

UNBIASED APPROXIMATION OF THE ERGODIC MEASURE FOR PIECEWISE α -STABLE ORNSTEIN-UHLENBECK PROCESSES ARISING IN QUEUEING NETWORKS

XINGHU JIN, GUODONG PANG, YU WANG, AND LIHU XU

ABSTRACT. Piecewise α -stable Ornstein-Uhlenbeck (OU) processes arising in queue networks usually do not have an explicit dissipation, which makes the related numerical methods such as Euler-Maruyama (EM) scheme more difficult to analyze. We develop an EM scheme with decreasing step size $\Lambda = (\eta_n)_{n \in \mathbb{N}}$ to approximate their ergodic measures. This approximation does not have a bias and has a rate $\eta_n^{1/\alpha}$ in Wasserstein-1 distance. We show by the classical OU process that our convergence rate is optimal. We further prove the central limit theorem (CLT) and moderate deviation principle (MDP) for the empirical measure of these piecewise α -stable Ornstein-Uhlenbeck processes. In addition, we use the Sinkhorn-Knopp algorithm to compute the Wasserstein-1 distance and conduct simulations for several concrete examples.

CONTENTS

1. Introduction	2
1.1. Related works	3
1.2. Our contributions and main results	4
1.3. Notations	6
2. Key ingredients for the proofs of the main results	7
2.1. Mollification of g and properties of the mollifier g_ε	7
2.2. Jacobi flow for the mollified process $(\mathbf{X}_t^{\varepsilon, \mathbf{x}})_{t \geq 0}$.	9
2.3. Semigroups, generators and their properties	10
2.4. Ergodicity in a weighted total variation distance and moment estimates	11
3. Proof of Theorem 1.3	11
3.1. The strategy of the proof	11
3.2. Proof of Theorem 1.3	12
4. Proofs of Theorems 1.4 and 1.5	13
4.1. Proof of Theorem 1.4	13
4.2. Proof of Theorem 1.5	17
5. Optimal error bound for the classical Ornstein-Uhlenbeck process	19
6. Sinkhorn-Knopp algorithm	21
6.1. Preliminary of Sinkhorn-Knopp algorithm	21
6.2. Simulations for two queue networks	22
Appendix A. Proofs of the lemmas in Section 2	25
A.1. Proof of Lemma 2.2	25
A.2. Proof of Lemma 2.3	26
A.3. Proof of Lemma 2.4	27
A.4. Proof of Lemma 2.6	28

2020 *Mathematics Subject Classification.* 60H10; 37M25; 60G51; 60H07; 60H35; 60G52.

Key words and phrases. Euler-Maruyama scheme; piecewise Ornstein-Uhlenbeck driven by α -stable motion; Wasserstein-1 distance; Stein's equation; Central limit theorem; Moderate deviation principle.

A.5. Proof of Lemma 2.5	32
Appendix B. Proofs of the lemmas in Section 3	32
B.1. Proof of Lemma 3.1	32
B.2. Proof of Lemma 3.2	33
B.3. Proof of Lemma 3.3	34
B.4. Proof of Lemma 3.4	36
Appendix C. Proofs of the lemmas in Section 4	36
Acknowledgements	39
References	40

1. INTRODUCTION

A large family of diffusion-scaled queueing processes in stochastic networks in heavy traffic converge to a universal limit, which is governed by the following stochastic differential equation (SDE):

$$(1.1) \quad d\mathbf{X}_t = g(\mathbf{X}_t) dt + \boldsymbol{\sigma} d\mathbf{Z}_t, \quad \mathbf{X}_0 = \mathbf{x} \in \mathbb{R}^d,$$

where \mathbf{X}_t is \mathbb{R}^d -valued for each $t \geq 0$, the coefficient $\boldsymbol{\sigma} = [\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_d]$ is a non-singular constant matrix in $\mathbb{R}^{d \times d}$ with each column vector $\boldsymbol{\sigma}_i \in \mathbb{R}^d$, and $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a Lipschitz function defined as

$$(1.2) \quad \begin{aligned} g(\mathbf{x}) &= \boldsymbol{\ell} - \mathbf{M}(\mathbf{x} - \langle \mathbf{e}_d, \mathbf{x} \rangle^+ \mathbf{v}) - \langle \mathbf{e}_d, \mathbf{x} \rangle^+ \boldsymbol{\Gamma} \mathbf{v} \\ &= \begin{cases} \boldsymbol{\ell} - (\mathbf{M} + (\boldsymbol{\Gamma} - \mathbf{M})\mathbf{v}\mathbf{e}'_d)\mathbf{x}, & \text{if } \mathbf{e}'_d \mathbf{x} > 0, \\ \boldsymbol{\ell} - \mathbf{M}\mathbf{x}, & \text{if } \mathbf{e}'_d \mathbf{x} \leq 0, \end{cases} \end{aligned}$$

where $\boldsymbol{\ell} \in \mathbb{R}^d$, the vector $\mathbf{v} \in \mathbb{R}_+^d$ satisfies $\langle \mathbf{e}_d, \mathbf{v} \rangle = \mathbf{e}'_d \mathbf{v} = 1$ with $\mathbf{e}_d = (1, \dots, 1)' \in \mathbb{R}^d$, the matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$ is a non-singular M-matrix (see, its definition below and more details in [7, Chapter 6]), the term $\langle \mathbf{e}_d, \mathbf{x} \rangle^+$ means the positive part of $\langle \mathbf{e}_d, \mathbf{x} \rangle$, and $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_d)$ with $\gamma_i \in \mathbb{R}_+, i = 1, \dots, d$. Here, we use notation z' for the transpose of any $z \in \mathbb{R}^d$. Furthermore, the α stable noise \mathbf{Z}_t is one of the below processes:

- (i) For $\alpha = 2$, the noise \mathbf{Z}_t is the standard d dimensional Brownian motion;
- (ii) For $\alpha \in (1, 2)$, the noise $\mathbf{Z}_t = (Z_t^1, \dots, Z_t^d)'$ is the cylindrical stable noise, and $Z_t^i, i = 1, \dots, d$ are i.i.d. symmetric α -stable processes in dimension 1 with Lévy measure $\nu_1(dz) = c_\alpha |z|^{-1-\alpha} dz$, where $c_\alpha = \alpha 2^{\alpha-1} \pi^{-1/2} \Gamma((1+\alpha)/2) / \Gamma(1-\alpha/2)$;
- (iii) For $\alpha \in (1, 2)$, the noise \mathbf{Z}_t is the d dimensional rotationally symmetric α stable noise with Lévy measure $\nu(d\mathbf{z}) = C_{d,\alpha} |\mathbf{z}|^{-\alpha-d} d\mathbf{z}$, where the constant $C_{d,\alpha}$ is given by

$$C_{d,\alpha} = \alpha 2^{\alpha-1} \pi^{-d/2} \frac{\Gamma((d+\alpha)/2)}{\Gamma(1-\alpha/2)}.$$

The cases in (i) and (ii) arise from multiclass parallel server networks and many-server queues with phase-type service times with light-tailed and heavy-tailed arrival processes, respectively. We also consider the case (iii) which have applications in data science (see, e.g., [15], [44]) and mathematical finance (see, e.g., [47]).

The SDE (1.1) is often called piecewise Ornstein Uhlenbeck (OU) process and has been widely used to model queueing networks (see, e.g., [4, 5, 9, 23]). It is often hard to find steady states for queueing networks directly, whereas the invariant measure of the limiting SDE can be used as an approximation in the appropriate scaling regimes. Therefore, it is

highly desirable to figure out the ergodic measure of SDE (1.1) and design an algorithm to generate random variables with this distribution.

Throughout this paper, we assume that the matrices \mathbf{M} , $\mathbf{\Gamma}$ and the vector \mathbf{v} satisfy the following assumption:

Assumption 1.1. *One of the following conditions holds:*

- (i) $\mathbf{M}\mathbf{v} \geq \mathbf{\Gamma}\mathbf{v} \geq \mathbf{0}$;
- (ii) $\mathbf{M} = \text{diag}(m_1, \dots, m_d)$ with $m_i > 0$, $i = 1, \dots, d$ and $\mathbf{\Gamma}\mathbf{v} \neq \mathbf{0}$,

where the notations \geq and \neq are defined in (1.7) and (1.8) respectively.

Under Assumption 1.1, the process $(\mathbf{X}_t)_{t \geq 0}$ is exponentially ergodic with the ergodic measure μ (see, [5, Theorem 3.5] or Lemma 2.5 below). We shall use the following Euler-Maruyama (EM) scheme to approximate the ergodic measure μ . Let $\Delta = (\eta_n)_{n \in \mathbb{N}}$ be the step size, for $n \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and

$$(1.3) \quad \mathbf{Y}_{t_{n+1}} = \mathbf{Y}_{t_n} + \eta_{n+1}g(\mathbf{Y}_{t_n}) + \sigma \Delta \mathbf{Z}_{\eta_{n+1}}, \quad \mathbf{Y}_0 = \mathbf{X}_0 = \mathbf{x},$$

where $t_n = \sum_{i=1}^n \eta_i$, $t_0 = 0$ and $\Delta \mathbf{Z}_{\eta_{n+1}} = \mathbf{Z}_{t_{n+1}} - \mathbf{Z}_{t_n}$. Then $(\mathbf{Y}_{t_n})_{n \in \mathbb{N}_0}$ is a non-homogeneous Markov chain.

1.1. Related works. *Diffusion approximations in stochastic networks.* The piecewise Ornstein–Uhlenbeck diffusion in (1.1) is the scaling limit of a large family of stochastic networks and queueing systems. In the Halfin-Whitt regime, such diffusions are established for multi-class $GI/Ph/n$ and $GI/Ph/n + GI$ queues in [20, 43], while their ergodic properties are studied in [23] and the related numerical analysis of the stationary distribution are studied in [19, 21]. In [5, 38], the α -stable (controlled) limiting diffusions are established for $G/M/n(+M)$ and multiclass $G/M/n(+M)$ queues, respectively, when the arrival processes are heavy-tailed (such as regularly varying interarrival times) and converge to an α -stable motion or independent α -stable motions, and the ergodicity properties of such processes are studied in [4, 5]. In fact, for (multiclass) $G/Ph/n$ and $G/Ph/n + GI$ queues, when the arrival processes are heavy-tailed and converge to an α -stable motion, one also obtains the limiting process as given in (1.1) (the proof follows from slight modifications of [20, 38, 43]).

Due to the non-Markovian and network complexities, the stationary distribution and ergodic properties of these queueing networks/systems are impossible to be derived directly. Therefore, their (jump) diffusion approximations become useful since the related techniques of studying invariant measures and ergodic properties have been well developed. We refer the reader to the recent studies on the relevant ergodic properties in these networks and queueing systems in [1, 3, 4, 5, 9, 10, 23, 25]. In this paper, we focus on the approximations of the ergodic measures of the limiting diffusions. This is in the same spirit with the Euler schemes for constrained diffusions as scaling limits of stochastic networks in [12] and for piecewise Ornstein-Uhlenbeck diffusions as scaling limits of $GI/Ph/n + GI$ queues and parallel server networks in [28].

EM scheme. Recently, Fang et al. [24] used the EM scheme to approximate ergodic measures for diffusions with help of Malliavin calculus and Stein’s method and obtained a convergence rate in Wasserstein-1 distance. There are also some references concerning on the computation of ergodic measures for SDEs, see [6, 39, 40] and references therein, but most of them are asymptotic type. Strong convergence for EM scheme can be found in [31] and references therein. Jin et al. [28] used the EM scheme to approximate the ergodic measure for the piecewise OU process driven by Brownian motion and considered the related asymptotics for the original process.

Central limit theorem (CLT) and Moderate deviation principle (MDP) for empirical measures. It follows from [5, Theorem 3.5] (or see Lemma 2.5 below) that the SDE (1.1) is exponentially ergodic with ergodic measure μ , which implies that the Birkhoff ergodic theorem holds true for the empirical measure \mathcal{E}_T (see the definition in (1.5) below) of the process $(\mathbf{X}_t)_{t \geq 0}$ (i.e., $\lim_{T \rightarrow \infty} \mathcal{E}_T = \mu$ almost surely); see, for instance, [18]. It is natural to consider its CLT and MDP with respect to \mathcal{E}_T . The recent work [28] considered the CLT and MDP for the SDE in (1.1) with $\alpha = 2$, it is natural to ask whether the CLT and MDP hold true for the case $\alpha \in (1, 2)$.

Sinkhorn-Knopp algorithm. The Sinkhorn-Knopp algorithm is widely used in approximating the Wasserstein type distances, see [16, 17, 29, 32, 33, 53] and references therein. Cuturi [17] introduced the Sinkhorn distance to study the optimal transportation problem between two probability measures, and used this distance with the entropy constraint to develop algorithms with a cheap cost. Such Sinkhorn distance with constraints can be computed by the Sinkhorn and Knopp's fixed point iteration. In addition, Luise et al. [33] gave regularity for the regularized Sinkhorn distance and a convergence rate for the approximation in Wasserstein distance. We refer the reader to [30] and references therein for more details about the Sinkhorn-Knopp algorithm.

1.2. Our contributions and main results. Jin et al. [28] develops an EM scheme for approximating the invariant measure of piecewise Ornstein-Uhlenbeck diffusions driven by Brownian motion. However, the numerical scheme in [28] has a constant step size, which leads to a bias in the approximation, see [28, Theorem 1]. In order to remove this bias, we will use a numerical scheme with decreasing step sizes, which converges to the ergodic measure exactly as the time tends to infinity.

The very recent works [13, 14] provide an approximation to the ergodic measure of SDEs driven by α -stable process and get convergence rates in Wasserstein-1 distance. However, they have to assume that the drift $g \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ satisfies the following dissipation condition: $\langle \mathbf{x} - \mathbf{y}, g(\mathbf{x}) - g(\mathbf{y}) \rangle \leq -c|\mathbf{x} - \mathbf{y}|^2 + K$ with some positive constants c and K . Most of the works on approximating the ergodic measures of SDEs need a similar assumption, see [27, 31, 35, 39, 40, 42]. However, the piecewise linear drift g in (1.2) does not satisfy such a dissipation condition, see (2.5) below, hence the known approximation methods do not work. Borrowing the mollification technique in [28] and the intermediate matrix \mathbf{Q} proposed in [5], we are able to establish a desirable dissipation property for the drift g , and surprisingly find that a straightforward strong approximation works.

Note that the SDE (1.1) driven by α -stable noise does not have finite second moment when $\alpha \in (1, 2)$, one may suspect whether the CLT and MDP will hold for the empirical measure \mathcal{E}_T . Our second main result confirms the both asymptotic properties are true. In their proofs, Stein's method with nonlocal operator plays an important role, the related analysis is much more delicate than the one in the case $\alpha = 2$ and has its own independent interest.

Besides proving the theorems about the approximation error and the CLT and MDP, we use the recently developed the Sinkhorn-Knopp algorithm to verify the convergence of our numerical scheme in Wasserstein-1 distance. This algorithm works very well for both the light-tailed models and the α -stable heavy tailed ones.

Before presenting our main results, we impose some assumptions on the step size sequence $\Lambda = (\eta_n)_{n \in \mathbb{N}}$. Conveniently, for some constant $\beta \in (0, 1]$, define ω as below:

$$(1.4) \quad \omega = \lim_{k \rightarrow \infty} \left\{ \frac{\eta_k^\beta - \eta_{k+1}^\beta}{\eta_{k+1}^{1+\beta}} \right\}.$$

Assumption 1.2. *The positive decreasing step size sequence $\Lambda = (\eta_n)_{n \in \mathbb{N}}$ satisfies the following conditions:*

- (i) $\eta_1 < 1$, $\sum_{n=1}^{\infty} \eta_n = \infty$ and $\sum_{n=1}^{\infty} \eta_n^2 < \infty$;
- (ii) *The above ω satisfies $\omega < \alpha\beta\theta$, where the constant θ is defined in (2.6) below.*

Our first main result is the following theorem about the EM scheme, which provides a nonasymptotic estimate for the error between the ergodic measures of the SDE and its EM scheme, whose proofs will be given in Section 3 below.

Theorem 1.3. *For $\alpha \in (1, 2]$, under Assumptions 1.1 and 1.2, let μ be the ergodic measure for the process $(\mathbf{X}_t)_{t \geq 0}$ of SDE (1.1), and let the EM scheme $(\mathbf{Y}_{t_n})_{n \in \mathbb{N}_0}$ be given by (1.3). Then there exists a positive constant C independent of $(\eta_n)_{n \in \mathbb{N}}$ such that for any $n \in \mathbb{N}$ and $\mathbf{x} \in \mathbb{R}^d$, one has*

$$\mathcal{W}_1(\mu, \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}})) \leq C(1 + |\mathbf{x}|)\eta_n^{1/\alpha},$$

where $\mathbf{Y}_{t_n}^{\mathbf{x}}$ means the marginal random variable of process $(\mathbf{Y}_{t_n})_{n \in \mathbb{N}_0}$ at time t_n with $\mathbf{Y}_0 = \mathbf{x}$ and \mathcal{W}_1 is the Wasserstein-1 distance defined in (1.9) below.

Our second set of main results includes the CLT and MDP for the long-term behavior of $(\mathbf{X}_t)_{t \geq 0}$ with $\alpha \in (1, 2)$. Recall that the CLT and MDP for the process $(\mathbf{X}_t)_{t \geq 0}$ with $\alpha = 2$ has been shown in [28, Theorems 2 and 3]. For any $\mathbf{x} \in \mathbb{R}^d$ and $T > 0$, the empirical measure $\mathcal{E}_T^{\mathbf{x}}$ of $(\mathbf{X}_t)_{t \geq 0}$ is defined by

$$(1.5) \quad \mathcal{E}_T^{\mathbf{x}}(A) = \frac{1}{T} \int_0^T \delta_{\mathbf{X}_s^{\mathbf{x}}}(A) ds, \quad A \in \mathcal{B}(\mathbb{R}^d),$$

where $\mathcal{B}(\mathbb{R}^d)$ is the collection of Borel sets on \mathbb{R}^d , and $\delta_{\mathbf{y}}(\cdot)$ is the delta measure defined as $\delta_{\mathbf{y}}(A) = 1$ if $\mathbf{y} \in A$ and $\delta_{\mathbf{y}}(A) = 0$ otherwise. It is easy to check that for any bounded Borel measurable function $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$,

$$\mathcal{E}_T^{\mathbf{x}}(h) = \frac{1}{T} \int_0^T h(\mathbf{X}_s^{\mathbf{x}}) ds.$$

Theorem 1.4 (CLT). *For $\alpha \in (1, 2)$ and under Assumption 1.1, let μ be the ergodic measure for the process $(\mathbf{X}_t)_{t \geq 0}$. For any $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$ and $\mathbf{x} \in \mathbb{R}^d$, the term $\sqrt{t}[\mathcal{E}_t^{\mathbf{x}}(h) - \mu(h)]$ converges weakly to Gaussian distribution $\mathcal{N}(0, \mathcal{V}(f_h))$ as $t \rightarrow \infty$, where $\mathcal{V}(f_h)$ is given as following:*

$$(1.6) \quad \mathcal{V}(f_h) = \begin{cases} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} [f_h(\mathbf{x} + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{x})]^2 \nu(d\mathbf{z}) \mu(d\mathbf{x}), & \text{rotationally symmetric noise,} \\ \sum_{i=1}^d \int_{\mathbb{R}^d} \int_{\mathbb{R} \setminus \{0\}} [f_h(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - f_h(\mathbf{x})]^2 \nu_1(dz_i) \mu(d\mathbf{x}), & \text{cylindrical stable noise,} \end{cases}$$

where $\boldsymbol{\sigma} = [\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_d]$ and f_h is the solution to the Stein's equation (4.1) below.

Theorem 1.5 (MDP). *For $\alpha \in (1, 2)$ and under Assumption 1.1, let μ be the ergodic measure for the process $(\mathbf{X}_t)_{t \geq 0}$. For any $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$, $\mathbf{x} \in \mathbb{R}^d$, and any measurable set $A \subseteq \mathbb{R}$,*

$$\begin{aligned} - \inf_{z \in A} \frac{z^2}{2\mathcal{V}(f_h)} &\leq \liminf_{t \rightarrow \infty} \frac{1}{a_t^2} \log \mathbb{P} \left(\frac{\sqrt{t}}{a_t} [\mathcal{E}_t^{\mathbf{x}}(h) - \mu(h)] \in A \right) \\ &\leq \limsup_{t \rightarrow \infty} \frac{1}{a_t^2} \log \mathbb{P} \left(\frac{\sqrt{t}}{a_t} [\mathcal{E}_t^{\mathbf{x}}(h) - \mu(h)] \in A \right) \leq - \inf_{z \in A} \frac{z^2}{2\mathcal{V}(f_h)}, \end{aligned}$$

where \bar{A} and $\overset{\circ}{A}$ are the closure and interior of the set A , respectively; the function a_t satisfies $a_t \rightarrow \infty$ and $t^{-1/2}a_t \rightarrow 0$ as $t \rightarrow \infty$, and $\mathcal{V}(f_h)$ is defined in (1.6).

1.3. Notations. Throughout this paper, the following notations will be used. We use normal font for scalars (e.g. a, A, \dots) and boldface for vectors and matrices (e.g. $\mathbf{x}, \mathbf{y}, \mathbf{A}, \mathbf{B}, \dots$). For any $a \in \mathbb{R}$, define $a^+ = \max\{a, 0\}$. For $x, y \in \mathbb{R}$, $x \vee y = \max\{x, y\}$, $x \wedge y = \min\{x, y\}$. And we use \mathbb{R}^d (and \mathbb{R}_+^d), $d \geq 1$, to denote real-valued d -dimensional (non-negative) vectors, and write \mathbb{R} (and \mathbb{R}_+) for $d = 1$. Additionally, $\mathbb{R}_0^d = \mathbb{R}^d \setminus \{\mathbf{0}\}$.

For any vector $\mathbf{u} = (u_1, \dots, u_d)'$ and $\mathbf{v} = (v_1, \dots, v_d)'$ in \mathbb{R}^d , let $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}'\mathbf{v} = \sum_{i=1}^d u_i v_i$ be the inner product. This induces the Euclidean norm $|\mathbf{u}| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2}$. Besides, denote $\mathbf{u}/\mathbf{v} = (u_1/v_1, \dots, u_d/v_d)'$. On the other hand, for any vector $\mathbf{z} = (z_1, \dots, z_d)'$ $\in \mathbb{R}^d$, let $\|\mathbf{z}\|_{\max} = \max\{|z_i|, i = 1, \dots, d\}$ be the maximum absolute value of complements of \mathbf{z} . We write

$$(1.7) \quad \mathbf{z} \geq \mathbf{0} \quad (\mathbf{z} > \mathbf{0})$$

to indicate that all the components of \mathbf{z} are non-negative (positive), and

$$(1.8) \quad \mathbf{z} \gneq \mathbf{0}$$

stands for $\mathbf{z} \geq \mathbf{0}$ and $\mathbf{z} \neq \mathbf{0}$, and analogously for a matrix in $\mathbb{R}^{d \times d}$.

For any matrices $\mathbf{A} = (A_{ij})_{d \times d}$ and $\mathbf{B} = (B_{ij})_{d \times d}$ in $\mathbb{R}^{d \times d}$, denote their Hilbert Schmidt inner product as $\langle \mathbf{A}, \mathbf{B} \rangle_{\text{HS}} = \sum_{i,j=1}^d A_{ij} B_{ij}$ and denote the Hilbert Schmidt norm as $\|\mathbf{A}\|_{\text{HS}} = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle_{\text{HS}}}$. The operator norm of matrix \mathbf{A} is defined as $\|\mathbf{A}\|_{\text{op}} = \sup_{\mathbf{x} \in \mathbb{R}^d, |\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|$, and we have relation $\|\mathbf{A}\|_{\text{op}} \leq \|\mathbf{A}\|_{\text{HS}} \leq \sqrt{d}\|\mathbf{A}\|_{\text{op}}$. Besides, let $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ be the minimum and maximum eigenvalues for symmetric matrix \mathbf{A} respectively. On the other hand, a matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$ is called a \mathbf{M} -matrix (see, [7, Definition 1.2, p. 133]) if its off-diagonal entries are less than or equal to zero and it can be expressed as $\mathbf{M} = s\mathbf{I} - \mathbf{N}$ for a constant $s > 0$ and a matrix $\mathbf{N} \in \mathbb{R}^{d \times d}$ with its entries being all nonnegative and $\rho(\mathbf{N}) \leq s$, where \mathbf{I} and $\rho(\mathbf{N})$ denote the identity matrix in $\mathbb{R}^{d \times d}$ and spectral radius of \mathbf{N} respectively.

