

# Theoretical framework for stochastic algorithms around symmetric $\alpha$ -stable processes

Jian Wang

Fujian Normal University

Based on joint work with Lujing Huang (Fujian Normal Uni.),  
and Mateusz B. Majka (Heriot-Watt Uni., Edinburgh).

September 24, 2022

## 1 Introduction

## 2 Main results

- Fractional Langevin Monte Carlo
- Strict Kantorovich contractions for Euler discretizations

# Background

- Suppose we are given a probability distribution  $\mu$  on  $\mathbb{R}^d$  defined via

$$\mu(dx) = Z^{-1} \exp(-V(x)) dx,$$

where  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  is the potential, and  $Z := \int_{\mathbb{R}^d} \exp(-V(x)) dx$  is the normalizing constant.

- A commonly used example of such process is the solution  $(X_t)_{t \geq 0}$  to the (overdamped) Langevin SDE

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t,$$

where  $(B_t)_{t \geq 0}$  is the standard Brownian motion in  $\mathbb{R}^d$ , and the potential  $V$  is assumed to be sufficiently regular. It can be easily shown that  $\mu$  given above is a stationary distribution of  $(X_t)_{t \geq 0}$ .

# Background

- Suppose we are given a probability distribution  $\mu$  on  $\mathbb{R}^d$  defined via

$$\mu(dx) = Z^{-1} \exp(-V(x)) dx,$$

where  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  is the potential, and  $Z := \int_{\mathbb{R}^d} \exp(-V(x)) dx$  is the normalizing constant.

- A commonly used example of such process is the solution  $(X_t)_{t \geq 0}$  to **the (overdamped) Langevin SDE**

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t,$$

where  $(B_t)_{t \geq 0}$  is the standard Brownian motion in  $\mathbb{R}^d$ , and the potential  $V$  is assumed to be sufficiently regular. It can be easily shown that  $\mu$  given above is a stationary distribution of  $(X_t)_{t \geq 0}$ .

# Langevin Monte Carlo methods

- The (overdamped) Langevin SDE

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t.$$

- Sampling algorithms based on Euler discretizations:

$$\begin{cases} X_{k+1} = X_k - \nabla V(X_k)h + \sqrt{2h}\xi_{k+1}, & k \geq 0, \\ X_0 \sim \mu_0, \end{cases}$$

where  $(\xi_k)_{k \geq 1}$  are i.i.d. with the standard normal distribution. (Langevin Monte Carlo, large-scale Bayesian machine learning, ...)

- Hence the analysis of convergence result for the SDE is an important first step towards evaluating performance of such algorithms, and it is difficult to expect fast convergence of the algorithm without fast convergence of the associated SDE, see Roberts-Tweedie ('96), Roberts-Stramer ('02).

# Langevin Monte Carlo methods

- The (overdamped) Langevin SDE

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t.$$

- Sampling algorithms based on Euler discretizations:

$$\begin{cases} X_{k+1} = X_k - \nabla V(X_k)h + \sqrt{2h}\xi_{k+1}, & k \geq 0, \\ X_0 \sim \mu_0, \end{cases}$$

where  $(\xi_k)_{k \geq 1}$  are i.i.d. with the standard normal distribution. (Langevin Monte Carlo, large-scale Bayesian machine learning, ...)

- Hence the analysis of convergence result for the SDE is an important first step towards evaluating performance of such algorithms, and it is difficult to expect fast convergence of the algorithm without fast convergence of the associated SDE, see Roberts-Tweedie ('96), Roberts-Stramer ('02).

# Langevin Monte Carlo methods

- The (overdamped) Langevin SDE

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t.$$

- Sampling algorithms based on Euler discretizations:

$$\begin{cases} X_{k+1} = X_k - \nabla V(X_k)h + \sqrt{2h}\xi_{k+1}, & k \geq 0, \\ X_0 \sim \mu_0, \end{cases}$$

where  $(\xi_k)_{k \geq 1}$  are i.i.d. with the standard normal distribution. (Langevin Monte Carlo, large-scale Bayesian machine learning, ...)

- Hence the analysis of convergence result for the SDE is an important first step towards evaluating performance of such algorithms, and it is difficult to expect fast convergence of the algorithm without fast convergence of the associated SDE, see Roberts-Tweedie ('96), Roberts-Stramer ('02).

# Langevin SDE

- The (overdamped) Langevin SDE

$$dX_t = -\nabla V(X_t) dt + \sqrt{2} dB_t.$$

- Standard approaches for exponential ergodicity: **Lyapunov-type drift condition by Meyn and Tweedie; functional inequalities; coupling methods**, and so on; see Chen (05') and Wang (05'). In particular, the operator  $L = \Delta - \nabla V$ , which is symmetric on  $L^2(\mathbb{R}^d; \mu)$ . Poincaré inequalities (equivalent to exponential ergodicity) essentially implies **exponential tails of  $\mu(dx) = Z^{-1} e^{-V(x)} dx$** .
- A very natural question is **how to approximate the heavy-tailed  $\mu$** ; for example,

$$\mu(dx) = \frac{C(d, \theta)}{(1 + |x|)^{d+\theta}} dx$$

for some  $\theta > 0$ .

