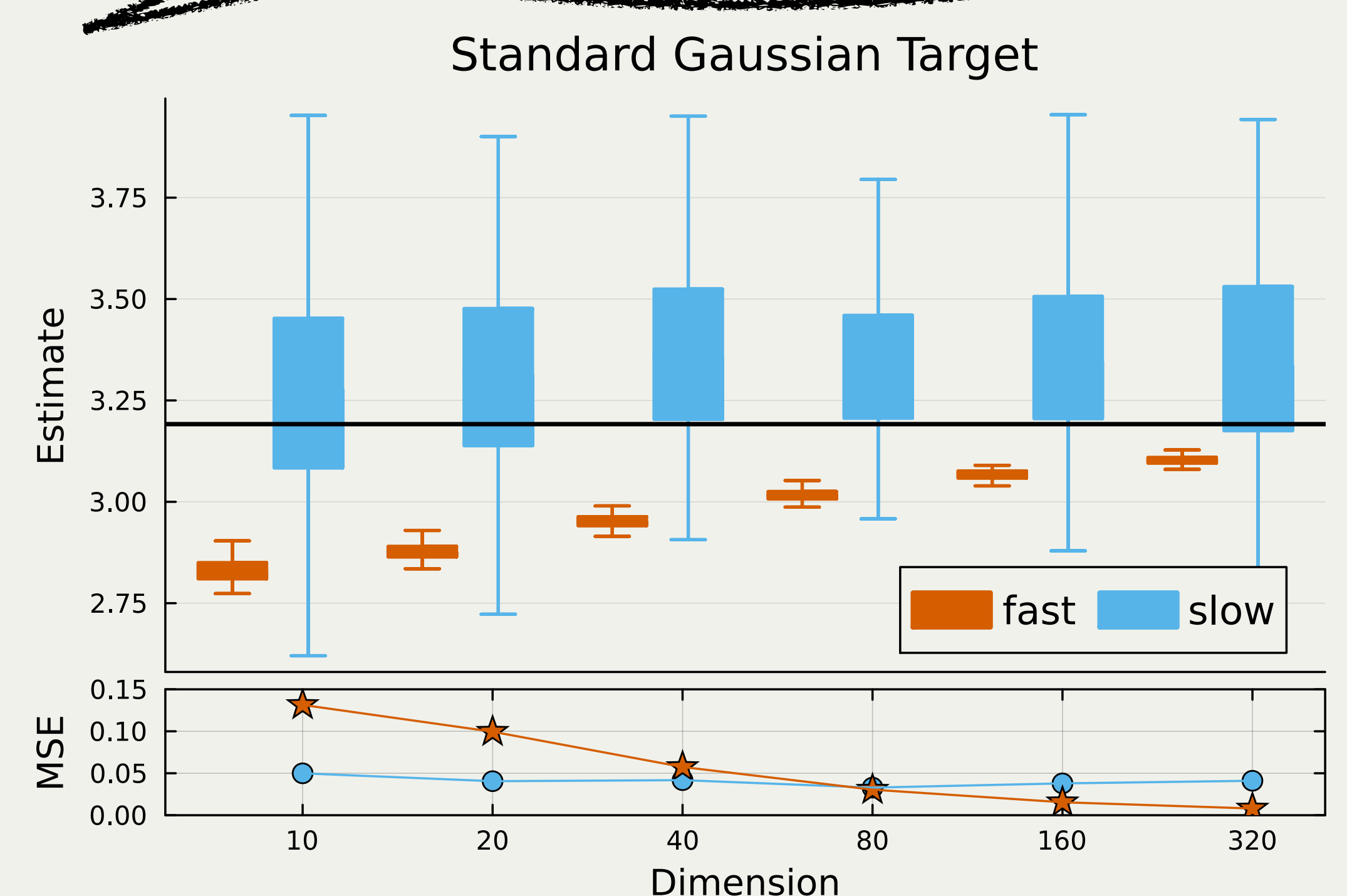
→ MSE reduction by an $O(d^{-1})$ scale

Proposition 4.5

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