For any two probability measures ν_1 and ν_2 defined on \mathbb{R}^d , the total variation distance $\|\nu_1 - \nu_2\|_{\text{TV}}$ is defined as below [48, p. 10]):

$$\|\nu_1 - \nu_2\|_{\text{TV}} = 2 \inf_{\Pi \in \mathcal{C}(\nu_1, \nu_2)} \{ \Pi(\{\mathbf{X} \neq \mathbf{Y}\}) : \mathcal{L}(\mathbf{X}) = \nu_1, \mathcal{L}(\mathbf{Y}) = \nu_2 \},$$

where the infimum runs over the set $\mathcal{C}(\nu_1, \nu_2)$ of all probability measures Π on with marginals ν_1 and ν_2 , and $\mathcal{L}(\mathbf{X})$ is the law of random vector \mathbf{X} . An alternative characterisation of the total variation distance is as the dual norm to the supremum norm on the space of bounded Borel measurable functions (see, [26, p. 57]):

$$\|\nu_1 - \nu_2\|_{\text{TV}} = \sup \left\{ \int_{\mathbb{R}^d} \psi(\mathbf{x}) \nu_1(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x}) \nu_2(d\mathbf{x}) : \sup_{\mathbf{x} \in \mathbb{R}^d} |\psi(\mathbf{x})| \leq 1 \right\}.$$

Given a weight function $V : \mathbb{R}^d \rightarrow [1, +\infty)$, define a weighted supremum norm on measurable functions by

$$\|\psi\|_V = \sup_{\mathbf{x} \in \mathbb{R}^d} \frac{|\psi(\mathbf{x})|}{V(\mathbf{x})},$$

as well as the weighted total variation distance is defined by (see, [26, (5.18)])

$$\|\nu_1 - \nu_2\|_{\text{TV}, V} = \sup \left\{ \int_{\mathbb{R}^d} \psi(\mathbf{x}) \nu_1(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x}) \nu_2(d\mathbf{x}) : \|\psi\|_V \leq 1 \right\}.$$

The L^p -Wasserstein distance $\mathcal{W}_p(\nu_1, \nu_2)$ with $p \geq 1$ between ν_1 and ν_2 is defined as

$$\mathcal{W}_p(\nu_1, \nu_2) = \inf_{\Pi \in \mathcal{C}(\nu_1, \nu_2)} \left(\int_{\mathbb{R}^{2d}} |\mathbf{x} - \mathbf{y}|^p d\Pi(\mathbf{x}, \mathbf{y}) \right)^{1/p}.$$

Using the Kantorovich duality (see, [48, Theorem 5.10]), the Wasserstein-1 distance $\mathcal{W}_1(\nu_1, \nu_2)$ also can be defined as

$$(1.9) \quad \mathcal{W}_1(\nu_1, \nu_2) = \sup_{h \in \text{Lip}(1)} \left(\int_{\mathbb{R}^d} h(\mathbf{x}) \nu_1(d\mathbf{x}) - \int_{\mathbb{R}^d} h(\mathbf{x}) \nu_2(d\mathbf{x}) \right),$$

where $\text{Lip}(1)$ is the set of Lipschitz functions with Lipschitz constant 1, that is, $\text{Lip}(1) = \{f : |f(\mathbf{x}) - f(\mathbf{y})| \leq |\mathbf{x} - \mathbf{y}|, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d\}$. For a probability measure $\pi(d\mathbf{x})$ and an integrable function f on \mathbb{R}^d , we write $\pi(f) = \int_{\mathbb{R}^d} f(\mathbf{x}) \pi(d\mathbf{x})$. For a given probability measure ν , we define $L^2(\nu)$ as the Hilbert space induced by ν with inner product

$$\langle f_1, f_2 \rangle_\nu = \int f_1(\mathbf{x}) f_2(\mathbf{x}) \nu(d\mathbf{x}) \quad \text{for } f_1, f_2 \in L^2(\nu).$$

The notation $\mathcal{N}(\mathbf{a}, \mathbf{A})$ denotes the Gaussian distribution with mean $\mathbf{a} \in \mathbb{R}^d$ and covariance matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$. A sequence of random variables $\{Y_n, n \geq 1\}$ is said to converge weakly or converge in distribution to a limit Y_∞ , denoted by $Y_n \Rightarrow Y_\infty$, if $\lim_{n \rightarrow \infty} \mathbb{E}f(Y_n) = \mathbb{E}f(Y_\infty)$ for all bounded continuous function f . In addition, $Y_n \xrightarrow{p} Y_\infty$ means convergence in probability, namely $\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - Y_\infty| > \delta) = 0$ for all $\delta > 0$.

We denote by $\mathcal{C}(\mathbb{R}^d)$ all the continuous functions from \mathbb{R}^d to \mathbb{R} , by $\mathcal{C}_b(\mathbb{R}^d)$ all the bounded continuous functions from \mathbb{R}^d to \mathbb{R} . For $f \in \mathcal{C}_b(\mathbb{R}^d)$, denote $\|f\|_\infty = \sup_{\mathbf{x} \in \mathbb{R}^d} |f(\mathbf{x})|$. $\mathcal{C}^k(\mathbb{R}^d)$ denotes the collection of k -th continuously differentiable functions with integers $k \geq 1$. For any $f \in \mathcal{C}^2(\mathbb{R}^d)$, let ∇f and $\nabla^2 f$ be the gradient and Hessian matrix for f , and denote $\|\nabla f\|_\infty = \sup_{\mathbf{x} \in \mathbb{R}^d} |\nabla f(\mathbf{x})|$ and $\|\nabla^2 f\|_{\text{HS}, \infty} = \sup_{\mathbf{x} \in \mathbb{R}^d} \|\nabla^2 f(\mathbf{x})\|_{\text{HS}}$. For any $f \in \mathcal{C}^1(\mathbb{R})$, we often denote by \dot{f} its derivative. We further introduce a function space $\mathcal{C}_{lin}(\mathbb{R}^d)$ as below:

$$\mathcal{C}_{lin}(\mathbb{R}^d) = \left\{ f \in \mathcal{C}(\mathbb{R}^d) : \sup_{\mathbf{x} \in \mathbb{R}^d} \frac{|f(\mathbf{x})|}{1 + |\mathbf{x}|} < \infty \right\}.$$

2. KEY INGREDIENTS FOR THE PROOFS OF THE MAIN RESULTS

In this section, we provide the main ingredients for the proofs of Theorems 1.3, 1.4 and 1.5, which include the mollification of g , Jacobi flow for mollified processes, the preliminaries for the semigroup theory with respect to Markov processes, and the ergodicities in the total variation type distance.

2.1. Mollification of g and properties of the mollifier g_ε . Because $g(\mathbf{x})$ is not differentiable, we need to consider its mollification in [28] as the following: for any $0 < \varepsilon < 1$,

$$g_\varepsilon(\mathbf{x}) = \ell - [\mathbf{M}\mathbf{x} + (\mathbf{\Gamma} - \mathbf{M})\mathbf{v}\rho_\varepsilon(\mathbf{e}'_d \mathbf{x})], \quad \mathbf{x} \in \mathbb{R}^d,$$

where the function $\rho_\varepsilon(y) : \mathbb{R} \rightarrow \mathbb{R}$ is defined as

$$\rho_\varepsilon(y) = y \mathbf{1}_{\{y > \varepsilon\}} + \left(\frac{3\varepsilon}{16} - \frac{1}{16\varepsilon^3} y^4 + \frac{3}{8\varepsilon} y^2 + \frac{1}{2} y \right) \mathbf{1}_{\{|y| \leq \varepsilon\}}.$$

It is easy to check that $\lim_{\varepsilon \rightarrow 0} \rho_\varepsilon(y) = y^+$, $\forall y \in \mathbb{R}$, $\rho_\varepsilon \in \mathcal{C}^2(\mathbb{R})$, and $\|\rho_\varepsilon\|_\infty \leq 1$ for all $\varepsilon \in (0, 1)$. Consequently, $g_\varepsilon \in \mathcal{C}^2(\mathbb{R}^d)$ and

$$(2.1) \quad \nabla g_\varepsilon(\mathbf{x}) = -\mathbf{M} - \dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x})(\mathbf{\Gamma} - \mathbf{M})\mathbf{v}'_d.$$

We shall consider the following auxiliary SDE:

$$(2.2) \quad d\mathbf{X}_t^\varepsilon = g_\varepsilon(\mathbf{X}_t^\varepsilon) dt + \boldsymbol{\sigma} d\mathbf{Z}_t, \quad \mathbf{X}_0^\varepsilon = \mathbf{X}_0 = \mathbf{x} \in \mathbb{R}^d.$$

The EM scheme for the SDE (2.2) with decreasing step-size $\Lambda = (\eta_n)_{n \in \mathbb{N}}$ is as the following:

$$(2.3) \quad \mathbf{Y}_{t_{n+1}}^\varepsilon = \mathbf{Y}_{t_n}^\varepsilon + \eta_{n+1} g_\varepsilon(\mathbf{Y}_{t_n}^\varepsilon) + \boldsymbol{\sigma} \Delta \mathbf{Z}_{\eta_{n+1}}, \quad \mathbf{Y}_0^\varepsilon = \mathbf{x},$$

where $t_n = \sum_{i=1}^n \eta_i$ with $t_0 = 0$, and $\Delta \mathbf{Z}_{\eta_{n+1}} = \mathbf{Z}_{t_{n+1}} - \mathbf{Z}_{t_n}$. Then $(\mathbf{Y}_{t_n}^\varepsilon)_{n \in \mathbb{N}_0}$ is a non-homogeneous Markov chain. For any $0 \leq s \leq t$, denote $\mathbf{X}_{s,t}^\mathbf{x}$ as the value of process $(\mathbf{X}_t)_{t \geq 0}$ at the time t with $\mathbf{X}_s = \mathbf{x}$. Similarly for $\mathbf{X}_{s,t}^{\varepsilon,\mathbf{x}}$, $\mathbf{Y}_{s,t}^\mathbf{x}$, and $\mathbf{Y}_{s,t}^{\varepsilon,\mathbf{x}}$. When $s = 0$, we will omit the subscript s and write $\mathbf{X}_t^{\varepsilon,\mathbf{x}} = \mathbf{X}_{0,t}^{\varepsilon,\mathbf{x}}$, $\mathbf{Y}_t^\mathbf{x} = \mathbf{Y}_{0,t}^\mathbf{x}$, and $\mathbf{Y}_t^{\varepsilon,\mathbf{x}} = \mathbf{Y}_{0,t}^{\varepsilon,\mathbf{x}}$ if there is no ambiguity.

It is easy to show that the coefficient $g_\varepsilon(\mathbf{x})$ is a global Lipschitz function, and its Lipschitz constant is independent of ε , that is, there exists some constant $C > 0$ independent of ε , such that

$$|g_\varepsilon(\mathbf{x}) - g_\varepsilon(\mathbf{y})| \leq C|\mathbf{x} - \mathbf{y}|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

Additionally, there exists some positive constant C independent of ε such that

$$(2.4) \quad |g_\varepsilon(\mathbf{x}) - g(\mathbf{x})| \leq C\varepsilon \mathbf{1}_{\{\|\mathbf{e}'_d \mathbf{x}\| \leq \varepsilon\}}, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

Moreover, $\lim_{\varepsilon \rightarrow 0} g_\varepsilon(\mathbf{x}) = g(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$. The coefficients $g_\varepsilon(\mathbf{x})$ and $g(\mathbf{x})$ do not satisfy a dissipation condition, that is, there exist some $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that for some constant $c > 0$,

$$(2.5) \quad \sup_{|\mathbf{x}-\mathbf{y}| \rightarrow \infty} \frac{\langle \mathbf{x} - \mathbf{y}, g_\varepsilon(\mathbf{x}) - g_\varepsilon(\mathbf{y}) \rangle}{|\mathbf{x} - \mathbf{y}|^2} \geq c, \quad \sup_{|\mathbf{x}-\mathbf{y}| \rightarrow \infty} \frac{\langle \mathbf{x} - \mathbf{y}, g(\mathbf{x}) - g(\mathbf{y}) \rangle}{|\mathbf{x} - \mathbf{y}|^2} \geq c.$$

The following lemma is from [5].

Lemma 2.1. ([5, Theorem 3.5]) *Under Assumption 1.1, there exists a positive definite matrix \mathbf{Q} such that matrices $\mathbf{M}'\mathbf{Q} + \mathbf{Q}\mathbf{M}$ and $(\mathbf{M}' - \mathbf{e}_d \mathbf{v}'(\mathbf{M}' - \boldsymbol{\Gamma}))\mathbf{Q} + \mathbf{Q}(\mathbf{M} - (\mathbf{M} - \boldsymbol{\Gamma})\mathbf{v}\mathbf{e}'_d)$ are both positive definite.*

For the positive definite matrix \mathbf{Q} in Lemma 2.1, denote

$$\begin{aligned} \lambda_1 &= \lambda_{\min}(\mathbf{M}'\mathbf{Q} + \mathbf{Q}\mathbf{M}), \\ \lambda_2 &= \lambda_{\min}([\mathbf{M}' - \mathbf{e}_d \mathbf{v}'(\mathbf{M}' - \boldsymbol{\Gamma})]\mathbf{Q} + \mathbf{Q}[\mathbf{M} - (\mathbf{M} - \boldsymbol{\Gamma})\mathbf{v}\mathbf{e}'_d]), \end{aligned}$$

and

$$(2.6) \quad \lambda = \lambda_1 \wedge \lambda_2, \quad \theta = \frac{1}{2} \lambda \lambda_{\max}^{-2}(\mathbf{Q}).$$

Under Assumption 1.1, it follows from Lemma 2.1 that we can get the following kind of dissipation condition and its proof is in Appendix A.1.

Lemma 2.2. *Under Assumption 1.1 and for the positive definite matrix \mathbf{Q} in Lemma 2.1, for any $\mathbf{x} \in \mathbb{R}^d$ and $\varepsilon \in (0, 1)$, the matrices*

$$[-\nabla g_\varepsilon(\mathbf{x})]'\mathbf{Q} + \mathbf{Q}[-\nabla g_\varepsilon(\mathbf{x})]$$

are positive definite. We also have

$$-\lambda = \sup_{\mathbf{x} \in \mathbb{R}^d} \left\{ \lambda_{\max}([\nabla g_\varepsilon(\mathbf{x})]'\mathbf{Q} + \mathbf{Q}[\nabla g_\varepsilon(\mathbf{x})]) \right\} < 0,$$

where λ is in (2.6). In addition, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$(2.7) \quad \langle \mathbf{Q}(\mathbf{x} - \mathbf{y}), g_\varepsilon(\mathbf{x}) - g_\varepsilon(\mathbf{y}) \rangle \leq -\frac{\lambda}{2} |\mathbf{x} - \mathbf{y}|^2.$$

Furthermore, for all $\mathbf{x} \in \mathbb{R}^d$,

$$\langle g(\mathbf{x}), \mathbf{Q}\mathbf{x} \rangle \leq -\frac{\lambda}{2}|\mathbf{x}|^2 + |\ell|\lambda_{\max}(\mathbf{Q})|\mathbf{x}|.$$

2.2. Jacobi flow for the mollified process $(\mathbf{X}_t^{\varepsilon, \mathbf{x}})_{t \geq 0}$. Recall that for $\alpha \in (1, 2]$,

$$d\mathbf{X}_t^\varepsilon = g_\varepsilon(\mathbf{X}_t^\varepsilon) dt + \boldsymbol{\sigma} d\mathbf{Z}_t, \quad \mathbf{X}_0^\varepsilon = \mathbf{X}_0 = \mathbf{x} \in \mathbb{R}^d.$$

It can be solved as

$$\mathbf{X}_t^{\varepsilon, \mathbf{x}} = \mathbf{x} + \int_0^t g_\varepsilon(\mathbf{X}_s^{\varepsilon, \mathbf{x}}) ds + \boldsymbol{\sigma} \mathbf{Z}_t.$$

For any $\mathbf{u} \in \mathbb{R}^d$, the corresponding Jacobi flow is defined as

$$\nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon, \mathbf{x}} = \lim_{\delta \rightarrow 0} \frac{\mathbf{X}_t^{\varepsilon, \mathbf{x} + \delta \mathbf{u}} - \mathbf{X}_t^{\varepsilon, \mathbf{x}}}{\delta},$$

which satisfies $\nabla_{\mathbf{u}} \mathbf{X}_0^{\varepsilon, \mathbf{x}} = \mathbf{u}$ and

$$d\nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon, \mathbf{x}} = \nabla g_\varepsilon(\mathbf{X}_t^{\varepsilon, \mathbf{x}}) \nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon, \mathbf{x}} dt.$$

Then, we have the following estimate which is proven in Appendix A.2.

Lemma 2.3. *For any $\mathbf{u} \in \mathbb{R}^d$, the following inequality holds*

$$|\nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon, \mathbf{x}}|^2 \leq \frac{\lambda_{\max}(\mathbf{Q})}{\lambda_{\min}(\mathbf{Q})} \exp\left\{-\frac{\lambda}{\lambda_{\max}(\mathbf{Q})}t\right\} |\mathbf{u}|^2,$$

where the matrix \mathbf{Q} is in Lemma 2.1 and λ is in (2.6).

Define

$$\mathbf{J}_{s,t}^{\varepsilon, \mathbf{x}} := \exp\left(\int_s^t \nabla g_\varepsilon(\mathbf{X}_r^{\varepsilon, \mathbf{x}}) dr\right), \quad 0 \leq s \leq t < \infty.$$

It is called the Jacobian between s and t . For notational simplicity, denote $\mathbf{J}_t^{\varepsilon, \mathbf{x}} = \mathbf{J}_{0,t}^{\varepsilon, \mathbf{x}}$. Then we have

$$\nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon, \mathbf{x}} = \mathbf{J}_t^{\varepsilon, \mathbf{x}} \mathbf{u}.$$

Define

$$\nabla \tilde{g}(\mathbf{x}) = -\mathbf{M} + \mathbf{1}_{\{\mathbf{e}'_d \mathbf{x} > 0\}} (\mathbf{M} - \boldsymbol{\Gamma}) \mathbf{v} \mathbf{e}'_d.$$

It is easy to see that $\lim_{\varepsilon \rightarrow 0} \nabla g_\varepsilon(\mathbf{x}) = \nabla \tilde{g}(\mathbf{x})$ for all $\mathbf{e}'_d \mathbf{x} \neq 0$. Because $g(\mathbf{x})$ is not differentiable for $\mathbf{e}'_d \mathbf{x} = 0$, it is necessary for us to define the above $\tilde{g}(\mathbf{x})$ which takes the same value as $\nabla g(\mathbf{x})$ for $\mathbf{e}'_d \mathbf{x} \neq 0$ and has a definition on $\mathbf{e}'_d \mathbf{x} = 0$. Define

$$\mathbf{J}_{s,t}^{\mathbf{x}} = \exp\left(\int_s^t \nabla \tilde{g}(\mathbf{X}_r^{\mathbf{x}}) dr\right), \quad \mathbf{x} \in \mathbb{R}^d, \quad 0 \leq s \leq t < \infty.$$

We also denote $\mathbf{J}_t^{\mathbf{x}} = \mathbf{J}_{0,t}^{\mathbf{x}}$. For any $\mathbf{u} \in \mathbb{R}^d$, define $\nabla_{\mathbf{u}} \mathbf{X}_t^{\mathbf{x}} = \mathbf{J}_t^{\mathbf{x}} \mathbf{u}$.

Then we have the following lemma which is proven in Appendix A.3.

Lemma 2.4. *For any $\mathbf{x} \in \mathbb{R}^d$, as $\varepsilon \rightarrow 0$, the following holds:*

$$\|\mathbf{J}_{s,t}^{\varepsilon, \mathbf{x}} - \mathbf{J}_{s,t}^{\mathbf{x}}\|_{\text{op}} \rightarrow 0, \quad 0 \leq s \leq t < \infty, \quad \text{a.s.}$$

Now we give estimates for $\|\mathbf{J}_{s,t}^{\varepsilon, \mathbf{x}}\|_{\text{op}}$ and $\|\mathbf{J}_{s,t}^{\mathbf{x}}\|_{\text{op}}$. By (2.1), we can easily see that

$$(2.8) \quad \|\nabla g_\varepsilon(\mathbf{x})\|_{\text{op}} \leq \|\mathbf{M}\|_{\text{op}} + \|(\mathbf{M} - \boldsymbol{\Gamma}) \mathbf{v} \mathbf{e}'_d\|_{\text{op}} = C_{\text{op}},$$

from which we obtain

$$\|\mathbf{J}_{s,t}^{\varepsilon, \mathbf{x}}\|_{\text{op}} \leq \exp\left(\int_s^t \|\nabla g_\varepsilon(\mathbf{X}_r^{\varepsilon, \mathbf{x}})\|_{\text{op}} dr\right) \leq e^{C_{\text{op}}(t-s)}.$$

So for all $0 \leq s \leq t < \infty$, we have

$$\|\mathbf{J}_{s,t}^{\mathbf{x}}\|_{\text{op}}, \|\mathbf{J}_{s,t}^{\varepsilon,\mathbf{x}}\|_{\text{op}} \leq e^{C_{\text{op}}(t-s)},$$

where the bound of $\|\mathbf{J}_{s,t}^{\mathbf{x}}\|_{\text{op}}$ comes from the same argument since the bound in (2.8) also holds for $\nabla \tilde{g}(\mathbf{x})$. Observe that the above estimates immediately imply that for any $\mathbf{u} \in \mathbb{R}^d$,

$$|\nabla_{\mathbf{u}} \mathbf{X}_t^{\mathbf{x}}|, |\nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon,\mathbf{x}}| \leq e^{C_{\text{op}}t} |\mathbf{u}|.$$

2.3. Semigroups, generators and their properties. In order to study the approximation in \mathcal{W}_1 distance, we consider the semigroups associated with the SDEs (1.1) and (2.2) in $\mathcal{C}_{lin}(\mathbb{R}^d)$ defined as: for a $f \in \mathcal{C}_{lin}(\mathbb{R}^d)$,

$$P_{s,t} f(\mathbf{x}) = \mathbb{E}[f(\mathbf{X}_t) | \mathbf{X}_s = \mathbf{x}], \quad 0 \leq s \leq t,$$

$$P_{s,t}^{\varepsilon} f(\mathbf{x}) = \mathbb{E}[f(\mathbf{X}_t^{\varepsilon}) | \mathbf{X}_s^{\varepsilon} = \mathbf{x}], \quad 0 \leq s \leq t.$$

By the same argument as in [22], we can show that $\{P_{s,t}\}_{0 \leq s \leq t < \infty}$ is a well defined semigroup in $\mathcal{C}_{lin}(\mathbb{R}^d)$. It is easy to see that both $P_{s,t}$ and $P_{s,t}^{\varepsilon}$ are time homogeneous semigroup, i.e., $P_{s,t} = P_{0,t-s}$ and $P_{s,t}^{\varepsilon} = P_{0,t-s}^{\varepsilon}$ for any $0 \leq s \leq t$. For simplicity, as $s = 0$, we drop the script 's' and write $P_t = P_{0,t}$ and $P_t^{\varepsilon} = P_{0,t}^{\varepsilon}$. The infinitesimal generators of P_t and P_t^{ε} are defined as

$$\mathcal{A}^{\alpha} f(\mathbf{x}) = \lim_{t \rightarrow 0} \frac{P_t f(\mathbf{x}) - f(\mathbf{x})}{t}, \quad f \in \mathcal{D}(\mathcal{A}^{\alpha}),$$

$$\mathcal{A}_{\varepsilon}^{\alpha} f(\mathbf{x}) = \lim_{t \rightarrow 0} \frac{P_t^{\varepsilon} f(\mathbf{x}) - f(\mathbf{x})}{t}, \quad f \in \mathcal{D}(\mathcal{A}_{\varepsilon}^{\alpha}),$$

where $\mathcal{D}(\mathcal{A}^{\alpha})$ and $\mathcal{D}(\mathcal{A}_{\varepsilon}^{\alpha})$ are the domains of \mathcal{A}^{α} and $\mathcal{A}_{\varepsilon}^{\alpha}$ respectively, which are both subsets of $\mathcal{C}_{lin}(\mathbb{R}^d)$. The forms of $\mathcal{D}(\mathcal{A}^{\alpha})$ and $\mathcal{D}(\mathcal{A}_{\varepsilon}^{\alpha})$ are not necessary to be exactly specified in this paper. Let us now give the explicit forms of the infinitesimal generators \mathcal{A}^{α} and $\mathcal{A}_{\varepsilon}^{\alpha}$:

(1) When $\alpha = 2$, i.e., the driven noise is Brownian motion, we have

$$\mathcal{A}_{\varepsilon}^2 f(\mathbf{x}) = \langle g_{\varepsilon}(\mathbf{x}), \nabla f(\mathbf{x}) \rangle + \frac{1}{2} \langle \nabla^2 f(\mathbf{x}), \boldsymbol{\sigma} \boldsymbol{\sigma}' \rangle_{\text{HS}},$$

$$\mathcal{A}^2 f(\mathbf{x}) = \langle g(\mathbf{x}), \nabla f(\mathbf{x}) \rangle + \frac{1}{2} \langle \nabla^2 f(\mathbf{x}), \boldsymbol{\sigma} \boldsymbol{\sigma}' \rangle_{\text{HS}}.$$

(2) When $\alpha \in (1, 2)$, i.e., the driven noise is α -stable noise, the generators \mathcal{A}^{α} can be written as the following according to the type of stable noise:

(2.i) When $\mathbf{Z}_t = (Z_t^1, \dots, Z_t^d)'$ is the cylindrical stable noise,

$$(2.9) \quad \mathcal{A}_{\varepsilon}^{\alpha} f(\mathbf{x}) = \langle g_{\varepsilon}(\mathbf{x}), \nabla f(\mathbf{x}) \rangle + \mathcal{G}_{\text{iid}}^{\alpha} f(\mathbf{x}),$$

$$(2.10) \quad \mathcal{A}^{\alpha} f(\mathbf{x}) = \langle g(\mathbf{x}), \nabla f(\mathbf{x}) \rangle + \mathcal{G}_{\text{iid}}^{\alpha} f(\mathbf{x}),$$

where

$$\mathcal{G}_{\text{iid}}^{\alpha} f(\mathbf{x}) = \sum_{i=1}^d \int_{\mathbb{R}_0} [f(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - f(\mathbf{x}) - z_i \langle \boldsymbol{\sigma}_i, \nabla f(\mathbf{x}) \rangle \mathbf{1}_{\{|z_i| \leq 1\}}] \frac{c_{\alpha}}{|z_i|^{1+\alpha}} dz_i,$$

where $\boldsymbol{\sigma}_i$ is the i -th column of matrix $\boldsymbol{\sigma}$.

(2.ii) When \mathbf{Z}_t is the d -dimensional rotationally symmetric α stable process, we have

$$(2.11) \quad \mathcal{A}_{\varepsilon}^{\alpha} f(\mathbf{x}) = \langle g_{\varepsilon}(\mathbf{x}), \nabla f(\mathbf{x}) \rangle + \mathcal{G}_{\text{sym}}^{\alpha} f(\mathbf{x}),$$

$$(2.12) \quad \mathcal{A}^{\alpha} f(\mathbf{x}) = \langle g(\mathbf{x}), \nabla f(\mathbf{x}) \rangle + \mathcal{G}_{\text{sym}}^{\alpha} f(\mathbf{x}),$$

where

$$\mathcal{G}_{\text{sym}}^{\alpha} f(\mathbf{x}) = \int_{\mathbb{R}_0^d} [f(\mathbf{x} + \boldsymbol{\sigma} \mathbf{z}) - f(\mathbf{x}) - \langle \boldsymbol{\sigma} \mathbf{z}, \nabla f(\mathbf{x}) \rangle \mathbf{1}_{\{|\mathbf{z}| \leq 1\}}] \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z}.$$

2.4. Ergodicity in a weighted total variation distance and moment estimates. Note that [5, Theorem 3.5] has shown the exponential ergodicity for the processes $(\mathbf{X}_t)_{t \geq 0}$ under the weighted total variation distance, but the weight function therein is so large that we cannot directly use this ergodicity result to prove the regularities for our Stein's equation. To solve this problem, we establish an exponential ergodicity for these two processes in another weighted total variation distance, whose weight function has the form as (2.13) below. This is proven in Appendix A.5.

