



Heriot-Watt University
Research Gateway

Exponential ergodicity for SDEs and McKean-Vlasov processes with Lévy noise

Citation for published version:

Liang, M, Majka, MB & Wang, J 2021, 'Exponential ergodicity for SDEs and McKean-Vlasov processes with Lévy noise', *Annales de l'institut Henri Poincaré (B) Probability and Statistics*, vol. 57, no. 3, pp. 1665-1701. <https://doi.org/10.1214/20-AIHP1123>

Digital Object Identifier (DOI):

[10.1214/20-AIHP1123](https://doi.org/10.1214/20-AIHP1123)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Annales de l'institut Henri Poincaré (B) Probability and Statistics

Publisher Rights Statement:

Copyright © 2021 Association des Publications de l'Institut Henri Poincaré

Ann. Inst. H. Poincaré Probab. Statist. 57 (3) 1665 - 1701, August 2021. <https://doi.org/10.1214/20-AIHP1123>

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

EXPONENTIAL ERGODICITY FOR SDES AND MCKEAN-VLASOV PROCESSES WITH LÉVY NOISE

MINGJIE LIANG MATEUSZ B. MAJKA JIAN WANG

ABSTRACT. We study exponential ergodicity of a broad class of stochastic processes whose dynamics are governed by pure jump Lévy noise. In the first part of the paper we focus on solutions of stochastic differential equations (SDEs) whose drifts satisfy general Lyapunov-type conditions. By applying techniques that combine couplings, appropriately constructed L^1 -Wasserstein distances and Lyapunov functions, we show exponential convergence of solutions of such SDEs to their stationary distributions, both in the total variation and Wasserstein distances. The second part of the paper is devoted to SDEs of McKean-Vlasov type with distribution dependent drifts. We prove a uniform in time propagation of chaos result, providing quantitative bounds on convergence rate of interacting particle systems with Lévy noise to the corresponding McKean-Vlasov SDE. Then, extending our techniques from the first part of the paper, we obtain results on exponential ergodicity of solutions of McKean-Vlasov SDEs, under general conditions on the drift and the driving noise.

Keywords: exponential ergodicity; Lévy noise; coupling; McKean-Vlasov process; mean-field SDE; propagation of chaos

MSC 2010: 60H10; 60J25; 60J75.

1. INTRODUCTION AND MAIN RESULTS

In recent years, the study of ergodicity for solutions of stochastic differential equations (SDEs) with Lévy noise has attracted considerable interest (see e.g. [37, 28, 50, 34, 33, 35, 3, 4]). The most widespread tools in this area of research are the Meyn-Tweedie approach based on showing irreducibility and aperiodicity of the associated process ([37, 3, 4]), and the coupling approach ([34, 33, 35]). For SDEs in \mathbb{R}^d of the form

$$(1.1) \quad dX_t = b(X_t)dt + dZ_t,$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the drift and $(Z_t)_{t \geq 0}$ is a d -dimensional pure jump Lévy process, the Meyn-Tweedie approach allows for studying exponential convergence under rather general conditions on the drift, both in the total variation (see e.g. [3, Corollary 5.2]) and the Wasserstein distances (see [4, Theorem 1.3] with the special case of $a(x) \equiv 0$ and $\nu(x, dy) = \nu(dy)$ for all $x \in \mathbb{R}^d$). Namely, the authors of [3, 4] considered conditions of the form

$$(1.2) \quad \langle x, b(x) \rangle \leq -\lambda|x|^2$$

M. Liang: College of Information Engineering, Sanming University, 365004, Sanming, P.R. China. liangmingjie@aliyun.com.

M. B. Majka: School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK. m.majka@hw.ac.uk.

J. Wang: School of Mathematics and Computer Science & Fujian Key Laboratory of Mathematical Analysis and Applications (FJKLMAA) & Center for Applied Mathematics of Fujian Province (FJNU), Fujian Normal University, 350007, Fuzhou, P.R. China. jianwang@fjnu.edu.cn.

for some constant $\lambda > 0$ and all $x \in \mathbb{R}^d$ with large enough $|x|$. The downside is that in order to show irreducibility and aperiodicity of solutions to (1.1), both [3] and [4] cover only pure jump Lévy noises that are either compound Poisson processes or contain a symmetric α -stable component; see [3, Section 4] and [4, Section 3.3] for details. Moreover, usually it is difficult to quantify convergence rates obtained via such methods (cf. [16, Section 3.1]). On the other hand, by applying the coupling technique, [33, 34, 35] considered SDEs with a much more general class of driving noises, while requiring that the drift $b(x)$ in (1.1) satisfies conditions of the form

$$(1.3) \quad \langle x - y, b(x) - b(y) \rangle \leq -K|x - y|^2$$

for all $x, y \in \mathbb{R}^d$ such that $|x - y| > R$, with some constants $K, R > 0$. Note that (1.3) obviously implies (1.2). In the present paper, we combine the coupling approach from [33] with techniques recently developed for diffusions in [16], in order to study exponential convergence of solutions to SDEs of the form (1.1) under conditions on the drift such as (1.2), and for a very general class of Lévy processes $(Z_t)_{t \geq 0}$, thus considerably extending the results on exponential ergodicity from all the papers [3, 4, 33, 34, 35]. To this end, we will use the L^1 -Wasserstein distance defined for any probability measures μ_1, μ_2 on \mathbb{R}^d by

$$W_1(\mu_1, \mu_2) = \inf_{\Pi \in \mathcal{C}(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| d\Pi(x, y),$$

where $\mathcal{C}(\mu_1, \mu_2)$ is the family of all measures on $\mathbb{R}^d \times \mathbb{R}^d$ having μ_1 and μ_2 as marginals. We will also denote by $\|\mu_1 - \mu_2\|_{\text{var}}$ the total variation distance between probability measures μ_1 and μ_2 . Everywhere in this paper we will assume that (1.1) has a unique strong solution. This is the case for any Lévy noise $(Z_t)_{t \geq 0}$ if we assume that $b(x)$ is continuous and that

$$(1.4) \quad \langle x - y, b(x) - b(y) \rangle \leq C|x - y|^2$$

for all $x, y \in \mathbb{R}^d$, with some constant $C > 0$ (see e.g. [18, 36]; see also the discussion in Section 2 for results under different assumptions on the drift). Moreover, for the Lévy measure ν of the process $(Z_t)_{t \geq 0}$ we define

$$(1.5) \quad J(r) := \inf_{|x| \leq r} (\nu \wedge (\delta_x * \nu))(\mathbb{R}^d), \quad r > 0.$$

We have the following result.

Theorem 1.1. *Let $(X_t)_{t \geq 0}$ be the unique strong solution to (1.1) with the drift $b(x)$ satisfying (1.2) and (1.4). Suppose that for the Lévy measure ν of $(Z_t)_{t \geq 0}$,*

$$(1.6) \quad \int_{\{|z| \geq 1\}} |z| \nu(dz) < \infty$$

and there is a constant $\kappa_0 \in (0, 1]$ such that $J(\kappa_0) > 0$. Let $P_t(x, \cdot)$ be the law of X_t for $t > 0$ when $X_0 = x$ for some $x \in \mathbb{R}^d$. Then there exist a unique invariant probability measure μ for $(X_t)_{t \geq 0}$ and a constant $\lambda > 0$ such that

$$\|P_t(x, \cdot) - \mu\|_{\text{var}} \leq c_1(x)e^{-\lambda t}$$

and

$$W_1(P_t(x, \cdot), \mu) \leq c_2(x)e^{-\lambda t}$$

hold for all $x \in \mathbb{R}^d$ and $t > 0$, where $c_1(x)$ and $c_2(x)$ are non-negative measurable functions on \mathbb{R}^d .

Note that the assumption that $J(\kappa_0) > 0$ for some $\kappa_0 \in (0, 1]$ is very weak (see e.g. [43, Proposition 1.5] and [34, Remark 1.7]), and hence Theorem 1.1 covers a much larger class of Lévy processes than [3, 4, 33, 34, 35]. On the other hand, we note that exponential ergodicity in terms of the total variation norm for solutions to (more general) SDEs with jump noise has been studied in [28] via the local Doeblin condition, which is expressed in terms of the transition probability of the associated process (see condition **(LD)** in [28]). It was shown in [28, Theorem 1.3] that the local Doeblin condition can be derived from non-degeneracy condition **N** and topological irreducibility condition **S**, both of which essentially require non-degenerate properties of the Lévy measure ν near zero; see [28, Section 4] for details. In view of this, it seems that our assumption that $J(\kappa_0) > 0$ for some $\kappa_0 \in (0, 1]$ in Theorem 1.1 (which can be interpreted as requiring a sufficient mass of the Lévy measure near zero) looks simpler and more direct. We also remark that conditions (1.2) and (1.4) can be weakened even further (cf. Theorem 2.2 below). All the technical details and the proof of Theorem 1.1 can be found in Section 2. Besides, we would like to mention that (1.6) implies that for any $x \in \mathbb{R}^d$ and $t > 0$, $\mathbb{E}^x |X_t| < \infty$, which is needed for proving the exponential ergodicity of the process $(X_t)_{t \geq 0}$ in terms of the L^1 -Wasserstein distance. However, (1.6) can be relaxed (for example, to the condition $\int_{\{|z| \geq 1\}} |z|^\alpha \nu(dz) < \infty$ with some $\alpha \in (0, 1)$; see [48, 49]) if one is interested in studying exponential ergodicity only in terms of the total variation norm.

The second part of this paper is devoted to stochastic differential equations on \mathbb{R}^d of McKean-Vlasov type, i.e.,

$$(1.7) \quad \begin{cases} dX_t = b(X_t, \mu_t) dt + dZ_t, & X_0 \sim \mu_0, \\ \mu_t = \text{Law}(X_t), \end{cases}$$

where $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is a measurable function and $(Z_t)_{t \geq 0}$ is a d -dimensional pure jump Lévy process. Here $\mathcal{P}(\mathbb{R}^d)$ denotes the family of all probability measures on \mathbb{R}^d . The solution $(X_t)_{t \geq 0}$ to (1.7) is a nonlinear Markov process [27] in the sense of McKean-Vlasov, i.e., the associated transition function depends both on the current state X_t and on the law of X_t . Equations of the form (1.7) driven by Brownian motion have recently attracted considerable attention, see e.g. [5, 14, 16, 21, 23, 47] and the references therein, as well as the monograph [9]. The case in which the driving process has jumps has been much less studied, however, with the foundations set by [25], several papers on the topic have appeared in recent years, see e.g. [1, 2, 6, 8, 24, 30, 38, 42]. By analogy to the relation between McKean-Vlasov SDEs with Brownian noise and nonlinear Fokker-Planck equations [5, 23], McKean-Vlasov SDEs with jumps can be related to nonlocal integral-PDEs [24, 25, 30]. They have also found applications in areas such as financial mathematics [6] and neural networks [38]. Such equations, regardless of the driving noise, are known in the literature under numerous different names, including McKean-Vlasov SDEs [2, 7, 16, 21, 17], mean-field SDEs [8, 24, 30, 42], distribution dependent SDEs (DDSDEs) [5, 47] and nonlinear SDEs [25]. The choice of the name depends on the preferred interpretation of the process $(X_t)_{t \geq 0}$ and different names are often used interchangeably.

A typical example of the drift in (1.7) is

$$(1.8) \quad b(x, \mu) = b_1(x) + \int b_2(x, z) \mu(dz)$$

for some measurable functions $b_1 : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $b_2 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. In such case the corresponding McKean-Vlasov SDE arises naturally as a marginal limit as $n \rightarrow \infty$ of the mean-field interacting particle system

$$(1.9) \quad dX_t^{i,n} = b_1(X_t^{i,n}) dt + \frac{1}{n} \sum_{j=1}^n b_2(X_t^{i,n}, X_t^{j,n}) dt + dZ_t^i, \quad i = 1, \dots, n$$

driven by independent Lévy processes $(Z_t^i)_{t \geq 0}$. This property was named *propagation of chaos* by Kac [26] and was further developed by Sznitman [44]. Propagation of chaos has been studied extensively in the Brownian setting, see e.g. [11, 15, 39] and the references therein. In the Lévy jump case, the result of such kind has been proved in [25] for McKean-Vlasov (nonlinear) SDEs even with distribution dependent noise coefficients. See also [2] for a recent study on nonlinear jump diffusions with finite jump activity. However, the bounds obtained in [2, 25] are uniform only on a finite time interval. In the present paper, we adapt the method from [15] to show uniform in time convergence, as well as its explicit quantitative rate, of the particle system (1.9) in the L^1 -Wasserstein distance.

We work under relatively weak assumptions on the drift and the driving Lévy process. Namely, for the term $b_1(x)$ in the drift $b(x, \mu)$ given by (1.8) we assume

(1-i) $b_1(x)$ is continuous on \mathbb{R}^d , and there are constants $K_{1,b_1}, r_{b_1} \geq 0$ and $K_{2,b_1} > 0$ such that for any $x, y \in \mathbb{R}^d$,

$$(1.10) \quad \langle b_1(x) - b_1(y), x - y \rangle \leq K_{1,b_1} |x - y|^2 \mathbb{1}_{\{|x-y| \leq r_{b_1}\}} - K_{2,b_1} |x - y|^2 \mathbb{1}_{\{|x-y| > r_{b_1}\}}.$$

Furthermore, for the term $b_2(x, \mu)$ in (1.8) we require

(1-ii) There exists a function $\tilde{b}_2 : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $\tilde{b}_2(0) = 0$ such that $b_2(x, z) = \tilde{b}_2(x - z)$ for any $x, z \in \mathbb{R}^d$, and there is a constant $K_{\tilde{b}_2} > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$(1.11) \quad |\tilde{b}_2(x) - \tilde{b}_2(y)| \leq K_{\tilde{b}_2} |x - y|.$$

Finally, for the Lévy measure ν of the Lévy process $(Z_t)_{t \geq 0}$ we consider the function $J(r)$ defined in (1.5) and we assume the following conditions:

(1-iii) There exists a constant $\alpha \in (0, 1)$ such that

$$(1.12) \quad \liminf_{r \rightarrow 0} \inf_{s \in (0, r]} J(s) s^\alpha > 0;$$

moreover,

$$\int_{\{|z| > 1\}} |z|^2 \nu(dz) < \infty.$$

We now consider the McKean-Vlasov equation (1.7) with the drift (1.8) and the corresponding mean-field particle system defined by (1.9) with initial values $X_0 \sim \mu_0$ and $X_0^{i,n} \sim \mu_0^n$, $1 \leq i \leq n$, respectively, which are i.i.d. random variables having finite second moments. Moreover, $(Z_t^i)_{t \geq 0}$, $1 \leq i \leq n$, are i.i.d. Lévy processes with the same law as $(Z_t)_{t \geq 0}$. Under the assumptions above, we can prove that both (1.7) and (1.9) have unique strong solutions, which we denote by $(X_t)_{t \geq 0}$ and $(X_t^{i,n})_{t \geq 0}$, $1 \leq i \leq n$, respectively; moreover, for any $t_0 > 0$ and $1 \leq i \leq n$,

$\sup_{0 < t \leq t_0} \mathbb{E}(|X_t|^2 + |X_t^{i,n}|^2) < \infty$. To see this, one can follow the standard arguments of [47, Theorem 2.1], [17, Theorem 2.1] and [25, Proposition 1.2], see the proof of Proposition 3.1 in Section 3 for details. Note that all the processes $(X_t^{i,n})_{t \geq 0}$, $1 \leq i \leq n$, have the same marginal laws due to the uniqueness of solutions to the SDE (1.9). In the present paper, we prove the following result on uniform in time propagation of chaos for weakly interacting mean-field particle systems with Lévy jumps.

Theorem 1.2. *Suppose that assumptions (1-i)–(1-iii) hold. Denote by $(X_t)_{t \geq 0}$ and $(X_t^{i,n})_{t \geq 0}$, $1 \leq i \leq n$, the unique strong solutions to (1.7) and (1.9) with initial distributions μ_0 and μ_0^n , respectively. Suppose that μ_0 and μ_0^n have finite first moments. Let μ_t be the marginal law of X_t , and μ_t^n the marginal law of $X_t^{i,n}$ for any $t > 0$ and any $1 \leq i \leq n$. Then there exists a constant $K_{b_2}^* > 0$ such that for any $K_{\tilde{b}_2} \in (0, K_{b_2}^*]$, $t > 0$ and $n \geq 2$,*

$$W_1(\mu_t, \mu_t^n) \leq C_0 e^{-\lambda t} W_1(\mu_0, \mu_0^n) + C_0 n^{-1/2}, \quad t > 0$$

holds for some positive constants λ and C_0 independent of t and n .

Remark 1.3. Condition (1.12) is an assumption about sufficient concentration of the Lévy measure ν around zero (i.e., sufficient small jump activity of the Lévy process $(Z_t)_{t \geq 0}$). It is, however, a weak requirement that can be satisfied even for singular measures. For example, if

$$\nu(dz) \geq \mathbb{1}_{\{0 < z_1 \leq 1\}} \frac{c}{|z|^{d+\alpha}} dz$$

holds for $z = (z_1, \dots, z_d) \in \mathbb{R}^d$ and some constants $c > 0$ and $\alpha \in (0, 2)$, then (1.12) is satisfied (see [33, Example 1.2]).

In order to study convergence of solutions of McKean-Vlasov SDEs to their invariant measures, we need weaker assumptions than for proving the uniform in time propagation of chaos in Theorem 1.2. Namely, we can consider (1.7) with a drift term $b : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ of general form. Here $\mathcal{P}_1(\mathbb{R}^d)$ denotes the family of all probability measures on \mathbb{R}^d with finite first moment, equipped with the topology of weak convergence metrized by W_1 (cf. [46, Theorem 6.9]). We impose the following continuity and contractivity at infinity conditions on the drift.

(2-i) $b(x, \mu)$ is continuous on $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ in the product topology, and there exist constants K_1 , $l_0 \geq 0$ and $K_2, K_3 > 0$ such that for any $x_1, x_2 \in \mathbb{R}^d$ with $x_1 \neq x_2$ and any $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$,

$$(1.13) \quad \frac{\langle b(x_1, \mu_1) - b(x_2, \mu_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \leq K_1 |x_1 - x_2| \mathbb{1}_{\{|x_1 - x_2| \leq l_0\}} - K_2 |x_1 - x_2| \mathbb{1}_{\{|x_1 - x_2| > l_0\}} + K_3 W_1(\mu_1, \mu_2).$$

Furthermore, there is a constant $C_1 > 0$ such that for all $\mu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$(1.14) \quad |b(0, \mu)| \leq C_1 \left(1 + \int_{\mathbb{R}^d} |z| \mu(dz) \right).$$

For the Lévy measure ν , as in condition (1.12), we assume

(2-ii) For the function $J(r)$ given by (1.5), there exists a constant $\alpha \in (0, 1)$ such that

$$(1.15) \quad \lim_{r \rightarrow 0} \inf_{s \in (0, r]} J(s) s^\alpha > 0;$$

moreover,

$$\int_{\{|z|>1\}} |z| \nu(dz) < \infty.$$

We have the following result on convergence of the solution of (1.7) to a stationary distribution.

Theorem 1.4. *Suppose that assumptions (2-i)–(2-ii) hold. Let μ_t denote the marginal law of a strong solution $(X_t)_{t \geq 0}$ of (1.7) with initial distribution μ_0 having finite first moment. Then there exists a constant $K_3^* > 0$ such that for any $K_3 \in (0, K_3^*]$, the associated McKean-Vlasov equation (1.7) has a unique invariant probability measure μ with finite first moment such that*

$$W_1(\mu_t, \mu) \leq C e^{-\lambda t}, \quad t > 0$$

for some positive constants λ and C independent of t .

Note that here, in contrast to condition (1-iii) required for Theorem 1.2, we only assume that the Lévy measure ν has a finite first (and not necessarily second) moment. Taking into account Remark 1.3, we see that our result covers McKean-Vlasov SDEs driven e.g. by nonsymmetric α -stable processes with $\alpha \in (1, 2)$, but also a much larger class of Lévy processes. Note also that existence of an invariant measure in Theorem 1.4 essentially requires the constant K_3 in (1.13) to be sufficiently small, and that condition (1.14) is only used to ensure existence of a non-explosive solution to (1.7).

Recently, Y. Song in [42] applied Malliavin calculus to obtain exponential ergodicity for McKean-Vlasov equations with Lévy jumps in the total variation distance, under an assumption similar to (1.13) but with $l_0 = 0$ and some additional regularity assumptions on the Lévy measure, see [42, Theorem 1.5]. For the proof of Theorem 1.4 we use different methods, and we are able to obtain exponential ergodicity in the L^1 -Wasserstein distance under relaxed assumptions, with an arbitrary $l_0 \geq 0$.