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

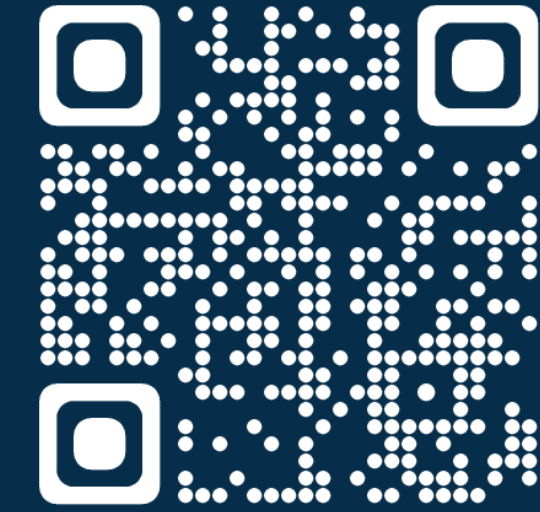
In general, σ^2 depends on the target and needs to be estimated on the fly in practice



Comparison of two batch means estimators

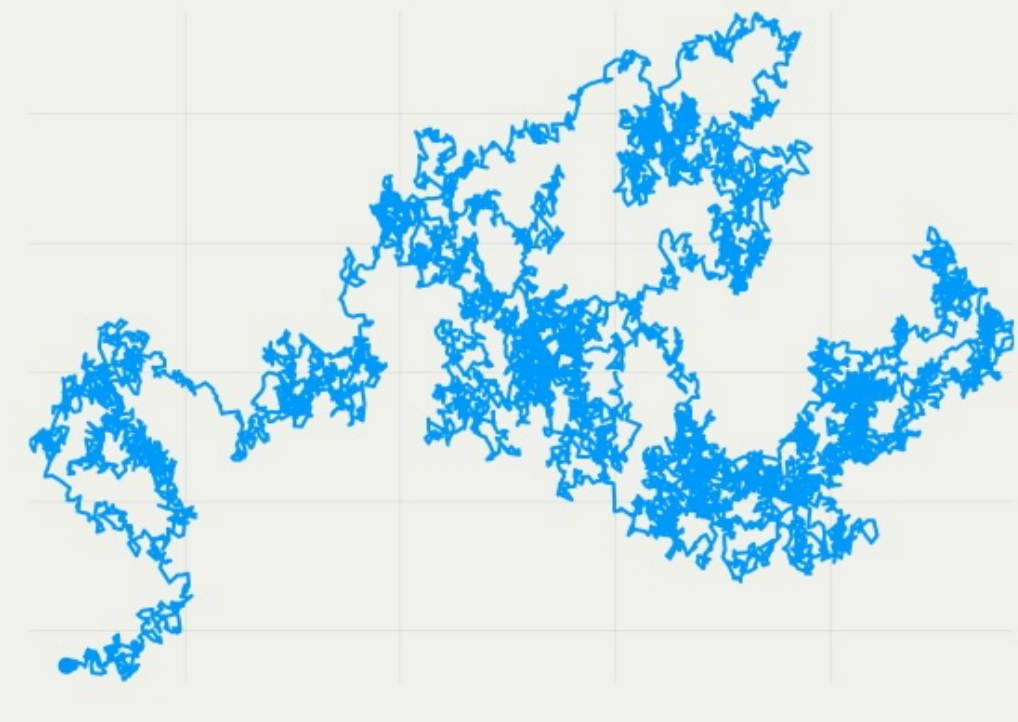
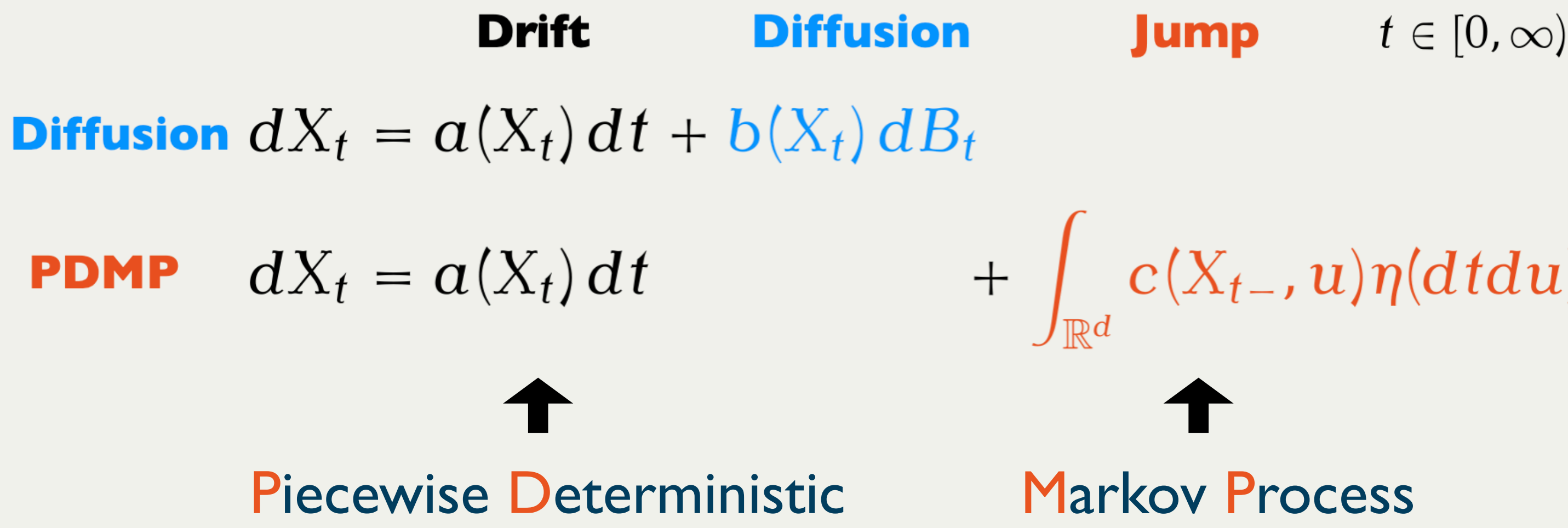
区分確定的モンテカルロ法の拡散極限と早期収束診断