# Question

- Motivating example: Ornstein-Uhlenbeck type processes

$$dX_t = -X_t dt + dZ_t,$$

where  $(Z_t)_{t \geq 0}$  is a symmetric (rotationally invariant)  $\alpha$ -stable process in  $\mathbb{R}^d$  with  $d \geq 1$  and  $\alpha \in (0, 2)$ . The associated invariant probability measure  $\mu(dx)$  is given by

$$\frac{\mu(dx)}{dx} \asymp \frac{1}{(1 + |x|)^{d+\alpha}}.$$

See Sato-Yamazato (84').

- Question** For the heavy-tailed  $\mu$ , how to find the drift term  $b$  such that  $\mu$  can be shown to be a stationary distribution of

$$dX_t = b(X_t) dt + dZ_t?$$

- Motivating example: Ornstein-Uhlenbeck type processes

$$dX_t = -X_t dt + dZ_t,$$

where  $(Z_t)_{t \geq 0}$  is a symmetric (rotationally invariant)  $\alpha$ -stable process in  $\mathbb{R}^d$  with  $d \geq 1$  and  $\alpha \in (0, 2)$ . The associated invariant probability measure  $\mu(dx)$  is given by

$$\frac{\mu(dx)}{dx} \asymp \frac{1}{(1 + |x|)^{d+\alpha}}.$$

See Sato-Yamazato (84').

- Question** For the **heavy-tailed**  $\mu$ , how to find the **drift term**  $b$  such that  $\mu$  can be shown to be a stationary distribution of

$$dX_t = b(X_t) dt + dZ_t?$$

## 1 Introduction

## 2 Main results

- Fractional Langevin Monte Carlo
- Strict Kantorovich contractions for Euler discretizations

# Special drift terms: $d > 2 - \alpha$ .



$$dX_t = b(X_t) dt + dZ_t,$$

where  $(Z_t)_{t \geq 0}$  is a symmetric  $\alpha$ -stable process on  $\mathbb{R}^d$  with  $d \geq 1$  and  $\alpha \in (0, 2)$ .



$$\mu(dx) = Z^{-1} e^{-V(x)} dx.$$

- When  $d > 2 - \alpha$ ,

$$b(x) = -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} \nabla V(y)}{|x-y|^{d-(2-\alpha)}} dy,$$

where  $V \in C^1(\mathbb{R}^d)$  such that  $e^{-V} |\nabla V| \in L^1(\mathbb{R}^d; dx) \cap B_b(\mathbb{R}^d)$ , and  $C_{d,2-\alpha}$  is the positive normalizing constant associated with  $(-\Delta)^{-(2-\alpha)/2}$ .

## Special drift terms: $d > 2 - \alpha$ .



$$dX_t = b(X_t) dt + dZ_t,$$

where  $(Z_t)_{t \geq 0}$  is a symmetric  $\alpha$ -stable process on  $\mathbb{R}^d$  with  $d \geq 1$  and  $\alpha \in (0, 2)$ .



$$\mu(dx) = Z^{-1} e^{-V(x)} dx.$$

- When  $d > 2 - \alpha$ ,

$$b(x) = -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} \nabla V(y)}{|x - y|^{d-(2-\alpha)}} dy,$$

where  $V \in C^1(\mathbb{R}^d)$  such that  $e^{-V} |\nabla V| \in L^1(\mathbb{R}^d; dx) \cap B_b(\mathbb{R}^d)$ , and  $C_{d,2-\alpha}$  is the positive normalizing constant associated with  $(-\Delta)^{-(2-\alpha)/2}$ .

# Special drift term: $d > 2 - \alpha$ .



$$\begin{aligned} b(x) &= -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} \nabla V(y)}{|x-y|^{d-(2-\alpha)}} dy \\ &= C_{d,2-\alpha} e^{V(x)} \nabla \left[ \int_{\mathbb{R}^d} \frac{e^{-V(y)}}{|\cdot-y|^{d-(2-\alpha)}} dy \right] (x) \\ &= e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x). \end{aligned}$$

- Letting  $\alpha \rightarrow 2$ ,  $b(x) \rightarrow -\nabla V(x)$  and  $Z_t \rightarrow \sqrt{2}B_t$ . Then,

$$dX_t = b(X_t) dt + dZ_t$$

becomes

$$dX_t = -\nabla V(X_t) dt + \sqrt{2}dB_t.$$

- For  $d \geq 1$  and  $\alpha \in (1, 2)$ , [Simsekli \('17\)](#); [Nguyen-Simsekli-Richard \('19\)](#). (Fractional Langevin Monte Carlo.)