The assumption (2-i) on the drift $b(x, \mu)$ in Theorem 1.4 can be further weakened, if we assume that $b(x, \mu)$ is of the special form (1.8). Namely, we assume

(3-i) $b(x, \mu)$ satisfies (1.8) and there exists a constant $\lambda > 0$ such that

$$(1.16) \quad \langle b_1(x), x \rangle \leq -\lambda |x|^2$$

for all $x \in \mathbb{R}^d$ with large enough $|x|$. Moreover, there are constants K_1, K_2 and $K_3 > 0$ such that for all $x_1, x_2 \in \mathbb{R}^d$ with $x_1 \neq x_2$ and $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\frac{\langle b_1(x_1) - b_1(x_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \leq K_1 |x_1 - x_2|$$

and

$$(1.17) \quad \frac{\langle b_2(x_1, \mu_1) - b_2(x_2, \mu_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \leq K_2 |x_1 - x_2| + K_3 W_1(\mu_1, \mu_2).$$

There is also a constant $B_0 > 0$ such that for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$(1.18) \quad |b_2(x, \mu)| \leq B_0 \left(1 + \int |z| \mu(dz) + |x| \right).$$

Furthermore, condition (1.15) on the concentration of the Lévy measure ν around zero can be only required to hold for a component of ν .

(3-ii) We have $\int_{\{|z|>1\}} |z| \nu(dz) < \infty$ and for any $\theta > 0$, there exists a measure $0 < \nu_\theta \leq \nu$ such that $\text{supp } \nu_\theta \subset B(0, 1)$, $\int_{\{|z|\leq 1\}} |z| \nu_\theta(dz) \leq \theta$, and

$$(1.19) \quad \lim_{r \rightarrow 0} \inf_{s \in (0, r]} J_{\nu_\theta}(s) s^\alpha > 0,$$

where $\alpha := \alpha(\theta) \in (0, 1)$ and

$$J_{\nu_\theta}(s) := \inf_{x \in \mathbb{R}^d: |x| \leq s} [\nu_\theta \wedge (\delta_x * \nu_\theta)](\mathbb{R}^d) > 0.$$

Then we have the following result.

Theorem 1.5. *Suppose that assumptions (3-i)–(3-ii) hold. Let μ_t denote the marginal law of a strong solution $(X_t)_{t \geq 0}$ of (1.7) with initial distribution μ_0 having finite first moment. Then there exist positive constants K_2^*, K_3^*, B_0^* such that for any $K_2 \in (0, K_2^*]$, $K_3 \in (0, K_3^*]$ and $B_0 \in (0, B_0^*]$, the associated McKean-Vlasov equation (1.7) has a unique invariant probability measure μ with finite first moment such that*

$$W_1(\mu_t, \mu) \leq C e^{-\lambda t}, \quad t > 0$$

for some positive constants λ and C independent of t .

At the end of this section, we give some specific examples of drift terms that satisfy the corresponding assumptions.

(1) Each $b_1(x)$ below satisfies (1.10) and (1.16).

(i) Let $b_1(x) = -\lambda x + U(x)$, where $U : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is bounded.

(ii) Let $b_1(x) = -\nabla V(x) + U(x)$, where $V(x) = |x|^{2\beta}$ with $\beta > 1$ and $U : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is Lipschitz continuous, i.e., there is a constant $c > 0$ such that for all $x, y \in \mathbb{R}^d$, $|U(x) - U(y)| \leq c|x - y|$.

It is obvious that both b_1 above satisfy (1.16). For b_1 given in (i), we have

$$\langle b_1(x) - b_1(y), x - y \rangle \leq -\lambda|x - y|^2 + \|U\|_\infty|x - y|, \quad x, y \in \mathbb{R}^d,$$

which implies that it fulfills (1.10). For $b_0(x) = -\nabla V(x)$ with $V(x) = |x|^{2\beta}$ for some $\beta > 1$, it has been proven in [10, Section 6, Example 1] that for any $x, y \in \mathbb{R}^d$,

$$\langle b_0(x) - b_0(y), x - y \rangle \leq -2\beta 2^{3(1-\beta)}|x - y|^{2\beta}.$$

With this at hand, we can see that b_1 given in (ii) also fulfills (1.10).

(2) For any $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, define $b_2(x, \mu) = \int_{\mathbb{R}^d} \bar{b}_2(x, z) \mu(dz)$, where $\bar{b}_2(x, z)$ satisfies that

$$|\bar{b}_2(x, z) - \bar{b}_2(y, z')| \leq K_{b_2}(|x - y| + |z - z'|)$$

holds for some constant $K_{b_2} > 0$ and for all x, y, z and $z' \in \mathbb{R}^d$. Then we can verify that (1) and (1.18) also hold true; see Remark 3.4 for more details.

Combining $b_1(x)$ and $b_2(x, \mu)$ above, one can obtain multiple examples of $b(x, \mu)$ which satisfy the corresponding assumptions in Theorems 1.2, 1.4 and 1.5.

The remaining part of the paper is organized as follows. In Section 2 we first present a general framework for studying Lévy-driven SDEs with distribution independent drifts via the coupling operator. Then we apply techniques based on

couplings and Lyapunov functions to prove results on convergence to stationary distributions for such equations, including Theorem 1.1. In particular, in Subsection 2.1, we also introduce a new coupling for SDEs with distribution independent drifts by combining the refined basic coupling with the synchronous coupling, which is crucial in the study of the exponential ergodicity for the SDE (1.7) with a distribution-dependent drift. In Section 3 we first prove a slightly more general version of Theorem 1.4 and then we discuss its further extensions using the Lyapunov function – based on methods of Section 2, thus proving Theorem 1.5. Finally, in Section 4 we prove Theorem 1.2.

2. EXPONENTIAL CONVERGENCE FOR SDEs WITH LÉVY NOISE

In this section, we will study the SDE (1.7) with distribution-independent drift, i.e.,

$$(2.1) \quad dX_t = b(X_t) dt + dZ_t, \quad X_0 = x \in \mathbb{R}^d,$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a measurable function, and $Z = (Z_t)_{t \geq 0}$ is a pure jump Lévy process on \mathbb{R}^d . Denote by ν the Lévy measure of the process Z . We always assume that *the SDE (2.1) has a unique strong solution*, which is true, for example, for any noise Z if the drift $b(x)$ is continuous and satisfies a one-sided Lipschitz condition, see [18, Theorem 2], or if $b(x)$ is Hölder continuous and Z is a Lévy process whose Lévy measure satisfies some integrability conditions at zero and at infinity and whose transition semigroup enjoys certain regularity properties, see e.g. [13, 29, 41, 51].

2.1. Coupling approach to exponential convergence: general framework.

In this part, we present a general approach, based on the probabilistic coupling method, to study the exponential convergence of the SDE given by (2.1). Let $X := (X_t)_{t \geq 0}$ be the (unique) strong solution to the SDE (2.1). Then, the infinitesimal generator of X acting on $C_b^2(\mathbb{R}^d)$ is given by

$$(2.2) \quad Lf(x) = \int (f(x+z) - f(x) - \langle \nabla f(x), z \rangle \mathbb{1}_{\{|z| \leq 1\}}) \nu(dz) + \langle b(x), \nabla f(x) \rangle.$$

To consider couplings of X , we will make use of the coupling operator of the generator L . A linear operator $\tilde{L} : C_b^2(\mathbb{R}^{2d}) \rightarrow B_b(\mathbb{R}^d)$ is called a coupling operator of the SDE given by (2.1) (or of the generator L given by (2.2)), if

$$\tilde{L}(f \otimes 1)(x, y) = Lf(x), \quad \tilde{L}(1 \otimes g)(x, y) = Lg(y), \quad f, g \in C_b^2(\mathbb{R}^d),$$

where $(f \otimes g)(x, y) = f(x)g(y)$ denotes the tensor product of two functions f and g . If the operator \tilde{L} can generate a Markov process $(X_t, Y_t)_{t \geq 0}$ on \mathbb{R}^{2d} , then we call $(X_t, Y_t)_{t \geq 0}$ a Markovian coupling process of the process X determined by (2.1). In this case, \tilde{L} is called a Markovian coupling operator of X (or L). In particular, let $(\tilde{P}_t)_{t \geq 0}$ be the semigroup of the Markovian coupling process $(X_t, Y_t)_{t \geq 0}$, and $(P_t)_{t \geq 0}$ be the semigroup of the process X . Then, for any $f, g \in C_b^2(\mathbb{R}^d)$ and $t > 0$,

$$\tilde{P}_t(f \otimes 1)(x, y) = P_t f(x), \quad \tilde{P}_t(1 \otimes g)(x, y) = P_t g(y).$$

Let Φ be a function on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\Phi(0, 0) = 0$ and Φ is strictly positive elsewhere. Given two probability measures μ_1 and μ_2 on \mathbb{R}^d , we define the following quantity (which can be called a Wasserstein-type distance or a Kantorovich distance)

$$W_\Phi(\mu_1, \mu_2) = \inf_{\Pi \in \mathcal{C}(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi(x, y) d\Pi(x, y),$$

where $\mathcal{C}(\mu_1, \mu_2)$ is the collection of all measures on $\mathbb{R}^d \times \mathbb{R}^d$ having μ_1 and μ_2 as marginals. In particular, when $\Phi(x, y) = |x - y|$, W_Φ is just the standard L^1 -Wasserstein distance, which is simply denoted by W_1 in the following; on the other hand, when $\Phi(x, y) = \mathbb{1}_{\{x \neq y\}}$, W_Φ leads to the total variation distance $W_\Phi(\mu_1, \mu_2) = \frac{1}{2} \|\mu_1 - \mu_2\|_{\text{Var}}$.

The following statement provides a general tool for showing exponential convergence in Wasserstein-type distances via the coupling method.

Proposition 2.1. *Suppose that \tilde{L} is a coupling operator for the SDE given by (2.1), generating a non-explosive Markovian coupling process $(X_t, Y_t)_{t \geq 0}$ such that $X_t = Y_t$ for all $t \geq T$, where $T := \inf\{t \geq 0 : X_t = Y_t\}$ is the coupling time of the process $(X_t, Y_t)_{t \geq 0}$. Assume that there exist a constant $\lambda > 0$ and a sequence of non-negative functions $\{\Phi_n(x, y)\}_{n \geq 1}$ on \mathbb{R}^{2d} such that for any $n \geq 1$, $\Phi_n(x, x) = 0$ for all $x \in \mathbb{R}^d$, $\Phi_n(x, y) > 0$ for all $x \neq y \in \mathbb{R}^d$, and $\tilde{L}\Phi_n(x, y)$ is pointwise well defined, and such that for $n \geq 1$ large enough and for all $x, y \in \mathbb{R}^d$ with $1/n \leq |x - y| \leq n$,*

$$(2.3) \quad \tilde{L}\Phi_n(x, y) \leq -\lambda\Phi_n(x, y).$$

Then for any $t > 0$ and $x, y \in \mathbb{R}^d$,

$$W_{\Phi_\infty}(P_t(x, \cdot), P_t(y, \cdot)) \leq \Phi_\infty(x, y)e^{-\lambda t},$$

where $\Phi_\infty = \liminf_{n \rightarrow \infty} \Phi_n$ and $P_t(x, \cdot)$ is the transition probability of the process $(X_t)_{t \geq 0}$ solving the SDE (2.1).

Proof. The proof follows step 2 of the proof of [33, Theorem 3.1] with some modifications. For the sake of completeness, we present all the details here. Let $(X_t, Y_t)_{t \geq 0}$ be the Markovian coupling process corresponding to the coupling operator \tilde{L} in the statement. To prove the desired assertion, it is enough to verify that for $x, y \in \mathbb{R}^d$ with $|x - y| > 0$ and any $t > 0$,

$$\tilde{\mathbb{E}}^{(x, y)} \Phi_\infty(X_t, Y_t) \leq \Phi_\infty(x, y)e^{-\lambda t},$$

where $\tilde{\mathbb{E}}^{(x, y)}$ is the expectation of $(X_t, Y_t)_{t \geq 0}$ starting from (x, y) .

For $n \geq 1$ define the stopping time

$$T_n = \inf\{t > 0 : |X_t - Y_t| \notin [1/n, n]\}.$$

Since the coupling process $(X_t, Y_t)_{t \geq 0}$ is non-explosive, we have $T_n \uparrow T$ a.s. as $n \rightarrow \infty$. For any $x, y \in \mathbb{R}^d$ with $|x - y| > 0$, we take $n \geq 1$ large enough such that $1/n < |x - y| < n$. Let $\{\Phi_n\}_{n \geq 1}$ be the sequence of non-negative functions, and λ be the constant given in the statement. Then, according to (2.3), for any $m \geq n$,

$$\begin{aligned} & \tilde{\mathbb{E}}^{(x, y)} [e^{\lambda(t \wedge T_n)} \Phi_m(X_{t \wedge T_n}, Y_{t \wedge T_n})] \\ &= \Phi_m(x, y) + \tilde{\mathbb{E}}^{(x, y)} \left(\int_0^{t \wedge T_n} e^{\lambda s} [\lambda \Phi_m(X_s, Y_s) + \tilde{L}\Phi_m(X_s, Y_s)] ds \right) \\ &\leq \Phi_m(x, y). \end{aligned}$$

Thus by Fatou's lemma, first letting $m \rightarrow \infty$ and then $n \rightarrow \infty$ in the above inequality, we find that

$$\tilde{\mathbb{E}}^{(x, y)} (e^{\lambda(t \wedge T)} \Phi_\infty(X_{t \wedge T}, Y_{t \wedge T})) \leq \Phi_\infty(x, y).$$

Thanks to the facts that $Y_t = X_t$ for $t \geq T$ and $\Phi_\infty(x, x) = \liminf_{n \rightarrow \infty} \Phi_n(x, x) = 0$ for all $x \in \mathbb{R}^d$, we have $\Phi_\infty(X_t, Y_t) = 0$ for all $t \geq T$, which implies

$$\tilde{\mathbb{E}}^{(x,y)}(e^{\lambda(t \wedge T)} \Phi_\infty(X_{t \wedge T}, Y_{t \wedge T})) = e^{\lambda t} \tilde{\mathbb{E}}^{(x,y)}(\Phi_\infty(X_t, Y_t) \mathbf{1}_{\{T > t\}}) = e^{\lambda t} \tilde{\mathbb{E}}^{(x,y)} \Phi_\infty(X_t, Y_t).$$

This completes the proof. \square

To apply Proposition 2.1, we need to construct a coupling operator \tilde{L} and a sequence of non-negative functions $\{\Phi_n(x, y)\}_{n \geq 1}$ on \mathbb{R}^{2d} which are strictly positive outside the diagonal such that (2.3) is satisfied. For the exponential convergence in terms of a total variation-type distance, we will take $\Phi_n(x, y) \in C^2(\mathbb{R}^{2d})$ such that $\Phi_n(x, x) = 0$ and

$$(2.4) \quad \Phi_n(x, y) = \psi_n(|x - y|) + \varepsilon(V(x) + V(y)), \quad |x - y| \geq 1/n,$$

where $\varepsilon > 0$, $V \in C^2(\mathbb{R}^d)$ is a Lyapunov function and $\{\psi_n(|x - y|)\}_{n \geq 1}$ is a sequence of functions approximating $a + \psi(|x - y|)$ for some constant $a > 0$ and a bounded concave function ψ on $[0, \infty)$ with $\psi(0) = 0$. It is obvious that $\lim_{n \rightarrow \infty} \Phi_n(x, y)$ is comparable to $(V(x) + V(y)) \mathbf{1}_{\{x \neq y\}}$ and so, with this choice of Φ_n , Proposition 2.1 could yield the exponential convergence in terms of the V -variation norm given for any probability measures μ_1 and μ_2 as $\|\mu_1 - \mu_2\|_{\text{var}, V} := \sup_{|f| \leq V} |\mu_1(f) - \mu_2(f)|$. The details will be provided in Subsection 2.2. As pointed out at the beginning of [16, Subsection 2.1], the choice of those additive metrics is partially motivated by the paper [20]. Concerning the exponential convergence in terms of L^1 -Wasserstein-type distances, we will take

$$(2.5) \quad \Phi_n(x, y) = \Phi(x, y) := \psi(|x - y|)(1 + \varepsilon(V(x) + V(y)))$$

for all $n \geq 1$ and $x, y \in \mathbb{R}^d$, with a bounded concave function ψ on $[0, \infty)$ satisfying $\psi(0) = 0$ and a Lyapunov function $V \in C^2(\mathbb{R}^d)$. Note that $\lim_{|x-y| \rightarrow 0} \Phi_n(x, y) = 0$, while the function Φ_n given by (2.4), associated with the variation-type distance above, is discontinuous near the diagonal. The details of the construction involving (2.5) will be provided in Subsection 2.3.

In order to construct the operator \tilde{L} , we will adopt the refined basic coupling for the SDE (2.1) from [33]. To describe the refined basic coupling, we start with a coupling for the Lévy measure ν . To this end, we introduce the notation

$$x \rightarrow x + z, \quad \nu(dz)$$

for a transition from a point $x \in \mathbb{R}^d$ to the point $x + z$, with the jump intensity $\nu(dz)$. Roughly speaking, the essential idea of the basic coupling is to make the two marginal processes jump to the same point with the biggest possible rate, where the biggest jump rate is the maximal common part of the jump intensities. In the Lévy setting, it takes the form

$$\mu_{y-x}(dz) := [\nu \wedge (\delta_{y-x} * \nu)](dz),$$

where x and y correspond to the positions of the two marginal processes before the jump. Note that for $x \neq 0$,

$$(2.6) \quad \mu_x(\mathbb{R}^d) \leq \int_{\{|z| \leq |x|/2\}} (\delta_x * \nu)(dz) + \int_{\{|z| > |x|/2\}} \nu(dz) \leq 2 \int_{\{|z| \geq |x|/2\}} \nu(dz) < \infty,$$

i.e., μ_x is a finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ for any $x \neq 0$. Let κ_0 be a fixed constant. For any $x, y \in \mathbb{R}^d$ and $\kappa \in (0, \kappa_0]$, define $(x - y)_\kappa = (1 \wedge \frac{\kappa}{|x-y|})(x - y)$. We use the

convention that $(x - x)_\kappa = 0$. Then the refined basic coupling of L , first introduced in [33, Section 2.1], is given as follows:

$$(2.7) \quad (x, y) \longrightarrow \begin{cases} (x + z, y + z + (x - y)_\kappa), & \frac{1}{2}\mu_{(y-x)_\kappa}(dz), \\ (x + z, y + z + (y - x)_\kappa), & \frac{1}{2}\mu_{(x-y)_\kappa}(dz), \\ (x + z, y + z), & \left(\nu - \frac{1}{2}\mu_{(y-x)_\kappa} - \frac{1}{2}\mu_{(x-y)_\kappa}\right)(dz). \end{cases}$$

We see that if $|x - y| \leq \kappa$, then (2.7) is reduced into

$$(x, y) \longrightarrow \begin{cases} (x + z, y + z + (x - y)), & \frac{1}{2}\mu_{y-x}(dz), \\ (x + z, y + z + (y - x)), & \frac{1}{2}\mu_{x-y}(dz), \\ (x + z, y + z), & \left(\nu - \frac{1}{2}\mu_{y-x} - \frac{1}{2}\mu_{x-y}\right)(dz). \end{cases}$$

The first row in the coupling above corresponds to the two marginal processes jumping to the same point. Note that the distance between the two marginals decreases from $|x - y|$ to $|(x + z) - (y + z + (x - y))| = 0$; this is indeed the idea of the *basic coupling* for Markov q -processes in [12, Example 2.10]. However, unlike in the basic coupling, here we send the marginal processes to the same point only with half the maximal possible probability, since we want to apply the synchronous (rather than the independent) coupling to the remaining jumps (see [33, Section 2.1] for more details). Hence we also need the second row, which corresponds to the change of the distance from $|x - y|$ to $2|x - y|$, and the last row, which is just a synchronous movement. If $|x - y| > \kappa$, then according to the first two rows in (2.7), the distances after the jump are $|x - y| - \kappa$ and $|x - y| + \kappa$, respectively. Hence the marginal processes can jump to the same point only if they are already close to each other before the jump (based on the threshold parameter $\kappa > 0$). Otherwise, they can only move slightly closer towards each other. Note that the introduction of κ prevents a situation in which the two marginal processes could never couple if they only have finite range jumps (i.e., the sizes of jumps are bounded). Since this construction can be interpreted as a modification of the basic coupling from [12, Example 2.10], we call the coupling given by (2.7) the *refined basic coupling* for pure jump Lévy processes.