司馬博文 総合研究大学院大学 5年一貫博士課程4年



PDMP とは何か？

運動論的なダイナミクスで速い収束を狙う



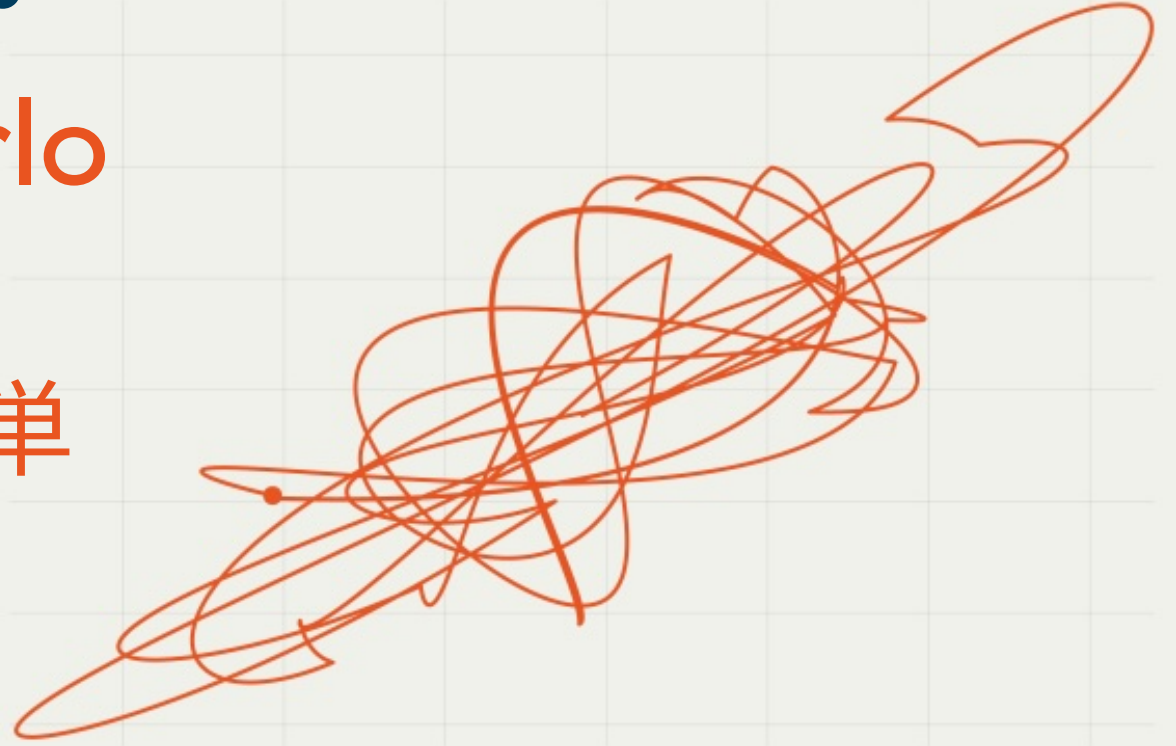
Langevin 拡散

- SDE で記述される過程
→ 数値離散化で誤差が入る
- 平衡状態のダイナミクス

VS.