For the small enough constant $\kappa > 0$ satisfying $16\kappa < 1$, and the positive definite matrix \mathbf{Q} in Lemma 2.1, we define a function $\tilde{V}(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$ as

$$(2.13) \quad \tilde{V}(\mathbf{x}) = (1 + \langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle)^\kappa.$$

Lemma 2.5. *Under Assumption 1.1, and for $\alpha \in (1, 2]$, the processes $(\mathbf{X}_t)_{t \geq 0}$ and $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ in (1.1) and (2.2) are both exponential ergodicity under the weighted total variation distance $\|\cdot\|_{\text{TV}, \tilde{V}}$ with ergodic measures μ and μ_ε respectively. More precisely, there exist some positive constants C and c , both independent of ε , such that*

$$\|\mathcal{L}(X_t^x) - \mu\|_{\text{TV}, \tilde{V}} \leq C\tilde{V}(\mathbf{x})e^{-ct} \quad \text{and} \quad \|\mathcal{L}(X_t^{\varepsilon, x}) - \mu_\varepsilon\|_{\text{TV}, \tilde{V}} \leq C\tilde{V}(\mathbf{x})e^{-ct}.$$

In addition, it holds that for all $\mathbf{x} \in \mathbb{R}^d$,

$$(2.14) \quad \mathbb{E}\tilde{V}(\mathbf{X}_t^{\mathbf{x}}) + \mathbb{E}\tilde{V}(\mathbf{X}_t^{\varepsilon, \mathbf{x}}) \leq C(1 + |\mathbf{x}|^{2\kappa}).$$

The moment estimates are given below and the corresponding proofs are in Appendix A.4.

Lemma 2.6. *For $\alpha \in (1, 2]$ and $\varepsilon \in (0, 1)$, under Assumptions 1.1 and 1.2, there exists some positive constant C , independent of ε , such that for all $\mathbf{x} \in \mathbb{R}^d$,*

$$(2.15) \quad \mathbb{E}|\mathbf{X}_t^{\varepsilon, \mathbf{x}}| + \mathbb{E}|\mathbf{X}_t^{\mathbf{x}}| \leq C(1 + |\mathbf{x}|), \quad \forall t \geq 0,$$

$$(2.16) \quad \mathbb{E}[|\mathbf{X}_{t_{n-1}, t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{n-1}, t_n}^{\varepsilon, \mathbf{x}}|] \leq C(1 + |\mathbf{x}|)\eta_n^{1+1/\alpha}, \quad \forall n \in \mathbb{N},$$

$$(2.17) \quad \mathbb{E}|\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}| \leq C(1 + |\mathbf{x}|), \quad \forall n \in \mathbb{N}_0,$$

$$(2.18) \quad \lim_{\varepsilon \rightarrow 0} \mathbb{E}|\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\mathbf{x}}| = 0, \quad \forall t \geq 0.$$

In particular, we have $\mu_\varepsilon(|\mathbf{x}|) \leq C$, where μ_ε is the ergodic measure of the process $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$.

3. PROOF OF THEOREM 1.3

3.1. The strategy of the proof. Let us give a brief strategy for the proof of the Theorem 1.3. Under Assumption 1.1, the process $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ in (2.2) has exponential ergodicity (see, Lemma 2.5 above) with ergodic measure μ_ε . By the triangle inequality,

$$(3.1) \quad \mathcal{W}_1(\mu, \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}})) \leq \mathcal{W}_1(\mu, \mu_\varepsilon) + \mathcal{W}_1(\mu_\varepsilon, \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})) + \mathcal{W}_1(\mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}})).$$

To estimate $\mathcal{W}_1(\mu, \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}}))$, it is sufficient to give estimates for these three Wasserstein-1 distances on the right hand of (3.1). $\mathcal{W}_1(\mu, \mu_\varepsilon)$ is bounded in Lemma 3.2 by a generator comparison argument and solving a Stein's equation, while $\mathcal{W}_1(\mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}}))$ is the distance between the two Markov chains and is bounded in Lemma 3.3. Finally, $\mathcal{W}_1(\mu_\varepsilon, \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}))$ is bounded by a generalized Lindeberg principle developed in [37].

The following contraction property of $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ will be proved in Appendix B.1.

Lemma 3.1. *For $\alpha \in (1, 2]$, under Assumption 1.1, for any $\varepsilon \in (0, 1)$, the processes $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ in (2.2) satisfy that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and $t \geq 0$,*

$$\mathcal{W}_1(\mathcal{L}(\mathbf{X}_t^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{X}_t^{\varepsilon, \mathbf{y}})) \leq \frac{\lambda_{\max}(\mathbf{Q})}{\lambda_{\min}(\mathbf{Q})} |\mathbf{x} - \mathbf{y}| e^{-\theta t},$$

where the constant θ is given in (2.6) and the positive definite matrix \mathbf{Q} is given in Lemma 2.1.

Next, we give bounds for $\mathcal{W}_1(\mu, \mu_\varepsilon)$ and $\mathcal{W}_1(\mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}}))$ in the following lemmas, and their proofs are postponed in Appendixes B.2 and B.3.

Lemma 3.2. *For $\alpha \in (1, 2]$ and $\varepsilon \in (0, 1)$, under Assumption 1.1, there exists some positive constant C independent of ε such that*

$$\mathcal{W}_1(\mu, \mu_\varepsilon) \leq C\varepsilon,$$

where μ and μ_ε are the ergodic measures of the processes $(\mathbf{X}_t)_{t \geq 0}$ and $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ respectively.

Lemma 3.3. *For $\alpha \in (1, 2]$ and $\varepsilon \in (0, 1)$, under Assumptions 1.1 and 1.2, for the two EM schemes $(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})_{n \in \mathbb{N}_0}$ and $(\mathbf{Y}_{t_n}^{\mathbf{x}})_{n \in \mathbb{N}_0}$ defined in (1.3) and (2.3) respectively, there exists some positive constant C independent of ε such that for all $\mathbf{x} \in \mathbb{R}^d$ and $n \in \mathbb{N}_0$,*

$$\mathcal{W}_1(\mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}})) \leq C\varepsilon^{1/2}.$$

The following Lemma is similar to that of [37, Lemma B.1] which is crucial to obtain the convergence rate, and we shall give the proof in Appendix B.4.

Lemma 3.4. *For $\alpha \in (1, 2]$, given that the step size $(\eta_n)_{n \in \mathbb{N}}$ satisfy Assumption 1.2, there exists some positive constant C such that*

$$\sum_{i=1}^n e^{-\theta(t_n - t_i)} \eta_i^{1+1/\alpha} \leq C\eta_n^{1/\alpha},$$

where the constant θ is in (2.6).

3.2. Proof of Theorem 1.3. The semigroups on $\mathcal{C}_{lin}(\mathbb{R}^d)$ for processes $(\mathbf{Y}_{t_n}^{\mathbf{x}})_{n \in \mathbb{N}_0}$, and $(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})_{n \in \mathbb{N}_0}$ can be defined as below: for any $f \in \mathcal{C}_{lin}(\mathbb{R}^d)$,

$$(3.2) \quad \begin{aligned} \mathbb{Q}_{t_i, t_j} f(\mathbf{x}) &= \mathbb{E}[f(\mathbf{Y}_{t_j}) | \mathbf{Y}_{t_i} = \mathbf{x}], \quad 0 \leq t_i \leq t_j, \\ \mathbb{Q}_{t_i, t_j}^\varepsilon f(\mathbf{x}) &= \mathbb{E}[f(\mathbf{Y}_{t_j}^\varepsilon) | \mathbf{Y}_{t_i}^\varepsilon = \mathbf{x}], \quad 0 \leq t_i \leq t_j. \end{aligned}$$

Proof of Theorem 1.3. It follows from (3.1) that

$$\mathcal{W}_1(\mu, \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}})) \leq \mathcal{W}_1(\mu, \mu_\varepsilon) + \mathcal{W}_1(\mu_\varepsilon, \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})) + \mathcal{W}_1(\mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}})).$$

The bounds for $\mathcal{W}_1(\mu, \mu_\varepsilon)$ and $\mathcal{W}_1(\mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}}))$ are given in Lemmas 3.2 and 3.3 respectively. Thus, it suffices to give the bound for $\mathcal{W}_1(\mu_\varepsilon, \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}))$.

By the generalized Lindeberg principle in [37], it follows from (1.9) that

$$\begin{aligned} \mathcal{W}_1(\mathcal{L}(\mathbf{X}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})) &= \sup_{h \in \text{Lip}(1)} \{ \mathbb{P}_{t_n}^\varepsilon h(\mathbf{x}) - \mathbb{Q}_{t_n}^\varepsilon h(\mathbf{x}) \} \\ &= \sup_{h \in \text{Lip}(1)} \left\{ \sum_{i=1}^n \mathbb{Q}_{0, t_{i-1}}^\varepsilon \circ (\mathbb{P}_{t_{i-1}, t_i}^\varepsilon - \mathbb{Q}_{t_{i-1}, t_i}^\varepsilon) \circ \mathbb{P}_{t_i, t_n}^\varepsilon h(\mathbf{x}) \right\}. \end{aligned}$$

We firstly give the estimate for $(\mathbb{P}_{t_{i-1}, t_i}^\varepsilon - \mathbb{Q}_{t_{i-1}, t_i}^\varepsilon) \circ \mathbb{P}_{t_i, t_n}^\varepsilon h(\mathbf{x})$ for $i = 1, 2, \dots, n$. Actually, for $\alpha \in (1, 2]$, we have that

$$\begin{aligned} |(\mathbb{P}_{t_{i-1}, t_i}^\varepsilon - \mathbb{Q}_{t_{i-1}, t_i}^\varepsilon) \circ \mathbb{P}_{t_i, t_n}^\varepsilon h(\mathbf{x})| &= |\mathbb{P}_{t_i, t_n}^\varepsilon h(\mathbf{X}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}}) - \mathbb{P}_{t_i, t_n}^\varepsilon h(\mathbf{Y}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}})| \\ &\leq C e^{-\theta(t_n - t_i)} \mathbb{E} |\mathbf{X}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}}| \\ &\leq C e^{-\theta(t_n - t_i)} \eta_i^{1+1/\alpha} (1 + |\mathbf{x}|), \end{aligned}$$

where the first inequality is due to Lemma 3.1 and the last is due to Lemma 2.6. Combining this with Lemmas 2.6 and 3.4, we can obtain that for $\alpha \in (1, 2]$,

$$|P_{t_n}^\varepsilon h(\mathbf{x}) - Q_{t_n}^\varepsilon h(\mathbf{x})| \leq C \sum_{i=1}^n e^{-\theta(t_n - t_i)} \eta_i^{1+1/\alpha} \mathbb{E} [1 + |\mathbf{Y}_{t_{i-1}}^{\varepsilon, \mathbf{x}}|] \leq C(1 + |\mathbf{x}|) \eta_n^{1/\alpha},$$

which immediately implies that

$$\mathcal{W}_1(\mathcal{L}(\mathbf{X}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})) \leq C(1 + |\mathbf{x}|) \eta_n^{1/\alpha}.$$

Moreover, combining this with Lemmas 2.6 and 3.1, we know

$$\begin{aligned} \mathcal{W}_1(\mu_\varepsilon, \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})) &\leq \mathcal{W}_1(\mu_\varepsilon, \mathcal{L}(\mathbf{X}_{t_n}^{\varepsilon, \mathbf{x}})) + \mathcal{W}_1(\mathcal{L}(\mathbf{X}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})) \\ &\leq C(1 + |\mathbf{x}|) e^{-\theta t_n} + C(1 + |\mathbf{x}|) \eta_n^{1/\alpha} \leq C(1 + |\mathbf{x}|) \eta_n^{1/\alpha}. \end{aligned}$$

This, together with Lemmas 3.2 and 3.3, implies that for $\alpha \in (1, 2]$,

$$\mathcal{W}_1(\mu, \mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}})) \leq C\varepsilon + C(1 + |\mathbf{x}|) \eta_n^{1/\alpha} + C\varepsilon^{1/2} \leq C(1 + |\mathbf{x}|) \eta_n^{1/\alpha},$$

where the last inequality holds by choosing $\varepsilon \leq \eta_n^{2/\alpha}$. The proof is complete. \square

4. PROOFS OF THEOREMS 1.4 AND 1.5

We split this section into two subsections, one for the proof of Theorem 1.4 and the other for that of Theorem 1.5.

To investigate the CLT and MDP for the process $(\mathbf{X}_t)_{t \geq 0}$, we consider the following first kind of Stein's equation: for any $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$,

$$(4.1) \quad \mathcal{A}^\alpha f_h(\mathbf{x}) = h(\mathbf{x}) - \mu(h),$$

where \mathcal{A}^α and μ are the infinitesimal generator and the ergodic measure for the process $(\mathbf{X}_t)_{t \geq 0}$. Then, we can get the following regularities for the solution f_h of Eq. (4.1).

Lemma 4.1. *Let $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$ and f_h be the solution of Eq. (4.1). Then it holds*

$$|f_h(\mathbf{x})| \leq C(1 + |\mathbf{x}|^{2\kappa}) \quad \text{and} \quad |\nabla f_h(\mathbf{x})| \leq C(1 + |\mathbf{x}|^{2\kappa}),$$

where the constant κ satisfies $16\kappa < 1$.

We will prove this lemma in Appendix C. With the help of this lemma, we can show the CLT and MDP for the process $(\mathbf{X}_t)_{t \geq 0}$ with α stable noise $(\mathbf{Z}_t)_{t \geq 0}$. Recall that the CLT and MDP for the process $(\mathbf{X}_t)_{t \geq 0}$ with $\alpha = 2$ have been shown in [28, Theorems 2 and 3].

4.1. Proof of Theorem 1.4. We prove Theorem 1.4 by Itô's formula and martingale CLT. The verification of the conditions in martingale CLT is highly non-trivial.

Proof of Theorem 1.4. We shall split the proof according to the different α stable noises: one is the d -dimensional rotationally symmetric α stable process, and the other is the cylindrical α -stable noise.

Case 1. For the rotationally symmetric stable noise. For $(\mathbf{X}_t)_{t \geq 0}$ in SDE (1.1) with $\mathbf{X}_0 = \mathbf{x}$, by using Itô's formula to f_h which is the solution of Stein's equation (4.1), we have

$$\begin{aligned} f_h(\mathbf{X}_t^{\mathbf{x}}) - f_h(\mathbf{x}) &= \int_0^t \mathcal{A}^\alpha f_h(\mathbf{X}_s^{\mathbf{x}}) ds + \int_0^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^{\mathbf{x}} + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{X}_{s-}^{\mathbf{x}})] \tilde{N}(ds, d\mathbf{z}) \\ &= \int_0^t [h(\mathbf{X}_s^{\mathbf{x}}) - \mu(h)] ds + \int_0^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^{\mathbf{x}} + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{X}_{s-}^{\mathbf{x}})] \tilde{N}(ds, d\mathbf{z}), \end{aligned}$$

where the infinitesimal operator \mathcal{A}^α is defined in (2.12). It implies that

$$\begin{aligned} & \sqrt{t} \left[\frac{1}{t} \int_0^t \delta_{\mathbf{X}_s^\mathbf{x}}(h) \, ds - \mu(h) \right] = \frac{1}{\sqrt{t}} \int_0^t [h(\mathbf{X}_s^\mathbf{x}) - \mu(h)] \, ds \\ & = \frac{1}{\sqrt{t}} [f_h(\mathbf{X}_t^\mathbf{x}) - f_h(\mathbf{x})] - \frac{1}{\sqrt{t}} \int_0^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - f_h(\mathbf{X}_{s-}^\mathbf{x})] \tilde{N}(ds, d\mathbf{z}). \end{aligned}$$

It follows from $\mathbb{E}f_h(\mathbf{X}_t^\mathbf{x}) \leq C(1 + |\mathbf{x}|^{2\kappa})$ in (2.14) and Lemma 4.1 that

$$\mathbb{E} \left| \frac{1}{\sqrt{t}} [f_h(\mathbf{X}_t^\mathbf{x}) - f_h(\mathbf{x})] \right| \rightarrow 0, \quad \text{as } t \rightarrow \infty.$$

It follows from Lemma 4.1 that for some $C > 0$

$$\begin{aligned} & \int_{\mathbb{R}_0^d} [f_h(\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - f_h(\mathbf{x})]^2 \nu(d\mathbf{z}) \\ & = \int_{0 < |\mathbf{z}| \leq 1} \left| \int_0^1 \langle \nabla f_h(\mathbf{x} + r\boldsymbol{\sigma}\mathbf{z}), \boldsymbol{\sigma}\mathbf{z} \rangle dr \right|^2 \nu(d\mathbf{z}) + \int_{|\mathbf{z}| > 1} [f_h(\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - f_h(\mathbf{x})]^2 \nu(d\mathbf{z}) \\ & \leq C \int_{0 < |\mathbf{z}| \leq 1} (1 + |\mathbf{x}|^{2\kappa} + |\mathbf{z}|^{2\kappa})^2 |\mathbf{z}|^2 \nu(d\mathbf{z}) + \int_{|\mathbf{z}| > 1} (1 + |\mathbf{x}|^{2\kappa} + |\mathbf{z}|^{2\kappa})^2 \nu(d\mathbf{z}), \end{aligned}$$

which implies that

$$\begin{aligned} (4.2) \quad & \mathcal{V}(f_h) := \int_{\mathbb{R}^d} \int_{\mathbb{R}_0^d} [f_h(\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - f_h(\mathbf{x})]^2 \nu(d\mathbf{z}) \mu(d\mathbf{x}) \\ & \leq C \int_{\mathbb{R}^d} \left[\int_{0 < |\mathbf{z}| \leq 1} (1 + |\mathbf{x}|^{2\kappa} + |\mathbf{z}|^{2\kappa})^2 |\mathbf{z}|^2 \nu(d\mathbf{z}) + \int_{|\mathbf{z}| > 1} (1 + |\mathbf{x}|^{2\kappa} + |\mathbf{z}|^{2\kappa})^2 \nu(d\mathbf{z}) \right] \mu(d\mathbf{x}) \\ & \leq C\mu(|\mathbf{x}|^{4\kappa}) < \infty, \end{aligned}$$

where the last two inequalities hold from the facts that $4\kappa < 1$ and $\int_{\mathbb{R}_0^d} (1 \wedge |\mathbf{z}|^2) \nu(d\mathbf{z}) < \infty$.

To make notations simple, we denote

$$\begin{aligned} U_i & = \int_{i-1}^i \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - f_h(\mathbf{X}_{s-}^\mathbf{x})] \tilde{N}(ds, d\mathbf{z}) \quad \text{for } i = 1, 2, \dots, [t], \\ U_{[t]+1} & = \int_{[t]}^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - f_h(\mathbf{X}_{s-}^\mathbf{x})] \tilde{N}(ds, d\mathbf{z}). \end{aligned}$$

Then, we have

$$\frac{1}{\sqrt{t}} \int_0^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - f_h(\mathbf{X}_{s-}^\mathbf{x})] \tilde{N}(ds, d\mathbf{z}) = \sum_{i=1}^{[t]+1} U_i.$$

We know that U_i 's are martingale differences and $\mathbb{E}U_i^2 < \infty$, which can be obtained similarly with inequality (4.2) for all $i = 1, 2, \dots, [t] + 1$. We claim that

$$(4.3) \quad \lim_{t \rightarrow \infty} \mathbb{E} \left(\max_{1 \leq i \leq [t]+1} \frac{1}{\mathcal{V}(f_h)t} |U_i|^2 \right) = 0,$$

$$(4.4) \quad \lim_{t \rightarrow \infty} \mathbb{E} \left| \sum_{i=1}^{[t]+1} \frac{1}{\mathcal{V}(f_h)t} |U_i|^2 - 1 \right|^2 = 0.$$

We know (4.3) and (4.4) imply

$$\mathbb{E} \left(\max_{1 \leq i \leq [t]+1} \frac{1}{\sqrt{\mathcal{V}(f_h)t}} |U_i| \right) \rightarrow 0 \quad \text{and} \quad \sum_{i=1}^{[t]+1} \frac{1}{\mathcal{V}(f_h)t} |U_i|^2 \xrightarrow{p} 1$$

as t tends to infinity. Applying the martingale CLT in [46, Theorem 2], due to [34], one has

$$\frac{1}{\sqrt{\mathcal{V}(f_h)t}} \sum_{i=1}^{[t]+1} U_i \Rightarrow \mathcal{N}(0, 1) \quad \text{as } t \rightarrow \infty.$$

Then, we know

$$\sqrt{t} \left[\frac{1}{t} \int_0^t \delta_{\mathbf{X}_s^x}(h) ds - \mu(h) \right] \Rightarrow \mathcal{N}(0, \mathcal{V}(f_h)) \quad \text{as } t \rightarrow \infty.$$

It remains to show (4.3) and (4.4). For (4.3), one has

$$\begin{aligned} \mathbb{E} \left(\max_{1 \leq i \leq [t]+1} |U_i|^2 \right) &= \mathbb{E} \left(\max_{1 \leq i \leq [t]+1} (|U_i|^2 \mathbf{1}_{\{|U_i|^2 \leq \sqrt{t}\}} + |U_i|^2 \mathbf{1}_{\{|U_i|^2 > \sqrt{t}\}}) \right) \\ &\leq \mathbb{E} \left(\max_{1 \leq i \leq [t]+1} |U_i|^2 \mathbf{1}_{\{|U_i|^2 \leq \sqrt{t}\}} \right) + \mathbb{E} \left(\max_{1 \leq i \leq [t]+1} |U_i|^2 \mathbf{1}_{\{|U_i|^2 > \sqrt{t}\}} \right) \\ &\leq \sqrt{t} + ([t] + 1) \max_{1 \leq i \leq [t]+1} \mathbb{E} \left[|U_i|^2 \mathbf{1}_{\{|U_i|^2 > \sqrt{t}\}} \right]. \end{aligned}$$

Thus, we obtain that as t tends to ∞

$$\mathbb{E} \left(\max_{1 \leq i \leq [t]+1} \frac{1}{\mathcal{V}(f_h)t} |U_i|^2 \right) \leq \frac{1}{\mathcal{V}(f_h)t} \left(\sqrt{t} + ([t] + 1) \max_{1 \leq i \leq [t]+1} \mathbb{E} \left[|U_i|^2 \mathbf{1}_{\{|U_i|^2 > \sqrt{t}\}} \right] \right) \rightarrow 0.$$

It is easy to verify (4.4) by proving the following equation,

$$(4.5) \quad \lim_{t \rightarrow \infty} \mathbb{E} \left| \sum_{i=1}^{[t]+1} \frac{1}{\mathcal{V}(f_h)t} |U_i|^2 - \frac{[t] + 1}{t} \right|^2 = 0.$$

Observe that

$$\begin{aligned} &\mathbb{E} \left| \sum_{i=1}^{[t]+1} \frac{1}{\mathcal{V}(f_h)t} |U_i|^2 - \frac{[t] + 1}{t} \right|^2 = \mathbb{E} \left| \frac{1}{t} \sum_{i=1}^{[t]+1} \left(\frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right) \right|^2 \\ &= \frac{1}{t^2} \sum_{i=1}^{[t]+1} \mathbb{E} \left(\frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right)^2 + \frac{2}{t^2} \sum_{i < j} \mathbb{E} \left[\left(\frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right) \left(\frac{1}{\mathcal{V}(f_h)} |U_j|^2 - 1 \right) \right] \\ &=: \text{I} + \text{II}. \end{aligned}$$

By using Kunita's inequality (see, [2, Theorem 4.4.23]), with similar calculations for (4.2), there exists some positive constant C such that

$$\begin{aligned} \mathbb{E} |U_i|^4 &\leq \mathbb{E} \left| \int_{i-1}^i \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^x + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{X}_{s-}^x)]^2 \nu(d\mathbf{z}) ds \right|^2 \\ &\quad + \mathbb{E} \int_{i-1}^i \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^x + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{X}_{s-}^x)]^4 \nu(d\mathbf{z}) ds \\ &\leq C \mathbb{E}(1 + |\mathbf{X}_{i-1}^x|^{8\kappa}) \leq C(1 + |\mathbf{x}|^{8\kappa}). \end{aligned}$$

Combining this with Lemma 4.1 and the fact $8\kappa < 1$, we know

$$\begin{aligned} \mathbb{E} \left(\frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right)^2 &= \mathbb{E} \left[\frac{1}{\mathcal{V}(f_h)} |U_i|^4 - \frac{2}{\mathcal{V}(f_h)} |U_i|^2 + 1 \right] \leq C(1 + \mathbb{E} |\mathbf{X}_i^{\mathbf{x}}|^{8\kappa}) \\ &\leq C(1 + |\mathbf{x}|^{8\kappa}), \end{aligned}$$

which implies that

$$I = \frac{1}{t^2} \sum_{i=1}^{\lfloor t \rfloor + 1} \mathbb{E} \left(\frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right)^2 \rightarrow 0, \quad \text{as } t \rightarrow \infty.$$

As for II, for each i , it holds that for some constant $C > 0$

$$\begin{aligned} &\sum_{j=i+1}^{\lfloor t \rfloor + 1} \mathbb{E} \left[\left(\frac{1}{\mathcal{V}(f_h)} |U_j|^2 - 1 \right) \middle| \mathcal{F}_i \right] = \frac{1}{\mathcal{V}(f_h)} \sum_{j=i+1}^{\lfloor t \rfloor + 1} \mathbb{E} [(|U_j|^2 - \mathcal{V}(f_h)) | \mathcal{F}_i] \\ &= \frac{1}{\mathcal{V}(f_h)} \int_i^{\lfloor t \rfloor + 1} \left(\mathbb{E} \int_{\mathbb{R}_0^d} |f_h(\mathbf{X}_s^{\mathbf{x}} + \sigma \mathbf{z}) - f_h(\mathbf{X}_s^{\mathbf{x}})|^2 \nu(d\mathbf{z}) - \mathcal{V}(f_h) \right) ds \\ &\leq \int_0^{\lfloor t \rfloor + 1} (1 + |\mathbf{X}_i^{\mathbf{x}}|^{4\kappa}) e^{-cs} ds \leq C(1 + |\mathbf{X}_i^{\mathbf{x}}|^{4\kappa}), \end{aligned}$$

where the last second inequality is by Lemma 2.5 and the fact $8\kappa < 1$. Hence, we can obtain

$$\begin{aligned} II &= \frac{2}{t^2} \sum_{i=1}^{\lfloor t \rfloor} \sum_{j=i+1}^{\lfloor t \rfloor + 1} \mathbb{E} \left\{ \left(\frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right) \mathbb{E} \left[\left(\frac{1}{\mathcal{V}(f_h)} |U_j|^2 - 1 \right) \middle| \mathcal{F}_i \right] \right\} \\ &\leq \frac{C}{t^2} \sum_{i=1}^{\lfloor t \rfloor} \mathbb{E} \left\{ \left| \frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right| (1 + |\mathbf{X}_i^{\mathbf{x}}|^{4\kappa}) \right\} \\ &\leq \frac{C}{t^2} \sum_{i=1}^{\lfloor t \rfloor} \left[\mathbb{E} \left| \frac{1}{\mathcal{V}(f_h)} |U_i|^2 - 1 \right|^2 \right]^{1/2} [1 + \mathbb{E} |\mathbf{X}_i^{\mathbf{x}}|^{8\kappa}]^{1/2} \rightarrow 0, \quad \text{as } t \rightarrow \infty. \end{aligned}$$

Combining the estimates for I and II, (4.5) holds.