Motivated by the intuitive description above, we define for any $F \in C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$,

$$(2.8) \quad \begin{aligned} & \tilde{L}F(x, y) \\ &= \frac{1}{2} \int \left(F(x + z, y + z + (x - y)_\kappa) - F(x, y) - \langle \nabla_x F(x, y), z \rangle \mathbf{1}_{\{|z| \leq 1\}} \right. \\ & \quad \left. - \langle \nabla_y F(x, y), z + (x - y)_\kappa \rangle \mathbf{1}_{\{|z + (x - y)_\kappa| \leq 1\}} \right) \mu_{(y-x)_\kappa}(dz) \\ & + \frac{1}{2} \int \left(F(x + z, y + z + (y - x)_\kappa) - F(x, y) - \langle \nabla_x F(x, y), z \rangle \mathbf{1}_{\{|z| \leq 1\}} \right. \\ & \quad \left. - \langle \nabla_y F(x, y), z + (y - x)_\kappa \rangle \mathbf{1}_{\{|z + (y - x)_\kappa| \leq 1\}} \right) \mu_{(x-y)_\kappa}(dz) \\ & + \int \left(F(x + z, y + z) - F(x, y) - \langle \nabla_x F(x, y), z \rangle \mathbf{1}_{\{|z| \leq 1\}} \right. \\ & \quad \left. - \langle \nabla_y F(x, y), z \rangle \mathbf{1}_{\{|z| \leq 1\}} \right) \left(\nu - \frac{1}{2}\mu_{(y-x)_\kappa} - \frac{1}{2}\mu_{(x-y)_\kappa} \right)(dz) \\ & + \langle \nabla_x F(x, y), b(x) \rangle + \langle \nabla_y F(x, y), b(y) \rangle. \end{aligned}$$

Due to the facts that μ_x is a finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ for all $x \neq 0$ and $\delta_x * \mu_{-x} = \mu_x$ (see [33, Corollary A.2]), we can check that \tilde{L} defined by (2.8) is a coupling operator for the generator L given by (2.2).

Below we briefly show that the coupling operator \tilde{L} generates a Markovian coupling process, which can be obtained as the unique strong solution to an SDE on \mathbb{R}^{2d} . First, by the Lévy-Itô decomposition, there exists a Poisson random measure N associated with $(Z_t)_{t \geq 0}$ such that

$$dZ_t = \int_{\{|z|>1\}} z N(dt, dz) + \int_{\{|z|\leq 1\}} z \tilde{N}(dt, dz),$$

where $\tilde{N}(dt, dz) = N(dt, dz) - dt \nu(dz)$ is the compensated Poisson random measure. Following the ideas from [34, Section 2.2] and [33, Section 2.2], we extend the Poisson random measure N from $\mathbb{R}_+ \times \mathbb{R}^d$ to $\mathbb{R}_+ \times \mathbb{R}^d \times [0, 1]$ in the following way

$$N(ds, dz, du) = \sum_{\{0 < s' \leq s, \Delta Z_{s'} \neq 0\}} \delta_{(s', \Delta Z_{s'})}(ds, dz) \mathbf{1}_{[0,1]}(du)$$

and we write

$$Z_t = \int_0^t \int_{\mathbb{R}^d \times [0,1]} z \bar{N}(ds, dz, du),$$

where

$$\bar{N}(ds, dz, du) = \mathbf{1}_{\{|z|>1\} \times [0,1]} N(ds, dz, du) + \mathbf{1}_{\{|z|\leq 1\} \times [0,1]} \tilde{N}(ds, dz, du).$$

We further define the control function ρ as follows: for any $x, z \in \mathbb{R}^d$,

$$\rho(x, z) = \frac{\mu_x(dz)}{\nu(dz)} = \frac{\nu \wedge (\delta_x * \nu)(dz)}{\nu(dz)} \in [0, 1].$$

Recall that for any $x \neq 0$, $(x)_\kappa = (1 \wedge (\kappa/|x|))x$. Fix any $x, y \in \mathbb{R}^d$ with $x \neq y$. We consider the system of equations:

$$(2.9) \quad \begin{cases} dX_t = b(X_t) dt + dZ_t, & X_0 = x, \\ dY_t = b(Y_t) dt + dZ_t + dL_t^*, & Y_0 = y, \end{cases}$$

where

$$dL_t^* = \int_{\mathbb{R}^d \times [0,1]} S(U_{t-}, z, u) N(dt, dz, du)$$

with $U_t = X_t - Y_t$ and

$$S(U_t, z, u) = (U_t)_\kappa \left[\mathbf{1}_{\{u \leq \frac{1}{2}\rho((-U_t)_\kappa, z)\}} - \mathbf{1}_{\{\frac{1}{2}\rho((-U_t)_\kappa, z) < u \leq \frac{1}{2}[\rho((-U_t)_\kappa, z) + \rho((U_t)_\kappa, z)]\}} \right].$$

According to [33, Propositions 2.2 and 2.3], the SDE (2.9) has a unique strong solution, which is a non-explosive coupling process $(X_t, Y_t)_{t \geq 0}$ of the SDE (2.1). Since we assume that (2.1) (i.e., the first equation in (2.9)) has a non-explosive and pathwise unique strong solution $(X_t)_{t \geq 0}$, the sample paths of $(Y_t)_{t \geq 0}$ can be obtained by repeatedly modifying those of the unique strong solution of the following equation:

$$d\tilde{Y}_t = b(\tilde{Y}_t) dt + dZ_t, \quad Y_0 = y$$

with aid of the so-called interlacing technique. Hence the uniqueness of the strong solution to (2.9) immediately follows; see the proof of [33, Proposition 2.2]. Note also that in this argument we do not need to assume that $b(x)$ is locally Lipschitz continuous if the SDE (2.1) has a unique strong solution (which can be the case for

non-locally Lipschitz drifts too, see e.g. [13, 29, 41, 51]). Moreover, the generator of $(X_t, Y_t)_{t \geq 0}$ is the refined basic coupling operator \tilde{L} defined by (2.8), and $X_t = Y_t$ for any $t \geq T$, where $T = \inf\{t \geq 0 : X_t = Y_t\}$ is the coupling time of the process $(X_t, Y_t)_{t \geq 0}$. In particular, the refined basic coupling \tilde{L} generates a Markovian coupling process $(X_t, Y_t)_{t \geq 0}$ which satisfies the assumptions in Proposition 2.1.

For the later use, we further show two properties for the coupling operator \tilde{L} and the coupling process $(X_t, Y_t)_{t \geq 0}$. First, due to (2.8) and the fact (see [33, Corollary A.2] again) that $\delta_x * \mu_{-x} = \mu_x$, it holds that, for any $\psi \in C_b^1([0, \infty))$ and $x, y \in \mathbb{R}^d$ with $x \neq y$,

$$\begin{aligned}
 (2.10) \quad & \tilde{L}\psi(|x - y|) \\
 &= \frac{1}{2} \mu_{(x-y)\kappa}(\mathbb{R}^d) \left[\psi(|x - y| + \kappa \wedge |x - y|) + \psi(|x - y| - \kappa \wedge |x - y|) \right. \\
 & \quad \left. - 2\psi(|x - y|) \right] + \frac{\psi'(|x - y|)}{|x - y|} \langle b(x) - b(y), x - y \rangle.
 \end{aligned}$$

Second, let $(X_t, Y_t)_{t \geq 0}$ be the coupling process constructed above. Recall that for any $t \geq 0$, $U_t = X_t - Y_t$. Then, it follows from the system (2.9) that

$$dU_t = (b(X_t) - b(Y_t)) dt - \int_{\mathbb{R}^d \times [0, 1]} S(U_{t-}, z, u) N(dt, dz, du).$$

Take $\psi \in C_b^1([0, \infty))$ with $\psi \geq 0$. By the Itô formula,

$$\begin{aligned}
 (2.11) \quad d\psi(|U_t|) &= \frac{\psi'(|U_t|)}{|U_t|} \langle U_t, b(X_t) - b(Y_t) \rangle dt \\
 & \quad + \int_{\mathbb{R}^d \times [0, 1]} (\psi(|U_{t-} - S(U_{t-}, z, u)|) - \psi(|U_{t-}|)) N(dt, dz, du) \\
 &= \frac{\psi'(|U_t|)}{|U_t|} \langle U_t, b(X_t) - b(Y_t) \rangle dt \\
 & \quad + \int_{\mathbb{R}^d \times [0, 1]} (\psi(|U_{t-} - S(U_{t-}, z, u)|) - \psi(|U_{t-}|)) \nu(dz) du dt \\
 & \quad + \int_{\mathbb{R}^d \times [0, 1]} (\psi(|U_{t-} - S(U_{t-}, z, u)|) - \psi(|U_{t-}|)) \tilde{N}(dt, dz, du).
 \end{aligned}$$

This along with (2.10) implies that we can rewrite (2.11) as follows

$$d\psi(|U_t|) = \tilde{L}\psi(|U_t|) dt + dM_t^\psi,$$

where

$$dM_t^\psi = \int_{\mathbb{R}^d \times [0, 1]} (\psi(|U_{t-} - S(U_{t-}, z, u)|) - \psi(|U_{t-}|)) \tilde{N}(dt, dz, du)$$

is a local martingale. We also note that (2.10) can be deduced from (2.11) directly. In the sequel the coupling operator for the Lévy process Z without a drift (i.e., with $b(x) = 0$ for all $x \in \mathbb{R}^d$ in (2.1)) will be denoted by \tilde{L}_Z .

In order to deal with the SDE (1.7) with distribution-dependent drift in Section 3, we will need to introduce a new coupling for (2.1) by combining the refined

basic coupling with the synchronous coupling. To this end, for any fixed $\delta > 0$, let $\phi_\delta : [0, \infty) \rightarrow [0, 1]$ be a smooth function such that

$$(2.12) \quad \phi_\delta(r) = \begin{cases} = 0, & 0 \leq r \leq \delta/2, \\ \in [0, 1], & \delta/2 \leq r \leq \delta, \\ = 1, & r \geq \delta. \end{cases}$$

We then define a coupling operator \tilde{L}_Z^δ given for any $F \in C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$ by

$$(2.13) \quad \tilde{L}_Z^\delta F(x, y) = \phi_\delta(|x - y|) \cdot \tilde{L}_Z F(x, y) + (1 - \phi_\delta(|x - y|)) \cdot \tilde{L}_{Z,*} F(x, y),$$

where $\tilde{L}_{Z,*}$ denotes the synchronous coupling operator for the Lévy process Z with the Lévy measure ν , i.e., for any $F \in C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\begin{aligned} \tilde{L}_{Z,*} F(x, y) = \int & \left(F(x + z, y + z) - F(x, y) - \langle \nabla_x F(x, y), z \rangle \mathbf{1}_{\{|z| \leq 1\}} \right. \\ & \left. - \langle \nabla_y F(x, y), z \rangle \mathbf{1}_{\{|z| \leq 1\}} \right) \nu(dz). \end{aligned}$$

Similarly as for \tilde{L}_Z , we can prove that \tilde{L}_Z^δ defined this way is indeed a coupling operator for the Lévy process Z , cf. [33, Section 2]. The coupling works as the refined basic coupling when the distance between the marginal processes is larger than δ , is the synchronous coupling when the distance between the marginal processes is smaller than $\delta/2$, and is a mixture of the refined basic coupling and the synchronous coupling in the remaining case. We can construct a coupling process $(X_t, Y_t^\delta)_{t \geq 0}$ with the generator \tilde{L}^δ defined for $F \in C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$ as

$$\tilde{L}^\delta F(x, y) = \tilde{L}_Z^\delta F(x, y) + \langle \nabla_x F(x, y), b(x) \rangle + \langle \nabla_y F(x, y), b(y) \rangle$$

exactly as described above by considering the system of SDEs

$$\begin{cases} dX_t = b(X_t) dt + dZ_t, & X_0 = x, \\ dY_t^\delta = b(Y_t^\delta) dt + dZ_t + dL_t^{*,\delta}, & Y_0^\delta = y, \end{cases}$$

where

$$(2.14) \quad dL_t^{*,\delta} = \int_{\mathbb{R}^d \times [0,1]} S^\delta(U_{t-}^\delta, z, u) N(dt, dz, du)$$

with

$$(2.15) \quad \begin{aligned} S^\delta(U_t^\delta, z, u) = & (U_t^\delta)_\kappa \left[\mathbf{1}_{\{u \leq \frac{1}{2}\rho((-U_t^\delta)_\kappa, z)\}} \right. \\ & \left. - \mathbf{1}_{\{\frac{1}{2}\rho((-U_t^\delta)_\kappa, z) < u \leq \frac{1}{2}[\rho((-U_t^\delta)_\kappa, z) + \rho((U_t^\delta)_\kappa, z)]\}} \right] \phi_\delta(|U_t^\delta|) \end{aligned}$$

and $U_t^\delta = X_t - Y_t^\delta$.

In the next two subsections, we will apply the refined basic coupling operator \tilde{L} to study the exponential ergodicity for the SDE (2.1). Hence, according to Proposition 2.1, the main task is to construct an appropriate sequence of non-negative functions $\{\Phi_n(x, y)\}_{n \geq 1}$ such that (2.3) is satisfied. By the remarks below the proof of Proposition 2.1, in Subsections 2.2 and 2.3 we will consider convergence in terms of additive distance and multiplicative distance, respectively.

2.2. Exponential ergodicity for the SDE (2.1): additive distance. In this part, we assume the following conditions.

Assumption (A)

(i) *There is a constant $K_1 > 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\langle b(x) - b(y), x - y \rangle \leq K_1(|x - y|^2 \vee |x - y|).$$

(ii) *There is a constant $0 < \kappa_0 \leq 1$ such that*

$$(2.16) \quad J(\kappa_0) := \inf_{x \in \mathbb{R}^d: |x| \leq \kappa_0} [\nu \wedge (\delta_x * \nu)](\mathbb{R}^d) > 0.$$

(iii) *There are a C^2 -function $V : \mathbb{R}^d \rightarrow [1, \infty)$ and constants $C, \lambda > 0$ such that $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$, and*

$$(2.17) \quad LV(x) \leq C - \lambda V(x), \quad x \in \mathbb{R}^d,$$

where L is the infinitesimal generator given by (2.2).

We emphasise that in this section we always assume that the SDE (2.1) has a unique strong solution. Note that condition (i) is weaker than the standard one-side Lipschitz condition (see Assumption (B) (i) below). In particular, it is easy to see that condition (i) is satisfied for any bounded measurable function b . Note that in the one-dimensional case the SDE (2.1) driven by a symmetric α -stable process with $\alpha \in (1, 2)$ has a strong solution, even when $b(x)$ is only bounded measurable; see [45]. However, in the general case some additional properties of the drift and the driving noise may be required in order to guarantee existence of a unique strong solution to (2.1), as discussed at the beginning of this section. Moreover, in order to construct a Lyapunov function V satisfying (2.17) we may have to impose some further assumptions on b and the driving Lévy process Z . For example, if $\int_{\{|z| \geq 1\}} |z| \nu(dz) < \infty$ and there are constants $c_0, l_0 > 0$ such that $\langle b(x), x \rangle \leq -c_0|x|^2$ for all $x \in \mathbb{R}^d$ with $|x| \geq l_0$, then (2.17) holds with a radial function $V \in C^2(\mathbb{R}^d)$ such that $V(x) \geq 1$ for all $x \in \mathbb{R}^d$ and $V(x) = 1 + |x|$ for $|x| \geq 1$, see Lemma 3.5 and Remark 3.6 for details. Alternatively, suppose that $\int_{\{|z| \geq 1\}} |z| \nu(dz) < \infty$ does not hold, but we assume the weaker condition $\int_{\{|z| \geq 1\}} |z|^\alpha \nu(dz) < \infty$ for some $\alpha \in (0, 1)$. In this case, if there are constants $c_0, l_0 > 0$ such that $\langle b(x), x \rangle \leq -c_0|x|^2$ for all $x \in \mathbb{R}^d$ with $|x| \geq l_0$, then (2.17) holds with $V(x) = (1 + |x|^2)^{\beta/2}$ for any $\beta \in (0, \alpha]$, cf. (the proof of) [22, Lemma 3.1].

Theorem 2.2. *Under Assumption (A), there are constants $C_0, \lambda_0 > 0$ such that for all $x, y \in \mathbb{R}^d$ and $t > 0$,*

$$W_\Phi(P_t(x, \cdot), P_t(y, \cdot)) \leq C_0 e^{-\lambda_0 t} \Phi(x, y),$$

where $\Phi(x, y) = (V(x) + V(y))\mathbf{1}_{\{x \neq y\}}$.

As a direct consequence of Theorem 2.2, we have

Corollary 2.3. *Under Assumption (A), the process $(X_t)_{t \geq 0}$ is exponentially ergodic; more explicitly, there are a unique invariant probability measure μ , a constant λ_0 and a measurable function $C_0(x)$ such that for all $x \in \mathbb{R}^d$ and $t > 0$,*

$$\|P_t(x, \cdot) - \mu\|_{\text{Var}, V} \leq C_0(x) e^{-\lambda_0 t},$$

where for any probability measure μ_1 and μ_2 ,

$$\|\mu_1 - \mu_2\|_{\text{Var}, V} = \sup_{|f| \leq V} |\mu_1(f) - \mu_2(f)|.$$

Proof. Assumption **(A)** (iii) is the well-known Foster-Lyapunov type condition in the study of stability of Markov processes, see [40] for more details. In particular, according to (2.17) and [40, Theorem 3.1], we know that $\mathbb{E}^x V(X_t) < \infty$ for all $x \in \mathbb{R}^d$ and $t > 0$. On the other hand, it was proven in [20, Lemma 2.1] that for any probability measures μ_1 and μ_2 ,

$$\|\mu_1 - \mu_2\|_{\text{Var}, V} = W_{\Phi}(\mu_1, \mu_2),$$

where $\Phi(x, y) = (V(x) + V(y))\mathbf{1}_{\{x \neq y\}}$. Then, combining these two conclusions with Theorem 2.2 and some standard arguments (see for example the proof of [34, Corollary 1.8]), we can prove the desired assertion. \square

Remark 2.4. We give some comments on Assumption **(A)** and Theorem 2.2.

- (i) We first make remarks on Assumption **(A)**(i) and (ii) respectively. When $b(x)$ is locally bounded, (i) holds locally, i.e., (i) holds for all $x, y \in B(0, R)$ and $R > 0$, cf. Remark 2.6 on how to extend Theorem 2.2 to hold under a local version of (i). Condition (ii) is very weak and can be true even for finite Lévy measures; see [43, Proposition 1.5] and [34, Remark 1.7].
- (ii) Theorem 2.2 and Corollary 2.3 give us a new way to yield the exponential ergodicity for SDEs with additive Lévy noises. We emphasize that the common approach to ergodicity is based on verifying the irreducibility and the strong Feller property of the associated Markov processes; however, we believe that such approach is not easy to apply only under Assumption **(A)**. We further note that an analogous result for diffusions was shown in [16, Theorem 2.1] under an additional growth condition [16, Assumption 2.3] for the Lyapunov function V (which essentially states that there is a constant $C_0 > 0$ such that $V(x) \geq C_0(1 + |x|)$ for large $|x|$, see [16, Lemma 2.1]); nevertheless, in the present setting we do not require this additional assumption.
- (iii) Explicit estimate for the (rate) constant λ_0 in Theorem 2.2 is available at the end of its proof.

To prove Theorem 2.2, we begin with the following lemma.

Lemma 2.5. *Let $\psi \in C^1([0, \infty))$. Then the following hold.*

- (i) *If ψ' is decreasing, then for any $0 \leq \delta \leq r$,*

$$\psi(r + \delta) + \psi(r - \delta) - 2\psi(r) \leq 0.$$

- (ii) *Suppose that $\psi \in C([0, 2l_0]) \cap C^4((0, 2l_0))$ for some $l_0 > 0$ such that $\psi' > 0$, $\psi'' \leq 0$, $\psi''' \geq 0$ and $\psi^{(4)} \leq 0$ on $(0, 2l_0)$. Then, for any $0 \leq \delta \leq r \leq l_0$,*

$$\psi(r + \delta) + \psi(r - \delta) - 2\psi(r) \leq \psi''(r)\delta^2.$$

Proof. The assertion (i) is trivial if $\delta = 0$, thus we assume $\delta > 0$. By the mean value formula, there exist constants $\xi_1 \in (r, r + \delta)$ and $\xi_2 \in (r - \delta, r)$ such that

$$\psi(r + \delta) - \psi(r) = \psi'(\xi_1)\delta$$

and

$$\psi(r - \delta) - \psi(r) = -\psi'(\xi_2)\delta.$$

Therefore,

$$\psi(r + \delta) + \psi(r - \delta) - 2\psi(r) = (\psi'(\xi_1) - \psi'(\xi_2))\delta \leq 0,$$

since ψ' is decreasing.