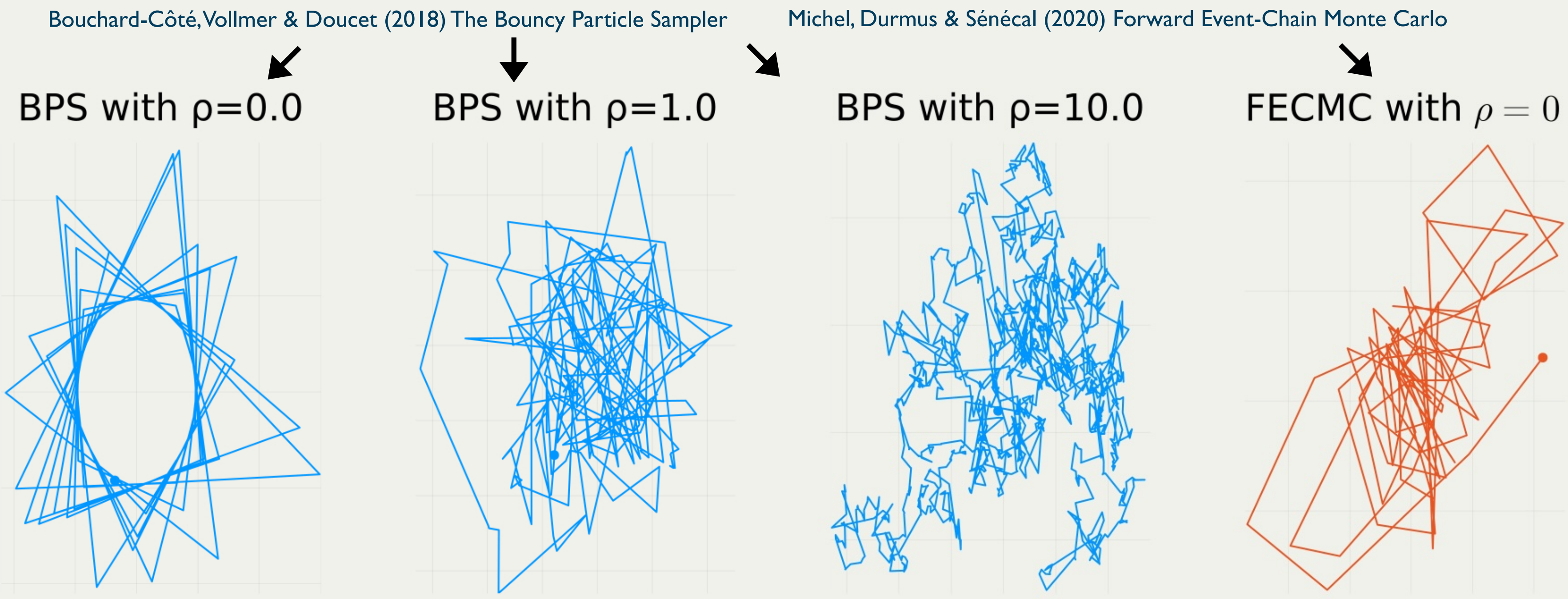
Randomized Hamiltonian Monte Carlo

- 分子運動論的なダイナミクス
→ ODE としての離散化はいくらか簡単
- 非平衡状態のダイナミクス
→ より速い緩和



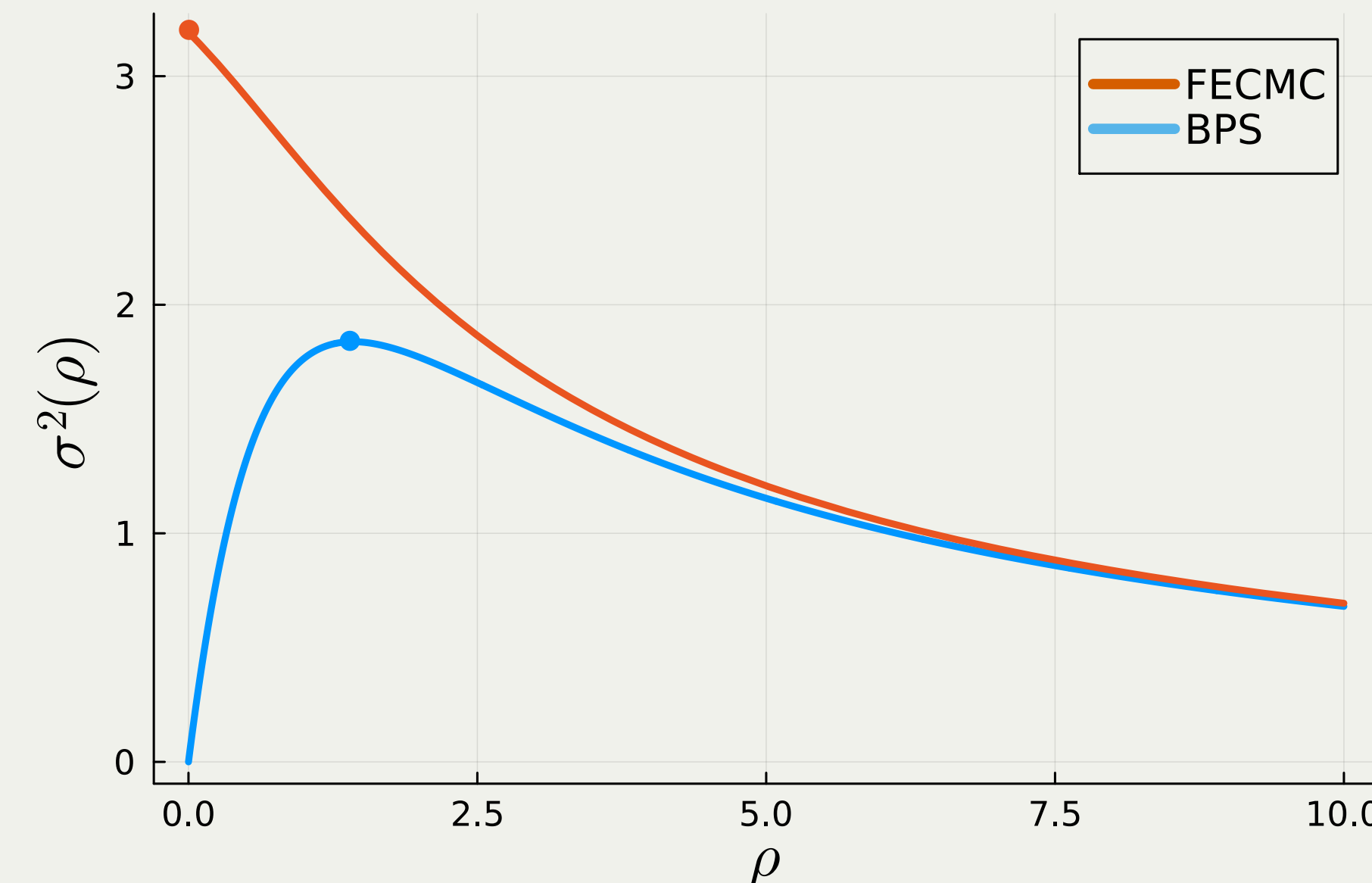
どの PDMP がより速い？

*今回は区分線型な手法に限る



ハイパーパラメータ ρ をどう選ぶ？ 本当に速くなってる？

標準 Gauss 分布
 $\pi(x) \propto e^{-U^{(d)}(x)}$ $U^{(d)}(x) = \frac{\|x\|_2^2}{2}$
について高次元極限 $d \rightarrow \infty$ を考える



$\sigma^2 \approx$ ポテンシャル $U^{(d)}(X_t)$ の極限速度

スケーリング解析: $d \rightarrow \infty$ 極限でのダイナミクスを比較

Theorem 3.7 & 3.9 BPS, FECMC のポテンシャルの過程は Ornstein-Uhlenbeck 拡散に収束する

$$Y_t^{(d)} := \frac{2U^{(d)}(X_{dt}^d) - d}{\sqrt{d}} \Rightarrow dY_t = -\frac{\sigma^2(\rho)}{4} Y_t dt + \sigma(\rho) dB_t$$

Theorem 4.1

$$\sigma_{\text{FECMC}}^2(\rho) = \sqrt{\frac{32}{\pi}} \left(1 - \frac{(\rho^2 - \rho\sqrt{\frac{\pi}{2}} + \Omega(\rho))^2}{\rho^4 \Omega(\rho)(2 - \Omega(\rho))} \right)$$
$$\sigma_{\text{BPS}}^2(\rho) = \frac{8}{\rho^4} \left(\rho^3 - \rho^2 \sqrt{\frac{8}{\pi}} + \rho - \sqrt{\frac{8}{\pi}} \frac{((1 + \rho^2)\Omega(\rho) - \rho^2)^2}{\Omega(2\rho)} \right)$$