## Special drift term: $d > 2 - \alpha$ .



$$\begin{aligned} b(x) &= -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} \nabla V(y)}{|x-y|^{d-(2-\alpha)}} dy \\ &= C_{d,2-\alpha} e^{V(x)} \nabla \left[ \int_{\mathbb{R}^d} \frac{e^{-V(y)}}{|\cdot-y|^{d-(2-\alpha)}} dy \right] (x) \\ &= e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x). \end{aligned}$$

- Letting  $\alpha \rightarrow 2$ ,  $b(x) \rightarrow -\nabla V(x)$  and  $Z_t \rightarrow \sqrt{2}B_t$ . Then,

$$dX_t = b(X_t) dt + dZ_t$$

becomes

$$dX_t = -\nabla V(X_t) dt + \sqrt{2}dB_t.$$

- For  $d \geq 1$  and  $\alpha \in (1, 2)$ , Simsekli ('17); Nguyen-Simsekli-Richard ('19). (Fractional Langevin Monte Carlo.)

## Special drift term: $d > 2 - \alpha$ .



$$\begin{aligned} b(x) &= -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} \nabla V(y)}{|x-y|^{d-(2-\alpha)}} dy \\ &= C_{d,2-\alpha} e^{V(x)} \nabla \left[ \int_{\mathbb{R}^d} \frac{e^{-V(y)}}{|\cdot-y|^{d-(2-\alpha)}} dy \right] (x) \\ &= e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x). \end{aligned}$$

- Letting  $\alpha \rightarrow 2$ ,  $b(x) \rightarrow -\nabla V(x)$  and  $Z_t \rightarrow \sqrt{2}B_t$ . Then,

$$dX_t = b(X_t) dt + dZ_t$$

becomes

$$dX_t = -\nabla V(X_t) dt + \sqrt{2}dB_t.$$

- For  $d \geq 1$  and  $\alpha \in (1, 2)$ , **Simsekli ('17); Nguyen-Simsekli-Richard ('19)**. (Fractional Langevin Monte Carlo.)

# Special drift term: $d > 2 - \alpha$ .



$$dX_t = b(X_t) dt + dZ_t.$$



$$b(x) = e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x).$$

- Heuristic proof Lévy-stable-Fokker-Planck equation (Duan-Sun, '12):

$$\partial_t u = -(-\Delta)^{\alpha/2} u - \operatorname{div}(bu).$$

Assume that  $u_\infty(x) = e^{-V(x)}$  is the steady state. Then,

$$(-\Delta)^{\alpha/2} e^{-V} = -\operatorname{div}(be^{-V}).$$

That is,

$$\begin{aligned} -\operatorname{div}(be^{-V}) &= (-\Delta)[(-\Delta)^{-(1-\alpha/2)} e^{-V}] \\ &= -\operatorname{div} \nabla [(-\Delta)^{-(1-\alpha/2)} e^{-V}]. \end{aligned}$$

# Special drift term: $d > 2 - \alpha$ .



$$dX_t = b(X_t) dt + dZ_t.$$



$$b(x) = e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x).$$

- **Heuristic proof** Lévy-stable-Fokker-Planck equation (Duan-Sun, '12):

$$\partial_t u = -(-\Delta)^{\alpha/2} u - \operatorname{div}(bu).$$

Assume that  $u_\infty(x) = e^{-V(x)}$  is the **steady state**. Then,

$$(-\Delta)^{\alpha/2} e^{-V} = -\operatorname{div}(b e^{-V}).$$

That is,

$$\begin{aligned} -\operatorname{div}(b e^{-V}) &= (-\Delta)[(-\Delta)^{-(1-\alpha/2)} e^{-V}] \\ &= -\operatorname{div} \nabla [(-\Delta)^{-(1-\alpha/2)} e^{-V}]. \end{aligned}$$

## Special drift term: $d > 2 - \alpha$ .



$$dX_t = b(X_t) dt + dZ_t.$$



$$b(x) = e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x).$$

- **Heuristic proof** Lévy-stable-Fokker-Planck equation (Duan-Sun, '12):

$$\partial_t u = -(-\Delta)^{\alpha/2} u - \operatorname{div}(bu).$$

Assume that  $u_\infty(x) = e^{-V(x)}$  is the **steady state**. Then,

$$(-\Delta)^{\alpha/2} e^{-V} = -\operatorname{div}(b e^{-V}).$$

That is,

$$\begin{aligned} -\operatorname{div}(b e^{-V}) &= (-\Delta)[(-\Delta)^{-(1-\alpha/2)} e^{-V}] \\ &= -\operatorname{div} \nabla [(-\Delta)^{-(1-\alpha/2)} e^{-V}]. \end{aligned}$$