To prove (ii), we will still assume $\delta > 0$. Similar to the proof of (i), by the Taylor formula, there exist constants $\xi_1 \in (r, r + \delta)$ and $\xi_2 \in (r - \delta, r)$ such that

$$\begin{aligned}\psi(r + \delta) &= \psi(r) + \psi'(r)\delta + \frac{1}{2}\psi''(r)\delta^2 + \frac{1}{6}\psi'''(\xi_1)\delta^3, \\ \psi(r - \delta) &= \psi(r) - \psi'(r)\delta + \frac{1}{2}\psi''(r)\delta^2 - \frac{1}{6}\psi'''(\xi_2)\delta^3.\end{aligned}$$

Therefore,

$$\psi(r + \delta) + \psi(r - \delta) - 2\psi(r) = \psi''(r)\delta^2 + \frac{\delta^3}{6}[\psi'''(\xi_1) - \psi'''(\xi_2)] \leq \psi''(r)\delta^2,$$

since ψ''' is decreasing due to $\psi^{(4)} \leq 0$. \square

Proof of Theorem 2.2. (i) For any $n \geq 1$, define $\Phi_n(x, y) \in C^2(\mathbb{R}^{2d})$ such that

$$\Phi_n(x, y) \begin{cases} = \psi(|x - y|), & 0 \leq |x - y| \leq 1/(n + 1), \\ \leq a + \psi(|x - y|) + \varepsilon(V(x) + V(y)), & 1/(n + 1) \leq |x - y| \leq 1/n, \\ = a + \psi(|x - y|) + \varepsilon(V(x) + V(y)), & |x - y| \geq 1/n. \end{cases}$$

Here,

$$\psi(r) = \begin{cases} 1 - e^{-cr}, & r \in (0, 2l_0], \\ 1 - e^{-2cl_0} + ce^{-2cl_0} \frac{r-2l_0}{1+r-2l_0}, & r > 2l_0 \end{cases}$$

with

$$l_0 = \sup_{(x,y) \in S_0} |x - y| + 1, \quad c = \frac{4K_1 l_0}{J(\kappa)\kappa^2} + 1$$

and

$$S_0 = \{(x, y) \in \mathbb{R}^{2d} : \lambda(V(x) + V(y)) \leq 16C\};$$

the function V and the constants $\lambda, C, K_1, J(\kappa)$ are given in Assumption **(A)**; and

$$a = \frac{4K_1 c}{J(\kappa)} + \kappa^2 c^2 e^{-cl_0}, \quad \varepsilon = \frac{1}{16C} J(\kappa) \kappa^2 c^2 e^{-cl_0}.$$

Note that $\sup_{(x,y) \in S_0} |x - y| < \infty$ and so $l_0 < \infty$, since $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$ by Assumption **(A)**(iii).

According to the definition of $\Phi_n(x, y)$, we know that for any $n \geq 1$ and $x, y \in \mathbb{R}^d$ with $|x - y| \geq 1/n$,

$$(2.18) \quad \tilde{L}\Phi_n(x, y) \leq \tilde{L}\psi_n(|x - y|) + \tilde{L}(\varepsilon(V(x) + V(y))),$$

where

$$\psi_n(r) \begin{cases} = \psi(r), & 0 \leq r \leq 1/(n + 1), \\ \leq a + \psi(r), & 1/(n + 1) \leq r \leq 1/n, \\ = a + \psi(r), & r \geq 1/n. \end{cases}$$

(ii) For any $n \geq 1$ and $r \in [1/n, \infty)$, we have $\psi_n(r) = a + \psi(r)$ and $\psi'_n(r) = \psi'(r)$. Therefore, for any $\kappa \in (0, \kappa_0]$ and $r > 0$,

$$\begin{aligned}\psi_n(r - \kappa \wedge r) &= \psi_n(r - \kappa \wedge r) \mathbf{1}_{\{r > \kappa\}} \leq (a + \psi(r - \kappa \wedge r)) \mathbf{1}_{\{r > \kappa\}} \\ &= (a + \psi(r - \kappa \wedge r)) - (a + \psi(r - \kappa \wedge r)) \mathbf{1}_{\{r \leq \kappa\}} \\ &= (a + \psi(r - \kappa \wedge r)) - a \mathbf{1}_{\{r \leq \kappa\}}.\end{aligned}$$

Recall that, under Assumption **(A)**(ii), for any $0 < \kappa \leq \kappa_0 \leq 1$, we have $J(\kappa) = \inf_{0 < s \leq \kappa} J(s) > 0$. For any $r \in (1/n, l_0]$,

$$(2.19) \quad \begin{aligned} & \frac{1}{2}J(\kappa \wedge r) [\psi_n(r + r \wedge \kappa) + \psi_n(r - r \wedge \kappa) - 2\psi_n(r)] \\ & \leq \frac{1}{2}J(\kappa \wedge r) [\psi(r + r \wedge \kappa) + \psi(r - r \wedge \kappa) - 2\psi(r)] - \frac{a}{2}J(\kappa \wedge r)\mathbb{1}_{\{r \leq \kappa \wedge l_0\}}. \end{aligned}$$

On the other hand, we can check that the function ψ above satisfies all the conditions in both statements of Lemma 2.5. Hence, according to (2.10), (2.19) and Assumption **(A)**(i), for any $x, y \in \mathbb{R}^d$ with $1/n \leq |x - y| \leq \kappa$,

$$\begin{aligned} \tilde{L}\psi_n(|x - y|) & \leq \frac{1}{2}J(|x - y|) [\psi_n(2|x - y|) - 2\psi_n(|x - y|)] + K_1\psi'_n(|x - y|) \\ & \leq -\frac{a}{2}J(\kappa) + \frac{1}{2}J(\kappa)|x - y|^2\psi''(|x - y|) + K_1\psi'_n(|x - y|) \\ & \leq -\frac{a}{2}J(\kappa) + K_1c \leq -\frac{a}{4}J(\kappa); \end{aligned}$$

for any $x, y \in \mathbb{R}^d$ with $\kappa < |x - y| \leq l_0$,

$$\begin{aligned} \tilde{L}\psi_n(|x - y|) & \leq \frac{1}{2}J(\kappa)\kappa^2\psi''(|x - y|) + K_1l_0\psi'(|x - y|) \\ & \leq -\frac{1}{4}J(\kappa)\kappa^2c^2e^{-c|x-y|} \leq -\frac{1}{4}J(\kappa)\kappa^2c^2e^{-cl_0}; \end{aligned}$$

and for any $x, y \in \mathbb{R}^d$ with $|x - y| > l_0$,

$$(2.20) \quad \begin{aligned} \tilde{L}\psi_n(|x - y|) & \leq K_1|x - y|\psi'(|x - y|) \\ & \leq K_1 \max \left\{ \sup_{l_0 \leq r \leq 2l_0} r\psi'(r), \psi'(2l_0) \sup_{r > 2l_0} r(1 + r - 2l_0)^{-2} \right\} \\ & = K_1l_0\psi'(l_0) \leq \frac{1}{4}J(\kappa)\kappa^2c^2e^{-cl_0}. \end{aligned}$$

Putting all the estimates together, we find that

$$\tilde{L}\psi_n(|x - y|) \leq \begin{cases} -\frac{a}{4}J(\kappa), & 1/n \leq |x - y| \leq \kappa, \\ -\frac{1}{4}J(\kappa)\kappa^2c^2e^{-cl_0}, & \kappa < |x - y| \leq l_0, \\ \frac{1}{4}J(\kappa)\kappa^2c^2e^{-cl_0}, & |x - y| > l_0, \end{cases}$$

On the other hand, by (2.17) in Assumption **(A)**(iii) and the definition of the coupling operator, for all $x, y \in \mathbb{R}^d$,

$$\tilde{L}[\varepsilon(V(x) + V(y))] = \varepsilon(LV(x) + LV(y)) \leq \begin{cases} 2\varepsilon C, & (x, y) \in S_0, \\ -\frac{\varepsilon\lambda}{2}(V(x) + V(y)), & (x, y) \notin S_0. \end{cases}$$

According to the choice of ε and (2.18), we conclude that for any $n \geq 1$ and $x, y \in \mathbb{R}^d$ with $|x - y| \geq 1/n$,

$$\begin{aligned} \tilde{L}\Phi_n(x, y) & \leq \tilde{L}\psi_n(|x - y|) + \tilde{L}(\varepsilon(V(x) + V(y))) \\ & \leq \begin{cases} -\frac{1}{8}J(\kappa)\kappa^2c^2e^{-cl_0}, & (x, y) \in S_0, \\ -\frac{\varepsilon\lambda}{4}(V(x) + V(y)), & (x, y) \notin S_0 \end{cases} \\ & \leq -\lambda_0[\psi_n(|x - y|) + \varepsilon(V(x) + V(y))] \\ & = -\lambda_0\Phi_n(x, y), \end{aligned}$$

where

$$\lambda_0 = \min \left\{ \frac{\lambda J(\kappa) \kappa^2 c^2 e^{-cl_0}}{8(\lambda + a\lambda + 16C\varepsilon)}, \frac{4\varepsilon\lambda C}{16C\varepsilon + \lambda(1 + a + ce^{-2cl_0})} \right\}.$$

This, along with Proposition 2.1 and the fact that for all $n \geq 1$ and $x, y \in \mathbb{R}^d$, $\varepsilon(V(x) + V(y)) \leq \psi_n(|x - y|) + \varepsilon(V(x) + V(y)) \leq [(1 + a + ce^{-2cl_0}) + \varepsilon](V(x) + V(y))$, where we used $V(x) \geq 1$ for all $x \in \mathbb{R}^d$ and $\psi(r) \leq 1 + ce^{-2cl_0}$ for all $r \geq 0$, proves the desired assertion. \square

Remark 2.6. Motivated by [52, Section 1.6.1], we can extend Theorem 2.2 by replacing the global condition, Assumption **(A)**(i) on $b(x)$, by a local one. More precisely, in the proof above, we replace K_1 by $K_1(l_0)$ which satisfies that for all $x, y \in \mathbb{R}^d$ with $|x - y| \leq l_0$,

$$(2.21) \quad \langle b(x) - b(y), x - y \rangle \leq K_1(l_0)(|x - y|^2 \vee |x - y|),$$

and change the argument for (2.20) as that for any $x, y \in \mathbb{R}^d$ with $|x - y| > l_0$,

$$\begin{aligned} \tilde{L}\psi_n(|x - y|) &\leq |b(x) - b(y)|\psi'(|x - y|) \\ &\leq |b(x) - b(y)| \max \left\{ \sup_{l_0 \leq r \leq 2l_0} \psi'(r), \psi'(2l_0) \sup_{r > 2l_0} (1 + r - 2l_0)^{-2} \right\} \\ &= |b(x) - b(y)|\psi'(l_0). \end{aligned}$$

Thus, we can follow the proof of Theorem 2.2 and obtain that, if for any $(x, y) \notin S_0$,

$$(2.22) \quad V(x) + V(y) \geq \frac{64C}{\lambda(J(\kappa)\kappa^2 + 4K_1(l_0)l_0)} |b(x) - b(y)|,$$

then, there exists $\lambda_0 > 0$ (which is different from that in the proof of Theorem 2.2) such that for any $n \geq 1$ and $x, y \in \mathbb{R}^d$ with $|x - y| \geq 1/n$,

$$\tilde{L}\Phi_n(x, y) \leq -\lambda_0\Phi_n(x, y).$$

In particular, the conclusion of Theorem 2.2 still holds true under Assumption **(A)**(ii)-(iii), (2.21) and (2.22).

Finally, we present the proof of Theorem 1.1.

Proof of Theorem 1.1. It is easy to see that condition (1.4) implies Assumption **(A)**(i). Moreover, it can be shown that condition (1.2), along with the assumption that $\int_{\{|z| \geq 1\}} |z| \nu(dz) < \infty$, implies Assumption **(A)**(iii) with a Lyapunov function V such that $V(x) \geq 1$ for all $x \in \mathbb{R}^d$ and $V(x) = |x| + 1$ for $|x| \geq 1$, cf. Lemma 3.5 and Remark 3.6. In particular, $V(x) \geq |x|$ for all $x \in \mathbb{R}^d$ and hence the weighted total variation distance $\|\cdot\|_{\text{var}, V}$ from Corollary 2.3 dominates both the standard total variation and L^1 -Wasserstein distances (see e.g. [46, Theorem 6.15] and [16, Remark 2.3]), which finishes the proof. \square

2.3. Exponential ergodicity for the SDE (2.1): multiplicative distance. In this subsection, we present an alternative way of studying exponential ergodicity for SDEs given by (2.1), via multiplicative Wasserstein pseudo-distances. The techniques that we will introduce here will also play a crucial role in Section 3 for studying exponential ergodicity of McKean-Vlasov equations and proving Theorem 1.5. We remark that for McKean-Vlasov SDEs we were not able to prove an analogue of Theorem 2.2 from the previous section, but only an analogue of a weaker Theorem 2.7 presented below.

We will work under the following set of conditions.

Assumption (B)

(i) *There is a constant $K_1 > 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\langle b(x) - b(y), x - y \rangle \leq K_1 |x - y|^2.$$

(ii) *For any $\theta > 0$, there exists a measure $0 < \nu_\theta \leq \nu$ such that $\text{supp } \nu_\theta \subset B(0, 1)$, $\int_{\{|z| \leq 1\}} |z| \nu_\theta(dz) \leq \theta$, and*

$$(2.23) \quad \lim_{r \rightarrow 0} \inf_{s \in (0, r]} J_{\nu_\theta}(s) s^\alpha > 0,$$

where $\alpha := \alpha(\theta) \in (0, 1)$ and

$$J_{\nu_\theta}(s) := \inf_{x \in \mathbb{R}^d: |x| \leq s} [\nu_\theta \wedge (\delta_x * \nu_\theta)](\mathbb{R}^d) > 0.$$

(iii) *There exists a C^2 -function $V : \mathbb{R}^d \rightarrow [1, \infty)$ such that the Lyapunov condition (2.17) holds and*

$$(2.24) \quad \sup_{z \in B(x, 2)} |\nabla V(z)| \leq C_0 V(x), \quad x \in \mathbb{R}^d,$$

where $C_0 > 0$ is a constant independent of $x \in \mathbb{R}^d$.

Assumption (2.24) is a growth condition on the Lyapunov function V . It is satisfied, e.g., if $1 \leq V \in C^2(\mathbb{R}^d)$ and $V(x) = |x|^\alpha$ with $\alpha > 0$ or $V(x) = \exp(|x|^\beta)$ with $\beta \in (0, 1]$ for large $|x|$. We note that, in the context of related results for diffusions, similar conditions (see [16, Assumptions 2.4 and 2.5]) are also imposed. Similarly as in Subsection 2.2, we remark that under additional assumptions that $\int_{\{|z| > 1\}} |z| \nu(dz) < \infty$ and that there are constants $c_0, l_0 > 0$ such that $\langle b(x), x \rangle \leq -c_0 |x|^2$ for all $x \in \mathbb{R}^d$ with $|x| \geq l_0$. Assumption (B) (iii) is satisfied with a radial function $V \in C^2(\mathbb{R}^d)$ such that $V(x) \geq 1$ for all $x \in \mathbb{R}^d$ and $V(x) = 1 + |x|$ for all $x \in \mathbb{R}^d$ with $|x| \geq 1$; see Lemma 3.5 and Remark 3.6.

Theorem 2.7. *Under Assumption (B), there are constants λ_0 and $C_0 > 0$ such that for all $x, y \in \mathbb{R}^d$ and $t > 0$,*

$$W_\Phi(P_t(x, \cdot), P_t(y, \cdot)) \leq C_0 e^{-\lambda_0 t} \Phi(x, y),$$

where $\Phi(x, y) = (|x - y| \wedge 1)(V(x) + V(y))$.

The use of the multiplicative distance W_Φ is inspired by the weak Harris theorem introduced in [19], see also [16, Section 2.2]. As mentioned in [19], the distance of multiplicative form is more applicable for SDEs with degenerate noises or infinite dimensional SDEs, where convergence in terms of the total variation norm (and so the additive metric in the previous subsection) does not hold. We note that the multiplicative distance W_Φ indeed is only a multiplicative semimetric, since the triangle inequality may be not true. See [19, Section 4] for more details. As shown in the proof of Theorem 2.7 below, we take the reference function corresponding to the multiplicative distance W_Φ of the form $\psi(|x - y|)(1 + \varepsilon(V(x) + V(y)))$, where $\psi(|x - y|)$ is a bounded concave function and is comparable to $|x - y|$ for all $x, y \in \mathbb{R}^d$ with $|x - y| \leq 1$. Note also that, since the function $\Phi(x, y)$ in Theorem 2.7 satisfies that $\Phi(x, y) \rightarrow 0$ as $x \rightarrow y$, we can get from Theorem 2.7 that the associated semigroup of the process $(X_t)_{t \geq 0}$ is Feller, i.e., for any $t > 0$ and $f \in C_b(\mathbb{R}^d)$, $P_t f \in C_b(\mathbb{R}^d)$; see the proof of [33, Proposition 1.5]. However, such an assertion can not be deduced from Theorem 2.2.

Similar to Corollary 2.3, we have the following statement.

Corollary 2.8. *Suppose that $\int_{\{|z|\geq 1\}} |z| \nu(dz) < \infty$ and Assumption **(B)**(iii) is satisfied for V with $V(x) \geq c_0|x|$ for all $|x|$ large enough and some constant $c_0 \in (0, 1)$. Under Assumption **(B)** (i) and (ii), the process $(X_t)_{t \geq 0}$ is exponentially ergodic in terms of W_1 -distance; more explicitly, there are a unique invariant probability measure μ with finite first moment, a constant λ_0 and a measurable function $C_0(x)$ such that for all $x \in \mathbb{R}^d$ and $t > 0$,*

$$W_1(P_t(x, \cdot), \mu) \leq C_0(x)e^{-\lambda_0 t}.$$

Proof. Note that here, unlike in Corollary 2.3, W_Φ is only a semimetric, as mentioned above. First, it can be verified that the condition $\int_{\{|z|\geq 1\}} |z| \nu(dz) < \infty$ along with Assumption **(B)**(i) yields that $\mathbb{E}^x|X_t| < \infty$ for all $x \in \mathbb{R}^d$ and $t > 0$. Hence, by Theorem 2.7 and the fact that $c_1 W_1(\mu_1, \mu_2) \leq W_\Phi(\mu_1, \mu_2)$ (which is implied by $c_1|x - y| \leq \Phi(x, y)$ for all $x, y \in \mathbb{R}^d$, due to the definition of $\Phi(x, y)$ and our assumption that $V(x) \geq c_0|x|$ for all $|x|$ large enough and some constant $c_0 \in (0, 1)$) one can obtain the existence of a unique invariant probability measure. With aid of this point, we can follow the argument of Corollary 2.3 to prove the desired assertion. \square

We now pass to the proof of Theorem 2.7. We begin with the following lemma.

Lemma 2.9. *Let $g \in C([0, 2l_0]) \cap C^3((0, 2l_0])$ be such that $g'(r) \geq 0$, $g''(r) \leq 0$ and $g'''(r) \geq 0$ for any $r \in (0, 2l_0]$. Then for all $c_1 > 0$ the function*

$$\psi(r) := \psi_{c_1}(r) = \begin{cases} c_1 r + \int_0^r e^{-g(s)} ds, & r \in [0, 2l_0], \\ \psi(2l_0) + \psi'(2l_0) \frac{r-2l_0}{1+r-2l_0}, & r \in (2l_0, \infty) \end{cases}$$

satisfies

- (i) $\psi \in C^1([0, \infty))$ and $c_1 r \leq \psi(r) \leq (c_1 + 1)r$ on $[0, 2l_0]$;
- (ii) $\psi' > 0$, $\psi'' \leq 0$, $\psi''' \geq 0$ and $\psi^{(4)} \leq 0$ on $(0, 2l_0]$;
- (iii) for any $0 \leq \delta \leq r$,

$$\psi(r + \delta) + \psi(r - \delta) - 2\psi(r) \leq 0;$$

- (iv) for any $0 \leq \delta \leq r \leq l_0$,

$$\psi(r + \delta) + \psi(r - \delta) - 2\psi(r) \leq \psi''(r)\delta^2.$$

Note that the points (i) and (ii) are easy to check, whereas (iii) and (iv) follow from Lemma 2.5.

Next, we present the proof of Theorem 2.7.