Case 2. For the cylindrical stable noise. For $(\mathbf{X}_t)_{t \geq 0}$ in SDE (1.1) with $\mathbf{X}_0 = \mathbf{x}$, by using Itô's formula for f_h which is the solution of Stein's equation (4.1), we have

$$\begin{aligned} f_h(\mathbf{X}_t^{\mathbf{x}}) - f_h(\mathbf{x}) &= \int_0^t \mathcal{A}^\alpha f_h(\mathbf{X}_s^{\mathbf{x}}) ds + \sum_{i=1}^d \int_0^t \int_{\mathbb{R}_0} [f_h(\mathbf{X}_{s-}^{\mathbf{x}} + \sigma_i z_i) - f_h(\mathbf{X}_{s-}^{\mathbf{x}})] \tilde{N}(ds, dz_i) \\ &= \int_0^t [h(\mathbf{X}_s^{\mathbf{x}}) - \mu(h)] ds + \sum_{i=1}^d \int_0^t \int_{\mathbb{R}_0} [f_h(\mathbf{X}_{s-}^{\mathbf{x}} + \sigma_i z_i) - f_h(\mathbf{X}_{s-}^{\mathbf{x}})] \tilde{N}(ds, dz_i), \end{aligned}$$

where the infinitesimal operator \mathcal{A}^α here is defined in (2.10). It follows from the Stein's equation (4.1) that

$$\begin{aligned} &\sqrt{t} \left[\frac{1}{t} \int_0^t \delta_{\mathbf{X}_s^{\mathbf{x}}}(h) ds - \mu(h) \right] \\ &= \frac{1}{\sqrt{t}} [f_h(\mathbf{X}_t^{\mathbf{x}}) - f_h(\mathbf{x})] + \sum_{i=1}^d \frac{1}{\sqrt{t}} \int_0^t \int_{\mathbb{R}_0} [f_h(\mathbf{X}_{s-}^{\mathbf{x}} + \sigma_i z_i) - f_h(\mathbf{X}_{s-}^{\mathbf{x}})] \tilde{N}(ds, dz_i). \end{aligned}$$

With similar calculations for the **Case 1**, we can get the following expression for $\mathcal{V}(f_h)$:

$$\mathcal{V}(f_h) = \sum_{i=1}^d \int_{\mathbb{R}^d} \int_{\mathbb{R}_0} [f_h(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - f_h(\mathbf{x})]^2 \nu_1(dz_i) \mu(d\mathbf{x}).$$

The proof is complete. \square

4.2. Proof of Theorem 1.5. We shall prove the MDP by the criterion in [51, Theorem 2.6].

Proof of Theorem 1.5. According to the criterion for MDP in [51, Theorem 2.6], we need to show that the semigroup P_t associated to the process $(\mathbf{X}_t)_{t \geq 0}$ is strong Feller, topological irreducible and satisfies a Lyapunov condition which is immediately implied by the second relation in (A.20). The strong Feller property and topological irreducibility hold from the Harnack inequality and log-Harnack inequality from [49, Theorem 2.1] for the d dimensional rotationally symmetric α stable noise and [50, Theorems 1.1 and 1.2] for the cylindrical stable noise respectively.

For $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$, it follows from Wu [52, Theorem 2.1] that

$$\mathbb{P} \left(\frac{1}{a_t \sqrt{t}} \int_0^t [h(\mathbf{X}_s^{\mathbf{x}}) - \mu(h)] ds \in \cdot \right)$$

satisfies the large deviation principle with speed a_t^{-2} and rate function $I_h(z) = \frac{z^2}{2\mathcal{H}(h)}$ with

$$\mathcal{H}(h) = 2 \int_0^\infty \langle P_t h, h - \mu(h) \rangle_\mu dt,$$

that is,

$$\begin{aligned} - \inf_{z \in \bar{A}} I_h(z) &\leq \liminf_{t \rightarrow \infty} \frac{1}{a_t^2} \log \mathbb{P} \left(\frac{\sqrt{t}}{a_t} [\mathcal{E}_t^{\mathbf{x}}(h) - \mu(h)] \in A \right) \\ &\leq \limsup_{t \rightarrow \infty} \frac{1}{a_t^2} \log \mathbb{P} \left(\frac{\sqrt{t}}{a_t} [\mathcal{E}_t^{\mathbf{x}}(h) - \mu(h)] \in A \right) \leq - \inf_{z \in \bar{A}} I_h(z), \end{aligned}$$

where \bar{A} and \mathring{A} are the closure and interior of set A respectively.

We claim that

$$(4.6) \quad \mathcal{H}(h) = \mathcal{V}(f_h),$$

where f_h is the solution to Stein's equation (4.1). Then the desired result holds.

Now, we show that the claim (4.6) holds. With some calculations, one has

$$\begin{aligned} \frac{1}{t} \mathbb{E}^\mu \left(\int_0^t [h(\mathbf{X}_s^{\mathbf{x}}) - \mu(h)] ds \right)^2 &= \frac{1}{t} \mathbb{E}^\mu \left(\int_0^t \int_0^t [h(\mathbf{X}_s^{\mathbf{x}}) - \mu(h)][h(\mathbf{X}_u^{\mathbf{x}}) - \mu(h)] du ds \right) \\ &= \frac{2}{t} \mathbb{E}^\mu \left(\int_0^t \int_0^u [h(\mathbf{X}_s^{\mathbf{x}}) - \mu(h)][h(\mathbf{X}_u^{\mathbf{x}}) - \mu(h)] ds du \right) \\ &= \frac{2}{t} \int_0^t \int_0^u \mathbb{E}^\mu \{ [h(\mathbf{X}_s^{\mathbf{x}}) - \mu(h)] \mathbb{E} \{ [h(\mathbf{X}_u^{\mathbf{x}}) - \mu(h)] | \mathbf{X}_s^{\mathbf{x}} \} \} ds du \\ &= \frac{2}{t} \int_0^t \int_0^u \mathbb{E}^\mu \{ h(\mathbf{X}_s^{\mathbf{x}}) \mathbb{E} \{ [h(\mathbf{X}_u^{\mathbf{x}}) - \mu(h)] | \mathbf{X}_s^{\mathbf{x}} \} \} ds du, \end{aligned}$$

where the third equality holds from conditional probability and the last equality holds from

$$\int_0^t \int_0^u \mathbb{E}^\mu \{ \mu(h) \mathbb{E} \{ [h(\mathbf{X}_u^{\mathbf{x}}) - \mu(h)] | \mathbf{X}_s^{\mathbf{x}} \} \} ds du = \mu(h) \int_0^t \int_0^u \mathbb{E}^\mu \{ h(\mathbf{X}_u^{\mathbf{x}}) - \mu(h) \} ds du = 0.$$

Furthermore, for all $0 \leq s \leq u < \infty$, one has

$$\begin{aligned} \mathbb{E}^\mu \{h(\mathbf{X}_s^x) \mathbb{E}\{[h(\mathbf{X}_u^x) - \mu(h)] | \mathbf{X}_s^x\}\} &= \mathbb{E}^\mu \{h(\mathbf{X}_s^x) [\mathbb{P}_{u-s} h(\mathbf{X}_s^x) - \mu(h)]\} \\ &= \int_{\mathbb{R}^d} h(\mathbf{y}) [\mathbb{P}_{u-s} h(\mathbf{y}) - \mu(h)] \mu(d\mathbf{y}), \end{aligned}$$

which implies that

$$\begin{aligned} &\frac{2}{t} \int_0^t \int_0^u \mathbb{E}^\mu \{h(\mathbf{X}_s^x) \mathbb{E}\{[h(\mathbf{X}_u^x) - \mu(h)] | \mathbf{X}_s^x\}\} ds du \\ &= \frac{2}{t} \int_0^t \int_0^u \int_{\mathbb{R}^d} h(\mathbf{y}) [\mathbb{P}_{u-s} h(\mathbf{y}) - \mu(h)] \mu(d\mathbf{y}) ds du \\ &= \int_{\mathbb{R}^d} h(\mathbf{y}) \frac{2}{t} \int_0^t \int_0^u [\mathbb{P}_{u-s} h(\mathbf{y}) - \mu(h)] ds du \mu(d\mathbf{y}) \\ &= \int_{\mathbb{R}^d} h(\mathbf{y}) \frac{2}{t} \int_0^t \int_0^u [\mathbb{P}_{\tilde{s}} h(\mathbf{y}) - \mu(h)] d\tilde{s} du \mu(d\mathbf{y}). \end{aligned}$$

Using the L'Hospital's rule, one has

$$\lim_{t \rightarrow \infty} \int_{\mathbb{R}^d} h(\mathbf{y}) \frac{2}{t} \int_0^t \int_0^u [\mathbb{P}_{\tilde{s}} h(\mathbf{y}) - \mu(h)] d\tilde{s} du \mu(d\mathbf{y}) = 2 \int_{\mathbb{R}^d} h(\mathbf{y}) \int_0^\infty [\mathbb{P}_{\tilde{s}} h(\mathbf{y}) - \mu(h)] d\tilde{s} \mu(d\mathbf{y}).$$

From the expression of $\mathcal{H}(h)$, we know

$$\mathcal{H}(h) = 2 \int_0^\infty \langle \mathbb{P}_t h, h - \mu(h) \rangle_\mu dt = 2 \int_0^\infty \langle \mathbb{P}_t h - \mu(h), h \rangle_\mu dt,$$

so that,

$$\frac{1}{t} \mathbb{E}^\mu \left(\int_0^t [h(\mathbf{X}_s^x) - \mu(h)] ds \right)^2 \rightarrow \mathcal{V}(f_h), \quad \text{as } t \rightarrow \infty.$$

Case 1. For the rotationally symmetric stable noise. Using Itô's formula and Stein's equation (4.1), we have

$$\begin{aligned} f_h(\mathbf{X}_t^x) - f_h(\mathbf{x}) &= \int_0^t \mathcal{A}^\alpha f_h(\mathbf{X}_s^x) ds + \int_0^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^x + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{X}_{s-}^x)] \tilde{N}(ds, d\mathbf{z}) \\ &= \int_0^t [h(\mathbf{X}_s^x) - \mu(h)] ds + \int_0^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^x + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{X}_{s-}^x)] \tilde{N}(ds, d\mathbf{z}). \end{aligned}$$

Combining this with ergodic theorem and the fact that $\mathbb{E}[f_h(\mathbf{X}_t^x) - f_h(\mathbf{x})]/\sqrt{t} \rightarrow 0$ as $t \rightarrow \infty$, one yields that

$$\begin{aligned} &\frac{1}{t} \mathbb{E}^\mu \left(\int_0^t [h(\mathbf{X}_s^x) - \mu(h)] ds \right)^2 \\ &= \mathbb{E}^\mu \left(\frac{1}{\sqrt{t}} [f_h(\mathbf{X}_t^x) - f_h(\mathbf{x})] - \int_0^t \int_{\mathbb{R}_0^d} [f_h(\mathbf{X}_{s-}^x + \boldsymbol{\sigma} \mathbf{z}) - f_h(\mathbf{X}_{s-}^x)] \tilde{N}(ds, d\mathbf{z}) \right)^2 \\ &\rightarrow \mathcal{V}(f_h), \quad \text{as } t \rightarrow \infty. \end{aligned}$$

Thus, we know $\mathcal{H}(h) = \mathcal{V}(f_h)$.

Case 2. For the cylindrical stable noise. Using Itô's formula and Stein's equation (4.1),

$$\begin{aligned} f_h(\mathbf{X}_t^x) - f_h(\mathbf{x}) &= \int_0^t \mathcal{A}^\alpha f_h(\mathbf{X}_s^x) ds + \sum_{i=1}^d \int_0^t \int_{\mathbb{R}_0} [f_h(\mathbf{X}_{s-}^x + \sigma_i z_i) - f_h(\mathbf{X}_{s-}^x)] \tilde{N}(ds, dz_i) \\ &= \int_0^t [h(\mathbf{X}_s^x) - \mu(h)] ds + \sum_{i=1}^d \int_0^t \int_{\mathbb{R}_0} [f_h(\mathbf{X}_{s-}^x + \sigma_i z_i) - f_h(\mathbf{X}_{s-}^x)] \tilde{N}(ds, dz_i), \end{aligned}$$

with similar discussions for **Case 1**, we also know that $\mathcal{H}(h) = \mathcal{V}(f_h)$. Combining above two cases, the claim (4.6) holds. The proof is complete. \square

5. OPTIMAL ERROR BOUND FOR THE CLASSICAL ORNSTEIN-UHLENBECK PROCESS

In this section, we shall show that the convergence rate obtained by our method is optimal for the classical Ornstein-Uhlenbeck process.

We consider the following Ornstein-Uhlenbeck process defined as

$$(5.1) \quad dX_t = -X_t dt + dZ_t, \quad X_0 = 0,$$

where $(Z_t)_{t \geq 0}$ is a one-dimensional symmetric α -stable process with $\alpha \in (1, 2]$. It is known that the solution to (5.1) is

$$X_t = e^{-t} X_0 + e^{-t} \int_0^t e^s dZ_s = \int_0^t e^{s-t} dZ_s, \quad \forall t \geq 0.$$

It follows from [45, Proposition 3.4.1] that the characteristic function for X_t^0 is

$$\begin{aligned} \mathbb{E} [\exp(iuX_t^0)] &= \mathbb{E} \left[\exp \left(i \int_0^t u e^{s-t} dZ_s \right) \right] = \exp \left(- \int_0^t |u e^{s-t}|^\alpha ds \right) \\ &= \exp \left(- \frac{1}{\alpha} |u|^\alpha [1 - e^{-\alpha t}] \right), \end{aligned}$$

for any $u \in \mathbb{R}$. It immediately leads to that $\mathbb{E}[\exp(iuX_t^0)] \rightarrow \exp(-|u|^\alpha/\alpha)$ as $t \rightarrow \infty$.

The EM scheme of (5.1) is given as follows: $Y_0 = X_0 = 0$ and

$$(5.2) \quad Y_{t_{n+1}} = Y_{t_n} - \eta_{n+1} Y_{t_n} + \Delta Z_{\eta_{n+1}}, \quad n \in \mathbb{N}_0,$$

where $t_n = \sum_{i=1}^n \eta_i$ with $t_0 = 0$ and $\Delta Z_{\eta_{n+1}} = Z_{t_{n+1}} - Z_{t_n}$.

Proposition 5.1. *Denote the invariant measure of Ornstein-Uhlenbeck process $(X_t^0)_{t \geq 0}$ in (5.1) by μ and let $(Y_{t_n}^0)_{n \in \mathbb{N}_0}$ be defined in (5.2). Under Assumption 1.2, if $\beta = \alpha^{-1}$ and $\omega = \alpha$ in (1.4) for any $\alpha \in (1, 2]$, we have*

$$0 < \liminf_{n \rightarrow \infty} \frac{\mathcal{W}_1(\mu, \mathcal{L}(Y_{t_n}^0))}{\eta_n^{1/\alpha}} < \limsup_{n \rightarrow \infty} \frac{\mathcal{W}_1(\mu, \mathcal{L}(Y_{t_n}^0))}{\eta_n^{1/\alpha}} < \infty.$$

Proof of Proposition 5.1. By induction, it follows from (5.2) that

$$\begin{aligned} Y_{t_n} &= (1 - \eta_n) Y_{t_{n-1}} + \Delta Z_{\eta_n} = (1 - \eta_{n-1})(1 - \eta_n) Y_{t_{n-2}} + (1 - \eta_n) \Delta Z_{\eta_{n-1}} + \Delta Z_{\eta_n} \\ &= \sum_{j=1}^n \Delta Z_{\eta_j} \prod_{k=j+1}^n (1 - \eta_k). \end{aligned}$$

Hence, the characteristic function for $Y_{t_n}^0$ is given by

$$(5.3) \quad \mathbb{E} [\exp(iuY_{t_n}^0)] = \exp \left(-|u|^\alpha \sum_{j=1}^n \eta_j \prod_{k=j+1}^n (1 - \eta_k)^\alpha \right)$$

for any $u \in \mathbb{R}$. Besides, the fact $1 - x \leq e^{-x}$ and $t_n = \sum_{k=1}^n \eta_k$ yield that

$$(5.4) \quad \sum_{j=1}^n \eta_j \prod_{k=j+1}^n (1 - \eta_k)^\alpha \leq \sum_{j=1}^n \eta_j e^{-\alpha(t_n - t_j)} \leq \frac{1}{\alpha}(1 - e^{-\alpha t_n}) + C\eta_n$$

holds for some constant $C > 0$ independent of n . The last inequality above is obtained by

$$\begin{aligned} \sum_{j=1}^n \eta_j e^{-\alpha(t_n - t_j)} - \frac{1}{\alpha} + \frac{1}{\alpha} e^{-\alpha t_n} &= e^{-\alpha t_n} \left[\sum_{j=1}^n \eta_j e^{\alpha t_j} - \int_0^{t_n} e^{\alpha s} ds \right] \\ &= e^{-\alpha t_n} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} (e^{\alpha t_j} - e^{\alpha s}) ds \\ &\leq e^{-\alpha t_n} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \alpha e^{\alpha t_j} (t_j - s) ds \\ &= \frac{\alpha}{2} \sum_{j=1}^n e^{-\alpha(t_n - t_j)} \eta_j^2 \leq C\eta_n, \end{aligned}$$

where the first inequality holds from the mean value theorem and the last inequality holds by Lemma 3.4.

Combining (5.3) and (5.4), we have

$$\mathbb{E} [\exp(iuY_{t_n}^0)] \geq \exp\left(-|u|^\alpha \left[\frac{1}{\alpha}(1 - e^{-\alpha t_n}) + C\eta_n\right]\right).$$

Hence, there exists a constant $C > 0$ independent of n such that

$$(5.5) \quad \begin{aligned} &\int_{-1}^1 [\mathbb{E} [\exp(iuY_{t_n}^0)] - \mathbb{E} [\exp(iu\alpha^{-1/\alpha} Z_1)]] du \\ &\geq \int_{-1}^1 \left[\exp\left(-|u|^\alpha \left[\frac{1}{\alpha}(1 - e^{-\alpha t_n}) + C\eta_n\right]\right) - \exp\left(-\frac{1}{\alpha}|u|^\alpha\right) \right] du \\ &= \int_{-1}^1 \exp\left(-\frac{1}{\alpha}|u|^\alpha\right) \left[\exp\left(\frac{1}{\alpha}|u|^\alpha e^{-\alpha t_n} + C|u|^\alpha \eta_n\right) - 1 \right] du \\ &\geq \int_{-1}^1 \exp\left(-\frac{1}{\alpha}|u|^\alpha\right) \left[\frac{1}{\alpha}|u|^\alpha e^{-\alpha t_n} + C|u|^\alpha \eta_n \right] du = C(e^{-\alpha t_n} + \eta_n). \end{aligned}$$

For constants ω and β in (1.4), we claim that

$$(5.6) \quad \frac{e^{-\alpha t_n}}{\eta_n^\beta} = \begin{cases} O(1), & \text{if } \omega = \alpha, \\ o(1), & \text{if } \omega < \alpha, \end{cases}$$

where $f(n) = O(g(n))$ as $n \rightarrow \infty$ means that $\lim_{n \rightarrow \infty} f(n)/g(n) < \infty$, $f(n) = o(g(n))$ as $n \rightarrow \infty$ means that $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$ for two sequences $\{f(n), n \in \mathbb{N}\}$ and $\{g(n), n \in \mathbb{N}\}$.

Furthermore, by the similar argument as in the proof of [13, Proposition B.1] and taking

$$h(y) = \frac{1}{M} \left(\frac{\sin y}{y} \mathbf{1}_{\{y \neq 0\}} + \mathbf{1}_{\{y=0\}} \right), \quad y \in \mathbb{R},$$

with constant $M = \sup_{z \in \mathbb{R}_0} |[z \cos z - \sin z]/z^2| \in (0, \infty)$, we can obtain that

$$\begin{aligned} \mathcal{W}_1(\mu, Y_{t_n}^0) &\geq |\mathbb{E}[h(Y_{t_n}^0)] - \mathbb{E}[h(\alpha^{-1/\alpha} Z_1)]| \\ &= \frac{1}{2M} \left| \int_{-1}^1 \left[\mathbb{E}[\exp(iuY_{t_n}^0)] - \mathbb{E}[\exp(iu\alpha^{-1/\alpha} Z_1)] \right] du \right| \geq C(e^{-\alpha t_n} + \eta_n) \\ &= O(\eta_n^{1/\alpha}), \end{aligned}$$

where the inequality arises from (5.5) and the last equality holds by taking $\beta = 1/\alpha$ and $\omega = \alpha$ in claim (5.6).

The reminder is to verify the claim (5.6). Recall that

$$\omega = \lim_{n \rightarrow \infty} \left\{ \frac{\eta_n^\beta - \eta_{n+1}^\beta}{\eta_{n+1}^{1+\beta}} \right\} < \infty,$$

which leads to $1 + \frac{\eta_{n-1}^\beta - \eta_n^\beta}{\eta_n^{1+\beta}} \eta_n \asymp \exp(\omega \eta_n)$, where $f(n) \asymp g(n)$ as $n \rightarrow \infty$ means that $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$ for two sequences $\{f(n), n \in \mathbb{N}\}$ and $\{g(n), n \in \mathbb{N}\}$. Let $\omega_n = \exp(-\alpha t_n) \eta_n^{-\beta}$, $\forall n \in \mathbb{N}_+$, then one has

$$\omega_n = \omega_{n-1} e^{-\alpha \eta_n} \cdot \frac{\eta_{n-1}^\beta}{\eta_n^\beta} = \omega_{n-1} e^{-\alpha \eta_n} \left(1 + \frac{\eta_{n-1}^\beta - \eta_n^\beta}{\eta_n^{1+\beta}} \eta_n \right) \asymp \omega_{n-1} e^{(-\alpha + \omega) \eta_n}.$$

By induction, one gets that

$$\omega_n \asymp \omega_1 e^{(-\alpha + \omega) \eta_1} e^{(-\alpha + \omega) t_n} = O(e^{(-\alpha + \omega) t_n}).$$

Thus, the claim (5.6) holds. The proof is complete. \square

Remark 5.2. As an example of the decreasing step size $\gamma = (\eta_n)_{n \in \mathbb{N}}$, we set

$$\eta_n = \frac{1}{\alpha^2 n}, \quad \forall n \in \mathbb{N}.$$

Then, taking $\beta = 1/\alpha$ in Assumption 1.2, we can verify that

$$\omega = \lim_{n \rightarrow \infty} \frac{\eta_n^\beta - \eta_{n+1}^\beta}{\eta_{n+1}^{1+\beta}} = \lim_{n \rightarrow \infty} \alpha^2 (n+1) \left[\left(1 + \frac{1}{n} \right)^{1/\alpha} - 1 \right] = \alpha.$$

Assumption 1.2 and $e^{-\alpha t_n} = \eta_n^{1/\alpha}$ hold, as long as θ in (2.6) is larger such that $\theta > \alpha$.

6. SINKHORN-KNOPP ALGORITHM

In this section, we shall use the Sinkhorn-Knopp algorithm to verify our convergence results in \mathcal{W}_1 distance by several concrete examples in queue networks. Let us first recall some preliminary of the Sinkhorn-Knopp algorithm.

6.1. Preliminary of Sinkhorn-Knopp algorithm. Let $\mathcal{P}(\mathbb{R}^d)$ be the set of probability measures on \mathbb{R}^d . Let

$$\nu_1 = \sum_{i=1}^n a_i \delta_{\mathbf{x}_i}, \quad \nu_2 = \sum_{j=1}^m b_j \delta_{\mathbf{y}_j}$$

be two discrete probabilities in $\mathcal{P}(\mathbb{R}^d)$ centered on $\{\mathbf{x}_i : 1 \leq i \leq n\} \subseteq \mathbb{R}^d$ and $\{\mathbf{y}_j : 1 \leq j \leq m\} \subseteq \mathbb{R}^d$ respectively, where $\mathbf{a} = (a_1, \dots, a_n)' \in \Delta_n$ and $\mathbf{b} = (b_1, \dots, b_m)' \in \Delta_m$ with

$$\Delta_n = \{\mathbf{p} \in \mathbb{R}_+^n : \mathbf{p}' \mathbf{e}_n = 1\}, \quad \Delta_m = \{\mathbf{q} \in \mathbb{R}_+^m : \mathbf{q}' \mathbf{e}_m = 1\},$$

where \mathbb{R}_+^n is the set of vectors $\mathbf{p} \in \mathbb{R}^n$ with non-negative entries and $\mathbf{e}_n = (1, \dots, 1)' \in \mathbb{R}^n$ and Δ_m is similarly defined. In this setting, the evaluation of the Wasserstein distance corresponds to solving a network flow problem in terms of the weight vectors \mathbf{a} and \mathbf{b} (see, [33, Eq. 4] or [8])

$$(6.1) \quad \mathcal{W}_1(\nu_1, \nu_2) = \min_{\mathbf{T} \in \mathcal{C}(\mathbf{a}, \mathbf{b})} \langle \mathbf{T}, \mathbf{M} \rangle_{\text{HS}},$$

where $\mathbf{M} = (M_{ij})_{n \times m}$ is the cost matrix with entries $M_{ij} = |\mathbf{x}_i - \mathbf{y}_j|$.