# Special drift term: $d \leq 2 - \alpha$ .



$$dX_t = b(X_t) dt + dZ_t.$$



$$b(x) = e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x).$$



$$b(x) = e^{V(x)} \nabla (-\Delta)^{-1} ((-\Delta)^{\alpha/2} e^{-V})(x).$$

- When  $d \leq 2 - \alpha$ ,

$$b(x) = e^{V(x)} \int_{-\infty}^x -(-\Delta)^{\alpha/2} e^{-V(z)} dz.$$

# Special drift term: $d \leq 2 - \alpha$ .



$$dX_t = b(X_t) dt + dZ_t.$$



$$b(x) = e^{V(x)} \nabla ((-\Delta)^{-(1-\alpha/2)} e^{-V})(x).$$



$$b(x) = e^{V(x)} \nabla (-\Delta)^{-1} ((-\Delta)^{\alpha/2} e^{-V})(x).$$

- When  $d \leq 2 - \alpha$ ,

$$b(x) = e^{V(x)} \int_{-\infty}^x -(-\Delta)^{\alpha/2} e^{-V(z)} dz.$$

**Assumptions:**  $d > 2 - \alpha$ .

$$b(x) = -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} \nabla V(y)}{|x-y|^{d-(2-\alpha)}} dy.$$

**Assumption (A)**  $V$  is a *radial function* on  $\mathbb{R}^d$ , i.e.,  $V(x) = V(|x|)$  such that

(A1)  $r_0 := \sup\{r > 0 : V'(r) \leq 0\} < \infty$ ;

(A2)

$$\int_0^\infty e^{-V(r)} V'(r) r^d dr > 0;$$

(A3)  $\limsup_{r \rightarrow \infty} [e^{-V(r)} r^{d+\alpha}] < \infty$ .

- The radial property implies that

$$b(x) = -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} V'(|y|) y}{|y| |x-y|^{d-(2-\alpha)}} dy = -b(-x).$$

**Assumptions:**  $d > 2 - \alpha$ .

$$b(x) = -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} \nabla V(y)}{|x-y|^{d-(2-\alpha)}} dy.$$

**Assumption (A)**  $V$  is a *radial function* on  $\mathbb{R}^d$ , i.e.,  $V(x) = V(|x|)$  such that

(A1)  $r_0 := \sup\{r > 0 : V'(r) \leq 0\} < \infty$ ;

(A2)

$$\int_0^\infty e^{-V(r)} V'(r) r^d dr > 0;$$

(A3)  $\limsup_{r \rightarrow \infty} [e^{-V(r)} r^{d+\alpha}] < \infty$ .

- The radial property implies that

$$b(x) = -C_{d,2-\alpha} e^{V(x)} \int_{\mathbb{R}^d} \frac{e^{-V(y)} V'(|y|) y}{|y| |x-y|^{d-(2-\alpha)}} dy = -b(-x).$$

# Main result

$$dX_t = b(X_t) dt + dZ_t.$$

## Theorem

Under Assumption (A), the SDE above has a unique non-explosive **strong solution**  $X := (X_t)_{t \geq 0}$ , and the process  $X$  is **exponentially ergodic** with the **unique** probability measure  $\mu(dx) = Z^{-1} e^{-V(x)} dx$ , such that, for any  $\beta \in [0, \alpha)$  there exists a constant  $\lambda > 0$  so that for any probability measure  $\mu_0$  on  $\mathbb{R}^d$  with finite  $\beta$ -moment and  $t > 0$ , if  $X_0 \sim \mu_0$  with finite first moment then we have

$$\|\mathcal{L}(X_t) - \mu\|_{V_0, \text{Var}} \leq C(\mu_0) e^{-\lambda t},$$

where  $V_0(x) = (1 + |x|)^\beta$ .

- $\limsup_{r \rightarrow \infty} [e^{-V(r)} r^{d+\alpha}] < \infty.$

- Estimate for the rate  $\lambda$ .

# Main result

$$dX_t = b(X_t) dt + dZ_t.$$

## Theorem

Under Assumption (A), the SDE above has a unique non-explosive *strong solution*  $X := (X_t)_{t \geq 0}$ , and the process  $X$  is *exponentially ergodic* with the *unique* probability measure  $\mu(dx) = Z^{-1} e^{-V(x)} dx$ , such that, for any  $\beta \in [0, \alpha)$  there exists a constant  $\lambda > 0$  so that for any probability measure  $\mu_0$  on  $\mathbb{R}^d$  with finite  $\beta$ -moment and  $t > 0$ , if  $X_0 \sim \mu_0$  with finite first moment then we have

$$\|\mathcal{L}(X_t) - \mu\|_{V_0, \text{Var}} \leq C(\mu_0) e^{-\lambda t},$$

where  $V_0(x) = (1 + |x|)^\beta$ .

- $\limsup_{r \rightarrow \infty} [e^{-V(r)} r^{d+\alpha}] < \infty$ .
- Estimate for the rate  $\lambda$ .