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

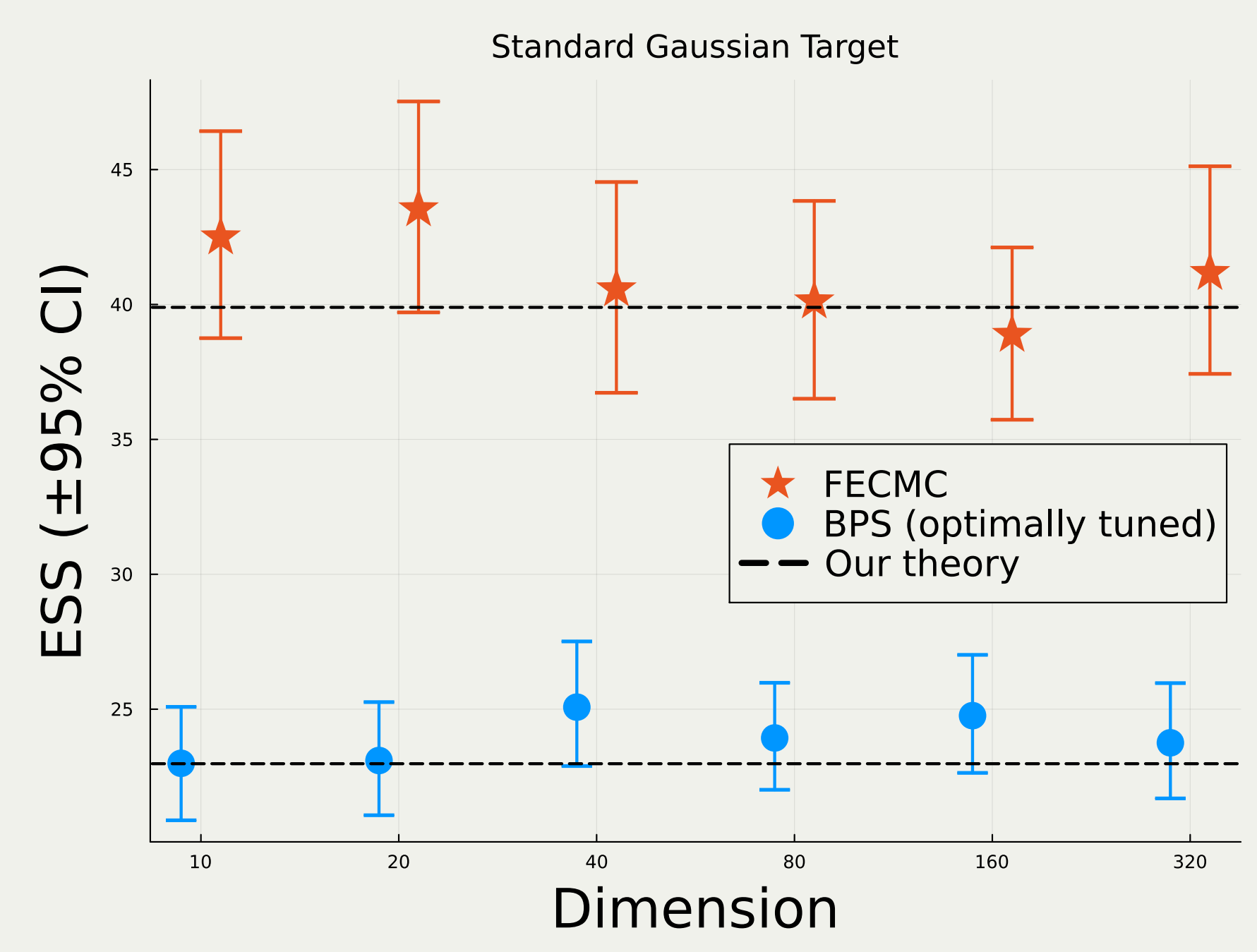
Corollary 4.3

$$\sigma_{\text{FECMC}}^2(0) = \sqrt{\frac{32}{\pi}}$$

ESS: A benchmark

$U^{(d)}$ の有効サンプルサイズ (ESS) は
 $\text{ESS}(U) \approx T \frac{\sigma^2(\rho)}{8}$

$\text{ESS} / T \approx$ 平均自乗誤差 (MSE: Mean Squared Error)



計算複雑性のスケーリング

Potential	$\nabla U(x)$ evals / ESS	$\nabla U(x)$	総計
BPS / FECMC	$O(d)$	$O(d^{0 \sim 1})$	$O(d^{1 \sim 2})$
Hamiltonian Monte Carlo		未解決	
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	$O(d^{1/4})$	$O(d)$	$O(d^{5/4})$

効率的な収束診断

時間微分の過程に注目する
 $\frac{d}{dt} h(X_t) = (\nabla h(X_t) | V_t) =: R_t$
→ MSE が $O(d^{-1})$ 低減可能！

Proposition 4.5

$$\frac{1}{\sqrt{T}} \int_0^T h(X_t) dt \xrightarrow{d, T \rightarrow \infty} N\left(0, \frac{8}{\sigma^2}\right)$$
$$\frac{1}{\sqrt{T}} \int_0^T R_t dt \xrightarrow{d, T \rightarrow \infty} N\left(0, \frac{\sigma^2}{4}\right)$$

一般のターゲットでは σ^2 は不明
→ シミュレーション中に on the fly で推定