A more practical approximation of (6.1) is Sinkhorn distance $S^\tau(\mathbf{a}, \mathbf{b})$, given by

$$(6.2) \quad S^\tau(\mathbf{a}, \mathbf{b}) = \langle \mathbf{T}^\tau, \mathbf{M} \rangle_{\text{HS}} \quad \text{with} \quad \mathbf{T}^\tau = \arg \min_{\mathbf{T} \in \mathcal{C}(\mathbf{a}, \mathbf{b})} \left\{ \langle \mathbf{T}, \mathbf{M} \rangle_{\text{HS}} - \frac{1}{\tau} h(\mathbf{T}) \right\},$$

where $\tau > 0$ is a regularization parameter, $h(\mathbf{T}) = -\sum_{i,j=1}^{n,m} T_{ij}(\log T_{ij} - 1)$ and matrix $\mathbf{T} = (T_{ij})_{n \times m}$. Then, we have the following approximation between $S^\tau(\mathbf{a}, \mathbf{b})$ and $\mathcal{W}_1(\nu_1, \nu_2)$:

Lemma 6.1. ([33, Proposition 1]) *For any pair of discrete measures $\nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^d)$ with respective weights $\mathbf{a} \in \Delta_n$ and $\mathbf{b} \in \Delta_m$, we have*

$$|S^\tau(\mathbf{a}, \mathbf{b}) - \mathcal{W}_1(\nu_1, \nu_2)| \leq C e^{-\tau},$$

where the constant C is independent of τ but depends on the supports of ν_1 and ν_2 .

It follows from [11, Eq. (3)] that there exist vectors $\mathbf{r} = (r_1, \dots, r_n)'$ and $\mathbf{c} = (c_1, \dots, c_m)'$ such that the matrix $\mathbf{T}^\tau = (T_{ij}^\tau)_{n \times m}$ satisfies

$$T_{ij}^\tau = r_i e^{-\tau M_{ij}} c_j, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m.$$

To make notations simple, we write $A_{ij}^\tau = e^{-\tau M_{ij}}$ for each i, j and denote by \mathbf{A}^τ the corresponding matrix. Then,

$$(6.3) \quad \mathbf{T}^\tau = \text{diag}(\mathbf{r}) \mathbf{A}^\tau \text{diag}(\mathbf{c}).$$

It follows from [30, Section 2] that the vectors \mathbf{c} and \mathbf{r} can be solved by

$$\mathbf{c} = \mathbf{b} ./ ((\mathbf{A}^\tau)' \mathbf{r}) \quad \text{and} \quad \mathbf{r} = \mathbf{a} ./ (\mathbf{A}^\tau \mathbf{c}).$$

The above expressions for vectors \mathbf{r} and \mathbf{c} suggest a fixed point iteration as the following:

$$\mathbf{c}_{k+1} = \mathbf{b} ./ ((\mathbf{A}^\tau)' \mathbf{r}_k), \quad \text{and} \quad \mathbf{r}_{k+1} = \mathbf{a} ./ (\mathbf{A}^\tau \mathbf{c}_{k+1}), \quad k \in \mathbb{N}.$$

Furthermore, it follows from (6.2) and (6.3) that $\mathcal{W}_1(\nu_1, \nu_2)$ can be approximated by $S^\tau(\mathbf{a}, \mathbf{b})$.

6.2. Simulations for two queue networks. We point out that the simulation for one dimensional symmetric α -stable random variables with $\alpha \in (1, 2)$ can be found in [36, Theorem 1.3] or [45, Proposition 1.7.1].

In the following examples, let the step size $\Lambda = (\eta_n)_{n \in \mathbb{N}}$ of the EM scheme be $\eta_n = 1/(10 + n)$, $\forall n \in \mathbb{N}$, and the corresponding time be $t_n = \sum_{k=1}^n \eta_k$, $\forall n \in \mathbb{N}$ with $t_0 = 0$. Besides, let the regularization parameter $\tau = 20$ in Sinkhorn-Knopp algorithm for 1-dimension. In the case of 2-dimension, we let the regularization parameter in the Sinkhorn-Knopp algorithm $\tau = 80$ to obtain a better estimation.

Example 6.2 (Stable piecewise Ornstein-Uhlenbeck process). As shown in [38], assuming that the arrival process is heavy-tailed and under the proper scaling, it converges to a symmetric α -stable motion, the scaling limit for the $G/M/N + M$ queues with such an arrival process is

$$X(t) = X(0) - \beta \rho t - \int_0^t [\rho(X(s) \wedge 0) + \gamma(X(s) \vee 0)] ds + Z_\alpha(t),$$

where $Z_\alpha(t)$ is a symmetric α -stable process with $\alpha \in (1, 2)$. Here we take $\gamma = 2$, $\varrho = 1$, $\beta = 1$, so that $\ell = -1$. The parameters in our SDE are: $\alpha \in (1, 2)$, $d = 1$, $\ell = -1$, $M = 1$, $(M + (\Gamma - M)v) = 2$ and $\sigma = 1$. Hence the coefficient g is

$$g(x) = \begin{cases} -1 - 2x, & \text{if } x > 0, \\ -1 - x, & \text{if } x \leq 0, \end{cases}$$

and the corresponding process $(X_t)_{t \geq 0}$ satisfies

$$dX_t = g(X_t) dt + dZ_t, \quad X_0 = x,$$

where $(Z_t)_{t \geq 0}$ is the symmetric α -stable process with Lévy measure $\nu_1(dz) = c_\alpha |z|^{-1-\alpha} dz$.

The ergodic measure μ in Theorem 1.3 does not have an explicit form, and we shall take the law of $Y_{t_n}^x$ with a sufficiently large n as its estimation, denoted by $\hat{\mu}$. We shall use the following four steps to estimate $\mathcal{W}_1(\mu, \mathcal{L}(Y_{t_n}^x))$. In our simulation, we take $\alpha \in \{1.25, 1.5, 1.75\}$.

Step 1. Empirical distribution of $\mathcal{L}(Y_{t_n}^x)$. Let $N_1 = 1000$ be iteration for the EM scheme and generate $N_2 = 2000$ trajectories for the Markov chain $(Y_{t_n})_{n \in \mathbb{N}_0}$, denoted by $\{Y_{t_n}(\omega_i)\}_{i=1}^{N_2}$, that is, for any $i \in \{1, 2, \dots, N_2\}$ and $n \in \{0, 1, 2, \dots, N_1 - 1\}$, one has

$$Y_{t_{n+1}}(\omega_i) = Y_{t_n}(\omega_i) + \eta_{n+1}g(Y_{t_n}(\omega_i)) + \frac{1}{\tilde{\sigma}}\eta_{n+1}^{1/\alpha}\xi_{n+1}(\omega_i), \quad Y_0(\omega_i) = x = 0,$$

where the constant $\tilde{\sigma} = (\frac{\alpha}{2c_\alpha})^{1/\alpha}$ and $(\xi_n)_{n \in \mathbb{N}}$ are stable random variables. Then, the law $\mathcal{L}(Y_{t_{N_1}}^0)$ can be approximated as $\mathcal{L}(Y_{t_{N_1}}^0) \approx N_2^{-1} \sum_{i=1}^{N_2} \delta_{Y_{t_{N_1}}^0(\omega_i)}$.

Step 2. Estimation of μ . Let $N_4 = 2N_1$ be iteration for the EM scheme and generate $N_2 = 2000$ trajectories for the Markov chain $(\tilde{Y}_{t_n})_{n \in \mathbb{N}_0}$ with different initial values $\mathbf{y} = \text{linspace}(-20, 20, N_2)$ which are all independent of those in Step 1, denote by $\{\tilde{Y}_{t_n}(\omega_j)\}_{j=1}^{N_2}$, that is, for any $j \in \{1, 2, \dots, N_2\}$ and $n \in \{0, 1, 2, \dots, N_4 - 1\}$, one knows

$$\tilde{Y}_{t_{n+1}}(\omega_j) = \tilde{Y}_{t_n}(\omega_j) + \eta_{n+1}g(\tilde{Y}_{t_n}(\omega_j)) + \frac{1}{\tilde{\sigma}}\eta_{n+1}^{1/\alpha}\xi_{n+1}(\omega_j), \quad \tilde{Y}_0(\omega_j) = \mathbf{y}(j),$$

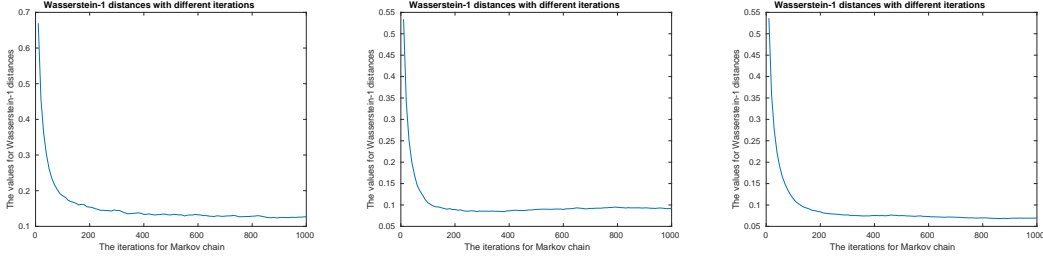
where $\mathbf{y}(j)$ means the j -th element of vector \mathbf{y} , and $(\xi_n)_{n \in \mathbb{N}}$ are α stable random variables. We use the empirical measures $\hat{\mu} = N_2^{-1} \sum_{j=1}^{N_2} \delta_{\tilde{Y}_{t_{N_4}}^{\mathbf{y}(j)}(\omega_j)}$ as an estimation of the ergodic measure μ .

Step 3. Run the Sinkhorn-Knopp algorithm. Let vectors $\mathbf{c}_0 = \mathbf{r}_0 = \mathbf{e}_{N_2}$, $\mathbf{a} = \mathbf{b} = N_2^{-1}\mathbf{e}_{N_2}$, two empirical measures $N_2^{-1} \sum_{i=1}^{N_2} \delta_{Y_{t_{N_1}}^0(\omega_i)}$ and $\hat{\mu}$, one obtains the value of S^τ which is the approximation of $\mathcal{W}_1(\mu, \mathcal{L}(Y_{t_{N_1}}^0))$.

Step 4. Obtain a better approximation. Repeat the above three steps for $N_3 = 20$ times, and we can obtain the values of $\mathcal{W}_1^{(i)}(\mu, \mathcal{L}(Y_{t_{N_1}}^0))$ with $i = 1, \dots, N_3$ in Step 3, and then take the mean of those $N_3 = 20$ distances to approximate the Wasserstein-1 distance.

In addition, we can get the following figures. Figures 1(a), 1(b) and 1(c) are the Wasserstein-1 distances for various iterations with $\alpha \in \{1.25, 1.5, 1.75\}$, that is, $\mathcal{W}_1(\mu, \mathcal{L}(Y_{t_k}^0))$ with $k \in \{10, 20, 30, \dots, 1000\}$.

Example 6.3 (The stable noise with independent components.) We consider a two-class $G/M/N + M$ queues (the ‘ V ’ network, see, [5, pp. 1086-1087]), where the limit will have $\mathbf{M} = \text{diag}(m_1, m_2)$ with $m_1 = 2, m_2 = 1$ and $\mathbf{\Gamma} = \text{diag}(\gamma_1, \gamma_2)$ with $\gamma_1 = 1, \gamma_2 = 2$, we can choose the control $\mathbf{v} = (1/2, 1/2)'$, and $\ell = (-1/2, -1/4)'$, then the coefficient g is



(a) Wasserstein-1 distances with different iterations and $\alpha = 1.25$ (b) Wasserstein-1 distances with different iterations and $\alpha = 1.5$ (c) Wasserstein-1 distances with different iterations and $\alpha = 1.75$

FIGURE 1. Wasserstein-1 distances for Example 6.2 with different iterations and $\alpha = 1.25, 1.5$ and 1.75

given as

$$g(\mathbf{x}) = \begin{cases} \begin{pmatrix} -1/2 \\ -1/4 \end{pmatrix} - \begin{pmatrix} 3/2 & -1/2 \\ 1/2 & 3/2 \end{pmatrix} \mathbf{x}, & \text{if } x_1 + x_2 > 0, \\ \begin{pmatrix} -1/2 \\ -1/4 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x}, & \text{if } x_1 + x_2 \leq 0, \end{cases}$$

where $\mathbf{x} = (x_1, x_2)' \in \mathbb{R}^2$. The corresponding process $(\mathbf{X}_t)_{t \geq 0}$ satisfies that

$$d\mathbf{X}_t = g(\mathbf{X}_t) dt + d\mathbf{Z}_t, \quad \mathbf{X}_0 = \mathbf{x},$$

where $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t})'$ and $Z_{i,t}$, $i = 1, 2$ are independent one-dimensional symmetric α -stable processes with Lévy measure $\nu_1(dz) = c_\alpha |z|^{-1-\alpha} dz$. We shall use the following four steps to estimate $\mathcal{W}_1(\mu, \mathcal{L}(\mathbf{Y}_{t_n}^x))$. In our simulation, we choose $\alpha = 1.8$.

Step 1. Empirical distribution of $\mathcal{L}(\mathbf{Y}_{t_n}^x)$. Let $N_1 = 800$ be iteration for the EM scheme and generate $N_2 = 6000$ trajectories for the Markov chain $(\mathbf{Y}_{t_n})_{n \in \mathbb{N}_0}$ denote by $\{\mathbf{Y}_{t_n}(\omega_i)\}_{i=1}^{N_2}$, that is, for any $i \in \{1, 2, \dots, N_2\}$ and $n \in \{0, 1, 2, \dots, N_1 - 1\}$, one has $\mathbf{Y}_0(\omega_i) = \mathbf{0}$ and

$$\begin{aligned} Y_{1,t_{n+1}}(\omega_i) &= Y_{1,t_n}(\omega_i) + \eta_{n+1}[g(\mathbf{Y}_{t_n}(\omega_i))](1) + \frac{1}{\tilde{\sigma}} \eta_{n+1}^{1/\alpha} \xi_{1,n+1}(\omega_i), \\ Y_{2,t_{n+1}}(\omega_i) &= Y_{2,t_n}(\omega_i) + \eta_{n+1}[g(\mathbf{Y}_{t_n}(\omega_i))](2) + \frac{1}{\tilde{\sigma}} \eta_{n+1}^{1/\alpha} \xi_{2,n+1}(\omega_i), \end{aligned}$$

where $\mathbf{Y}_{t_n} = (Y_{1,t_n}, Y_{2,t_n})'$, and $[a](i)$ means the i -th element of vector \mathbf{a} and the constant $\tilde{\sigma} = (\frac{\alpha}{2c_\alpha})^{1/\alpha}$ and $(\xi_{i,n})_{n \in \mathbb{N}}$, $i = 1, 2$ are stable random variables. Then, the law $\mathcal{L}(\mathbf{Y}_{t_{N_1}}^0)$ can be approximated as $\mathcal{L}(\mathbf{Y}_{t_{N_1}}^0) \approx N_2^{-1} \sum_{i=1}^{N_2} \delta_{\mathbf{Y}_{t_{N_1}}^0(\omega_i)}$.

Step 2. Estimation of μ . Let $N_4 = 10N_1$ be iteration for the EM scheme and generate N_2 trajectories for the Markov chain $(\mathbf{Y}_{t_n})_{n \in \mathbb{N}_0}$ with different initial values which are all independent of those in Step 1, denote by $\{\tilde{\mathbf{Y}}_{t_n}(\omega_j)\}_{j=1}^{N_2}$, that is, for any $j \in \{1, 2, \dots, N_2\}$ and $n \in \{0, 1, 2, \dots, N_4 - 1\}$, one knows $\tilde{\mathbf{Y}}_0(\omega_j) = (\mathbf{y}(j), \mathbf{y}(j))$ and

$$\begin{aligned} \tilde{Y}_{1,t_{n+1}}(\omega_j) &= \tilde{Y}_{1,t_n}(\omega_j) + \eta_{n+1}[g(\tilde{\mathbf{Y}}_{t_n}(\omega_j))](1) + \frac{1}{\tilde{\sigma}} \eta_{n+1}^{1/\alpha} \xi_{1,n+1}(\omega_j), \\ \tilde{Y}_{2,t_{n+1}}(\omega_j) &= \tilde{Y}_{2,t_n}(\omega_j) + \eta_{n+1}[g(\tilde{\mathbf{Y}}_{t_n}(\omega_j))](2) + \frac{1}{\tilde{\sigma}} \eta_{n+1}^{1/\alpha} \xi_{2,n+1}(\omega_j), \end{aligned}$$

where $\mathbf{y}(j)$ means the j -th element of vector $\mathbf{y} = \text{linspace}(-20, 20, N_2)$, and the random variables $(\xi_{i,n})_{n \in \mathbb{N}}$, $i = 1, 2$ have stable distribution. We use the empirical measures $\hat{\mu} = N_2^{-1} \sum_{j=1}^{N_2} \delta_{\tilde{\mathbf{Y}}_{t_{N_4}}(\omega_j)}$ as an estimation of the ergodic measure μ .

Step 3. Run the Sinkhorn-Knopp algorithm. Let vectors $\mathbf{c}_0 = \mathbf{r}_0 = \mathbf{e}_{N_2}$, $\mathbf{a} = \mathbf{b} = N_2^{-1}\mathbf{e}_{N_2}$, two empirical measures $N_2^{-1} \sum_{i=1}^{N_2} \delta_{\mathbf{Y}_{t_{N_1}}^0(\omega_i)}$ and $\hat{\mu}$, one obtains the value of S^τ which is the approximation of $\mathcal{W}_1(\mu, \mathcal{L}(\mathbf{Y}_{t_{N_1}}^0))$.

Step 4. Obtain a better approximation. Repeat the above three steps for $N_3 = 20$ times, and we can obtain the values of $\mathcal{W}_1^{(i)}(\mu, \mathcal{L}(\mathbf{Y}_{t_{N_1}}^0))$ with $i = 1, \dots, N_3$ in Step 3, and then take the mean of those $N_3 = 20$ distances to approximate the Wasserstein-1 distance.

In addition, we can get the following figure for the Wasserstein-1 distances for various iterations with $\alpha = 1.8$. Figure 2 is for the number of sample $N_2 = 6000$ and the iteration of EM scheme $N_1 = 800$.

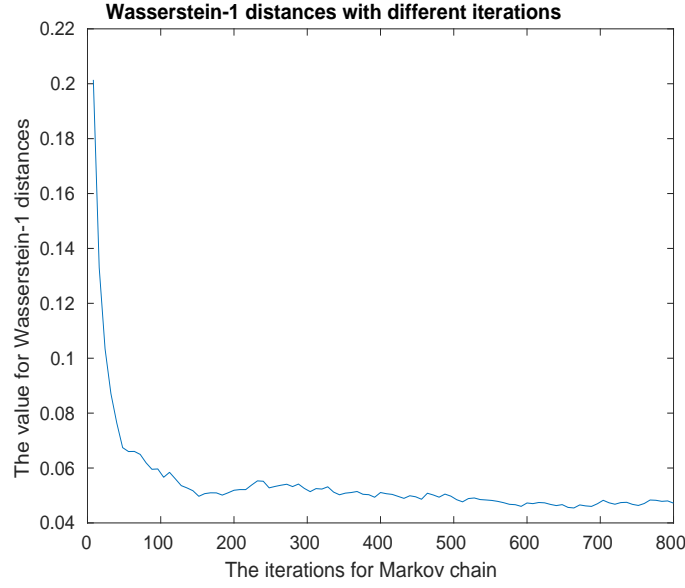


FIGURE 2. Wasserstein-1 distances for Example 6.3 with different iterations and $\alpha = 1.8$, $N_2 = 6000$.

APPENDIX A. PROOFS OF THE LEMMAS IN SECTION 2

A.1. Proof of Lemma 2.2.

Proof of Lemma 2.2. (i) Recall that

$$-\nabla g_\varepsilon(\mathbf{x}) = \mathbf{M} - (\mathbf{M} - \mathbf{\Gamma})\mathbf{v}\mathbf{e}'_d \dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x}).$$

Case 1: If $\mathbf{e}'_d \mathbf{x} < -\varepsilon$, we know $\dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x}) = 0$ and $-\nabla g_\varepsilon(\mathbf{x}) \equiv \mathbf{M}$, it follows from Lemma 2.1 that matrices

$$[-\nabla g_\varepsilon(\mathbf{x})]'\mathbf{Q} + \mathbf{Q}[-\nabla g_\varepsilon(\mathbf{x})] \equiv \mathbf{M}'\mathbf{Q} + \mathbf{Q}\mathbf{M}$$

are positive definite.

Case 2: If $\mathbf{e}'_d \mathbf{x} > \varepsilon$, $\dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x}) = 1$. Then one has

$$-\nabla g_\varepsilon(\mathbf{x}) = \mathbf{M} - (\mathbf{M} - \mathbf{\Gamma})\mathbf{v}\mathbf{e}'_d,$$

and the result can be obtained similarly from Lemma 2.1.

Case 3: If $|\mathbf{e}'_d \mathbf{x}| \leq \varepsilon$, we know $\dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x}) \in [0, 1]$ and

$$\begin{aligned}
& [-\nabla g_\varepsilon(\mathbf{x})]' \mathbf{Q} + \mathbf{Q} [-\nabla g_\varepsilon(\mathbf{x})] \\
&= [\mathbf{M} - (\mathbf{M} - \mathbf{\Gamma}) \mathbf{v} \mathbf{e}'_d \dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x})]' \mathbf{Q} + \mathbf{Q} [\mathbf{M} - (\mathbf{M} - \mathbf{\Gamma}) \mathbf{v} \mathbf{e}'_d \dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x})] \\
\text{(A.1)} \quad &= \dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x}) \{ [\mathbf{M} - (\mathbf{M} - \mathbf{\Gamma}) \mathbf{v} \mathbf{e}'_d]' \mathbf{Q} + \mathbf{Q} [\mathbf{M} - (\mathbf{M} - \mathbf{\Gamma}) \mathbf{v} \mathbf{e}'_d] \} \\
& \quad + (1 - \dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x})) [\mathbf{M}' \mathbf{Q} + \mathbf{Q} \mathbf{M}].
\end{aligned}$$

It follows from Lemma 2.1 that the matrices

$$[\mathbf{M} - (\mathbf{M} - \mathbf{\Gamma}) \mathbf{v} \mathbf{e}'_d]' \mathbf{Q} + \mathbf{Q} [\mathbf{M} - (\mathbf{M} - \mathbf{\Gamma}) \mathbf{v} \mathbf{e}'_d] \quad \text{and} \quad [\mathbf{M}' \mathbf{Q} + \mathbf{Q} \mathbf{M}]$$

are both positive definite. By the fact $\dot{\rho}_\varepsilon(\mathbf{e}'_d \mathbf{x}) \in [0, 1]$, we know that the matrix $[-\nabla g_\varepsilon(\mathbf{x})]' \mathbf{Q} + \mathbf{Q} [-\nabla g_\varepsilon(\mathbf{x})]$ with $|\mathbf{e}'_d \mathbf{x}| \leq \varepsilon$ is also positive definite.

Combining above Cases 1, 2 and 3, one can derive the desired result.

(ii) According to the definitions of λ_1 and λ_2 , we know that

$$\begin{aligned}
-\lambda_1 &= \lambda_{\max}([-\mathbf{M}' \mathbf{Q} + \mathbf{Q} [-\mathbf{M}])), \\
-\lambda_2 &= \lambda_{\max}((-\mathbf{M}' + \mathbf{e}_d \mathbf{v}' (\mathbf{M}' - \mathbf{\Gamma})) \mathbf{Q} + \mathbf{Q} (-\mathbf{M} + (\mathbf{M} - \mathbf{\Gamma}) \mathbf{v} \mathbf{e}'_d)).
\end{aligned}$$

By the fact $\lambda = \lambda_1 \wedge \lambda_2$ and (A.1), one has

$$-\lambda = (-\lambda_1) \vee (-\lambda_2) = \sup_{\mathbf{x} \in \mathbb{R}^d} \{ \lambda_{\max}([\nabla g_\varepsilon(\mathbf{x})]' \mathbf{Q} + \mathbf{Q} [\nabla g_\varepsilon(\mathbf{x})]) \}.$$

(iii) Applying $-\lambda$ above, one can obtain that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\begin{aligned}
\langle \mathbf{Q}(\mathbf{x} - \mathbf{y}), g_\varepsilon(\mathbf{x}) - g_\varepsilon(\mathbf{y}) \rangle &= \langle \mathbf{Q}(\mathbf{x} - \mathbf{y}), \int_0^1 \nabla g_\varepsilon(\mathbf{y} + r(\mathbf{x} - \mathbf{y})) [\mathbf{x} - \mathbf{y}] dr \rangle \\
&= \int_0^1 [\mathbf{x} - \mathbf{y}]' \mathbf{Q} \nabla g_\varepsilon(\mathbf{y} + r(\mathbf{x} - \mathbf{y})) [\mathbf{x} - \mathbf{y}] dr \leq -\frac{\lambda}{2} |\mathbf{x} - \mathbf{y}|^2.
\end{aligned}$$

(iv) Recall that

$$g(\mathbf{x}) = \boldsymbol{\ell} - \mathbf{M}(\mathbf{x} - \langle \mathbf{e}_d, \mathbf{x} \rangle^+ \mathbf{v}) - \langle \mathbf{e}_d, \mathbf{x} \rangle^+ \mathbf{\Gamma} \mathbf{v}.$$

It follows from Lemma 2.1 that for $\mathbf{e}'_d \mathbf{x} > 0$

$$\langle g(\mathbf{x}), \mathbf{Q} \mathbf{x} \rangle = \langle \boldsymbol{\ell} - (\mathbf{M} + (\mathbf{\Gamma} - \mathbf{M}) \mathbf{v} \mathbf{e}'_d) \mathbf{x}, \mathbf{Q} \mathbf{x} \rangle \leq -\frac{\lambda_2}{2} |x|^2 + |\boldsymbol{\ell}| \lambda_{\max, \mathbf{Q}} |x|.$$

In addition, it follows from Lemma 2.1 that for $\mathbf{e}'_d \mathbf{x} \leq 0$

$$\langle g(\mathbf{x}), \mathbf{Q} \mathbf{x} \rangle = \langle \boldsymbol{\ell} - \mathbf{M} \mathbf{x}, \mathbf{Q} \mathbf{x} \rangle \leq -\frac{\lambda_1}{2} |x|^2 + |\boldsymbol{\ell}| \lambda_{\max, \mathbf{Q}} |x|.$$

Combining above two cases and the fact $\lambda = \lambda_1 \wedge \lambda_2$, the desired inequality holds. The proof is complete. \square

A.2. Proof of Lemma 2.3.