# Remark

**Assumption (A)**  $V$  is a *radial function* on  $\mathbb{R}^d$ , i.e.,  $V(x) = V(|x|)$  such that

$$(A1) \quad r_0 := \sup\{r > 0 : V'(r) \leq 0\} < \infty;$$

(A2)

$$\int_0^\infty e^{-V(r)} V'(r) r^d dr > 0.$$

- $\mu(dx) = \frac{C_{d,\theta}}{(1+|x|)^{d+\theta}} dx = Z^{-1} e^{-V(x)} dx.$

- 

$$\langle x, b(x) \rangle \leq 0 \quad \text{for } |x| \gg 1.$$

- 

$$\langle x, b(x) \rangle \leq -c_0 |x|^2 \quad \text{for } |x| \gg 1.$$

# Remark

**Assumption (A)**  $V$  is a *radial function* on  $\mathbb{R}^d$ , i.e.,  $V(x) = V(|x|)$  such that

$$(A1) \quad r_0 := \sup\{r > 0 : V'(r) \leq 0\} < \infty;$$

(A2)

$$\int_0^\infty e^{-V(r)} V'(r) r^d dr > 0.$$

- $\mu(dx) = \frac{C_{d,\theta}}{(1+|x|)^{d+\theta}} dx = Z^{-1} e^{-V(x)} dx.$

- $\langle x, b(x) \rangle \leq 0 \quad \text{for } |x| \gg 1.$

- $\langle x, b(x) \rangle \leq -c_0 |x|^2 \quad \text{for } |x| \gg 1.$

# Remark

**Assumption (A)**  $V$  is a *radial function* on  $\mathbb{R}^d$ , i.e.,  $V(x) = V(|x|)$  such that

$$(A1) \quad r_0 := \sup\{r > 0 : V'(r) \leq 0\} < \infty;$$

(A2)

$$\int_0^\infty e^{-V(r)} V'(r) r^d dr > 0.$$

- $\mu(dx) = \frac{C_{d,\theta}}{(1+|x|)^{d+\theta}} dx = Z^{-1} e^{-V(x)} dx.$

- $\langle x, b(x) \rangle \leq 0 \quad \text{for } |x| \gg 1.$

- $\langle x, b(x) \rangle \leq -c_0 |x|^2 \quad \text{for } |x| \gg 1.$

# Remark

**Assumption (A)**  $V$  is a *radial function* on  $\mathbb{R}^d$ , i.e.,  $V(x) = V(|x|)$  such that

$$(A1) \quad r_0 := \sup\{r > 0 : V'(r) \leq 0\} < \infty;$$

(A2)

$$\int_0^\infty e^{-V(r)} V'(r) r^d dr > 0.$$

- $\mu(dx) = \frac{C_{d,\theta}}{(1+|x|)^{d+\theta}} dx = Z^{-1} e^{-V(x)} dx.$

- 

$$\langle x, b(x) \rangle \leq 0 \quad \text{for } |x| \gg 1.$$

- 

$$\langle x, b(x) \rangle \leq -c_0 |x|^2 \quad \text{for } |x| \gg 1.$$

# Remark

**Assumption (A)**  $V$  is a *radial function* on  $\mathbb{R}^d$ , i.e.,  $V(x) = V(|x|)$  such that

$$(A1) \quad r_0 := \sup\{r > 0 : V'(r) \leq 0\} < \infty;$$

(A2)

$$\int_0^\infty e^{-V(r)} V'(r) r^d dr > 0.$$

- $\mu(dx) = \frac{C_{d,\theta}}{(1+|x|)^{d+\theta}} dx = Z^{-1} e^{-V(x)} dx.$

- 

$$\langle x, b(x) \rangle \leq 0 \quad \text{for } |x| \gg 1.$$

- 

$$\langle x, b(x) \rangle \leq -c_0 |x|^2 \quad \text{for } |x| \gg 1.$$

# Multidimensional case

SDE driven by a symmetric  $\alpha$ -stable process  $(Z_t)_{t \geq 0}$  on  $\mathbb{R}^d$ :

$$dX_t = b(X_t) dt + \sigma(X_{t-}) dZ_t.$$

## Theorem (Zhang-Zhang,'22)

*The process  $(X_t)_{t \geq 0}$  is exponentially ergodic, if there are  $c_0 > 0$  large enough and  $c_1, c_2 > 0$  such that*

$$\langle b(x), x \rangle + c_0 \|\sigma(x)\|^\alpha |x|^{2-\alpha} \leq -c_1 |x|^2 + c_2.$$

# One-dimensional case

one-dimensional SDE driven by a symmetric  $\alpha$ -stable process  $(Z_t)_{t \geq 0}$  on  $\mathbb{R}$  with  $\alpha \in (0, 2]$

$$dX_t = b(X_t) dt + \sigma(X_{t-}) dZ_t.$$

## Theorem

*If one of the following conditions holds:*

(i)  $\alpha = 1$  and

$$\liminf_{r \rightarrow \infty} \inf_{|x| \geq r} \frac{C_{1,1}(\log 2)(2 + \log 2)\sigma(x) - b(x)(\log |x|) \operatorname{sgn}(x)}{|x|(\log |x|)^2(\log \log |x|)} > 0;$$