Proof of Theorem 2.7. Under Assumption **(B)**(ii), we will apply the refined basic coupling for the component ν_θ of the Lévy measure ν , and couple the remaining mass synchronously, where $\theta > 0$ is determined later. Let ψ be the function given in Lemma 2.9, where

$$g(r) = C_*(2K_1 + 1)l_0^{2-\alpha}r^\alpha, \quad l_0 = \sup_{(x,y) \in S_0} |x - y| + 2L_1, \quad c_1 = e^{-g(2l_0)}$$

and

$$S_0 = \{(x, y) \in \mathbb{R}^{2d} : \lambda(V(x) + V(y)) \leq L_2 C\}.$$

Here, $\alpha \in (0, 1)$ is given in Assumption **(B)**(ii), and C_* , L_1 and L_2 are large enough (their exact values will be determined later).

By the definition of g , we know that $g'(r) \geq 0$, $g''(r) \leq 0$ and $g'''(r) \geq 0$ for any $r \in (0, 2l_0]$. In particular, we can use Lemma 2.9. On the other hand, by (2.23), there are constants $\kappa := \kappa(\theta) \in (0, \kappa_0]$ where $\kappa_0 \leq 1$ is given by (2.16) and $C_* := C_*(\kappa, \theta) > 0$ (both are independent of l_0) such that for all $r \in (0, 2l_0]$,

$$(2.25) \quad \sigma(r) := (\alpha C_*)^{-1} l_0^{\alpha-2} r^{1-\alpha} \leq \frac{1}{2r} J(\kappa \wedge r) (\kappa \wedge r)^2.$$

In particular, $g'(r) = \frac{2K_1+1}{\sigma(r)}$.

In the following, let $F(x, y) = \psi(|x - y|)(1 + \varepsilon(V(x) + V(y)))$ for any $x, y \in \mathbb{R}^d$, where $\varepsilon > 0$ is determined later. For any $x \in \mathbb{R}^d$, set $\mu_{\theta, x} = \nu_\theta \wedge (\delta_x * \nu_\theta)$. Then, by (2.8) and some elementary calculations, we can find that for any $x, y \in \mathbb{R}^d$,

$$(2.26) \quad \begin{aligned} & \tilde{L}F(x, y) \\ &= \tilde{L}\psi(|x - y|) \cdot (1 + \varepsilon(V(x) + V(y))) + \psi(|x - y|) \cdot \varepsilon \tilde{L}(V(x) + V(y)) \\ & \quad + \frac{1}{2} \varepsilon (\psi(|x - y - (x - y)_\kappa|) - \psi(|x - y|)) \\ & \quad \times \int [(V(x + z) - V(x)) + (V(y + z + (x - y)_\kappa) - V(y))] \mu_{\theta, (y-x)_\kappa}(dz) \\ & \quad + \frac{1}{2} \varepsilon (\psi(|x - y + (x - y)_\kappa|) - \psi(|x - y|)) \\ & \quad \times \int [(V(x + z) - V(x)) + (V(y + z - (x - y)_\kappa) - V(y))] \mu_{\theta, (x-y)_\kappa}(dz). \end{aligned}$$

For any $x, y \in \mathbb{R}^d$ with $|x - y| \leq l_0$, by (2.10), Lemma 2.9(iii), (2.25) and (2.17),

$$\begin{aligned} & \tilde{L}F(x, y) \\ & \leq \left(-\frac{1}{2} J(|x - y| \wedge \kappa) (|x - y| \wedge \kappa)^2 e^{-g(|x-y|)} g'(|x - y|) + K_1 |x - y| (c_1 + e^{-g(|x-y|)}) \right) \\ & \quad \times (1 + \varepsilon(V(x) + V(y))) \\ & \quad - \varepsilon \lambda (V(x) + V(y)) \cdot \psi(|x - y|) + 2C\varepsilon \cdot \psi(|x - y|) \\ & \quad + \varepsilon \psi(|x - y|) \\ & \quad \times \left(\sup_{z \in B(x, 1)} \nabla V(z) \int |z| \nu_\theta(dz) + \sup_{z \in B(y, 2)} \nabla V(z) \int |z + (x - y)_\kappa| \mu_{\theta, (y-x)_\kappa}(dz) \right) \\ & \quad + \frac{1}{2} \varepsilon (\psi(2|x - y|) + \psi(|x - y|)) \\ & \quad \times \left(\sup_{z \in B(x, 1)} \nabla V(z) \int |z| \nu_\theta(dz) + \sup_{z \in B(y, 2)} \nabla V(z) \int |z + (y - x)_\kappa| \mu_{\theta, (x-y)_\kappa}(dz) \right) \\ & \leq \left(-\frac{1}{2} J(|x - y| \wedge \kappa) (|x - y| \wedge \kappa)^2 e^{-g(|x-y|)} \frac{2K_1 + 1}{\sigma(|x - y|)} + K_1 |x - y| (c_1 + e^{-g(|x-y|)}) \right) \\ & \quad \times (1 + \varepsilon(V(x) + V(y))) \\ & \quad - \varepsilon \lambda (V(x) + V(y)) \cdot \psi(|x - y|) + 2C\varepsilon \cdot \psi(|x - y|) \\ & \quad + 2C_0 \varepsilon (\psi(2|x - y|) + \psi(|x - y|)) (V(x) + V(y)) \int |z| \nu_\theta(dz) \\ & \leq -|x - y| e^{-g(|x-y|)} \cdot (1 + \varepsilon(V(x) + V(y))) \end{aligned}$$

$$\begin{aligned}
 & -\varepsilon\lambda(V(x) + V(y)) \cdot \psi(|x - y|) + 2C\varepsilon \cdot \psi(|x - y|) \\
 & + C_1\theta\varepsilon\psi(|x - y|)(V(x) + V(y)),
 \end{aligned}$$

where the first inequality follows from the mean value theorem, in the second inequality we used Assumption **(B)**(iii), and in the last inequality we used the facts that $\psi(2r) \leq 2\psi(r)$ for $0 < r \leq l_0$, $\int_{\{|z| \leq 1\}} |z| \nu_\theta(dz) \leq \theta$ and Assumption **(B)**(iii) together with the fact that $c_1 \leq e^{-g(|x-y|)}$ for $|x-y| \leq l_0$. Note that C_1 is a constant independent of θ , ε , λ and l_0 , and that the argument for the estimates of the last two terms in (2.26) works for all $x, y \in \mathbb{R}^d$. Now let us choose $\theta > 0$ small enough so that $C_1\theta \leq \lambda/4$ and take $\varepsilon > 0$ small enough so that $2C\varepsilon\psi(r) \leq re^{-g(r)}$ for all $0 < r \leq l_0$. More precisely, we can take $\varepsilon > 0$ such that

$$(2.27) \quad 2C\varepsilon = \inf_{0 \leq r \leq l_0} re^{-g(r)}\psi(r)^{-1}.$$

Therefore, for all $x, y \in \mathbb{R}^d$ with $|x - y| \leq l_0$,

$$(2.28) \quad \tilde{L}F(x, y) \leq -(\varepsilon\lambda/2)\psi(|x - y|)(V(x) + V(y)).$$

For any $(x, y) \notin S_0$, following the argument above and (2.20), we can get that

$$\begin{aligned}
 (2.29) \quad \tilde{L}F(x, y) & \leq 4K_1l_0e^{-g(l_0)}\mathbf{1}_{\{|x-y| \geq l_0\}} \cdot (1 + \varepsilon(V(x) + V(y))) \\
 & \quad - (\varepsilon\lambda/2)(V(x) + V(y)) \cdot \psi(|x - y|) \\
 & \quad + C_1\theta\varepsilon\psi(|x - y|)(V(x) + V(y)) \\
 & \leq 4K_1l_0e^{-g(l_0)}\mathbf{1}_{\{|x-y| \geq l_0\}} \cdot (1 + \varepsilon(V(x) + V(y))) \\
 & \quad - (\varepsilon\lambda/4)(V(x) + V(y)) \cdot \psi(|x - y|),
 \end{aligned}$$

where in the first inequality we used the definition of S_0 and the last inequality follows from the choice of θ above. Next, we choose $L_1 \geq 1$ large enough so that

$$4K_1e^{-C_*(1+2K_1)r^2} \leq \lambda e^{-C_*(1+2K_1)r^{-2/\alpha}}/16, \quad r \geq L_1$$

and hence, since $l_0 > L_1$, we have

$$4K_1l_0e^{-g(l_0)} \leq \lambda e^{-C_*(1+2K_1)l_0^{(\alpha-2)/\alpha}}/16.$$

In particular,

$$(2.30) \quad 4K_1l_0e^{-g(l_0)} \leq \frac{\lambda}{16} \int_0^{l_0^{-(2-\alpha)/\alpha}} e^{-C_*(2K_1+1)l_0^{2-\alpha}s^\alpha} ds \leq \frac{\lambda}{16}\psi(l_0).$$

Furthermore, we choose

$$L_2 \geq \max\left(4, \frac{64K_1l_0e^{-g(l_0)}}{\varepsilon C\psi(l_0)}\right).$$

Then for all $x, y \in \mathbb{R}^d$ with $|x - y| > l_0$ we have

$$(2.31) \quad 4K_1l_0e^{-g(l_0)} \leq \frac{\lambda\varepsilon}{16}\psi(|x - y|)(V(x) + V(y))$$

(note that $|x - y| > l_0$ implies $(x, y) \notin S_0$). Hence, combining (2.30) with (2.31) and (2.29), we see that for any $(x, y) \notin S_0$,

$$(2.32) \quad \tilde{L}F(x, y) \leq -(\varepsilon\lambda/8)(V(x) + V(y)) \cdot \psi(|x - y|).$$

This along with (2.28) shows that (2.32) holds for all $x, y \in \mathbb{R}^d$, which proves the desired assertion by Proposition 2.1. \square

Remark 2.10. The results discussed in the present paper can also be obtained by applying other coupling operators. For instance, one could apply the coupling studied in [34, 35] to obtain inequalities such as (2.3) for SDEs driven by Lévy processes with rotationally symmetric Lévy measures which are not required to satisfy the concentration around zero property (1.12), cf. [34, Remark 1.6]. This would allow us to prove e.g. an analogue of Theorem 2.7 under different (neither strictly weaker nor stronger) assumptions on the noise. Yet another possibility would be to use the coupling from [32, 50]. See [31] for a discussion on different couplings for Lévy processes and Lévy-driven SDEs. In the present paper we choose to work with the refined basic coupling given by (2.8) since it can apply to a very large class of non-symmetric Lévy measures.

Remark 2.11. We present two further remarks on the assumptions on the drift in Theorems 2.2 and 2.7.

- (i) Assumption **(B)** (i) in Theorem 2.7 is the well known one-sided Lipschitz condition on the drift. Note that it is stronger than the assumption we needed in Theorem 2.2. By carefully checking the proof above, we can see that Theorem 2.7 still holds true if Assumption **(B)** (i) is weakened into the following condition: *there are constants $K_1 > 0$ and $\beta \in (0, 1 - \alpha)$ with $\alpha \in (0, 1)$ in Assumption **(B)** (ii) such that for all $x, y \in \mathbb{R}^d$,*

$$\langle b(x) - b(y), x - y \rangle \leq K_1(|x - y|^{1+\beta} \vee |x - y|^2).$$

For the modification of the proof to adjust to this weaker assumption, one can refer to the proof of [33, Theorem 4.2].

- (ii) Following the argument of [52, Section 1.6.2], one can also prove a more general version of Theorem 2.7 by replacing the one-sided Lipschitz condition on the drift in Assumption **(B)** by a local counterpart, and by imposing some growth condition on ψV similar to (2.22). The details are omitted here.

3. EXPONENTIAL CONVERGENCE FOR MCKEAN-VLASOV SDEs WITH LÉVY NOISE

In this section, we are concerned with the McKean-Vlasov (distribution dependent) SDE with jumps given by (1.7). Let $\mathcal{P}_1(\mathbb{R}^d)$ be the set of probability measures on \mathbb{R}^d with finite first moment. Throughout this section, we always assume that the following conditions hold.

Assumption **(H)**

- (i) *The drift term $b(x, \mu)$ is continuous on $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$, and that it satisfies the following one-sided Lipschitz condition:*

$$(3.1) \quad \frac{\langle b(x_1, \mu_1) - b(x_2, \mu_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \leq K(|x_1 - x_2| + W_1(\mu_1, \mu_2))$$

for all $x_1, x_2 \in \mathbb{R}^d$ with $x_1 \neq x_2$ and all $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$ for some constant $K > 0$, as well as the following growth condition: there is a constant $C_1 > 0$ such that for all $\mu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$(3.2) \quad |b(0, \mu)| \leq C_1 \left(1 + \int_{\mathbb{R}^d} |z| \mu(dz) \right).$$

(ii) The Lévy measure ν of the Lévy process $Z := (Z_t)_{t \geq 0}$ satisfies

$$\int_{\{|z| \geq 1\}} |z| \nu(dz) < \infty.$$

We start with the following statement.

Proposition 3.1. *Under Assumption **(H)**, the McKean-Vlasov SDE (1.7) has a unique non-explosive strong solution $(X_t)_{t \geq 0}$ such that $\mathbb{E}|X_t| < \infty$ for all $t > 0$.*

Proof. This follows from (the proofs of) [17, Theorem 2.1] and [47, Theorem 2.1], and we sketch the argument here. Equation (1.7) is a special case of the McKean-Vlasov stochastic differential equation given in [17, (2.1)] with

$$\sigma(x, \mu) \equiv 0, \quad f(x, \mu, u) = \tilde{f}(x, \mu, u) = u$$

and $N(du, dt)$ being the Poisson random measure associated with the Lévy process $(Z_t)_{t \geq 0}$, with the intensity measure $\nu(du)$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then, according to (the proof of) [17, Theorem 2.1] and Assumption **(H)**, there exists a unique strong solution for equation (1.7). Note that the proof of [17, Theorem 2.1] is based on the parametrized approach by considering a sequence of SDEs $(X_t^{(n)})_{t \geq 0}$ given by

$$dX_t^{(n)} = b(X_t^{(n)}, \mu_t^{(n-1)}) dt + dZ_t, \quad t \geq 0, n \geq 1$$

with $X_t^{(0)} = x$ for all $t \geq 0$ and $\mu_t^{(n-1)}$ being the law of $X_t^{(n-1)}$. This idea is the same as that for [47, Theorem 2.1]. In particular, according to the argument of [47, Theorem 2.1], the Lipschitz condition

$$|b(x_1, \mu_1) - b(x_2, \mu_2)| \leq K(|x_1 - x_2| + W_1(\mu_1, \mu_2))$$

assumed in [17, Theorem 2.1] can be replaced by Assumption **(H)**. Moreover, one can show that the unique strong solution $(X_t)_{t \geq 0}$ to the SDE (1.7) satisfies $\mathbb{E}|X_t| < \infty$ for all $t > 0$, by following again the proof of [47, Theorem 2.1]. \square

The existence and uniqueness of a strong solution to the SDE (1.7) clearly implies the existence of a weak solution to (1.7); see [47, Definition 1.1]. Let μ_{X_t} be the distribution of the time marginal X_t of the process $(X_t)_{t \geq 0}$ with initial distribution μ_{X_0} . Then, it follows from [25, Propositions 1.6 and 1.7] that for any $f \in C_b^2(\mathbb{R}^d)$,

$$\left\{ f(X_t) - f(X_0) - \int_0^t L[\mu_{X_s}] f(X_s) ds, t \geq 0 \right\}$$

is a \mathbb{P} -martingale, where

$$(3.3) \quad \begin{aligned} L[\mu]f(x) &= \langle b(x, \mu), \nabla f(x) \rangle \\ &+ \int (f(x+z) - f(x) - \langle \nabla f(x), z \rangle \mathbf{1}_{\{|z| \leq 1\}}) \nu(dz). \end{aligned}$$

That is, $L[\mu]f$ can be interpreted as the infinitesimal generator of the process $(X_t)_{t \geq 0}$.

In the rest of this section, we will always assume that Assumption **(H)** holds.

3.1. Convergence in W_1 : the contractivity at infinity approach. We say that the drift term $b(x, \mu)$ in (1.7) satisfies $\mathbf{B}(K_1, K_2, l_0; K_3)$, if (1.13) holds, i.e., for any $x_1, x_2 \in \mathbb{R}^d$ with $x_1 \neq x_2$ and $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$,

$$(3.4) \quad \frac{\langle b(x_1, \mu_1) - b(x_2, \mu_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \leq K_1 |x_1 - x_2| \mathbf{1}_{\{|x_1 - x_2| \leq l_0\}} - K_2 |x_1 - x_2| \mathbf{1}_{\{|x_1 - x_2| > l_0\}} + K_3 W_1(\mu_1, \mu_2),$$

where $l_0 \in [0, \infty)$, K_1, K_2 and $K_3 \geq 0$.

Theorem 3.2. *Suppose that Assumption (H) holds. Assume that*

$$(3.5) \quad J(\kappa_0) := \inf_{x \in \mathbb{R}^d: |x| \leq \kappa_0} [\nu \wedge (\delta_x * \nu)](\mathbb{R}^d) > 0$$

for some $0 < \kappa_0 \leq 1$, and that the drift $b(x, \mu)$ satisfies $\mathbf{B}(K_1, K_2, l_0; K_3)$. Suppose that there exists a nondecreasing and concave function $\sigma \in C([0, 2l_0]) \cap C^2((0, 2l_0))$ such that $g_1(r) := \int_0^r \frac{1}{\sigma(s)} ds$ is well defined for all $r \in [0, 2l_0]$, and for some $\kappa \in (0, \kappa_0]$,

$$(3.6) \quad \sigma(r) \leq \frac{1}{2r} J(\kappa \wedge r) (\kappa \wedge r)^2, \quad r \in (0, 2l_0],$$

where $J(r)$ is defined by (1.5). Let μ_{X_t} (resp. μ_{Y_t}) be the distribution of the time marginal X_t (resp. Y_t) of a solution to (1.7) with initial distribution μ_{X_0} (resp. μ_{Y_0}). Then for any $t > 0$,

$$W_1(\mu_{X_t}, \mu_{Y_t}) \leq C e^{-\lambda t} W_1(\mu_{X_0}, \mu_{Y_0}),$$

where

$$\lambda = \frac{c_1 c_2}{1 + c_1} - \frac{(1 + c_1) K_3}{2c_1}, \quad C = \frac{1 + c_1}{2c_1}$$

with $c_2 = (2K_2) \wedge g_1(2l_0)^{-1}$, $c_1 = e^{-c_2 g(2l_0)}$ and $g(r) = (1 + \frac{2K_1}{c_2}) g_1(r)$ for all $r \in [0, 2l_0]$.

Remark 3.3. We make some comments on Theorem 3.2 and its proof.

- (i) Theorem 3.2 extends [33, Theorem 4.2], where the drift term is distribution independent. As pointed out in [33, Remark 4.3(1)], when $b(x, \mu) = b(x)$ satisfies the uniformly dissipative condition in the sense that there is a constant $K_2 > 0$ such that for any $x, y \in \mathbb{R}^d$,

$$\langle b(x) - b(y), x - y \rangle \leq -K_2 |x - y|^2;$$

(that is, $l_0 = 0$ and $K_3 = 0$ in Theorem 3.2), then the constant λ in the statement is equal to K_2 , which is optimal.

- (ii) When K_3 is small enough so that $\lambda > 0$, using Theorem 3.2 and the fact that $\mathbb{E}|X_t| < \infty$ for all $t > 0$, we can obtain exponential ergodicity of $(X_t)_{t \geq 0}$ in terms of W_1 -distance, e.g. see the proof of [47, Theorem 3.1(2)]
- (iii) Note that for proving convergence of solutions to McKean-Vlasov SDEs we cannot use the classical argument involving a coupling $(X_t, Y_t)_{t \geq 0}$ such that $Y_t = X_t$ for all $t \geq T$, where T is the coupling time, see [47, Remarks on pages 598–599] or [16, Remarks between Assumption 2.7 and Theorem 2.4].

Note also that for any stopping time τ , if $(X_t)_{t \geq 0}$ is a solution to (1.7), then the stopped process $(Y_t)_{t \geq 0}$ defined by $Y_t := X_{t \wedge \tau}$ for all $t \geq 0$ solves

$$Y_t = Y_0 + \int_0^{t \wedge \tau} b(Y_s, \mu_s) ds + Z_{t \wedge \tau}, \quad t \geq 0,$$

which is not the same SDE since in general $\text{Law}(Y_t) \neq \text{Law}(X_t) = \mu_t$, see [21, Remarks on page 3]. This means that applying the standard argument via Proposition 2.1 cannot give us the exponential convergence for $W_{\Phi_\infty}(\mu_{X_t}, \mu_{Y_t})$ as desired, see its proof for more details. Hence to consider convergence of McKean-Vlasov SDEs we will use a different approach, which is based on the combination of the refined basic coupling and the synchronous coupling defined by (2.13).