Proof of Lemma 2.3. The inequality holds obviously for $\mathbf{u} = \mathbf{0}$ due to $\nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon, \mathbf{x}} = \mathbf{0}, \forall t \geq 0$. Thus, it suffices to show the inequality for $\mathbf{u} \in \mathbb{R}_0^d$. For the matrix \mathbf{Q} in Lemma 2.2, define a function $\widehat{V}(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$ as

$$\widehat{V}(\mathbf{x}) = \frac{1}{2} \langle \mathbf{x}, \mathbf{Q} \mathbf{x} \rangle.$$

Using Itô's formula, it follows from Lemma 2.2 that

$$\begin{aligned} d\widehat{V}(\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}) &= \langle \mathbf{Q}\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}, \nabla g_{\varepsilon}(\mathbf{X}_t^{\varepsilon,\mathbf{x}})\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}} \rangle dt \\ &= [\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}]' \frac{\mathbf{Q}\nabla g_{\varepsilon}(\mathbf{X}_t^{\varepsilon,\mathbf{x}}) + [\nabla g_{\varepsilon}(\mathbf{X}_t^{\varepsilon,\mathbf{x}})]'\mathbf{Q}}{2} \nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}} dt \\ &\leq -\frac{\lambda}{2} \frac{|\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}|^2}{\widehat{V}(\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}})} \widehat{V}(\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}) dt \\ &\leq -\frac{\lambda}{\lambda_{\max}(\mathbf{Q})} \widehat{V}(\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}) dt, \end{aligned}$$

where the first inequality holds from the fact $\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}} \neq \mathbf{0}$, therefore $\widehat{V}(\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}) > 0$ for all $\mathbf{u} \neq \mathbf{0}$. Hence, one derives that

$$\widehat{V}(\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}) \leq \exp\left\{-\frac{\lambda}{\lambda_{\max}(\mathbf{Q})}t\right\} \widehat{V}(\mathbf{u}).$$

The definition of $\widehat{V}(\mathbf{x})$ leads that $\lambda_{\min}(\mathbf{Q})|\mathbf{x}|/2 \leq \widehat{V}(\mathbf{x}) \leq \lambda_{\max}(\mathbf{Q})|\mathbf{x}|/2$. Thus, one can obtain

$$\frac{1}{2}\lambda_{\min}(\mathbf{Q})|\nabla_{\mathbf{u}}\mathbf{X}_t^{\varepsilon,\mathbf{x}}|^2 \leq \frac{1}{2}\lambda_{\max}(\mathbf{Q}) \exp\left\{-\frac{\lambda}{\lambda_{\max}(\mathbf{Q})}t\right\} |u|^2,$$

which implies the desired result immediately. \square

A.3. Proof of Lemma 2.4.

Proof of Lemma 2.4. We claim that

$$(A.2) \quad \mathbb{P}(N) = 0, \quad \text{with } N = \left\{ \int_0^{\infty} \|\nabla g_{\varepsilon}(\mathbf{X}_s^{\varepsilon,\mathbf{x}})\|_{\text{op}} \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_s^{\varepsilon,\mathbf{x}} = 0\}} ds \neq 0 \right\}.$$

Indeed, for any $T > 0$, by $\|\nabla g_{\varepsilon}(\mathbf{x})\|_{\text{op}} \leq C_{\text{op}}$ for all $\mathbf{x} \in \mathbb{R}^d$ and ε , we have

$$(A.3) \quad \begin{aligned} \mathbb{E} \int_0^T \|\nabla g_{\varepsilon}(\mathbf{X}_s^{\varepsilon,\mathbf{x}})\|_{\text{op}} \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_s^{\varepsilon,\mathbf{x}} = 0\}} ds &= \int_0^T \mathbb{E}[\|\nabla g_{\varepsilon}(\mathbf{X}_s^{\varepsilon,\mathbf{x}})\|_{\text{op}} \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_s^{\varepsilon,\mathbf{x}} = 0\}}] ds \\ &\leq \int_0^T C_{\text{op}} \mathbb{E} \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_s^{\varepsilon,\mathbf{x}} = 0\}} ds = 0. \end{aligned}$$

Since (A.3) holds for all $T > 0$, we see that

$$\mathbb{E} \int_0^{\infty} \|\nabla g_{\varepsilon}(\mathbf{X}_s^{\varepsilon,\mathbf{x}})\|_{\text{op}} \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_s^{\varepsilon,\mathbf{x}} = 0\}} ds = 0,$$

hence (A.2) holds.

Recall the definition of $\mathbf{J}_{s,t}^{\varepsilon,\mathbf{x}}$ and define

$$\widehat{\mathbf{J}}_{s,t}^{\varepsilon,\mathbf{x}} := \exp\left(\int_s^t \nabla g_{\varepsilon}(\mathbf{X}_r^{\varepsilon,\mathbf{x}}) \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_r^{\varepsilon,\mathbf{x}} \neq 0\}} dr\right).$$

It is easy to verify that

$$(A.4) \quad \lim_{\varepsilon \rightarrow 0} \widehat{\mathbf{J}}_{s,t}^{\varepsilon,\mathbf{x}} = \mathbf{J}_{s,t}^{\mathbf{x}}, \quad 0 \leq s \leq t < \infty.$$

For any $\omega \notin N$, we know $\int_0^{\infty} \nabla g_{\varepsilon}(\mathbf{X}_s^{\varepsilon,\mathbf{x}}) \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_s^{\varepsilon,\mathbf{x}} = 0\}} ds = 0$ and thus

$$\exp\left(\int_s^t \nabla g_{\varepsilon}(\mathbf{X}_r^{\varepsilon,\mathbf{x}}) \mathbf{1}_{\{\mathbf{e}'_d \mathbf{X}_r^{\varepsilon,\mathbf{x}} = 0\}} dr\right) = \mathbf{I}, \quad 0 \leq s \leq t < \infty.$$

Since \mathbf{I} commutes with any matrix, for all $\omega \notin N$, we get

$$\hat{\mathbf{J}}_{s,t}^{\varepsilon, \mathbf{x}} = \hat{\mathbf{J}}_{s,t}^{\varepsilon, \mathbf{x}} \exp \left(\int_s^t \nabla g_\varepsilon(\mathbf{X}_r^{\varepsilon, \mathbf{x}}) \mathbb{1}_{\{e'_d \mathbf{X}_r^{\varepsilon, \mathbf{x}} = 0\}} dr \right) = \mathbf{J}_{s,t}^{\varepsilon, \mathbf{x}}, \quad 0 \leq s \leq t < \infty.$$

This, combining with (A.4), implies that for all $\omega \notin N$,

$$\lim_{\varepsilon \rightarrow 0} \mathbf{J}_{s,t}^{\varepsilon, \mathbf{x}} = \mathbf{J}_{s,t}^{\mathbf{x}}, \quad 0 \leq s \leq t < \infty.$$

Note that $\mathbf{J}_{s,t}^{\varepsilon, \mathbf{x}}$ and $\mathbf{J}_{s,t}^{\mathbf{x}}$ are matrices, hence the above pointwise convergence implies the convergence in operator. \square

A.4. Proof of Lemma 2.6.

Proof of Lemma 2.6. (i) For the positive definite matrix \mathbf{Q} in Lemma 2.1, define a function $V : \mathbb{R}^d \rightarrow [1, \infty)$ as

$$V(\mathbf{x}) = (1 + \langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle)^{1/2}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Then, we obtain

$$(A.5) \quad \lambda_{\min}^{1/2}(\mathbf{Q})|\mathbf{x}| \vee 1 \leq V(\mathbf{x}) \leq 1 + \lambda_{\max}^{1/2}(\mathbf{Q})|\mathbf{x}|$$

and

$$\nabla V(\mathbf{x}) = \frac{1}{V(\mathbf{x})} \mathbf{Q}\mathbf{x}, \quad \nabla^2 V(\mathbf{x}) = \frac{-1}{V^3(\mathbf{x})} \mathbf{Q}\mathbf{x}\mathbf{x}'\mathbf{Q} + \frac{1}{V(\mathbf{x})} \mathbf{Q}, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

It is easy to check that $\|\nabla V\|_\infty \leq 1$ and $\|\nabla^2 V\|_{\text{HS}, \infty} \leq 1$. We will find a differential inequality of $\mathbb{E}V(\mathbf{X}_t^{\varepsilon, \mathbf{x}})$ firstly, and split the proof into three parts based on the type of the noises $(\mathbf{Z}_t)_{t \geq 0}$.

We shall prove the following Lyapunov's condition in three cases: there exists some positive constant C independent of ε such that

$$(A.6) \quad \mathcal{A}_\varepsilon^\alpha V(\mathbf{x}) \leq -\frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} V(\mathbf{x}) + C.$$

Case 1: $\alpha = 2$. Equation (2.9) yields that

$$(A.7) \quad \mathcal{A}_\varepsilon^\alpha V(\mathbf{x}) = \langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle + \frac{1}{2} \langle \nabla^2 V(\mathbf{x}), \boldsymbol{\sigma}\boldsymbol{\sigma}' \rangle_{\text{HS}}.$$

For the first term, by a straightforward calculation and (2.7), there exists some positive constant C independent of ε such that

$$(A.8) \quad \begin{aligned} \langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle &= \frac{1}{V(\mathbf{x})} \langle \mathbf{Q}\mathbf{x}, g_\varepsilon(\mathbf{x}) - g_\varepsilon(\mathbf{0}) \rangle + \frac{1}{V(\mathbf{x})} \langle \mathbf{Q}\mathbf{x}, g_\varepsilon(\mathbf{0}) \rangle \\ &\leq -\frac{\lambda}{2} V^{-1}(\mathbf{x}) |\mathbf{x}|^2 + \lambda_{\max}(\mathbf{Q}) |g_\varepsilon(\mathbf{0})| \frac{|\mathbf{x}|}{V(\mathbf{x})} \\ &\leq -\frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} V^{-1}(\mathbf{x}) \langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle + \lambda_{\max}(\mathbf{Q}) |g_\varepsilon(\mathbf{0})| \frac{|\mathbf{x}|}{V(\mathbf{x})} \\ &\leq -\frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} V(\mathbf{x}) + C. \end{aligned}$$

As for the second term, it is easy to check that by Cauchy inequality

$$(A.9) \quad \langle \nabla^2 V(\mathbf{x}), \boldsymbol{\sigma}\boldsymbol{\sigma}' \rangle_{\text{HS}} \leq \|\nabla^2 V\|_{\text{HS}, \infty} \cdot \|\boldsymbol{\sigma}\boldsymbol{\sigma}'\|_{\text{HS}} \leq \sqrt{d} \|\boldsymbol{\sigma}\|_{\text{op}} \leq C.$$

Combining (A.7), (A.8) and (A.9), the Lyapunov condition (A.6) immediately follows.

Case 2: $(Z_t)_{t \geq 0}$ being the cylindrical stable noise with $\alpha \in (1, 2)$. Combining with (2.9), one knows

(A.10)

$$\begin{aligned} \mathcal{A}_\varepsilon^\alpha V(\mathbf{x}) &= \langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle + \sum_{i=1}^d \int_{\mathbb{R}_0} [V(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - V(\mathbf{x}) - \langle \boldsymbol{\sigma}_i, \nabla V(\mathbf{x}) \rangle z_i \mathbb{1}_{\{|z_i| \leq 1\}}] \frac{c_\alpha}{|z_i|^{1+\alpha}} dz_i \\ &= \langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle + \sum_{i=1}^d \int_{|z_i| > 1} [V(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - V(\mathbf{x})] \frac{c_\alpha}{|z_i|^{1+\alpha}} dz_i \\ &\quad + \sum_{i=1}^d \int_{0 < |z_i| \leq 1} [V(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - V(\mathbf{x}) - \langle \boldsymbol{\sigma}_i, \nabla V(\mathbf{x}) \rangle z_i] \frac{c_\alpha}{|z_i|^{1+\alpha}} dz_i. \end{aligned}$$

The first term $\langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle$ is bounded by (A.8). For the second term of equality (A.10) and any $i = 1, \dots, d$, by the Taylor's expansion and absolute value inequality, one has

$$\begin{aligned} \left| \int_{|z_i| > 1} [V(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - V(\mathbf{x})] \frac{c_\alpha}{|z_i|^{1+\alpha}} dz_i \right| &= \left| \int_{|z_i| > 1} \int_0^1 \langle \nabla V(\mathbf{x} + r \boldsymbol{\sigma}_i z_i), \boldsymbol{\sigma}_i z_i \rangle dr \frac{c_\alpha}{|z_i|^{1+\alpha}} dz_i \right| \\ &\leq c_\alpha \|\nabla V\|_\infty \|\boldsymbol{\sigma}_i\|_{\max} \int_{|z_i| > 1} \frac{|z_i|}{|z_i|^{1+\alpha}} dz_i \leq C. \end{aligned}$$

Similarly, for the third term of equality (A.10), we know

$$\begin{aligned} &\left| \int_{0 < |z_i| \leq 1} [V(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - V(\mathbf{x}) - \langle \boldsymbol{\sigma}_i, \nabla V(\mathbf{x}) \rangle z_i] \frac{c_\alpha}{|z_i|^{1+\alpha}} dz_i \right| \\ &= \left| c_\alpha \int_{0 < |z_i| \leq 1} \int_0^1 \int_0^r \langle \boldsymbol{\sigma}_i \boldsymbol{\sigma}_i', \nabla^2 V(\mathbf{x} + s \boldsymbol{\sigma}_i z) \rangle_{\text{HS}} \frac{z_i^2}{|z_i|^{1+\alpha}} ds dr dz_i \right| \\ &\leq c_\alpha \|\nabla^2 V\|_{\text{HS}, \infty} \|\boldsymbol{\sigma}_i\|_{\max}^2 \int_{0 < |z_i| \leq 1} |z_i|^{1-\alpha} dz_i \leq C. \end{aligned}$$

Hence, we can obtain that

$$(A.11) \quad \sum_{i=1}^d \left| \int_{\mathbb{R}_0} [V(\mathbf{x} + \boldsymbol{\sigma}_i z_i) - V(\mathbf{x}) - \langle \boldsymbol{\sigma}_i, \nabla V(\mathbf{x}) \rangle z_i \mathbb{1}_{\{|z_i| \leq 1\}}] \frac{c_\alpha}{|z_i|^{1+\alpha}} dz_i \right| \leq Cd.$$

Combining (A.8), (A.10) and (A.11), the Lyapunov condition (A.6) immediately follows.

Case 3: $(Z_t)_{t \geq 0}$ being the rotationally symmetric stable noise with $\alpha \in (1, 2)$. By (2.11), we obtain

(A.12)

$$\begin{aligned} \mathcal{A}_\varepsilon^\alpha V(\mathbf{x}) &= \langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle + \int_{\mathbb{R}_0^d} [V(\mathbf{x} + \boldsymbol{\sigma} \mathbf{z}) - V(\mathbf{x}) - \langle \boldsymbol{\sigma} \mathbf{z}, \nabla V(\mathbf{x}) \rangle \mathbb{1}_{\{|\mathbf{z}| \leq 1\}}] \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z} \\ &= \langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle + \int_{|\mathbf{z}| > 1} [V(\mathbf{x} + \boldsymbol{\sigma} \mathbf{z}) - V(\mathbf{x})] \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z} \\ &\quad + \int_{0 < |\mathbf{z}| \leq 1} [V(\mathbf{x} + \boldsymbol{\sigma} \mathbf{z}) - V(\mathbf{x}) - \langle \boldsymbol{\sigma} \mathbf{z}, \nabla V(\mathbf{x}) \rangle] \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z}. \end{aligned}$$

The first term $\langle g_\varepsilon(\mathbf{x}), \nabla V(\mathbf{x}) \rangle$ is bounded by (A.8). For the second term, by the Taylor's expansion, we have

$$(A.13) \quad \left| \int_{|\mathbf{z}|>1} [V(\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - V(\mathbf{x})] \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z} \right| \leq \int_{|\mathbf{z}|>1} \int_0^1 |\langle \nabla V(\mathbf{x} + r\boldsymbol{\sigma}\mathbf{z}), \boldsymbol{\sigma}\mathbf{z} \rangle| dr \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z} \\ \leq \|\nabla V\|_\infty \|\boldsymbol{\sigma}\|_{\text{op}} \int_{|\mathbf{z}|>1} \frac{C_{d,\alpha} |\mathbf{z}|}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z} \leq C.$$

Similarly, for the third term, we obtain

$$(A.14) \quad \left| \int_{0<|\mathbf{z}|\leq 1} [V(\mathbf{x} + \boldsymbol{\sigma}\mathbf{z}) - V(\mathbf{x}) - \langle \boldsymbol{\sigma}\mathbf{z}, \nabla V(\mathbf{x}) \rangle] \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} d\mathbf{z} \right| \\ \leq \int_{0<|\mathbf{z}|\leq 1} \int_0^1 \int_0^r |\langle \boldsymbol{\sigma}\mathbf{z}, \nabla^2 V(\mathbf{x} + s\boldsymbol{\sigma}\mathbf{z}) \boldsymbol{\sigma}\mathbf{z} \rangle| \frac{C_{d,\alpha}}{|\mathbf{z}|^{d+\alpha}} ds dr d\mathbf{z} \\ \leq \|\nabla^2 V\|_{\text{HS},\infty} \|\boldsymbol{\sigma}\|_{\text{op}}^2 \int_{0<|\mathbf{z}|\leq 1} \int_0^1 \int_0^r \frac{C_{d,\alpha} |\mathbf{z}|^2}{|\mathbf{z}|^{d+\alpha}} ds dr d\mathbf{z} \leq C.$$

Combining (A.8), (A.12), (A.13) and (A.14), the Lyapunov condition (A.6) immediately follows.

Using the Itô's formula, we have

$$\mathbb{E}V(\mathbf{X}_t^{\varepsilon,\mathbf{x}}) = V(\mathbf{x}) + \mathbb{E} \int_0^t \mathcal{A}_\varepsilon^\alpha V(\mathbf{X}_s^{\varepsilon,\mathbf{x}}) ds \leq V(\mathbf{x}) + \int_0^t \left[-\frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} \mathbb{E}V(\mathbf{X}_s^{\varepsilon,\mathbf{x}}) + C \right] ds.$$

It follows from [25, pp. 32-33] that

$$(A.15) \quad \mathbb{E}V(\mathbf{X}_t^{\varepsilon,\mathbf{x}}) \leq e^{-\frac{\lambda t}{2\lambda_{\max}(\mathbf{Q})}} V(\mathbf{x}) + \frac{C\lambda_{\max}(\mathbf{Q})}{\lambda} \left(1 - e^{-\frac{\lambda t}{2\lambda_{\max}(\mathbf{Q})}} \right).$$

Combining this with the fact in (A.5), we get the estimate (2.15) for $\mathbb{E}|\mathbf{X}_t^{\varepsilon,\mathbf{x}}|$. With similar discussions, we also get the moment estimate for $\mathbb{E}|\mathbf{X}_t^{\mathbf{x}}|$ in (2.15).

Let \mathbf{X}_0^ε follow the ergodic measure μ_ε , then \mathbf{X}_t^ε follows measure μ_ε for all $t \geq 0$ and the inequality (A.15) implies that for all $t > 0$,

$$\mu_\varepsilon(V) \leq e^{-\frac{\lambda t}{2\lambda_{\max}(\mathbf{Q})}} \mu_\varepsilon(V) + \frac{C\lambda_{\max}(\mathbf{Q})}{\lambda} \left(1 - e^{-\frac{\lambda t}{2\lambda_{\max}(\mathbf{Q})}} \right).$$

Thus, we obtain $\mu_\varepsilon(V) \leq C$, and so does $\mu_\varepsilon(|\mathbf{x}|)$, while the constant C is independent of ε .

(ii) For any $s \in (t_{i-1}, t_i]$ and $\alpha \in (1, 2]$, one has

$$\mathbf{X}_{t_{i-1},s}^{\varepsilon,\mathbf{x}} = \mathbf{x} + \int_{t_{i-1}}^s g_\varepsilon(\mathbf{X}_{t_{i-1},u}^{\varepsilon,\mathbf{x}}) du + \boldsymbol{\sigma}\mathbf{Z}_{s-t_{i-1}}.$$

It follows from the estimate for $\mathbb{E}|\mathbf{X}_t^{\varepsilon,\mathbf{x}}|$ above and the linear growth for function g_ε , we know for any $s \in (t_{i-1}, t_i]$,

$$(A.16) \quad \mathbb{E}|\mathbf{X}_{t_{i-1},s}^{\varepsilon,\mathbf{x}} - \mathbf{x}| \leq C \int_{t_{i-1}}^s \mathbb{E}|\mathbf{X}_{t_{i-1},u}^{\varepsilon,\mathbf{x}}| du + \|\boldsymbol{\sigma}\|_{\text{op}} (s - t_{i-1})^{1/\alpha} \mathbb{E}|\mathbf{Z}_1| \\ \leq C(1 + |\mathbf{x}|) [(s - t_{i-1})^{1/\alpha} \vee (s - t_{i-1})] \leq C(1 + |\mathbf{x}|) \eta_i^{1/\alpha},$$

where the last inequality holds since $s - t_{i-1} \leq \eta_i < 1$.

On the other hand, we have

$$\mathbf{Y}_{t_{i-1},t_i}^{\varepsilon,\mathbf{x}} = \mathbf{x} + \int_{t_{i-1}}^{t_i} g_\varepsilon(\mathbf{x}) ds + \boldsymbol{\sigma}\mathbf{Z}_{t_i-t_{i-1}},$$

which leads to

$$\mathbf{X}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}} = \int_{t_{i-1}}^{t_i} [g_\varepsilon(\mathbf{X}_{t_{i-1}, s}^{\varepsilon, \mathbf{x}}) - g_\varepsilon(\mathbf{x})] ds.$$

Combining (A.16) and the global Lipschitz property for function g_ε , we know

$$(A.17) \quad \mathbb{E}|\mathbf{X}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{i-1}, t_i}^{\varepsilon, \mathbf{x}}| \leq C \int_{t_{i-1}}^{t_i} \mathbb{E}|\mathbf{X}_{t_{i-1}, s}^{\varepsilon, \mathbf{x}} - \mathbf{x}| ds \leq C(1 + |\mathbf{x}|)\eta_i^{1+1/\alpha},$$

which is the estimate (2.16) for $\mathbb{E}|\mathbf{X}_{t_{n-1}, t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{n-1}, t_n}^{\varepsilon, \mathbf{x}}|$.

(iii) Similar to estimating $\mathbb{E}|\mathbf{X}_t^{\varepsilon, \mathbf{x}}|$, we give the estimate for $\mathbb{E}|\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}|$ with $n \in \mathbb{N}_0$. Let us now estimate $\mathbb{E}V(\mathbf{Y}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}})$. We have

$$\begin{aligned} \mathbb{E}V(\mathbf{Y}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}) &= \mathbb{E}V(\mathbf{Y}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}) - \mathbb{E}V(\mathbf{X}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}) + \mathbb{E}V(\mathbf{X}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}) \\ &\leq \|\nabla V\|_\infty \mathbb{E}|\mathbf{X}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}} - \mathbf{Y}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}| + \mathbb{E}V(\mathbf{X}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}). \end{aligned}$$

Using (A.15) and (A.17), there exist some positive constants c and C such that

$$(A.18) \quad \begin{aligned} \mathbb{E}V(\mathbf{Y}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}) &\leq C(1 + |\mathbf{y}|)\eta_{k+1}^{1+1/\alpha} + e^{-c\eta_{k+1}}V(\mathbf{y}) + C(1 - e^{-c\eta_{k+1}}) \\ &\leq (C\eta_{k+1}^{1+1/\alpha} + e^{-c\eta_{k+1}})V(\mathbf{y}) + C\eta_{k+1}. \end{aligned}$$

Due to Assumption 1.2, $(\eta_n)_{n \in \mathbb{N}}$ decreases to 0, and there exists a positive integer k_0 such that for all $k \geq k_0$, one has $C\eta_{k+1}^{1+1/\alpha} + e^{-c\eta_{k+1}} \leq 1 - c\eta_{k+1}/2$. Thus,

$$\mathbb{E}V(\mathbf{Y}_{t_k, t_{k+1}}^{\varepsilon, \mathbf{y}}) \leq \left(1 - \frac{c}{2}\eta_{k+1}\right)V(\mathbf{y}) + C\eta_{k+1},$$

which implies that

$$\mathbb{E}V(\mathbf{Y}_{t_{k+1}}^{\varepsilon, \mathbf{x}}) = \mathbb{E}\{\mathbb{E}[V(\mathbf{Y}_{t_{k+1}}^{\varepsilon, \mathbf{x}})|\mathbf{Y}_{t_k}^{\varepsilon, \mathbf{x}}]\} \leq \left(1 - \frac{c}{2}\eta_{k+1}\right)\mathbb{E}V(\mathbf{Y}_{t_k}^{\varepsilon, \mathbf{x}}) + C\eta_{k+1}.$$

By induction, we obtain for any $k \geq k_0$, there exists some positive constant C independent of k satisfying

$$(A.19) \quad \begin{aligned} \mathbb{E}V(\mathbf{Y}_{t_{k+1}}^{\varepsilon, \mathbf{x}}) &\leq \left(1 - \frac{c}{2}\eta_k\right)\left(1 - \frac{c}{2}\eta_{k+1}\right)\mathbb{E}V(\mathbf{Y}_{t_{k-1}}^{\varepsilon, \mathbf{x}}) + C\eta_k\left(1 - \frac{c}{2}\eta_{k+1}\right) + C\eta_{k+1} \\ &\leq \prod_{i=k_0}^k \left(1 - \frac{c}{2}\eta_i\right)\mathbb{E}V(\mathbf{Y}_{t_{k_0}}^{\varepsilon, \mathbf{x}}) + C \sum_{j=k_0}^k \eta_j \prod_{i=j+1}^k \left(1 - \frac{c}{2}\eta_i\right) \\ &\leq C\mathbb{E}[1 + V(\mathbf{Y}_{t_{k_0}}^{\varepsilon, \mathbf{x}})], \end{aligned}$$

where the last inequality holds from the facts that $\prod_{i=k_0}^k (1 - c\eta_i/2)$ and $\sum_{j=k_0}^k \eta_j \prod_{i=j+1}^k (1 - c\eta_i/2)$ are bounded.