(ii)  $\alpha \in (0, 2)$  and

$$\liminf_{r \rightarrow \infty} \inf_{|x| \geq r} \frac{-C_{1,\alpha}\pi \cot(\alpha\pi/2)\sigma^\alpha(x)/(2\alpha) - b(x)|x|^{\alpha-1}\operatorname{sgn}(x)}{|x|^\alpha} > 0,$$

*then the process  $X$  is exponentially ergodic.*

## 1 Introduction

## 2 Main results

- Fractional Langevin Monte Carlo
- Strict Kantorovich contractions for Euler discretizations

# Background: Euler discretizations

- Euler discretizations of the following SDE

$$dX_t = b(X_t) dt + dZ_t,$$

where  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is Lipschitz continuous, and  $(Z_t)_{t \geq 0}$  is a stochastic process on  $\mathbb{R}^d$ .

- When  $(Z_t)_{t \geq 0}$  is a rotationally invariant symmetric  $\alpha$ -stable process with  $\alpha = (0, 2]$  (when  $\alpha = 2$ ,  $(Z_t)_{t \geq 0}$  is a Brownian motion), the transitions of the Markov chain for the Euler schemes corresponding to the SDE above with step size  $h > 0$  are given by

$$x \mapsto x + hb(x) + h^{1/\alpha} \xi, \quad x \in \mathbb{R}^d,$$

where  $\xi$  is a random variable with  $\alpha$ -stable law (when  $\alpha = 2$ ,  $\xi$  is a random variable with normal distribution).

# Result

Let  $h > 0$ ,  $g$  be a continuous and strictly increasing function on  $[0, \infty)$  with  $g(0) = 0$ , and  $\xi$  be a random variable whose distribution is given by  $\mu$ . We will consider the Markov chain  $X$  on  $\mathbb{R}^d$  whose transitions are of the following form:

$$x \mapsto x + hb(x) + g(h)\xi, \quad x \in \mathbb{R}^d.$$

We always assume that the coefficient  $b(x)$  satisfies the following assumption.

**Assumption (C):**

- (c1) *There is a constant  $L \geq 0$  such that  $|b(x) - b(y)| \leq L|x - y|$  for  $x, y \in \mathbb{R}^d$ ;*
- (c2) *There are constants  $K > 0$  and  $\mathcal{R} \geq 0$  such that  $\langle x - y, b(x) - b(y) \rangle \leq -K|x - y|^2$  for  $x, y \in \mathbb{R}^d$  with  $|x - y| \geq \mathcal{R}$ .*

# Result

Let  $h > 0$ ,  $g$  be a continuous and strictly increasing function on  $[0, \infty)$  with  $g(0) = 0$ , and  $\xi$  be a random variable whose distribution is given by  $\mu$ . We will consider the Markov chain  $X$  on  $\mathbb{R}^d$  whose transitions are of the following form:

$$x \mapsto x + hb(x) + g(h)\xi, \quad x \in \mathbb{R}^d.$$

We always assume that the coefficient  $b(x)$  satisfies the following assumption.

**Assumption (C):**

- (c1) *There is a constant  $L \geq 0$  such that  $|b(x) - b(y)| \leq L|x - y|$  for  $x, y \in \mathbb{R}^d$ ;*
- (c2) *There are constants  $K > 0$  and  $\mathcal{R} \geq 0$  such that  $\langle x - y, b(x) - b(y) \rangle \leq -K|x - y|^2$  for  $x, y \in \mathbb{R}^d$  with  $|x - y| \geq \mathcal{R}$ .*

# Strict Kantorovich contractivity

A Markov chain with the transition kernel  $p(x, dz)$  is called strictly Kantorovich contractive with respect to the metric  $d(x, y)$ , if there exists a constant  $c \in (0, 1)$  such that

$$W_d(p(x, \cdot), p(y, \cdot)) \leq (1 - c)d(x, y), \quad x, y \in \mathbb{R}^d.$$

The strict Kantorovich contractivity goes back to the famous work by Dobrushin, and is also known as the “Dobrushin uniqueness condition”.

For any fixed  $x, y \in S$  and  $l \geq 0$ , define

$$\pi(x, y) = \mathbb{P}_{x,y}(d(X, Y) = 0), \quad \beta(x, y) = \mathbb{E}_{x,y}[d(X, Y) - d(x, y)],$$

$$\alpha_l(x, y) = \frac{1}{2} \mathbb{E}_{x,y}[(d(X, Y) - d(x, y))^2 1_{\{d(X, Y) < d(x, y) + l\}}].$$

$$\begin{aligned} -a1_{\{R=0\}} + f_0(R) - f_0(r) &\leq -a1_{\{R=0\}} + f'_0(r)(R - r) + \frac{1}{2} \left( \max_{\xi \in [r \wedge R, r \vee R]} f''_0(\xi) \right) \\ &\leq -a1_{\{R=0\}} + f'_0(r)(R - r) + \frac{1}{2} f''_0(r)(R - r)^2 1_{\{R < r\}} \end{aligned}$$

where the last inequality follows from  $f_0^{(3)} > 0$  and  $f_0'' \leq 0$ .