- (iv) We further remark that the typical way of constructing couplings of Lévy-driven SDEs by using the interlacing technique (see e.g. [33, Proposition 2.2] or [34, Section 2.4]) is non-applicable here due to the lack of the strong Markov property of the solution $(X_t)_{t \geq 0}$, see the discussions in [17]. Hence we apply here a different approach based on the results on non-linear martingale problems from [25].

Proof of Theorem 3.2. (i) For any $r > 0$, define

$$\psi(r) = \begin{cases} c_1 r + \int_0^r e^{-c_2 g(s)} ds, & r \in [0, 2l_0], \\ \psi(2l_0) + \psi'(2l_0)(r - 2l_0), & r \in (2l_0, \infty). \end{cases}$$

Note that the function ψ is the same as in the proof of [33, Theorem 4.2]. Similarly as there, we can show by a simple calculation that

$$(3.7) \quad \begin{aligned} & \psi'(|x - y|) \left[K_1 |x - y| \mathbf{1}_{\{|x - y| \leq l_0\}} - K_2 |x - y| \mathbf{1}_{\{|x - y| \geq l_0\}} \right] + \tilde{L}_Z \psi(|x - y|) \\ & \leq -\lambda_0 \psi(|x - y|), \end{aligned}$$

where \tilde{L}_Z is the refined basic coupling operator for the Lévy process Z , and hence

$$(3.8) \quad \begin{aligned} \tilde{L}_Z \psi(|x - y|) &= \frac{1}{2} \mu_{(x-y)\kappa}(\mathbb{R}^d) \left[\psi(|x - y| + \kappa \wedge |x - y|) \right. \\ & \quad \left. + \psi(|x - y| - \kappa \wedge |x - y|) - 2\psi(|x - y|) \right] \end{aligned}$$

with $\lambda_0 = c_1 c_2 / (1 + c_1)$.

(ii) As mentioned in Remark 3.3 (iii) and (iv), from this point onwards the proof deviates substantially from that of [33, Theorem 4.2]. For any $\delta > 0$, consider the equation as follows

$$(3.9) \quad \begin{cases} dX_t = b(X_t, \mu_{X_t}) dt + dZ_t, & X_0 \sim \mu_{X_0}, \\ dY_t^\delta = b(Y_t^\delta, \mu_{Y_t^\delta}) dt + dZ_t + dL_t^{*,\delta}, & Y_0^\delta \sim \mu_{Y_0^\delta}, \end{cases}$$

where $(L_t^{*,\delta})_{t \geq 0}$ is given by (2.14). To prove the existence of a weak solution $(X_t, Y_t^\delta)_{t \geq 0}$ to the system (3.9), we consider the following nonlinear operator acting on $f \in C_b^2(\mathbb{R}^{2d})$ (for any fixed μ_1 and $\mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$),

$$\tilde{L}[\mu_1, \mu_2] f(x, y) = \langle b(x, \mu_1), \nabla_x f(x, y) \rangle + \langle b(y, \mu_2), \nabla_y f(x, y) \rangle + \tilde{L}_Z^\delta f(|x - y|),$$

where \tilde{L}_Z^δ is defined by (2.13). It is obvious that $\tilde{L}[\mu_1, \mu_2]$ is a coupling operator for the operators $L[\mu_1]$ and $L[\mu_2]$ given by (3.3). Due to the continuity of $b(x, \mu)$, the

drift coefficient of the coupling operator $\tilde{L}[\mu_1, \mu_2]$ is also continuous (with respect to the product metric). On the other hand, as shown in [33, Proposition A.5], (3.5) implies that there is a non-negative measurable function ρ on \mathbb{R}^d such that $\nu(dz) \geq \rho(z) dz$ and

$$\inf_{x \in \mathbb{R}^d: |x| \leq \kappa_0} \int_{\mathbb{R}^d} \rho(z) \wedge \rho(x+z) dz > 0.$$

Moreover, by [33, (A.3) in the proof of Proposition A.5], the function

$$(3.10) \quad x \mapsto \int_{\mathbb{R}^d} \rho(z) \wedge \rho(x+z) dz$$

is continuous on $\{x \in \mathbb{R}^d : 0 < |x| \leq \kappa_0\}$. Hence without loss of generality, under condition (3.5) we can consider the refined basic coupling applied only to the component $\rho(z) dz$ of the Lévy measure. Continuity of (3.10) together with the fact that $\phi_\delta \in C_b^1([0, \infty))$ yields that the coefficients of the operator \tilde{L}_Z^δ are continuous. Besides, by (2.6) and the definition of \tilde{L}_Z^δ , we can see that its coefficients are bounded. This follows from the fact that we only consider the refined basic coupling in the coupling operator \tilde{L}_Z^δ when the distance of the two marginal processes is larger than $\delta/2$. Hence, according to [25, Proposition 1.10], there exists a solution, belonging to the set of probability measures on $\bar{\mathbb{R}}^{2d}$ (the standard one point compactification of \mathbb{R}^{2d}), to the nonlinear martingale problem for the operator $\tilde{L}[\mu_1, \mu_2]$. (Note that, since the drift coefficient of the coupling operator $\tilde{L}[\mu_1, \mu_2]$ is not necessarily bounded, the argument of [25, Proposition 1.10] only guarantees the existence of a solution to the martingale problem for $\tilde{L}[\mu_1, \mu_2]$ on $\bar{\mathbb{R}}^{2d}$.) This further along with [25, Proposition 1.7] yields the existence of a weak solution $(X_t, Y_t^\delta)_{t \geq 0}$, taking values in $\bar{\mathbb{R}}^{2d}$, to the system (3.9). The explosion time of the process $(X_t, Y_t^\delta)_{t \geq 0}$ is

$$e = \lim_{n \rightarrow \infty} \tau_n, \quad \tau_n = \inf\{t \geq 0 : |X_t| + |Y_t^\delta| \geq n\}.$$

As mentioned above, $\tilde{L}[\mu_1, \mu_2]$ is a coupling operator for the operators $L[\mu_1]$ and $L[\mu_2]$ given by (3.3), and so the marginal processes $(X_t)_{t \geq 0}$ and $(Y_t^\delta)_{t \geq 0}$ enjoy the same law as the solution to the SDE given by (1.7). Since under assumptions in the beginning of this section any (weak) solution to (1.7) is non-explosive (e.g., see [47, Theorem 1.2]), we have $e = \infty$. In particular, there exists a weak solution to (3.9), which indeed takes values on \mathbb{R}^{2d} and is a coupling of the process determined by (1.7).

(iii) Recalling that $U_t^\delta = X_t - Y_t^\delta$, we have

$$dU_t^\delta = (b(X_t, \mu_{X_t}) - b(Y_t^\delta, \mu_{Y_t^\delta})) dt - dL_t^{*, \delta},$$

and hence, arguing as in (2.11), we see that

$$(3.11) \quad d\psi(|U_t^\delta|) = \left[\frac{\psi'(|U_t^\delta|)}{|U_t^\delta|} \langle U_t^\delta, b(X_t, \mu_{X_t}) - b(Y_t^\delta, \mu_{Y_t^\delta}) \rangle + \tilde{L}_Z^\delta \psi(|U_t^\delta|) \right] dt + dM_t^{\psi, \delta},$$

where

$$dM_t^{\psi, \delta} := \int_{\mathbb{R}^d \times [0, 1]} (\psi(|U_{t-}^\delta - S^\delta(U_{t-}^\delta, z, u)|) - \psi(|U_{t-}^\delta|)) \tilde{N}(dt, dz, du)$$

is a martingale. To see this, note that

$$\psi(|U_{t-}^\delta - S^\delta(U_{t-}^\delta, z, u)|) - \psi(|U_{t-}^\delta|) \leq \|\psi'\|_\infty |S^\delta(U_{t-}^\delta, z, u)|,$$

and observe that $|(U_t^\delta)_\kappa| \leq \kappa$ for all $t > 0$ and that for any $\delta > 0$ the measure $\phi_\delta(|U_{t-}^\delta|)\rho((U_t^\delta)_\kappa, z)\nu(dz)$ is finite.

Moreover, according to (2.13), we have

$$\tilde{L}_Z^\delta \psi(|x - y|) = \tilde{L}_Z \psi(|x - y|) \cdot \phi_\delta(|x - y|).$$

Hence, using the fact that ψ' is decreasing and the definition of $\phi_\delta(r)$, we see that (3.7) implies

$$\begin{aligned} & \psi'(|x - y|) \left[K_1|x - y| \mathbf{1}_{\{|x-y| \leq l_0\}} - K_2|x - y| \mathbf{1}_{\{|x-y| \geq l_0\}} \right] + \tilde{L}_Z^\delta \psi(|x - y|) \\ & \leq -\lambda_0 \psi(|x - y|) + \lambda_0 \psi(|x - y|) \cdot (1 - \phi_\delta(|x - y|)) \\ (3.12) \quad & + \psi'(|x - y|) \left[K_1|x - y| \mathbf{1}_{\{|x-y| \leq l_0\}} - K_2|x - y| \mathbf{1}_{\{|x-y| \geq l_0\}} \right] \\ & \quad \times (1 - \phi_\delta(|x - y|)) \\ & \leq -\lambda_0 \psi(|x - y|) + \lambda_0 \psi(\delta) + K_1 \psi'(0)(l_0 \wedge \delta). \end{aligned}$$

Combining the inequality above with $\mathbf{B}(K_1, K_2, l_0; K_3)$ and (3.11), we obtain

$$\begin{aligned} d\psi(|U_t^\delta|) & \leq \left[\psi'(|U_t^\delta|) \left(K_1|U_t^\delta| \mathbf{1}_{\{|U_t^\delta| \leq l_0\}} - K_2|U_t^\delta| \mathbf{1}_{\{|U_t^\delta| \geq l_0\}} \right) + \tilde{L}_Z^\delta \psi(|U_t^\delta|) \right] dt \\ & \quad + K_3 \psi'(|U_t^\delta|) W_1(\mu_{X_t}, \mu_{Y_t^\delta}) dt + dM_t^{\psi, \delta} \\ (3.13) \quad & \leq -\lambda_0 \psi(|U_t^\delta|) dt + \lambda_0 \psi(\delta) dt + K_1 \psi'(0)(l_0 \wedge \delta) dt \\ & \quad + K_3 \psi'(0) \mathbb{E}|U_t^\delta| dt + dM_t^{\psi, \delta} \\ & \leq -(\lambda_0 - (K_3(1 + c_1)/(2c_1))) \psi(|U_t^\delta|) dt + \lambda_0 \psi(\delta) dt \\ & \quad + K_1 \psi'(0)(l_0 \wedge \delta) dt + K_3(1 + c_1)((\mathbb{E}|U_t^\delta|) - |U_t^\delta|) dt + dM_t^{\psi, \delta} \\ & = -\lambda \psi(|U_t^\delta|) dt + \lambda_0 \psi(\delta) dt + K_1 \psi'(0)(l_0 \wedge \delta) dt \\ & \quad + K_3(1 + c_1)((\mathbb{E}|U_t^\delta|) - |U_t^\delta|) dt + dM_t^{\psi, \delta}, \end{aligned}$$

where in the second inequality we again used the fact that ψ' is decreasing on $[0, \infty)$, and the last inequality follows from the facts that $\psi'(0) = 1 + c_1$ and

$$\sup_{r>0} \frac{r}{\psi(r)} \leq \sup_{r>0} \frac{1}{\psi'(r)} = \frac{1}{\psi'(2l_0)} = \frac{1}{2c_1}.$$

Note that in the argument above we also used the assumption that $\mathbb{E}|X_t| < \infty$ for all $t > 0$ to ensure the finiteness of $\mathbb{E}|U_t^\delta|$.

(iv) As a consequence of (3.13), we find that

$$\begin{aligned}
& \mathbb{E}[e^{\lambda t} \psi(|U_t^\delta|)] \\
&= \mathbb{E}\psi(|U_0^\delta|) + \mathbb{E}\left(\int_0^t [\lambda e^{\lambda s} \psi(|U_s^\delta|) + e^{\lambda s} d\psi(|U_s^\delta|)/ds] ds\right) \\
&\leq \mathbb{E}\psi(|U_0^\delta|) + \mathbb{E}\left(\int_0^t e^{\lambda s} \left[\lambda \psi(|U_s^\delta|) \right. \right. \\
&\quad \left. \left. - \lambda \psi(|U_s^\delta|) + K_3(1+c_1)((\mathbb{E}|U_s^\delta|) - |U_s^\delta|) + \lambda_0 \psi(\delta) + K_1 \psi'(0)(l_0 \wedge \delta)\right] ds\right) \\
&= \mathbb{E}\psi(|U_0^\delta|) + \int_0^t e^{\lambda s} (\lambda_0 \psi(\delta) + K_1 \psi'(0)(l_0 \wedge \delta)) ds \\
&= \mathbb{E}\psi(|U_0^\delta|) + \left(\lambda_0 \psi(\delta) + K_1 \psi'(0)(l_0 \wedge \delta)\right) \frac{1}{\lambda} (e^{\lambda t} - 1).
\end{aligned}$$

Hence we obtain

$$\mathbb{E}\psi(|U_t^\delta|) \leq e^{-\lambda t} \mathbb{E}\psi(|U_0^\delta|) + e^{-\lambda t} (\lambda_0 \psi(\delta) + K_1 \psi'(0)(l_0 \wedge \delta)) \frac{1}{\lambda} (e^{\lambda t} - 1)$$

for any $\delta > 0$. Recall, however, from the discussion in step (ii), that for any $\delta > 0$ the process $(X_t, Y_t)_{t \geq 0}$ is a coupling of two copies of $(X_t)_{t \geq 0}$ with initial distributions μ_{X_0} and μ_{Y_0} , respectively. Hence

$$W_\psi(\mu_{X_t}, \mu_{Y_t}) \leq e^{-\lambda t} \mathbb{E}\psi(|X_0 - Y_0|) + e^{-\lambda t} (\lambda_0 \psi(\delta) + K_1 \psi'(0)(l_0 \wedge \delta)) \frac{1}{\lambda} (e^{\lambda t} - 1).$$

Since the estimate above holds for any $\delta > 0$, we can now take the limit $\delta \rightarrow 0$ and, using the continuity of ψ , we obtain

$$W_\psi(\mu_{X_t}, \mu_{Y_t}) \leq e^{-\lambda t} \mathbb{E}\psi(|X_0 - Y_0|),$$

which, along with the fact that $2c_1 r \leq \psi(r) \leq (1+c_1)r$ for all $r > 0$, yields the desired assertion. \square

To conclude this subsection, we explain how to apply Theorem 3.2 to prove Theorem 1.4 and then we discuss how to apply Theorem 3.2 to McKean-Vlasov SDEs with the drift of the form (1.8).

Proof of Theorem 1.4. It is easily seen from condition (2-ii) that assumption (3.6) of Theorem 3.2 holds with $\sigma(r) = c_0 r^{1-\alpha}$. In particular, $\sigma(r) \in C([0, 2l_0]) \cap C^2((0, 2l_0])$ is nondecreasing and concave, and $\int_0^r \sigma(s)^{-1} ds$ is well defined for all $r \in [0, 2l_0]$. Hence, using Theorem 3.2 and Remark 3.3(ii), we obtain the assertion of Theorem 1.4. \square

Remark 3.4. For McKean-Vlasov SDEs with the drift of the form (1.8), we can easily check that condition (1-i), combined with the assumption that

$$(3.14) \quad |b_2(x, z) - b_2(y, z')| \leq K_{b_2}(|x - y| + |z - z'|)$$

holds for some constant $K_{b_2} > 0$ and for all x, y, z and $z' \in \mathbb{R}^d$, implies (2-i). First, we claim that (3.14) implies that for any $x, y \in \mathbb{R}^d$ and $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$,

$$(3.15) \quad \left| \int b_2(x, z) \mu_1(dz) - \int b_2(y, z) \mu_2(dz) \right| \leq K_{b_2}(|x - y| + W_1(\mu_1, \mu_2)).$$

Indeed, let π be the coupling of μ_1 and μ_2 for which the infimum in the definition of $W_1(\mu_1, \mu_2)$ is attained. Then, for any $x, y \in \mathbb{R}^d$ and $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\begin{aligned} & \left| \int b_2(x, z) \mu_1(dz) - \int b_2(y, z) \mu_2(dz) \right| \\ &= \left| \iint b_2(x, z) \pi(dz, dz') - \iint b_2(y, z') \pi(dz, dz') \right| \\ &\leq \int \int |b_2(x, z) - b_2(y, z')| \pi(dz, dz') \\ &\leq K_{b_2}(|x - y| + W_1(\mu_1, \mu_2)). \end{aligned}$$

Now, according to (3.15) and the continuity of $b_1(x)$ on \mathbb{R}^d , we can see that $b(x, \mu) = b_1(x) + \int b_2(x, z) \mu(dz)$ is continuous on $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$. On the other hand, it immediately follows from condition (1-i) and (3.15) that (1.13) is satisfied. Moreover, for any $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, due to (3.14) again,

$$\begin{aligned} |b(0, \mu)| &\leq |b_1(0)| + \int |b_2(0, z)| \mu(dz) \\ &\leq |b_1(0)| + \int (|b_2(0, 0)| + K_{b_2}|z|) \mu(dz) \leq C_0 \left(1 + \int |z| \mu(dz) \right). \end{aligned}$$

Hence, (2-i) holds true. We shall note that (3.14) holds true for any $b_2(x, z)$ which satisfies condition (1-ii).

3.2. Convergence in W_1 : the Lyapunov function approach. To study exponential convergence of the McKean-Vlasov SDE (1.7) without assuming the contractivity at infinity condition $\mathbf{B}(K_1, K_2, l_0; K_3)$ on the drift, we will make use of the Lyapunov function approach instead. Though the arguments below are partly motivated by these in Subsection 2.3, we will see that the case for the distribution-dependent drift is much more complex and delicate than that for the distribution-independent drift. To this end, we will assume that the drift term $b(x, \mu)$ is of the form

$$(3.16) \quad b(x, \mu) = b_1(x) + b_2(x, \mu),$$

where $b_1(x)$ is such that there exist constants λ and $C_0 > 0$ such that for all $x \in \mathbb{R}^d$,

$$(3.17) \quad \langle b_1(x), x \rangle \leq -\lambda|x|^2 + C_0.$$

We assume (3.17) in order to be able to construct a Lyapunov function for (1.7). Note that the general Lyapunov condition (2.17) that we used in the distribution-independent case does not have a straightforward counterpart in the McKean-Vlasov case, cf. the proof of Lemma 3.5 below, and hence we work directly with (3.17). Moreover, (3.17) seems to be a natural condition in the study of the exponential ergodicity of McKean-Vlasov SDEs (see [16, Assumption 2.7]). We will also need the following assumption:

Assumption (C)

- (i) *There are constants K_1, K_2 and $K_3 > 0$ such that for all $x_1, x_2 \in \mathbb{R}^d$ with $x_1 \neq x_2$ and $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$,*

$$\frac{\langle b_1(x_1) - b_1(x_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \leq K_1|x_1 - x_2|$$

and

$$\frac{\langle b_2(x_1, \mu_1) - b_2(x_2, \mu_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \leq K_2|x_1 - x_2| + K_3W_1(\mu_1, \mu_2).$$

(ii) There is a constant $B_0 > 0$ such that for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$|b_2(x, \mu)| \leq B_0 \left(1 + \int |z| \mu(dz) + |x| \right).$$

(iii) For any $\theta > 0$, there exists a measure $0 < \nu_\theta \leq \nu$ such that $\text{supp } \nu_\theta \subset B(0, 1)$, $\int_{\{|z| \leq 1\}} |z| \nu_\theta(dz) \leq \theta$, and

$$\liminf_{r \rightarrow 0} \inf_{s \in (0, r]} J_{\nu_\theta}(s) s^\alpha > 0,$$

where $\alpha := \alpha(\theta) \in (0, 1)$ and

$$J_{\nu_\theta}(s) := \inf_{x \in \mathbb{R}^d: |x| \leq s} [\nu_\theta \wedge (\delta_x * \nu_\theta)](\mathbb{R}^d) > 0.$$

It is easy to see that under Assumption **(C)**(i)–(ii), both (3.1) and (3.2) are satisfied. Thus, according to Proposition 3.1, if $\int_{\{|z| > 1\}} |z| \nu(dz) < \infty$, then the SDE (1.7) has a unique non-explosive strong solution $(X_t)_{t \geq 0}$ such that $\mathbb{E}|X_t| < \infty$ for all $t > 0$.