Due to Assumption 1.2 and for any $k \leq k_0$, we know $C\eta_k^{1+1/\alpha} + e^{-c\eta_k} \leq 1 - c\eta_k + C\eta_k^{3/2}$. Combining (A.18) and with a similar argument for (A.19), we obtain

$$\begin{aligned} \mathbb{E}V(\mathbf{Y}_{t_{k_0}}^{\varepsilon, \mathbf{x}}) &\leq \left(1 - c\eta_{k_0} + C\eta_{k_0}^{3/2}\right)\mathbb{E}V(\mathbf{Y}_{t_{k_0-1}}^{\varepsilon, \mathbf{x}}) + C\eta_{k_0} \\ &\leq \left(1 - c\eta_{k_0-1} + C\eta_{k_0-1}^{3/2}\right)\left(1 - c\eta_{k_0} + C\eta_{k_0}^{3/2}\right)\mathbb{E}V(\mathbf{Y}_{t_{k_0-2}}^{\varepsilon, \mathbf{x}}) \\ &\quad + C\left(1 - c\eta_{k_0-1} + C\eta_{k_0-1}^{3/2}\right)\eta_{k_0} + C\eta_{k_0} \\ &\leq C(1 + V(\mathbf{x})). \end{aligned}$$

Combining this with (A.19), for any $n \in \mathbb{N}_0$, there exists some positive constant C independent of n satisfying

$$\mathbb{E}V(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) \leq C(1 + V(\mathbf{x})),$$

which implies the estimate (2.17) for $\mathbb{E}|\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}|$ by combining the relationship between $V(\mathbf{x})$ and $|\mathbf{x}|$.

(iv) Consider $\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\mathbf{x}}$, which satisfies the following equation

$$\begin{aligned} \frac{d}{dt} (\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\mathbf{x}}) &= g_\varepsilon(\mathbf{X}_t^{\varepsilon, \mathbf{x}}) - g(\mathbf{X}_t^{\mathbf{x}}) = g_\varepsilon(\mathbf{X}_t^{\varepsilon, \mathbf{x}}) - g_\varepsilon(\mathbf{X}_t^{\mathbf{x}}) + g_\varepsilon(\mathbf{X}_t^{\mathbf{x}}) - g(\mathbf{X}_t^{\mathbf{x}}) \\ &= \nabla g_\varepsilon(\boldsymbol{\theta}_t) (\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\mathbf{x}}) + g_\varepsilon(\mathbf{X}_t^{\mathbf{x}}) - g(\mathbf{X}_t^{\mathbf{x}}), \end{aligned}$$

where $\boldsymbol{\theta}_t$ is between $\mathbf{X}_t^{\mathbf{x}}$ and $\mathbf{X}_t^{\varepsilon, \mathbf{x}}$. The above equation can be solved by

$$\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\mathbf{x}} = \int_0^t \exp\left(\int_s^t \nabla g_\varepsilon(\boldsymbol{\theta}_r) dr\right) (g_\varepsilon(\mathbf{X}_s^{\mathbf{x}}) - g(\mathbf{X}_s^{\mathbf{x}})) ds.$$

Since $\|\nabla g_\varepsilon(\mathbf{x})\|_{\text{op}} \leq C_{\text{op}}$ for all $\mathbf{x} \in \mathbb{R}^d$, the relation $|g_\varepsilon(\mathbf{x}) - g(\mathbf{x})| \leq C_{\text{op}}\varepsilon$ for all $\mathbf{x} \in \mathbb{R}^d$ immediately gives (2.18). The proof is complete. \square

A.5. Proof of Lemma 2.5.

Proof of Lemma 2.5. This proof is standard, and we just give some sketch of the proof here. With similar calculations for inequality (A.6) in the proof of Lemma 2.6, there exist some positive constants c and C both independent of ε such that

$$(A.20) \quad \mathcal{A}_\varepsilon^\alpha \tilde{V}(x) \leq -c\tilde{V}(x) + C \quad \text{and} \quad \mathcal{A}^\alpha \tilde{V}(x) \leq -c\tilde{V}(x) + C,$$

which implies that the processes $(\mathbf{X}_t)_{t \geq 0}$ and $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ are both exponential ergodicity under $\|\cdot\|_{\text{TV}, \tilde{V}}$. With similar discussions for inequality (A.15), one also gets the moment estimate in inequality (2.14). The proof is complete. \square

APPENDIX B. PROOFS OF THE LEMMAS IN SECTION 3

B.1. Proof of Lemma 3.1.

Proof of Lemma 3.1. (i) For $\alpha \in (1, 2]$ and any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have

$$d(\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}}) = [g_\varepsilon(\mathbf{X}_t^{\varepsilon, \mathbf{x}}) - g_\varepsilon(\mathbf{X}_t^{\varepsilon, \mathbf{y}})] dt, \quad \mathbf{X}_0^{\varepsilon, \mathbf{x}} - \mathbf{X}_0^{\varepsilon, \mathbf{y}} = \mathbf{x} - \mathbf{y}.$$

By the Itô's formula and (2.7), we obtain

$$\begin{aligned} d|\mathbf{Q}(\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}})|^2 &= 2\langle \mathbf{Q}(\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}}), g_\varepsilon(\mathbf{X}_t^{\varepsilon, \mathbf{x}}) - g_\varepsilon(\mathbf{X}_t^{\varepsilon, \mathbf{y}}) \rangle dt \\ &\leq -\lambda|\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}}|^2 dt = -\lambda|\mathbf{Q}^{-1}\mathbf{Q}(\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}})|^2 dt \\ &\leq -\lambda\lambda_{\max}^{-2}(\mathbf{Q})|\mathbf{Q}(\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}})|^2 dt, \end{aligned}$$

which implies that

$$|\mathbf{Q}(\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}})|^2 \leq \exp\{-\lambda\lambda_{\max}^{-2}(\mathbf{Q})t\} |\mathbf{Q}(\mathbf{x} - \mathbf{y})|^2.$$

Due to $\lambda_{\min}(\mathbf{Q})|\mathbf{x} - \mathbf{y}| \leq |\mathbf{Q}(\mathbf{x} - \mathbf{y})| \leq \lambda_{\max}(\mathbf{Q})|\mathbf{x} - \mathbf{y}|$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we get

$$(B.1) \quad |\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}}|^2 \leq \frac{\lambda_{\max}^2(\mathbf{Q})}{\lambda_{\min}^2(\mathbf{Q})} \exp\{-\lambda\lambda_{\max}^{-2}(\mathbf{Q})t\} |\mathbf{x} - \mathbf{y}|^2.$$

From the definition of Wasserstein-1 distance in (1.9) and (B.1), we obtain

$$\begin{aligned} \mathcal{W}_1(\mathcal{L}(\mathbf{X}_t^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{X}_t^{\varepsilon, \mathbf{y}})) &= \sup_{h \in \text{Lip}(1)} \{\mathbb{E}h(\mathbf{X}_t^{\varepsilon, \mathbf{x}}) - \mathbb{E}h(\mathbf{X}_t^{\varepsilon, \mathbf{y}})\} \leq \mathbb{E}|\mathbf{X}_t^{\varepsilon, \mathbf{x}} - \mathbf{X}_t^{\varepsilon, \mathbf{y}}| \\ &\leq \frac{\lambda_{\max}(\mathbf{Q})}{\lambda_{\min}(\mathbf{Q})} \exp\left\{-\frac{1}{2}\lambda\lambda_{\max}^{-2}(\mathbf{Q})t\right\} |\mathbf{x} - \mathbf{y}|. \end{aligned}$$

The proof is complete. \square

B.2. Proof of Lemma 3.2. For $\alpha \in (1, 2]$, $\varepsilon \in (0, 1)$, and any function $h \in \text{Lip}(1)$, we consider the second kind of Stein's equations:

$$(B.2) \quad \mathcal{A}_\varepsilon^\alpha f_{\varepsilon,h}(\mathbf{x}) = h(\mathbf{x}) - \mu_\varepsilon(h),$$

where $\mathcal{A}_\varepsilon^\alpha$ and μ_ε are the infinitesimal generators and ergodic measures for the processes $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ in (2.2) respectively. Then, we can get the following expression and estimates for the solution to the Stein's equations in (B.2).

Lemma B.1. *Let $\alpha \in (1, 2]$ and $\varepsilon \in (0, 1)$. Under Assumptions 1.1 and 1.2, the Stein's equations (B.2) admits a solution $f_{\varepsilon,h}$ with the following representation:*

$$(B.3) \quad f_{\varepsilon,h}(\mathbf{x}) = - \int_0^\infty P_t^\varepsilon [h(\mathbf{x}) - \mu_\varepsilon(h)] dt,$$

where P_t^ε are semigroups defined in (3.2). Furthermore, there exists some positive constant C independent of ε such that

$$|f_{\varepsilon,h}(\mathbf{x})| \leq C(1 + |\mathbf{x}|), \quad \text{and} \quad \|\nabla f_{\varepsilon,h}\|_\infty \leq C.$$

We will prove this lemma at the end of this section. Applying this lemma, we can prove Lemma 3.2 as follow.

Proof of Lemma 3.2. In this proof, the expectation \mathbb{E} is with respect to measure μ , that is, \mathbb{E}^μ , and we omit the superscript. For g in the domain of \mathcal{A}^α , i.e., $\mathcal{A}^\alpha g \in \mathcal{C}_{lin}(\mathbb{R}^d)$, we have

$$\begin{aligned} \mu(\mathcal{A}^\alpha g) &= \mu \left(\lim_{t \rightarrow 0} \frac{P_t g - g}{t} \right) = \lim_{t \rightarrow 0} \mu \left(\frac{P_t g - g}{t} \right) \\ &= \lim_{t \rightarrow 0} \frac{\mathbb{E}[g(\mathbf{X}_t^\alpha)] - \mu(g)}{t} = \lim_{t \rightarrow 0} \frac{\mu(g) - \mu(g)}{t} = 0, \end{aligned}$$

where the second equality is by dominated convergence theorem and the fourth one is by the stationarity of \mathbf{X}_t . Hence,

$$(B.4) \quad \mathbb{E}[\mathcal{A}^\alpha f_{\varepsilon,h}(\mathbf{X}_0)] = 0.$$

From the definition of Wasserstein-1 distance, Stein's equations in (B.2) and (B.4), we obtain that for $\alpha \in (1, 2]$,

$$\begin{aligned} \mathcal{W}_1(\mu, \mu_\varepsilon) &= \sup_{h \in \text{Lip}(1)} \{\mathbb{E}h(\mathbf{X}_0) - \mu_\varepsilon(h)\} = \sup_{h \in \text{Lip}(1)} \{\mathbb{E}\mathcal{A}_\varepsilon^\alpha f_{\varepsilon,h}(\mathbf{X}_0)\} \\ &= \sup_{h \in \text{Lip}(1)} \{\mathbb{E}[(\mathcal{A}_\varepsilon^\alpha - \mathcal{A}^\alpha)f_{\varepsilon,h}(\mathbf{X}_0)]\} \\ &= \sup_{h \in \text{Lip}(1)} \{\mathbb{E}[\langle g_\varepsilon(\mathbf{X}_0) - g(\mathbf{X}_0), \nabla f_{\varepsilon,h}(\mathbf{X}_0) \rangle]\} \\ &\leq C\varepsilon \mathbb{E}\mathbb{1}_{\{|\mathbf{e}'_d \mathbf{X}_0| \leq \varepsilon\}} \leq C\varepsilon, \end{aligned}$$

where the first inequality holds from the estimate for $\|\nabla f_{\varepsilon,h}\|_\infty$ in Lemma B.1 and (2.4). The proof is complete. \square

Proof of Lemma B.1. (i) For any $h \in \text{Lip}(1)$, it follows from Lemma 3.1 that

$$\left| \int_0^\infty P_t^\varepsilon [h(\mathbf{x}) - \mu_\varepsilon(h)] dt \right| \leq \int_0^\infty \mathcal{W}_1(\mathcal{L}(\mathbf{X}_t^{\varepsilon,\mathbf{x}}), \mu_\varepsilon) dt$$

is well defined and

$$|f_{\varepsilon,h}(\mathbf{x})| \leq \int_0^\infty |P_t^\varepsilon [h(\mathbf{x}) - \mu_\varepsilon(h)]| dt \leq \int_0^\infty C(1 + |\mathbf{x}|)e^{-\theta t} dt \leq C(1 + |\mathbf{x}|).$$

The expression for $f_{\varepsilon,h}$ in (B.3) can be proved similarly as that in [24, Proposition 6.1]. We omit the details here.

(ii) For any $\mathbf{u} \in \mathbb{R}_0^d$, we have

$$\nabla_{\mathbf{u}} f_{\varepsilon,h}(\mathbf{x}) = \int_0^\infty \nabla_{\mathbf{u}} \mathbb{E} h(\mathbf{X}_t^{\varepsilon,\mathbf{x}}) dt = \int_0^\infty \mathbb{E} [\nabla h(\mathbf{X}_t^{\varepsilon,\mathbf{x}}) \nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon,\mathbf{x}}] dt.$$

It follows from the estimate for $|\nabla_{\mathbf{u}} \mathbf{X}_t^{\varepsilon,\mathbf{x}}|$ in Lemma 2.3 and $\|\nabla h\|_\infty \leq 1$ that

$$|\nabla_{\mathbf{u}} f_{\varepsilon,h}(\mathbf{x})| \leq \int_0^\infty \left(\frac{\lambda_{\max}(\mathbf{Q})}{\lambda_{\min}(\mathbf{Q})} \right)^{1/2} \exp \left\{ -\frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} t \right\} |\mathbf{u}| dt \leq 2 \left(\frac{\lambda_{\max}^3(\mathbf{Q})}{\lambda^2 \lambda_{\min}(\mathbf{Q})} \right)^{1/2} |\mathbf{u}|,$$

which implies the desired result. \square

B.3. Proof of Lemma 3.3.

Proof of Lemma 3.3. For positive definite matrix \mathbf{Q} in Lemma 2.1, recall that

$$\widehat{V}(\mathbf{x}) = \frac{1}{2} \langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle, \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

Then, one has

$$(B.5) \quad \frac{1}{2} \lambda_{\min}(\mathbf{Q}) |\mathbf{x}|^2 \leq \widehat{V}(\mathbf{x}) \leq \frac{1}{2} \lambda_{\max}(\mathbf{Q}) |\mathbf{x}|^2.$$

It follows from those two EM schemes that for any $n \in \mathbb{N}_0$

$$\mathbf{Y}_{t_{n+1}}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}} = \mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}} + \eta_{n+1} [g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}})].$$

Then, by a straightforward calculation, one has

$$(B.6) \quad \begin{aligned} \widehat{V}(\mathbf{Y}_{t_{n+1}}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}}) &= \frac{1}{2} \langle \mathbf{Y}_{t_{n+1}}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}}, \mathbf{Q}[\mathbf{Y}_{t_{n+1}}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}}] \rangle \\ &= \frac{1}{2} \langle \mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}, \mathbf{Q}[\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}] \rangle + \eta_{n+1} \langle \mathbf{Q}[\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}], g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}}) \rangle \\ &\quad + \frac{1}{2} \eta_{n+1}^2 \langle g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}}), \mathbf{Q}[g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}})] \rangle. \end{aligned}$$

It suffices to give estimates for the last two terms.

For $\langle \mathbf{Q}[\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}], g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}}) \rangle$, it follows from (2.4), (2.7) and (B.5) that there exists some positive constant C independent of ε such that

$$(B.7) \quad \begin{aligned} &|\langle \mathbf{Q}[\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}], g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}}) \rangle| \\ &= |\langle \mathbf{Q}[\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}], g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}}) - g_\varepsilon(\mathbf{Y}_{t_n}^{\mathbf{x}}) + g_\varepsilon(\mathbf{Y}_{t_n}^{\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}}) \rangle| \\ &\leq -\frac{\lambda}{2} |\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}|^2 + C\varepsilon |\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}| \\ &\leq -\frac{\lambda}{2} |\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}|^2 + C\varepsilon |\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}|^2 + C\varepsilon \\ &\leq \left(-\frac{\lambda}{\lambda_{\max}(\mathbf{Q})} + C\varepsilon \right) \widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) + C\varepsilon, \end{aligned}$$

where the second inequality holds from the Young's inequality, that is,

$$\varepsilon |\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}| = \varepsilon^{1/2} \varepsilon^{1/2} |\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}| \leq (\varepsilon + \varepsilon |\mathbf{Y}_{t_n}^{\varepsilon,\mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}|^2) / 2.$$

For $\langle g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}}), \mathbf{Q}[g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}})] \rangle$, it follows from (2.4) and (B.5) that there exists some positive constant C independent of ε such that

$$\begin{aligned}
 & | \langle g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}}), \mathbf{Q}[g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}})] \rangle | \\
 & \leq \lambda_{\max}(\mathbf{Q}) |g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}})|^2 \\
 (B.8) \quad & = \lambda_{\max}(\mathbf{Q}) |g_\varepsilon(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) - g_\varepsilon(\mathbf{Y}_{t_n}^{\mathbf{x}}) + g_\varepsilon(\mathbf{Y}_{t_n}^{\mathbf{x}}) - g(\mathbf{Y}_{t_n}^{\mathbf{x}})|^2 \\
 & \leq C |\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}|^2 + C\varepsilon^2 \\
 & \leq C\widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) + C\varepsilon^2.
 \end{aligned}$$

Combining (B.6), (B.7) and (B.8), we obtain

$$\widehat{V}(\mathbf{Y}_{t_{n+1}}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}}) \leq \left(1 - \frac{\lambda}{\lambda_{\max}(\mathbf{Q})} \eta_{n+1} + C\eta_{n+1}\varepsilon + C\eta_{n+1}^2 \right) \widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) + C\eta_{n+1}\varepsilon.$$

Due to Assumption 1.2, $(\eta_n)_{n \in \mathbb{N}}$ decreases to 0. Hence, there exists some positive integer k_0 such that for any $n \leq k_0$, we obtain

$$\widehat{V}(\mathbf{Y}_{t_{n+1}}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}}) \leq C\widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) + C\eta_{n+1}\varepsilon.$$

Since $\widehat{V}(\mathbf{Y}_0^{\varepsilon, \mathbf{x}} - \mathbf{Y}_0^{\mathbf{x}}) = \widehat{V}(\mathbf{0}) = 0$, this inequality implies that there exists some positive constant C (depending on k_0) independent of n and ε such that

$$(B.9) \quad \widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) \leq C\varepsilon, \quad \forall n \leq k_0.$$

In addition, for all $n > k_0$, one has

$$\widehat{V}(\mathbf{Y}_{t_{n+1}}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}}) \leq \left(1 - \frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} \eta_{n+1} \right) \widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) + C\eta_{n+1}\varepsilon.$$

By induction, for any $n \geq k_0$,

$$\begin{aligned}
 \widehat{V}(\mathbf{Y}_{t_{n+1}}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{n+1}}^{\mathbf{x}}) & \leq \prod_{i=k_0+1}^{n+1} \left(1 - \frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} \eta_i \right) \widehat{V}(\mathbf{Y}_{t_{k_0}}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{k_0}}^{\mathbf{x}}) \\
 & \quad + C\varepsilon \sum_{j=k_0+1}^{n+1} \eta_j \prod_{i=j+1}^{n+1} \left(1 - \frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} \eta_i \right).
 \end{aligned}$$

Combining this with (B.9) and the fact $1 - x \leq \exp(-x)$ for all $x \in \mathbb{R}$, there exists some positive constant C independent of n and ε such that for $n > k_0$,

$$\widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) \leq \exp \left\{ -\frac{\lambda}{2\lambda_{\max}(\mathbf{Q})} (t_n - t_{k_0}) \right\} \widehat{V}(\mathbf{Y}_{t_{k_0}}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_{k_0}}^{\mathbf{x}}) + C\varepsilon \leq C\varepsilon.$$

Thus, there exists some positive constant C independent of n and ε such that

$$\widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}) \leq C\varepsilon, \quad \forall n \in \mathbb{N}_0,$$

which implies that

$$\begin{aligned}
 \mathcal{W}_1(\mathcal{L}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}), \mathcal{L}(\mathbf{Y}_{t_n}^{\mathbf{x}})) & = \sup_{h \in \text{Lip}(1)} \{ \mathbb{E}h(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}}) - \mathbb{E}h(\mathbf{Y}_{t_n}^{\mathbf{x}}) \} \\
 & \leq \mathbb{E} |\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}| \leq C(\mathbb{E}\widehat{V}(\mathbf{Y}_{t_n}^{\varepsilon, \mathbf{x}} - \mathbf{Y}_{t_n}^{\mathbf{x}}))^{1/2} \leq C\varepsilon^{1/2}.
 \end{aligned}$$

The proof is complete. \square

B.4. Proof of Lemma 3.4.

Proof of Lemma 3.4. For convenience, we denote

$$u_n = \sum_{i=1}^n e^{-\theta(t_n - t_i)} \eta_i^{1+1/\alpha} \eta_n^{-1/\alpha}.$$

Then, u_{n+1} and u_n have the following relationship

$$u_{n+1} = \lambda_n u_n + \eta_{n+1},$$

where

$$\lambda_n = e^{-\theta \eta_{n+1}} \left(\frac{\eta_n}{\eta_{n+1}} \right)^{1/\alpha} = e^{-\theta \eta_{n+1}} \left(1 + \frac{\eta_n^\beta - \eta_{n+1}^\beta}{\eta_{n+1}^{1+\beta}} \eta_{n+1} \right)^{1/(\alpha\beta)}.$$

By Assumption 1.2, we obtain

$$\lambda_n \leq e^{-\theta \eta_{n+1}} (1 + \omega \eta_{n+1})^{1/(\alpha\beta)} \leq \exp \left\{ \left(-\theta + \frac{\omega}{\alpha\beta} \right) \eta_{n+1} \right\},$$

which implies that

$$u_{n+1} \leq \exp \left\{ \left(-\theta + \frac{\omega}{\alpha\beta} \right) \eta_{n+1} \right\} u_n + \eta_{n+1}.$$

Thus, multiplying $e^{(\theta - \frac{\omega}{\alpha\beta})t_{n+1}}$ both sides, we obtain

$$\exp \left\{ \left(\theta - \frac{\omega}{\alpha\beta} \right) t_{n+1} \right\} u_{n+1} \leq \exp \left\{ \left(\theta - \frac{\omega}{\alpha\beta} \right) t_n \right\} u_n + \exp \left\{ \left(\theta - \frac{\omega}{\alpha\beta} \right) t_{n+1} \right\} \eta_{n+1}.$$

By induction, we know

$$\exp \left\{ \left(\theta - \frac{\omega}{\alpha\beta} \right) t_{n+1} \right\} u_{n+1} \leq \exp \left\{ \left(\theta - \frac{\omega}{\alpha\beta} \right) t_1 \right\} u_1 + \sum_{k=2}^{n+1} \exp \left\{ \left(\theta - \frac{\omega}{\alpha\beta} \right) t_k \right\} \eta_k.$$

To simplify notation, we denote $J = \theta - \omega/(\alpha\beta)$. Notice that $\omega < \alpha\beta\theta$ and then

$$\begin{aligned} \sum_{k=2}^{n+1} e^{J t_k} \eta_k &= \sum_{k=2}^{n+1} e^{J t_{k-1} + J \eta_k} \eta_k \leq e^{J \eta_1} \sum_{k=2}^{n+1} e^{J t_{k-1}} \eta_k \\ &\leq e^{J \eta_1} \int_{t_1}^{t_{n+1}} e^{J t} dt \leq J^{-1} e^{J(\eta_1 + t_{n+1})}. \end{aligned}$$

Hence, there is a positive constant C such that

$$u_{n+1} \leq C, \quad \forall n \in \mathbb{N}_0.$$

Consequently, we complete the proof according to the definition of u_n . \square

APPENDIX C. PROOFS OF THE LEMMAS IN SECTION 4

Recall that the Stein's equation (4.1), for any $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$,

$$\mathcal{A}^\alpha f_h(\mathbf{x}) = h(\mathbf{x}) - \mu(h).$$

It follows from [24, Proposition 6.1] and Lemma 2.5 that the solution f_h can be expressed as

$$(C.1) \quad f_h(\mathbf{x}) = - \int_0^\infty \mathbb{P}_t[h(\mathbf{x}) - \mu(h)] dt.$$

The solution f_h can be also expressed as (see, [28, Lemma 2])

$$(C.2) \quad f_h(\mathbf{x}) = \int_0^\infty e^{-at} P_t[af_h(\mathbf{x}) - h(\mathbf{x}) + \mu(h)] dt, \quad \forall a > 0.$$

If the test function $h \in \mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$, denote \tilde{f}_h as the corresponding solution to the Stein's equation (4.1) which also can be expressed as those in (C.1) and (C.2).