# Strict Kantorovich contractivity

A Markov chain with the transition kernel  $p(x, dz)$  is called strictly Kantorovich contractive with respect to the metric  $d(x, y)$ , if there exists a constant  $c \in (0, 1)$  such that

$$W_d(p(x, \cdot), p(y, \cdot)) \leq (1 - c)d(x, y), \quad x, y \in \mathbb{R}^d.$$

The strict Kantorovich contractivity goes back to the famous work by Dobrushin, and is also known as the “Dobrushin uniqueness condition”. For any fixed  $x, y \in S$  and  $l \geq 0$ , define

$$\pi(x, y) = \mathbb{P}_{x,y}(d(X, Y) = 0), \quad \beta(x, y) = \mathbb{E}_{x,y}[d(X, Y) - d(x, y)],$$

$$\alpha_l(x, y) = \frac{1}{2} \mathbb{E}_{x,y}[(d(X, Y) - d(x, y))^2 1_{\{d(X, Y) < d(x, y) + l\}}].$$

$$\begin{aligned} -a1_{\{R=0\}} + f_0(R) - f_0(r) &\leq -a1_{\{R=0\}} + f'_0(r)(R - r) + \frac{1}{2} \left( \max_{\xi \in [r \wedge R, r \vee R]} f''_0(\xi) \right) \\ &\leq -a1_{\{R=0\}} + f'_0(r)(R - r) + \frac{1}{2} f''_0(r)(R - r)^2 1_{\{R < r\}} \end{aligned}$$

where the last inequality follows from  $f_0^{(3)} > 0$  and  $f_0'' \leq 0$ .

# Strict Kantorovich contractivity

A Markov chain with the transition kernel  $p(x, dz)$  is called strictly Kantorovich contractive with respect to the metric  $d(x, y)$ , if there exists a constant  $c \in (0, 1)$  such that

$$W_d(p(x, \cdot), p(y, \cdot)) \leq (1 - c)d(x, y), \quad x, y \in \mathbb{R}^d.$$

The strict Kantorovich contractivity goes back to the famous work by Dobrushin, and is also known as the “Dobrushin uniqueness condition”. For any fixed  $x, y \in S$  and  $l \geq 0$ , define

$$\pi(x, y) = \mathbb{P}_{x,y}(d(X, Y) = 0), \quad \beta(x, y) = \mathbb{E}_{x,y}[d(X, Y) - d(x, y)],$$

$$\alpha_l(x, y) = \frac{1}{2} \mathbb{E}_{x,y}[(d(X, Y) - d(x, y))^2 1_{\{d(X, Y) < d(x, y) + l\}}].$$

$$\begin{aligned} -a 1_{\{R=0\}} + f_0(R) - f_0(r) &\leq -a 1_{\{R=0\}} + f'_0(r)(R - r) + \frac{1}{2} \left( \max_{\xi \in [r \wedge R, r \vee R]} f''_0(\xi) \right) \\ &\leq -a 1_{\{R=0\}} + f'_0(r)(R - r) + \frac{1}{2} f''_0(r)(R - r)^2 1_{\{R < r\}} \end{aligned}$$

where the last inequality follows from  $f_0^{(3)} > 0$  and  $f_0'' < 0$ .

# General cases

(c3) *There is a constant  $\kappa_0 > 0$  such that*

$$J_{\kappa_0} := \inf_{|z| \leq \kappa_0} (\mu \wedge (\delta_z * \mu))(\mathbb{R}^d) > 0.$$

Let  $\kappa > 0$  be a constant, and

$$(x)_\kappa := \left(1 \wedge \frac{\kappa}{|x|}\right)x, \quad \mu_x(dz) := (\mu \wedge (\delta_x * \mu))(dz),$$

- Write  $\hat{x} = \hat{x}_h := x + hb(x)$ . We will adopt the following Markov coupling (**refined basic coupling (Luo-W., 19')**)  $((X, Y), \mathbb{P}_{x,y})$  of the chain  $X$ :

$$\begin{cases} X = \hat{x} + g(h)z, & \mu(dz); \\ Y = \begin{cases} \hat{y} + g(h)(z + g(h)^{-1}(\hat{x} - \hat{y})_\kappa), & \frac{1}{2}\mu_{g(h)^{-1}(\hat{y}-\hat{x})_\kappa}(dz), \\ \hat{y} + g(h)(z + g(h)^{-1}(\hat{y} - \hat{x})_\kappa), & \frac{1}{2}\mu_{g(h)^{-1}(\hat{x}-\hat{y})_\kappa}(dz), \\ \hat{y} + g(h)z, & \mu(dz) - \dots \end{cases} \end{cases}$$