Let us begin by discussing how to construct a Lyapunov function for the McKean-Vlasov SDE (1.7) under Assumption **(C)**.

Lemma 3.5. *Let the drift term $b(x, \mu)$ be of the form (3.16), and satisfy (3.17) and Assumption **(C)**(i) and (ii). Let $\int_{\{|z| > 1\}} |z| \nu(dz) < \infty$, and let $V \in C^2(\mathbb{R}^d)$ be a radial function such that $V(x) \geq 1$ for all $x \in \mathbb{R}^d$, $V(x) = 1 + |x|$ for all $|x| \geq 1$ and $\|\nabla V\|_\infty + \|\nabla^2 V\|_\infty < \infty$. Then, there is a constant $C > 0$ (independent of B_0) such that for all $t > 0$,*

$$dV(X_t) \leq [-(\lambda - 2\|\nabla V\|_\infty B_0)V(X_t) + \|\nabla V\|_\infty B_0((\mathbb{E}|X_t|) - |X_t|) + C(1 + B_0)] dt + dM_t^V,$$

where $(M_t^V)_{t \geq 0}$ is a martingale, λ is given in (3.17) and B_0 is the constant in Assumption **(C)**(ii). In particular, when $0 < 2\|\nabla V\|_\infty B_0 < \lambda$, for any X_0 such that $\mathbb{E}V(X_0) < \infty$ and for any $t > 0$,

$$(3.18) \quad \mathbb{E}V(X_t) \leq \mathbb{E}V(X_0)e^{-(\lambda - 2\|\nabla V\|_\infty B_0)t} + C(1 + B_0)/(\lambda - 2\|\nabla V\|_\infty B_0).$$

Proof. Let $V \in C^2(\mathbb{R}^d)$ be a radial function such that $V(x) \geq 1$ for all $x \in \mathbb{R}^d$, $V(x) = 1 + |x|$ for all $|x| \geq 1$, and $\|\nabla V\|_\infty + \|\nabla^2 V\|_\infty < \infty$. Let μ_{X_t} be the distribution of X_t . Then, by the Itô formula, it holds that

$$dV(X_t) = L[\mu_{X_t}]V(X_t) dt + dM_t^V,$$

where $L[\mu]$ is defined in (3.3) and

$$M_t^V := \int_0^t \int_{\mathbb{R}^d} (V(X_{s-} + z) - V(X_{s-})) \tilde{N}(ds, dz)$$

is a martingale, thanks to the assumption that $\int_{\{|z| \geq 1\}} |z| \nu(dz) < \infty$.

Using the mean value theorem and the fact that $\int_{\{|z|\geq 1\}} |z| \nu(dz) < \infty$ again, we find that

$$\begin{aligned} & \int (V(x+z) - V(x) - \langle \nabla V(x), z \rangle \mathbf{1}_{\{|z|\leq 1\}}) \nu(dz) \\ &= \int_{\{|z|\leq 1\}} (V(x+z) - V(x) - \langle \nabla V(x), z \rangle) \nu(dz) + \int_{\{|z|\geq 1\}} (V(x+z) - V(x)) \nu(dz) \\ &\leq \frac{1}{2} \|\nabla^2 V\|_\infty \int_{\{|z|\leq 1\}} |z|^2 \nu(dz) + \|\nabla V\|_\infty \int_{\{|z|\geq 1\}} |z| \nu(dz) =: C_1, \end{aligned}$$

where in the last step we used the definition of V (i.e., the function $V \in C^2(\mathbb{R}^d)$ such that $\|\nabla V\|_\infty + \|\nabla^2 V\|_\infty < \infty$).

On the other hand, by Assumption **(C)**(ii),

$$\langle b_2(x, \mu), \nabla V(x) \rangle \leq |b_2(x, \mu)| \|\nabla V\|_\infty \leq \|\nabla V\|_\infty B_0 \left(1 + \int |z| \mu(dz) + |x| \right).$$

In order to deal with the term $\langle b_1(x), \nabla V(x) \rangle$, we write

$$\langle b_1(x), \nabla V(x) \rangle = \langle b_1(x), \nabla V(x) \rangle \mathbf{1}_{\{|x|\leq 1\}} + \langle b_1(x), \nabla V(x) \rangle \mathbf{1}_{\{|x|>1\}}$$

and recall that $V(x)$ is radial, which means that there exists a function $\bar{V} : [0, \infty) \rightarrow [1, \infty)$ such that $V(x) = \bar{V}(|x|)$ for all $x \in \mathbb{R}^d$. Since by Assumption **(C)**(i),

$$\langle b_1(x), x \rangle \leq K_1 |x|^2 + |b_1(0)| \cdot |x|,$$

we have

$$\langle b_1(x), \nabla V(x) \rangle \mathbf{1}_{\{|x|\leq 1\}} = \frac{\langle b_1(x), x \rangle}{|x|} \bar{V}'(|x|) \mathbf{1}_{\{|x|\leq 1\}} \leq (K_1 + |b_1(0)|) \|\nabla V\|_\infty$$

where we used $\|\bar{V}'\|_\infty = \|\nabla V\|_\infty$. Moreover, according to (3.17),

$$\langle b_1(x), \nabla V(x) \rangle \mathbf{1}_{\{|x|>1\}} \leq -\lambda |x| \mathbf{1}_{\{|x|>1\}} + C_0,$$

where we used the facts that $V(x) = |x| + 1$ for $|x| > 1$, which implies $\bar{V}'(|x|) = 1$ for $|x| > 1$, and $C_0/|x| \leq C_0$ for $|x| > 1$.

Hence we get

$$\begin{aligned} L[\mu]V(x) &\leq C_1 + \|\nabla V\|_\infty B_0 \left(1 + \int |z| \mu(dz) + |x| \right) + (K_1 + |b_1(0)|) \|\nabla V\|_\infty \\ &\quad - \lambda(|x| + 1) \mathbf{1}_{\{|x|>1\}} + \lambda + C_0 \\ &= C_1 + \|\nabla V\|_\infty B_0 \left(1 + \int |z| \mu(dz) + |x| \right) + (K_1 + |b_1(0)|) \|\nabla V\|_\infty \\ &\quad - \lambda V(x) + \lambda V(x) \mathbf{1}_{\{|x|\leq 1\}} + \lambda + C_0 \\ &\leq C_1 + \|\nabla V\|_\infty B_0 \left(\int |z| \mu(dz) + V(x) \right) + (K_1 + |b_1(0)|) \|\nabla V\|_\infty \\ &\quad - \lambda V(x) + \lambda \sup_{x \in B(0,1)} V(x) + \lambda + C_0 + \|\nabla V\|_\infty B_0, \end{aligned}$$

where in the second step we used the fact that $V(x) = |x| + 1$ for $|x| > 1$ and the last inequality follows from $V(x) \geq |x|$ for all $x \in \mathbb{R}^d$. We conclude that there is a

constant $C > 0$ given by

$$C := \max \left(\|\nabla V\|_\infty, C_1 + C_0 + (K_1 + |b_1(0)|)\|\nabla V\|_\infty + \lambda \left(1 + \sup_{x \in B(0,1)} V(x) \right) \right)$$

such that for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$L[\mu]V(x) \leq -(\lambda - \|\nabla V\|_\infty B_0)V(x) + \|\nabla V\|_\infty B_0 \int |z| \mu(dz) + C(1 + B_0).$$

With this inequality at hand, we obtain

$$\begin{aligned} dV(X_t) &= L[\mu_{X_t}]V(X_t) dt + dM_t^V \\ &\leq [-(\lambda - \|\nabla V\|_\infty B_0)V(X_t) + \|\nabla V\|_\infty B_0 \mathbb{E}|X_t| + C(1 + B_0)] dt + dM_t^V \\ &\leq [-(\lambda - 2\|\nabla V\|_\infty B_0)V(X_t) + \|\nabla V\|_\infty B_0(\mathbb{E}|X_t|) - |X_t| + C(1 + B_0)] dt \\ &\quad + dM_t^V, \end{aligned}$$

where in the last step we used the fact that $|x| \leq V(x)$ for all $x \in \mathbb{R}^d$. This proves the first assertion. Furthermore, since $\mathbb{E}|X_t| < \infty$ for all $t > 0$, we can obtain the second one. \square

Remark 3.6. Note that when $b_2(x, \mu) = 0$ for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$ in (3.16), the proof of Lemma 3.5 still works in the same way. Hence, assuming $\int_{\{|z|>1\}} |z| \nu(dz) < \infty$ and $\langle b(x), x \rangle \leq -\lambda|x|^2 + C_0$ for all $x \in \mathbb{R}^d$ with some $\lambda, C_0 > 0$, we obtain a Lyapunov function V as stated in Lemma 3.5 for the distribution independent SDE (2.1). In particular, this Lyapunov function satisfies both Assumption **(A)**(iii) and Assumption **(B)**(iii).

We will now extend Theorem 2.7 to the McKean-Vlasov SDE given by (1.7).

Theorem 3.7. *Suppose that the drift $b(x, \mu)$ is of the form (3.16) such that (3.17) and Assumption **(C)** are satisfied. Suppose also that $\int_{\{|z|>1\}} |z| \nu(dz) < \infty$. Let V be the Lyapunov function from Lemma 3.5, and let μ_{X_0} and μ_{Y_0} be probability measures such that the integrals $\mu_{X_0}(V)$ and $\mu_{Y_0}(V)$ are finite (which is equivalent to saying that both μ_{X_0} and μ_{Y_0} have finite first moment). Then there are constants K_2^* , K_3^* and $B_0^* > 0$ such that for all $K_2 \in (0, K_2^*)$, $K_3 \in (0, K_3^*)$, $B_0 \in (0, B_0^*)$ and $t > 0$,*

$$W_\Phi(\mu_{X_t}, \mu_{Y_t}) \leq C_0 e^{-\lambda_0 t} (\mu_{X_0}(V) + \mu_{Y_0}(V))^2,$$

where $\Phi(x, y) = (|x - y| \wedge 1)(V(x) + V(y))$, and $C_0, \lambda_0 > 0$ are constants independent of μ_{X_0} , μ_{Y_0} and t .

Proof. (i) Let V be the Lyapunov function from Lemma 3.5. Choose ψ as that in Lemma 2.9, and define $F(x, y) = \psi(|x - y|)(1 + \varepsilon(V(x) + V(y)))$ for any $x, y \in \mathbb{R}^d$ and some $\varepsilon > 0$. Then, following the proof of Theorem 2.7 and using (3.17) and Assumption **(C)**(i) and (iii), we can find constants $\varepsilon, \lambda_0 > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$(3.19) \quad \tilde{L}F(x, y) \leq -\lambda_0 F(x, y),$$

where \tilde{L} is the operator given by (2.8) with b replaced by b_1 . Note that in the argument above, in order to verify Assumption **(B)**(iii), we used (3.17) and the properties of the function V , cf. Remark 3.6. Note also that the constructions of functions ψ and F are independent of $b_2(x, \mu)$, and, in particular, the constants ε, λ_0 are independent of K_2, K_3 and B_0 in Assumptions **(C)**(i) and (ii).

Furthermore, for any $x_1, x_2 \in \mathbb{R}^d$ and $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\begin{aligned}
 & \frac{\langle b_2(x_1, \mu_1) - b_2(x_2, \mu_2), x_1 - x_2 \rangle}{|x_1 - x_2|} \psi'(|x_1 - x_2|) (1 + \varepsilon(V(x_1) + V(x_2))) \\
 & + \varepsilon \psi(|x_1 - x_2|) (\langle b_2(x_1, \mu_1), \nabla V(x_1) \rangle + \langle b_2(x_2, \mu_2), \nabla V(x_2) \rangle) \\
 & \leq K_2 \psi'(|x_1 - x_2|) |x_1 - x_2| (1 + \varepsilon(V(x_1) + V(x_2))) \\
 & + K_3 \psi'(|x_1 - x_2|) W_1(\mu_1, \mu_2) (1 + \varepsilon(V(x_1) + V(x_2))) \\
 & + \varepsilon B_0 \|\nabla V\|_\infty \psi(|x_1 - x_2|) \left(2 + \int |z| \mu_1(dz) + \int |z| \mu_2(dz) + |x_1| + |x_2| \right) \\
 & \leq (K_2 + 2B_0 \|\nabla V\|_\infty) \psi(|x_1 - x_2|) \\
 & \quad \times \left[1 + \varepsilon(V(x_1) + V(x_2)) + \varepsilon \left(\int |z| \mu_1(dz) + \int |z| \mu_2(dz) \right) \right] \\
 & + K_3 \psi'(0) W_1(\mu_1, \mu_2) (1 + \varepsilon(V(x_1) + V(x_2))),
 \end{aligned}$$

where we used Assumptions **(C)**(i) and (ii) in the first inequality, and in the second inequality we used $V(x) \geq |x|$ and the fact that $\psi'(r)r \leq \psi(r)$ (since $\psi(0) = 0$ and ψ' is decreasing).

(ii) In the following, let L be the operator given by (2.2) with b replaced by b_1 . In particular, \tilde{L} above is a coupling operator of L . Motivated by the proof of Theorem 3.2, we want to replace the component \tilde{L}_Z of \tilde{L} in (2.8) with the generator \tilde{L}_Z^δ corresponding to the combination of the refined basic coupling and the synchronous coupling, defined by (2.13). Note that, for the coupling operator \tilde{L}^δ , the equality (2.26) in the proof of Theorem 2.7 becomes

$$\begin{aligned}
 & \tilde{L}^\delta F(x, y) \\
 & = \tilde{L}^\delta \psi(|x - y|) \cdot (1 + \varepsilon(V(x) + V(y))) + \psi(|x - y|) \cdot \varepsilon L(V(x) + V(y)) \\
 & + \phi_\delta(|x - y|) \cdot \frac{1}{2} \varepsilon (\psi(|x - y - (x - y)_\kappa|) - \psi(|x - y|)) \\
 & \quad \times \int [(V(x + z) - V(x)) + (V(y + z + (x - y)_\kappa) - V(y))] \mu_{\theta, (y-x)_\kappa}(dz) \\
 & + \phi_\delta(|x - y|) \cdot \frac{1}{2} \varepsilon (\psi(|x - y + (x - y)_\kappa|) - \psi(|x - y|)) \\
 & \quad \times \int [(V(x + z) - V(x)) + (V(y + z - (x - y)_\kappa) - V(y))] \mu_{\theta, (x-y)_\kappa}(dz) \\
 & = \tilde{L} F(x, y) \cdot \phi_\delta(|x - y|) + \varepsilon \psi(|x - y|) L(V(x) + V(y)) \cdot (1 - \phi_\delta(|x - y|)).
 \end{aligned}$$

Hence, using (3.19) and repeating the reasoning from the proof of Theorem 3.2, we obtain

$$\begin{aligned}
 \tilde{L}^\delta F(x, y) & \leq -\lambda_0 F(x, y) + \lambda_0 F(x, y) \cdot (1 - \phi_\delta(|x - y|)) \\
 & + \varepsilon \psi(|x - y|) L(V(x) + V(y)) \cdot (1 - \phi_\delta(|x - y|)) \\
 & + K_1 \psi'(|x - y|) |x - y| \cdot (1 - \phi_\delta(|x - y|)),
 \end{aligned}$$

where for the last term we used Assumption **(C)**(i), i.e., b_1 is one-sided Lipschitz with the constant K_1 . Thus, due to the continuity of ψ at zero and the fact that both ψ'

and LV are bounded, we see that for any $x, y \in \mathbb{R}^d$,

$$\tilde{L}^\delta F(x, y) \leq -\lambda_0 F(x, y) + \lambda_0 \psi(\delta)(1 + \varepsilon(V(x) + V(y))) + C(\delta),$$

where $C(\delta)$ is independent of K_2, K_3 and B_0 , and such that $C(\delta) \rightarrow 0$ as $\delta \rightarrow 0$.

(iii) Now, as in part (ii) of the proof of Theorem 3.2, we can define a coupling process $(X_t, Y_t^\delta)_{t \geq 0}$ of the process $(X_t)_{t \geq 0}$, by using the system of SDEs (3.9). Similarly as in the proof of Theorem 2.7, we apply the refined basic coupling only to the component ν_θ of the Lévy measure ν . Under Assumption **(C)**, we can verify the existence of a weak solution to (3.9) in the present setting. Furthermore, by the Itô formula,

$$\begin{aligned} & dF(X_t, Y_t^\delta) \\ &= \tilde{L}^\delta F(X_t, Y_t^\delta) dt \\ &+ \left(\frac{\psi'(|X_t - Y_t^\delta|)}{|X_t - Y_t^\delta|} \langle b_2(X_t, \mu_{X_t}) - b_2(Y_t^\delta, \mu_{Y_t^\delta}), X_t - Y_t^\delta \rangle \right) \\ &\quad \times (1 + \varepsilon V(X_t) + \varepsilon V(Y_t^\delta)) dt \\ &+ \varepsilon \psi(|X_t - Y_t^\delta|) \left(\langle b_2(X_t, \mu_{X_t}), \nabla V(X_t) \rangle + \langle b_2(Y_t^\delta, \mu_{Y_t^\delta}), \nabla V(Y_t^\delta) \rangle \right) dt \\ &+ (1 + \varepsilon V(X_t) + \varepsilon V(Y_t^\delta)) dM_t^{\psi, \delta} + 2\varepsilon \psi(|X_t - Y_t^\delta|) dM_t^V, \end{aligned}$$

where both $(M_t^{\psi, \delta})_{t \geq 0}$ and $(M_t^V)_{t \geq 0}$ are martingales. Using all the estimates in parts (i) and (ii), we can bound the drift component in the inequality above by

$$\begin{aligned} & -\lambda_0 F(X_t, Y_t^\delta) + \lambda_0 \psi(\delta)(1 + \varepsilon(V(X_t) + V(Y_t^\delta))) + C(\delta) \\ &+ (K_2 + 2B_0 \|\nabla V\|_\infty) \psi(|X_t - Y_t^\delta|) (1 + \varepsilon(V(X_t) + V(Y_t^\delta)) + \mathbb{E}|X_t| + \mathbb{E}|Y_t^\delta|) \\ &+ K_3 \psi'(0) (\mathbb{E}|X_t - Y_t^\delta|) \cdot (1 + \varepsilon(V(X_t) + V(Y_t^\delta))) \\ &\leq -\lambda_0 F(X_t, Y_t^\delta) + \lambda_0 \psi(\delta)(1 + \varepsilon(V(X_t) + V(Y_t^\delta))) + C(\delta) \\ &+ (K_2 + 2B_0 \|\nabla V\|_\infty) F(X_t, Y_t^\delta) \\ &+ (K_2 + 2B_0 \|\nabla V\|_\infty) \psi(|X_t - Y_t^\delta|) \varepsilon (\mathbb{E}|X_t| + \mathbb{E}|Y_t^\delta|) \\ &+ K_3 \psi'(0) (\mathbb{E}|X_t - Y_t^\delta|) \cdot (1 + \varepsilon(V(X_t) + V(Y_t^\delta))) \\ &\leq -[\lambda_0 - (K_2 + 2B_0 \|\nabla V\|_\infty)] F(X_t, Y_t^\delta) \\ &+ (K_2 + 2B_0 \|\nabla V\|_\infty) \psi'(0) |X_t - Y_t^\delta| (1 + \varepsilon \mathbb{E}V(X_t) + \varepsilon \mathbb{E}V(Y_t^\delta)) \\ &+ (\lambda_0 \psi(\delta) + K_3 \psi'(0) \mathbb{E}|X_t - Y_t^\delta|) (1 + \varepsilon(V(X_t) + V(Y_t^\delta))) + C(\delta), \end{aligned}$$

where in the second inequality we used the facts that $V(x) \geq |x|$ for all $x \in \mathbb{R}^d$ and $\psi(r) \leq \psi'(0)r$ for all $r \geq 0$.