To discuss the regularities for the Stein's equation (4.1), we consider the third kind of Stein's equation: for $h \in \mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$,

$$\mathcal{A}_\varepsilon \tilde{f}_{\varepsilon,h}(\mathbf{x}) = h(\mathbf{x}) - \mu_\varepsilon(h),$$

where \mathcal{A}_ε and μ_ε are the infinitesimal operator and the invariant measure for the process $(\mathbf{X}_t^\varepsilon)_{t \geq 0}$ in (2.2) respectively. It follows from [24, Proposition 6.1] and Lemma 2.5 that the solution $\tilde{f}_{\varepsilon,h}$ can be expressed as

$$\tilde{f}_{\varepsilon,h}(\mathbf{x}) = - \int_0^\infty P_t^\varepsilon[h(\mathbf{x}) - \mu_\varepsilon(h)] dt,$$

which implies that there exist some positive constants c and C , both independent of ε , such that

$$(C.3) \quad |\tilde{f}_{\varepsilon,h}(\mathbf{x})| \leq C(1 + |\mathbf{x}|^{2\kappa}).$$

In addition, it follows from Lemma 3.2 that $P_t^\varepsilon h(\mathbf{x}) \rightarrow P_t h(\mathbf{x})$ and $\mu_\varepsilon(h) \rightarrow \mu(h)$ as $\varepsilon \rightarrow 0$, hence, we obtain that

$$(C.4) \quad \lim_{\varepsilon \rightarrow 0} \tilde{f}_{\varepsilon,h}(\mathbf{x}) = - \lim_{\varepsilon \rightarrow 0} \int_0^\infty P_t^\varepsilon[h(\mathbf{x}) - \mu_\varepsilon(h)] dt = - \int_0^\infty P_t[h(\mathbf{x}) - \mu(h)] dt = \tilde{f}_h(\mathbf{x}).$$

Furthermore, the solution $\tilde{f}_{\varepsilon,h}$ also can be expressed as (see, [28, Lemma 2])

$$\tilde{f}_{\varepsilon,h}(\mathbf{x}) = \int_0^\infty e^{-at} P_t^\varepsilon[a\tilde{f}_{\varepsilon,h}(\mathbf{x}) - h(\mathbf{x}) + \mu_\varepsilon(h)] dt, \quad \forall a > 0.$$

It follows from Lemma B.1 that $\tilde{f}_{\varepsilon,h} \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$. Then it follows from [54, Theorem 1.1] that for any $\mathbf{u} \in \mathbb{R}^d$,

$$\nabla_{\mathbf{u}} \tilde{f}_{\varepsilon,h}(\mathbf{x}) = \int_0^\infty e^{-at} \mathbb{E} \left[\frac{1}{S_t} \left(a\tilde{f}_{\varepsilon,h}(\mathbf{X}_s^{\varepsilon,\mathbf{x}}) - h(\mathbf{X}_s^{\varepsilon,\mathbf{x}}) + \mu_\varepsilon(h) \right) \right] \int_0^t \boldsymbol{\sigma}^{-1} \nabla_{\mathbf{u}} \mathbf{X}_s^{\varepsilon,\mathbf{x}} dW_{S_s} ds,$$

where the subordinated $(W_{S_t})_{t \geq 0}$ is the rotationally symmetric α stable process. Using [54, Theorem 1.1] again, there exists some positive constant C independent of ε such that

$$(C.5) \quad \begin{aligned} & |\nabla_{\mathbf{u}} \tilde{f}_{\varepsilon,h}(\mathbf{x})| \\ &= \int_0^\infty e^{-at} \left[a \left(\mathbb{E} |\tilde{f}_{\varepsilon,h}(\mathbf{X}_s^{\varepsilon,\mathbf{x}})|^2 \right)^{1/2} + \left(\mathbb{E} |h(\mathbf{X}_s^{\varepsilon,\mathbf{x}}) - \mu_\varepsilon(h)|^2 \right)^{1/2} \right] \|\boldsymbol{\sigma}^{-1}\|_{\text{op}} e^{\|\nabla g_\varepsilon\|_{\text{op}} t} t^{-1/\alpha} dt \\ &\leq C(1 + |\mathbf{x}|^{2\beta}), \end{aligned}$$

where the last inequality holds from taking $a > \|\nabla g_\varepsilon\|_{\text{op}} + 1$ and moment estimate for $\mathbb{E} |\tilde{f}_{\varepsilon,h}(\mathbf{X}_s^{\varepsilon,\mathbf{x}})|^2$, that is,

$$\mathbb{E} |\tilde{f}_{\varepsilon,h}(\mathbf{X}_s^{\varepsilon,\mathbf{x}})|^2 \leq C \mathbb{E}(1 + |\mathbf{X}_s^{\varepsilon,\mathbf{x}}|^{4\kappa}) \leq C(1 + |x|^{4\kappa}),$$

where the first inequality holds from the estimate for $\tilde{f}_{\varepsilon,h}$ in (C.3), the second inequality holds from the moment estimate for $\mathbb{E} |\mathbf{X}_s^{\varepsilon,\mathbf{x}}|$ in (2.15) and the fact $4\kappa < 1$.

By the convergence results in (C.4), Lemmas 2.6 and 2.4, we know that

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} \nabla_{\mathbf{u}} \tilde{f}_{\varepsilon, h}(\mathbf{x}) \\
\text{(C.6)} \quad &= \lim_{\varepsilon \rightarrow 0} \int_0^\infty e^{-at} \mathbb{E} \left[\frac{1}{S_t} \left(a \tilde{f}_{\varepsilon, h}(\mathbf{X}_s^{\varepsilon, \mathbf{x}}) - h(\mathbf{X}_s^{\varepsilon, \mathbf{x}}) + \mu_\varepsilon(h) \right) \right] \int_0^t \boldsymbol{\sigma}^{-1} \nabla_{\mathbf{u}} X_s^{\varepsilon, \mathbf{x}} dW_{S_s} dt \\
&= \int_0^\infty e^{-at} \mathbb{E} \left[\frac{1}{S_t} \left(a \tilde{f}_h(\mathbf{X}_s^{\mathbf{x}}) - h(\mathbf{X}_s^{\mathbf{x}}) + \mu(h) \right) \right] \int_0^t \boldsymbol{\sigma}^{-1} \nabla_{\mathbf{u}} X_s^{\mathbf{x}} dW_{S_s} dt.
\end{aligned}$$

Since the operator ∇ is closed (see Partington [41, Theorem 2.2.6]), we know

$$\lim_{\varepsilon \rightarrow 0} \nabla_{\mathbf{u}} \tilde{f}_{\varepsilon, h}(\mathbf{x}) = \nabla_{\mathbf{u}} \tilde{f}_h(\mathbf{x}).$$

Proof of Lemma 4.1. (i) We shall use the expression for f_h in (C.1) and exponential ergodicity in Lemma 2.5 to get the estimate for $|f_h(\mathbf{x})|$ for those two kinds of $\alpha \in (1, 2)$ stable noises $(\mathbf{Z}_t)_{t \geq 0}$.

Since $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$, one knows that $|h(\mathbf{x})| \leq C \tilde{V}(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^d$. It follows from Lemma 2.5 that

$$\begin{aligned}
\text{(C.7)} \quad |f_h(\mathbf{x})| &\leq \int_0^\infty |P_t h(\mathbf{x}) - \mu(h)| dt \leq \|h\|_\infty \int_0^\infty \|P_t^* \delta_{\mathbf{x}} - \mu\|_{\text{TV}, \tilde{V}} dt \\
&\leq C \|h\|_\infty (1 + \tilde{V}(\mathbf{x})) \leq C \|h\|_\infty (1 + |\mathbf{x}|^{2\kappa}).
\end{aligned}$$

(ii) We need to discuss the estimate for $|\nabla f_h(\mathbf{x})|$ separately with respect to different α stable noises.

Case 1. For the rotationally symmetric α stable noise. Let $h \in \mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$. It follows from (C.5) and (C.6) that

$$\text{(C.8)} \quad |\nabla_{\mathbf{u}} \tilde{f}_h(\mathbf{x})| \leq C \|h\|_\infty (1 + |\mathbf{x}|^{2\kappa}) |\mathbf{u}|.$$

Secondly, we extend $h \in \mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$ to $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$ above, by the standard approximation (see Fang et al. [24, pp. 968-969]). Let $h \in \mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$, and define

$$h_\delta(\mathbf{x}) = \int_{\mathbb{R}^d} \varphi_\delta(\mathbf{y}) h(\mathbf{x} - \mathbf{y}) d\mathbf{y}, \quad \delta > 0,$$

where φ_δ is the density function of the normal distribution $\mathcal{N}(\mathbf{0}, \delta^2 \mathbf{I})$. Thus h_δ is smooth, $\|h_\delta\|_\infty \leq \|h\|_\infty$ and the solution to the Stein's equation (4.1) with h replaced by h_δ , is

$$f_{h, \delta}(\mathbf{x}) = - \int_0^\infty P_t [h_\delta(\mathbf{x}) - \mu(h_\delta)] dt.$$

Denote $\hat{h}_\delta = -h_\delta + \mu(h_\delta)$. Since $h_\delta \in \mathcal{C}_b^1(\mathbb{R}^d, \mathbb{R})$, one has $h_\delta(\mathbf{x}) \leq \|h_\delta\|_\infty (1 + \tilde{V}(\mathbf{x}))$ for all $\mathbf{x} \in \mathbb{R}^d$. From Lemma 2.5, one has

$$|P_t h_\delta(\mathbf{x}) - \mu(h_\delta)| \leq \|\hat{h}_\delta\|_\infty \|P_t^* \delta_{\mathbf{x}} - \mu\|_{\text{TV}, \tilde{V}} \leq C \|h_\delta\|_\infty (1 + \tilde{V}(\mathbf{x})) e^{-ct}.$$

With similar calculations for (C.7), one has

$$|f_{h, \delta}(\mathbf{x})| \leq C \|h_\delta\|_\infty (1 + |\mathbf{x}|^{2\kappa}).$$

By the dominated convergence theorem, one has

$$\lim_{\delta \rightarrow 0} f_{h, \delta}(\mathbf{x}) = - \int_0^\infty P_t [h(\mathbf{x}) - \mu(h)] dt = f_h(\mathbf{x}).$$

From (C.7) and $\|h_\delta\|_\infty \leq \|h\|_\infty$, one has

$$|f_{h, \delta}(\mathbf{x})| \leq C \|h_\delta\|_\infty (1 + |\mathbf{x}|^{2\kappa}) \leq C \|h\|_\infty (1 + |\mathbf{x}|^{2\kappa}).$$

Letting $\delta \rightarrow 0$, one obtains

$$(C.9) \quad |f_h(\mathbf{x})| \leq C \|h\|_\infty (1 + |\mathbf{x}|^{2\kappa}).$$

From calculations for (C.6), one has

$$\nabla_{\mathbf{u}} f_{h,\delta}(\mathbf{x}) = \int_0^\infty e^{-at} \mathbb{E} \left[\frac{1}{S_t} (af_{h,\delta}(\mathbf{X}_s^{\mathbf{x}}) - h_\delta(\mathbf{X}_s^{\mathbf{x}}) + \mu(h_\delta)) \right] \int_0^t \sigma^{-1} \nabla_{\mathbf{u}} X_s^{\mathbf{x}} dW_{S_s} dt$$

With similar calculations for the proof of (C.8), one has

$$(C.10) \quad |\nabla_{\mathbf{u}} f_{h,\delta}(\mathbf{x})| \leq C \|h_\delta\|_\infty (1 + |\mathbf{x}|^{2\kappa}) |\mathbf{u}| \leq C \|h\|_\infty (1 + |\mathbf{x}|^{2\kappa}) |\mathbf{u}|.$$

Since the operator ∇ is closed (see Partington [41, Theorem 2.2.6]), it follows from the dominated convergence theorem that

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \nabla_{\mathbf{u}} f_{h,\delta}(\mathbf{x}) \\ &= \lim_{\delta \rightarrow 0} \int_0^\infty e^{-at} \mathbb{E} \left[\frac{1}{S_t} (af_{h,\delta}(\mathbf{X}_s^{\mathbf{x}}) - h_\delta(\mathbf{X}_s^{\mathbf{x}}) + \mu(h_\delta)) \right] \int_0^t \sigma^{-1} \nabla_{\mathbf{u}} \mathbf{X}_s^{\mathbf{x}} dW_{S_s} dt \\ &= \int_0^\infty e^{-at} \mathbb{E} \left[\frac{1}{S_t} (af_h(\mathbf{X}_s^{\mathbf{x}}) - h(\mathbf{X}_s^{\mathbf{x}}) + \mu(h)) \right] \int_0^t \sigma^{-1} \nabla_{\mathbf{u}} \mathbf{X}_s^{\mathbf{x}} dW_{S_s} dt = \nabla_{\mathbf{u}} f_h(\mathbf{x}). \end{aligned}$$

Let $\delta \rightarrow 0$, it follows from (C.10) that

$$|\nabla_{\mathbf{u}} f_h(\mathbf{x})| \leq C \|h\|_\infty (1 + |\mathbf{x}|^{2\kappa}) |\mathbf{u}|.$$

Case 2. For the cylindrical α stable noise. Since g is Lipschitz continuous, it follows from (C.2) and [50, Theorem 1.2] that for any $\mathbf{u} \in \mathbb{R}^d$,

$$\begin{aligned} |\nabla_{\mathbf{u}} f_h(\mathbf{x})| &= \left| \int_0^\infty \nabla_{\mathbf{u}} [ae^{-at} \mathbb{E} f_h(\mathbf{X}_t^{\mathbf{x}}) - e^{-at} \mathbb{E} (h(\mathbf{X}_t^{\mathbf{x}}) - \mu(h))] dt \right| \\ &\leq C \int_0^{+\infty} e^{-at} (1 + te^{d\|g\|_{\text{Lip}} t}) t^{1-1/\alpha} [a(\mathbb{E} |f_h(\mathbf{X}_t^{\mathbf{x}})|^2)^{1/2} + (\mathbb{E} |h(\mathbf{X}_t^{\mathbf{x}}) - \mu(h)|^2)^{1/2}] dt \\ &\leq C(1 + |\mathbf{x}|^{2\kappa}), \end{aligned}$$

where the last inequality holds from taking $a > d\|g\|_{\text{Lip}}$ with $\|g\|_{\text{Lip}} = \sup_{\mathbf{x}, \mathbf{y} \in \mathbb{R}^d} \frac{|g(\mathbf{x}) - g(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|}$ and the moment estimate for $\mathbb{E} |f_h(\mathbf{X}_t^{\mathbf{x}})|^2$ as below

$$\mathbb{E} |f_h(\mathbf{X}_s^{\mathbf{x}})|^2 \leq C \mathbb{E} (1 + |\mathbf{X}_s^{\mathbf{x}}|^{4\kappa}) \leq C(1 + |\mathbf{x}|^{4\kappa}),$$

where the first inequality holds from the estimate for $|f_h|$ in (C.9), the second inequality holds from the moment estimate for $\mathbb{E} |\mathbf{X}_t^{\mathbf{x}}|$ in (2.15) and the fact $4\kappa < 1$. The proof is complete. \square

ACKNOWLEDGEMENTS

L. Xu is supported by the National Natural Science Foundation of China No. 12071499, the Science and Technology Development Fund (FDCT) of Macau S.A.R. FDCT 0074/2023/RIA2, and the University of Macau grants MYRG2020-00039-FST, MYRG-GRG2023-00088-FST. G. Pang is supported in part by the US National Science Foundation grant DMS-2216765. X. Jin is supported in part by the National Natural Science Foundation of China No. 12401176, and the Fundamental Research Funds for the Central Universities grant No. JZ2023HGTA0170.

REFERENCES

- [1] AGHAJANI, R., AND RAMANAN, K. The limit of stationary distributions of many-server queues in the Halfin-Whitt regime. *Mathematics of Operations Research* 45, 3 (2020), 1016–1055. 3
- [2] APPLEBAUM, D. *Lévy processes and stochastic calculus*. Cambridge university press, 2009. 15
- [3] ARAPOSTATHIS, A., HMEDI, H., AND PANG, G. On uniform exponential ergodicity of Markovian multiclass many-server queues in the Halfin-Whitt regime. *Mathematics of Operations Research* 46, 2 (2021), 772–796. 3
- [4] ARAPOSTATHIS, A., HMEDI, H., PANG, G., AND SANDRIĆ, N. Uniform polynomial rates of convergence for a class of Lévy-driven controlled SDEs arising in multiclass many-server queues. In *modeling, stochastic control, optimization, and applications*, vol. 164 of *IMA Vol. Math. Appl.* Springer, Cham, 2019, pp. 1–20. 2, 3
- [5] ARAPOSTATHIS, A., PANG, G., AND SANDRIĆ, N. Ergodicity of a Lévy-driven SDE arising from multiclass many-server queues. *The Annals of Applied Probability* 29, 2 (2019), 1070–1126. 2, 3, 4, 8, 11, 23
- [6] BENAÏM, M., BOUGUET, F., AND CLOEZ, B. Ergodicity of inhomogeneous Markov chains through asymptotic pseudotrajectories. *The Annals of Applied Probability* 27, 5 (2017), 3004–3049. 3
- [7] BERMAN, A., AND PLEMMONS, R. J. *Nonnegative matrices in the Mathematical sciences*, vol. 9 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. Revised reprint of the 1979 original. 2, 6
- [8] BERTSIMAS, D., AND TSITSIKLIS, J. N. *Introduction to linear optimization*, vol. 6 of *Athena Scientific books*. Athena Scientific Belmont, MA, 1997. 22
- [9] BRAVERMAN, A., AND DAI, J. G. Stein’s method for steady-state diffusion approximations of $M/Ph/n + M$ systems. *The Annals of Applied Probability* 27, 1 (2017), 550–581. 2, 3
- [10] BRAVERMAN, A., DAI, J. G., AND FENG, J. Stein’s method for steady-state diffusion approximations: An introduction through the Erlang-A and Erlang-C models. *Stochastic Systems* 6, 2 (2016), 301–366. 3
- [11] BROWNE, S., AND WHITT, W. Piecewise-linear diffusion processes. In *Advances in Queueing Theory, Methods, and Open Problems*. CRC Press, 1995, pp. 463–480. 22
- [12] BUDHIRAJA, A., CHEN, J., AND RUBENTHALER, S. A numerical scheme for invariant distributions of constrained diffusions. *Mathematics of Operations Research* 39, 2 (2014), 262–289. 3
- [13] CHEN, P., DENG, C.-S., SCHILLING, R. L., AND XU, L. Approximation of the invariant measure of stable SDEs by an Euler-Maruyama scheme. *Stochastic Processes and their Applications* 163 (2023), 136–167. 4, 20
- [14] CHEN, P., JIN, X., XIAO, Y., AND XU, L. Approximation of the invariant measure for stable SDE by the Euler-Maruyama scheme with decreasing step-sizes, 2023. 4
- [15] CHEN, X., YANG, L., DUAN, J., AND KARNIADAKIS, G. E. Solving inverse stochastic problems from discrete particle observations using the Fokker–Planck equation and Physics-Informed Neural networks. *SIAM Journal on Scientific Computing* 43, 3 (2021), B811–B830. 2
- [16] CHIZAT, L., ROUSSILLON, P., LÉGER, F., VIALARD, F.-X., AND PEYRÉ, G. Faster Wasserstein distance estimation with the Sinkhorn divergence. In *Advances in Neural Information Processing Systems* (2020), H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, Curran Associates, Inc., pp. 2257–2269. 4
- [17] CUTURI, M. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* (2013), C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26, Curran Associates, Inc. 4
- [18] DA PRATO, G., AND ZABCZYK, J. *Ergodicity for infinite dimensional systems*, vol. 229. Cambridge university press, 1996. 4
- [19] DAI, J., AND HE, S. Many-server queues with customer abandonment: Numerical analysis of their diffusion model. *Stochastic Systems* 3, 1 (2013), 96–146. 3
- [20] DAI, J., HE, S., AND TEZCAN, T. Many-server diffusion limits for $G/Ph/n + GI$ queues. *Annals of Applied Probability* 20, 5 (2010), 1854–1890. 3
- [21] DAI, J. G., AND HE, S. Numerical analysis for diffusion models of $GI/Ph/n + GI$ queues. *ACM SIGMETRICS Performance Evaluation Review* 39, 4 (2012), 37–37. 3
- [22] DENG, C., SCHILLING, R. L., AND XU, L. Wasserstein-1 distance between SDEs driven by Brownian motion and stable processes. *arXiv preprint arXiv:2302.03372* (2023). 10

- [23] DIEKER, A. B., AND GAO, X. Positive recurrence of piecewise Ornstein-Uhlenbeck processes and common quadratic Lyapunov functions. *The Annals of Applied Probability* 23, 4 (2013), 1291–1317. 2, 3
- [24] FANG, X., SHAO, Q.-M., AND XU, L. Multivariate approximations in Wasserstein distance by Stein’s method and Bismut’s formula. *Probability Theory and Related Fields* 174, 3-4 (2019), 945–979. 3, 34, 36, 37, 38
- [25] GURVICH, I. Diffusion models and steady-state approximations for exponentially ergodic Markovian queues. *The Annals of Applied Probability* 24, 6 (2014), 2527–2559. 3, 30
- [26] HAIRER, M. An introduction to stochastic PDEs. *arXiv preprint arXiv:0907.4178* (2009). 6
- [27] HUANG, X., AND LIAO, Z.-W. The Euler-Maruyama method for S(F)DEs with Hölder drift and α -stable noise. *Stochastic Analysis and Applications* 36, 1 (2018), 28–39. 4
- [28] JIN, X., PANG, G., XU, L., AND XU, X. An approximation to the invariant measure of the limiting diffusion of $G/Ph/n + GI$ queues in the Halfin–Whitt regime and related asymptotics. *Mathematics of Operations Research* (2024). 3, 4, 5, 7, 13, 37
- [29] KHAMLICH, M., PICHI, F., AND ROZZA, G. Optimal transport-inspired deep learning framework for slow-decaying problems: Exploiting Sinkhorn loss and Wasserstein kernel, 2023. 4
- [30] KNIGHT, P. A. The Sinkhorn-Knopp algorithm: Convergence and applications. *SIAM Journal on Matrix Analysis and Applications* 30, 1 (2008), 261–275. 4, 22
- [31] KÜHN, F., AND SCHILLING, R. L. Strong convergence of the Euler-Maruyama approximation for a class of Lévy-driven SDEs. *Stochastic Processes and their Applications* 129, 8 (2019), 2654–2680. 3, 4
- [32] LIAO, Q., CHEN, J., WANG, Z., BAI, B., JIN, S., AND WU, H. Fast Sinkhorn I: An $O(N)$ algorithm for the Wasserstein-1 metric. *arXiv preprint arXiv:2202.10042* (2022). 4
- [33] LUISE, G., RUDI, A., PONTIL, M., AND CILIBERTO, C. Differential properties of Sinkhorn approximation for learning with Wasserstein distance. In *Advances in Neural Information Processing Systems* (2018), S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31, Curran Associates, Inc. 4, 22
- [34] MCLEISH, D. L. Dependent central limit theorems and invariance principles. *the Annals of Probability* 2, 4 (1974), 620–628. 15
- [35] MENOUCHEU PAMEN, O., AND TAGUCHI, D. Strong rate of convergence for the Euler-Maruyama approximation of SDEs with Hölder continuous drift coefficient. *Stochastic Processes and their Applications* 127, 8 (2017), 2542–2559. 4
- [36] NOLAN, J. P. *Univariate stable distributions: Models for heavy tailed data*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2020. 22
- [37] PAGÈS, G., AND PANLOUP, F. Unadjusted Langevin algorithm with multiplicative noise: total variation and Wasserstein bounds. *Ann. Appl. Probab.* 33, 1 (2023), 726–779. 11, 12
- [38] PANG, G., AND WHITT, W. Continuity of a queueing integral representation in the M_1 topology. *The Annals of Applied Probability* 20, 1 (2010), 214 – 237. 3, 22
- [39] PANLOUP, F. Computation of the invariant measure for a Lévy driven SDE: Rate of convergence. *Stochastic Processes and their Applications* 118, 8 (2008), 1351–1384. 3, 4
- [40] PANLOUP, F. Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *The Annals of Applied Probability* 18, 2 (2008), 379–426. 3, 4
- [41] PARTINGTON, J. R. *Linear operators and linear systems*, vol. 60 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2004. An analytical approach to control theory. 38, 39
- [42] PROTTER, P., AND TALAY, D. The Euler scheme for Lévy driven stochastic differential equations. *The Annals of Probability* 25, 1 (1997), 393–423. 4
- [43] PUHALSKII, A. A., AND REIMAN, M. I. The multiclass $GI/PH/N$ queue in the Halfin-Whitt regime. *Advances in Applied Probability* 32, 2 (2000), 564–595. 3
- [44] RAJ, A., ZHU, L., GURBUZBALABAN, M., AND SIMSEKLI, U. Algorithmic stability of heavy-tailed SGD with general loss functions. In *International Conference on Machine Learning* (2023), PMLR, pp. 28578–28597. 2
- [45] SAMORODNITSKY, G., AND TAQQU, M. S. *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York, 1994. Stochastic models with infinite variance. 19, 22
- [46] SETHURAMAN, S. A martingale central limit theorem. *PDF accessible on author’s website: https://www.math.arizona.edu/sethuram/notes/wi_mart1.pdf* 17 (2002). 15
- [47] TANKOV, P. *Financial modelling with jump processes*. Chapman and Hall/CRC, 2003. 2

- [48] VILLANI, C. *Optimal transport*, vol. 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new. [6](#), [7](#)
- [49] WANG, F.-Y., AND WANG, J. Harnack inequalities for stochastic equations driven by Lévy noise. *J. Math. Anal. Appl.* *410*, 1 (2014), 513–523. [17](#)
- [50] WANG, L., AND ZHANG, X. Harnack inequalities for sdes driven by cylindrical α -stable processes. *Potential Analysis* *42* (2015), 657–669. [17](#), [39](#)
- [51] WU, L. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic processes and their applications* *91*, 2 (2001), 205–238. [17](#)
- [52] WU, L. M. Moderate deviations of dependent random variables related to CLT. *Ann. Probab.* *23*, 1 (1995), 420–445. [17](#)
- [53] XIE, Y., WANG, X., WANG, R., AND ZHA, H. A fast proximal point method for computing exact Wasserstein distance. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference (22–25 Jul 2020)*, R. P. Adams and V. Gogate, Eds., vol. 115 of *Proceedings of Machine Learning Research*, PMLR, pp. 433–453. [4](#)
- [54] ZHANG, X. Derivative formulas and gradient estimates for SDEs driven by α -stable processes. *Stochastic Processes and their Applications* *123*, 4 (2013), 1213–1228. [37](#)

XINGHU JIN: SCHOOL OF MATHEMATICS, HEFEI UNIVERSITY OF TECHNOLOGY, HEFEI, ANHUI, 230601, CHINA.

Email address: xinghujin@hfut.edu.cn

GUODONG PANG: DEPARTMENT OF COMPUTATIONAL APPLIED MATHEMATICS AND OPERATIONS RESEARCH, GEORGE R. BROWN SCHOOL OF ENGINEERING, RICE UNIVERSITY, HOUSTON, TX 77005

Email address: gdpang@rice.edu

YU WANG: 1. DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE AND TECHNOLOGY, UNIVERSITY OF MACAU, TAIPA, MACAU, 999078, CHINA; 2. UM ZHUHAI RESEARCH INSTITUTE, ZHUHAI, GUANGDONG, 519000, CHINA.

Email address: yc17447@um.edu.mo

LIHU XU: 1. DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE AND TECHNOLOGY, UNIVERSITY OF MACAU, TAIPA, MACAU, 999078, CHINA; 2. UM ZHUHAI RESEARCH INSTITUTE, ZHUHAI, GUANGDONG, 519000, CHINA.

Email address: lihuxu@umac.mo