- Total variation distance

# General cases

(c3) There is a constant  $\kappa_0 > 0$  such that

$$J_{\kappa_0} := \inf_{|z| \leq \kappa_0} (\mu \wedge (\delta_z * \mu))(\mathbb{R}^d) > 0.$$

Let  $\kappa > 0$  be a constant, and

$$(x)_\kappa := \left(1 \wedge \frac{\kappa}{|x|}\right)x, \quad \mu_x(dz) := (\mu \wedge (\delta_x * \mu))(dz),$$

- Write  $\hat{x} = \hat{x}_h := x + hb(x)$ . We will adopt the following Markov coupling (refined basic coupling (Luo-W., 19'))  $((X, Y), \mathbb{P}_{x,y})$  of the chain  $X$ :

$$\begin{cases} X = \hat{x} + g(h)z, & \mu(dz); \\ Y = \begin{cases} \hat{y} + g(h)(z + g(h)^{-1}(\hat{x} - \hat{y})_\kappa), & \frac{1}{2}\mu_{g(h)^{-1}(\hat{y}-\hat{x})_\kappa}(dz), \\ \hat{y} + g(h)(z + g(h)^{-1}(\hat{y} - \hat{x})_\kappa), & \frac{1}{2}\mu_{g(h)^{-1}(\hat{x}-\hat{y})_\kappa}(dz), \\ \hat{y} + g(h)z, & \mu(dz) - \dots \end{cases} \end{cases}$$

- Total variation distance

# General cases

(c3) There is a constant  $\kappa_0 > 0$  such that

$$J_{\kappa_0} := \inf_{|z| \leq \kappa_0} (\mu \wedge (\delta_z * \mu))(\mathbb{R}^d) > 0.$$

Let  $\kappa > 0$  be a constant, and

$$(x)_\kappa := \left(1 \wedge \frac{\kappa}{|x|}\right)x, \quad \mu_x(dz) := (\mu \wedge (\delta_x * \mu))(dz),$$

- Write  $\hat{x} = \hat{x}_h := x + hb(x)$ . We will adopt the following Markov coupling (**refined basic coupling (Luo-W., 19')**)  $((X, Y), \mathbb{P}_{x,y})$  of the chain  $X$ :

$$\begin{cases} X = \hat{x} + g(h)z, & \mu(dz); \\ Y = \begin{cases} \hat{y} + g(h)(z + g(h)^{-1}(\hat{x} - \hat{y})_\kappa), & \frac{1}{2}\mu_{g(h)^{-1}(\hat{y}-\hat{x})_\kappa}(dz), \\ \hat{y} + g(h)(z + g(h)^{-1}(\hat{y} - \hat{x})_\kappa), & \frac{1}{2}\mu_{g(h)^{-1}(\hat{x}-\hat{y})_\kappa}(dz), \\ \hat{y} + g(h)z, & \mu(dz) - \dots \end{cases} \end{cases}$$

- Total variation distance

# Special cases: $\mu$ is rotationally invariant

(c4) The density function  $m(x) = m(|x|)$  of the distribution  $\mu$  satisfies that  $m(r)$  is non-increasing in  $(0, \infty)$ , and has the finite first moment, i.e.,  $\int_{\mathbb{R}^d} |z| m(|z|) dz < \infty$ .

- For this special case, we use the following Markov coupling (W., 16'; Majka, 17')  $((X, Y^{(1)}), \mathbb{P}_{x,y})$  of the chain  $X$ :

$$\begin{cases} X = \hat{x} + g(h)z, & \mu(dz); \\ Y^{(1)} = \begin{cases} \hat{y} + g(h)(z + g(h)^{-1}(\hat{x} - \hat{y})_{\kappa}), & \mu_{g(h)^{-1}(\hat{y} - \hat{x})_{\kappa}}(dz), \\ \hat{y} + g(h)R_{\hat{x}, \hat{y}}(z), & \mu(dz) - \mu_{g(h)^{-1}(\hat{y} - \hat{x})_{\kappa}}(dz). \end{cases} \end{cases}$$

Here,

$$R_{x,y}(z) := \begin{cases} z - \frac{2\langle x-y, z \rangle}{|x-y|^2}(x-y), & x \neq y, \\ z, & x = y. \end{cases}$$

- Total variation distance and  $W_1$ -distance

# Remark

- In particular, we observe that for isotropic noise distributions  $\mu$  (for which both couplings discussed above are applicable), the result based on the reflection coupling can lead to a slightly better contractivity constant than the result based on the refined basic coupling.
- We show that chains driven by an  $\alpha$ -stable noise can have better contraction rates than corresponding chains driven by a Gaussian noise, due to the heavy tails of the  $\alpha$ -stable distribution.

*Thank you for your attention!*