Hence, applying all the estimates above, we find that for any $\lambda_*, t > 0$ and for any $X_0 \sim \mu_{X_0}$ and $Y_0^\delta \sim \mu_{Y_0}$,

$$\begin{aligned} & e^{\lambda_* t} \mathbb{E}F(X_t, Y_t^\delta) \\ &\leq \mathbb{E}F(X_0, Y_0^\delta) + [\lambda_* - (\lambda_0 - (K_2 + 2B_0 \|\nabla V\|_\infty))] \int_0^t e^{\lambda_* s} \mathbb{E}F(X_s, Y_s^\delta) ds \\ &+ \lambda_0 \psi(\delta) \int_0^t e^{\lambda_* s} (1 + \varepsilon(\mathbb{E}V(X_s) + \mathbb{E}V(Y_s^\delta))) ds \\ &+ (K_3 + K_2 + 2B_0 \|\nabla V\|_\infty) \psi'(0) \end{aligned}$$

$$\begin{aligned}
 & \times \int_0^t e^{\lambda_* s} (\mathbb{E}|X_s - Y_s^\delta|) (1 + \varepsilon(\mathbb{E}V(X_s) + \mathbb{E}V(Y_s^\delta))) ds \\
 & + C(\delta) \int_0^t e^{\lambda_* s} ds \\
 & =: \mathbb{E}F(X_0, Y_0^\delta) + \sum_{i=1}^4 I_i.
 \end{aligned}$$

Next, we will give bounds for I_2 and I_3 . We first choose $B_0 \in (0, \lambda/(2\|\nabla V\|_\infty))$, where λ is the constant given in (3.17). Then, according to Lemma 3.5,

$$\begin{aligned}
 I_2 & \leq \lambda_0 \psi(\delta) \int_0^t e^{\lambda_* s} \\
 & \quad \times \left(1 + \varepsilon(\mathbb{E}V(X_0) + \mathbb{E}V(Y_0^\delta)) e^{-(\lambda - 2\|\nabla V\|_\infty B_0)s} + \frac{2\varepsilon C(1 + B_0)}{\lambda - 2\|\nabla V\|_\infty B_0} \right) ds \\
 & \leq \lambda_0 \psi(\delta) \left(1 + \frac{2\varepsilon C(1 + B_0)}{\lambda - 2\|\nabla V\|_\infty B_0} + \varepsilon \mu_{X_0}(V) + \varepsilon \mu_{Y_0}(V) \right) \int_0^t e^{\lambda_* s} ds \\
 & =: \tilde{C}(\delta) \int_0^t e^{\lambda_* s} ds,
 \end{aligned}$$

where $\tilde{C}(\delta)$ satisfies $\tilde{C}(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. We will now proceed with bounding I_3 . In the arguments below, the constants C_i ($i = 1, 2, \dots, 4$) are all independent of $X_0, Y_0^\delta, t, \delta, K_2, K_3$ and B_0 . By the definition of $F(x, y)$ and the properties of V , there is a constant $C_1 \geq 1$ such that for all $x, y \in \mathbb{R}^d$,

$$C_1^{-1}|x - y| \leq F(x, y) \leq C_1(V(x) + V(y)).$$

According to Lemma 3.5 again, for all $s > 0$,

$$\begin{aligned}
 & 1 + \varepsilon(\mathbb{E}V(X_s) + \mathbb{E}V(Y_s^\delta)) \\
 (3.20) \quad & \leq C_2 \left(1 + \frac{1 + B_0}{\lambda - 2\|\nabla V\|_\infty B_0} \right) + C_2 \mathbb{E}(V(X_0) + V(Y_0^\delta)) e^{-(\lambda - 2\|\nabla V\|_\infty B_0)s}.
 \end{aligned}$$

Hence,

$$\begin{aligned}
 I_3 & \leq C_3 (K_3 + K_2 + B_0) \left(1 + \frac{1 + B_0}{\lambda - 2\|\nabla V\|_\infty B_0} \right) \int_0^t e^{\lambda_* s} \mathbb{E}F(X_s, Y_s^\delta) ds \\
 & \quad + C_3 (K_3 + K_2 + B_0) \mathbb{E}(V(X_0) + V(Y_0^\delta)) \int_0^t e^{[\lambda_* - (\lambda - 2\|\nabla V\|_\infty B_0)]s} \mathbb{E}F(X_s, Y_s^\delta) ds \\
 & \leq C_3 (K_3 + K_2 + B_0) \left(1 + \frac{1 + B_0}{\lambda - 2\|\nabla V\|_\infty B_0} \right) \int_0^t e^{\lambda_* s} \mathbb{E}F(X_s, Y_s^\delta) ds \\
 & \quad + C_4 (K_3 + K_2 + B_0) \left(1 + \frac{1 + B_0}{\lambda - 2\|\nabla V\|_\infty B_0} \right) \\
 & \quad \times (\mathbb{E}(V(X_0) + V(Y_0^\delta)))^2 \int_0^t e^{[\lambda_* - (\lambda - 2\|\nabla V\|_\infty B_0)]s} ds,
 \end{aligned}$$

where in the first inequality we used the fact that $|x - y| \leq C_1 F(x, y)$ for all $x, y \in \mathbb{R}^d$ and (3.20), while in the second inequality we used the fact that $F(x, y) \leq$

$C_1(V(x) + V(y))$ for all $x, y \in \mathbb{R}^d$ and Lemma 3.5 again, as well as

$$\mathbb{E}(V(X_0) + V(Y_0^\delta)) \leq (\mathbb{E}(V(X_0) + V(Y_0^\delta)))^2$$

since $V \geq 1$.

Combining all the estimates above, we arrive at

$$\begin{aligned} & e^{\lambda_* t} \mathbb{E}F(X_t, Y_t^\delta) \\ & \leq \mathbb{E}F(X_0, Y_0^\delta) \\ & \quad + \left[\lambda_* + K_2 + 2B_0 \|\nabla V\|_\infty + C_3(K_3 + K_2 + B_0) \left(1 + \frac{1 + B_0}{\lambda - 2\|\nabla V\|_\infty B_0} \right) \right. \\ (3.21) \quad & \quad \left. - \lambda_0 \right] \int_0^t e^{\lambda_* s} \mathbb{E}F(X_s, Y_s^\delta) ds \\ & \quad + C_5(K_2, K_3, B_0) (\mu_{X_0}(V) + \mu_{Y_0}(V))^2 \int_0^t e^{(\lambda_* + 2\|\nabla V\|_\infty B_0 - \lambda)s} ds \\ & \quad + (C(\delta) + \tilde{C}(\delta)) \int_0^t e^{\lambda_* s} ds. \end{aligned}$$

Here and in what follows, the constants $C_i(K_2, K_3, B_0)$ ($i = 5, 6, 7$) may depend on K_2, K_3 and B_0 . Choosing K_2^*, K_3^* and B_0^* small enough so that

$$K_2^* + 2B_0^* \|\nabla V\|_\infty + C_3(K_3^* + K_2^* + B_0^*) \left(1 + \frac{1 + B_0^*}{\lambda - 2\|\nabla V\|_\infty B_0^*} \right) \leq \lambda_0/2,$$

and then letting $\lambda_* \leq \lambda_0/2$, we can bound the second term on the right hand side of (3.21) by zero. Moreover, if we choose $\lambda_* < \lambda - 2\|\nabla V\|_\infty B_0$, then we can bound the integral in the third term by a constant independent of t . As a consequence, we find that for any $K_2 \in (0, K_2^*)$, $K_3 \in (0, K_3^*)$, $B_0 \in (0, B_0^*)$, $\lambda_* < \min\{\lambda_0/2, \lambda - 2\|\nabla V\|_\infty B_0^*\}$ and $t > 0$,

$$\begin{aligned} \mathbb{E}F(X_t, Y_t^\delta) & \leq C_6(K_2, K_3, B_0) e^{-\lambda_* t} (\mathbb{E}F(X_0, Y_0^\delta) + (\mu_{X_0}(V) + \mu_{Y_0}(V))^2) \\ & \quad + (C(\delta) + \tilde{C}(\delta))/\lambda_* \\ & \leq C_7(K_2, K_3, B_0) e^{-\lambda_* t} (\mu_{X_0}(V) + \mu_{Y_0}(V))^2 + (C(\delta) + \tilde{C}(\delta))/\lambda_*, \end{aligned}$$

where in the second inequality we used the fact that

$$\mathbb{E}F(X_0, Y_0^\delta) \leq C_1 (\mu_{X_0}(V) + \mu_{Y_0}(V)) \leq C_1 (\mu_{X_0}(V) + \mu_{Y_0}(V))^2.$$

Letting $\delta \rightarrow 0$ in the inequality above, we can prove the desired assertion by following the arguments from the final part of the proof of Theorem 3.2. \square

At the end of this part, we give the proof of Theorem 1.5.

Proof of Theorem 1.5. Combining Theorem 3.7 with the fact that for all probability measures $\mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)$ we have $W_1(\mu_1, \mu_2) \leq CW_\Phi(\mu_1, \mu_2)$ for some constant $C > 0$ (cf. the proof of Corollary 2.8), we obtain

$$(3.22) \quad W_1(\mu_{X_t}, \mu_{Y_t}) \leq C_1 e^{-\lambda_0 t} (\mu_{X_0}(V) + \mu_{Y_0}(V))^2$$

for some constant $C_1 > 0$. Here the constant $\lambda_0 > 0$, the function V and the measures $\mu_{X_0}, \mu_{Y_0}, \mu_{X_t}$ and μ_{Y_t} are as in the statement of Theorem 3.7. Similarly as in the proof of Theorem 1.4, we can follow e.g. the proof of [47, Theorem 3.1(2)] and use (3.18) to conclude that there exists a unique invariant measure μ for the solution

$(X_t)_{t \geq 0}$ to (1.7). Hence we can consider our bound (3.22) with $(Y_t)_{t \geq 0}$ initiated at $\mu_{Y_0} = \mu$, which concludes the proof. \square

4. PROPAGATION OF CHAOS

This section is devoted to the proof of Theorem 1.2.

Proof of Theorem 1.2. The proof is based on the idea of the proof for [15, Theorem 2]. For convenience, we denote $\tilde{b}_2 = b_2$, i.e., we assume that the drift $b(x, \mu)$ is given by

$$b(x, \mu) = b_1(x) + \int b_2(x - z) \mu(dz) = b_1(x) + b_2 * \mu(x).$$

In particular, we denote $K_{\tilde{b}_2} = K_{b_2}$ in condition (1.11). We first follow part (i) of the proof of Theorem 3.2. Based on the contractivity at infinity condition (1.10) for $b_1(x)$, we construct a function ψ as in Theorem 3.2, with constants K_1, K_2, l_0 replaced by $K_{1,b_1}, K_{2,b_1}, r_{b_1}$, respectively. Now observe that due to (1.12), for $l_0 = r_{b_1}$ we can find a function $\sigma(r)$ satisfying condition (3.6) for $r \in (0, 2r_{b_1}]$, cf. the proof of [33, Theorem 1.1]. Hence, proceeding as in (3.7), we have

$$(4.1) \quad \begin{aligned} & \psi'(|x - y|) \left[K_{1,b_1} |x - y| \mathbf{1}_{\{|x-y| \leq r_{b_1}\}} - K_{2,b_1} |x - y| \mathbf{1}_{\{|x-y| \geq r_{b_1}\}} \right] + \tilde{L}_Z \psi(|x - y|) \\ & \leq -\lambda_0 \psi(|x - y|), \end{aligned}$$

where \tilde{L}_Z is the refined basic coupling operator for Z and $\lambda_0 > 0$ is an explicit constant defined as in the proof of Theorem 3.2, depending only on $K_{1,b_1}, K_{2,b_1}, r_{b_1}$ and properties of the Lévy measure ν . In particular, the construction of ψ and the constant λ_0 is independent of K_{b_2} .

Motivated by part (ii) of the proof of Theorem 3.2, for any fixed $\delta > 0$ we consider the following coupling between the independent McKean-Vlasov processes and the mean-field particle system

$$\begin{cases} d\bar{X}_t^i = b_1(\bar{X}_t^i) + b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i) dt + dZ_t^i, \\ dX_t^{i,n,\delta} = b_1(X_t^{i,n,\delta}) + \frac{1}{n} \sum_{j=1}^n b_2(X_t^{i,n,\delta} - X_t^{j,n,\delta}) dt + dZ_t^i + dL_t^{*,i,\delta}, \quad 1 \leq i \leq n, \end{cases}$$

where $(L_t^{*,i,\delta})_{t \geq 0}$ is defined by analogy to (2.14), with U_t^δ replaced by $U_t^{i,\delta} := \bar{X}_t^i - X_t^{i,n,\delta}$. Here we assume that $(\bar{X}_0^i, X_0^{i,n,\delta})$ are independent random variables with the same law having finite second moments, and $(Z_t^i)_{t \geq 0}$ are independent Lévy processes associated with the same Lévy measure ν . We can follow step (ii) in the proof of Theorem 3.2 to show that the system above has a weak solution $(\bar{X}_t^i, X_t^{i,n,\delta})_{t \geq 0}$ for all $1 \leq i \leq n$.

Below, let $U_t^{i,\delta} = \bar{X}_t^i - X_t^{i,n,\delta}$ for all $t > 0$. Then, by (4.1) and (3.12), for all $t > 0$ and $1 \leq i \leq n$,

$$\begin{aligned} d\psi(|U_t^{i,\delta}|) & \leq \psi'(|U_t^{i,\delta}|) \frac{\langle b_1(\bar{X}_t^i) - b_1(X_t^{i,n,\delta}), U_t^{i,\delta} \rangle}{|U_t^{i,\delta}|} dt + \tilde{L}_Z^\delta \psi(|U_t^{i,\delta}|) dt \\ & \quad + \psi'(|U_t^{i,\delta}|) \frac{\langle b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i) - \frac{1}{n} \sum_{j=1}^n b_2(X_t^{i,n,\delta} - X_t^{j,n,\delta}), U_t^{i,\delta} \rangle}{|U_t^{i,\delta}|} dt + dM_t^{\psi,i,\delta} \\ & \leq -\lambda_0 \psi(|U_t^{i,\delta}|) + \psi'(|U_t^{i,\delta}|) A_t^{i,\delta} dt + dM_t^{\psi,i,\delta} \\ & \quad + \lambda_0 \psi(|U_t^{i,\delta}|) \left(1 - \phi_\delta(|U_t^{i,\delta}|) \right) + \psi'(0) K_{1,b_1} |U_t^{i,\delta}| \left(1 - \phi_\delta(|U_t^{i,\delta}|) \right), \end{aligned}$$

where ψ and λ_0 are as described above, \widetilde{L}_Z^δ is defined by (2.13), $(M_t^{\psi,i,\delta})_{t \geq 0}$ is a martingale, and

$$A_t^{i,\delta} := \left| b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i) - \frac{1}{n} \sum_{j=1}^n b_2(X_t^{i,n,\delta} - X_t^{j,n,\delta}) \right|.$$

In the argument above we used the facts that ψ' is decreasing and that $b_1(x)$ is one-sided Lipschitz with the constant K_{1,b_1} . We note again that the construction of ψ and the constant λ_0 is independent of K_{b_2} .

For notational convenience, let us drop the superscript δ for now (i.e., denote $A_t^i := A_t^{i,\delta}$, $X_t^{i,n} := X_t^{i,n,\delta}$ and $U_t^i := U_t^{i,\delta}$) and estimate A_t^i as follows

$$\begin{aligned} A_t^i &\leq \frac{1}{n} \sum_{j=1}^n |b_2(\bar{X}_t^i - \bar{X}_t^j) - b_2(X_t^{i,n} - X_t^{j,n})| + \left| b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i) - \frac{1}{n} \sum_{j=1}^n b_2(\bar{X}_t^i - \bar{X}_t^j) \right| \\ &\leq \frac{\eta}{n} \sum_{j=1}^n (\psi(|U_t^i|) + \psi(|U_t^j|)) + \left| b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i) - \frac{1}{n} \sum_{j=1}^n b_2(\bar{X}_t^i - \bar{X}_t^j) \right| \\ &=: \frac{\eta}{n} \sum_{j=1}^n (\psi(|U_t^i|) + \psi(|U_t^j|)) + A_t^{i,*}, \end{aligned}$$

where $\eta := \eta(K_{b_2}) = K_{b_2}/c_1$ with the constant c_1 defined as in Theorem 3.2 (independent of K_{b_2}), and in the second inequality we used condition (1-ii) and the fact that $c_1 r \leq \psi(r)$ for all $r > 0$. Hence, according to both estimates above and the fact that $\psi'(r) \leq 1 + c_1$ for all $r > 0$, we arrive at

$$\frac{1}{n} \sum_{i=1}^n \frac{d\mathbb{E}\psi(|U_t^i|)}{dt} \leq -\frac{\lambda_0 - 2(1 + c_1)\eta}{n} \sum_{i=1}^n \mathbb{E}\psi(|U_t^i|) dt + \frac{(1 + c_1)}{n} \sum_{i=1}^n \mathbb{E}A_t^{i,*} + C(\delta),$$

where $C(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. In particular, choosing

$$(4.2) \quad 0 < K_{b_2} < \lambda_0 c_1 / (2(1 + c_1)),$$

we find that

$$(4.3) \quad \frac{1}{n} \sum_{i=1}^n \frac{d\mathbb{E}\psi(|U_t^i|)}{dt} \leq -\frac{\lambda}{n} \sum_{i=1}^n \mathbb{E}\psi(|U_t^i|) dt + \frac{(1 + c_1)}{n} \sum_{i=1}^n \mathbb{E}A_t^{i,*} + C(\delta),$$

where $\lambda := \lambda_0 - 2(1 + c_1)\eta > 0$ due to (4.2).

Furthermore, given \bar{X}_t^i , the random variables \bar{X}_t^j , $j \neq i$, are i.i.d. with the same law $\mu_{\bar{X}_t^i} =: \mu_{X_t}$. In particular,

$$\mathbb{E}(b_2(\bar{X}_t^i - \bar{X}_t^j) | \bar{X}_t^i) = b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i).$$

Due to condition (1-ii) (particularly, $b_2(0) = 0$ and $|b_2(z)| \leq K_{b_2}|z|$ for all $z \in \mathbb{R}^d$),

$$\begin{aligned} \mathbb{E}A_t^{i,*} &\leq \mathbb{E} \left| b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i) - \frac{1}{n-1} \sum_{j=1}^n b_2(\bar{X}_t^i - \bar{X}_t^j) \right| \\ &\quad + \left(\frac{1}{n-1} - \frac{1}{n} \right) \sum_{j=1}^n \mathbb{E}|b_2(\bar{X}_t^i - \bar{X}_t^j)| \end{aligned}$$

$$\begin{aligned}
 &\leq \left(\mathbb{E} \left(\mathbb{E} \left(\left| b_2 * \mu_{\bar{X}_t^i}(\bar{X}_t^i) - \frac{1}{n-1} \sum_{j=1}^n b_2(\bar{X}_t^i - \bar{X}_t^j) \right|^2 \middle| \bar{X}_t^i \right) \right) \right)^{1/2} \\
 &\quad + \sqrt{2} \left(\frac{1}{n-1} - \frac{1}{n} \right) nK_{b_2} \left(\int_{\mathbb{R}^d} |z|^2 \mu_{X_t}(dz) \right)^{1/2} \\
 &\leq \left(\mathbb{E} \left(\frac{1}{n-1} \text{Var}_{\mu_{X_t}}(b_2(\bar{X}_t^i - \cdot)) \right) \right)^{1/2} \\
 &\quad + \sqrt{2} \left(\frac{1}{n-1} - \frac{1}{n} \right) nK_{b_2} \left(\int_{\mathbb{R}^d} |z|^2 \mu_{X_t}(dz) \right)^{1/2} \\
 &\leq K_{b_2} \left(\frac{1}{\sqrt{n-1}} + \frac{\sqrt{2}}{n-1} \right) \left(\int_{\mathbb{R}^d} |z|^2 \mu_{X_t}(dz) \right)^{1/2}.
 \end{aligned}$$

We now need to bound the second moment of $(X_t)_{t \geq 0}$. To this end, consider

$$\begin{aligned}
 (4.4) \quad L[\mu]f(x) &= \langle b_1(x), \nabla f(x) \rangle + \left\langle \int b_2(x-z) \mu(dz), \nabla f(x) \right\rangle \\
 &\quad + \int (f(x+z) - f(x) - \langle \nabla f(x), z \rangle \mathbf{1}_{\{|z| \leq 1\}}) \nu(dz),
 \end{aligned}$$

(cf. (3.3)) and apply it to $f(x) = |x|^2$. According to condition (1-i), there is a constant $C_1 > 0$ such that

$$\langle b_1(x), \nabla f(x) \rangle = 2\langle b_1(x), x \rangle \leq -2K_{2,b_1}|x|^2 + 2|b_1(0)||x| + C_1, \quad x \in \mathbb{R}^d.$$

Moreover, for the second term in (4.4), due to the fact that $|b_2(z)| \leq K_{b_2}|z|$ for all $z \in \mathbb{R}^d$, we get

$$\begin{aligned}
 \left\langle \int b_2(x-z) \mu(dz), \nabla f(x) \right\rangle &= 2 \int \langle b_2(x-z), x \rangle \mu(dz) \\
 &\leq 2K_{b_2} \int |x-z| \cdot |x| \mu(dz) \\
 &\leq 2K_{b_2}|x|^2 + 2K_{b_2}|x| \int |z| \mu(dz).
 \end{aligned}$$

Finally, since $\int_{\mathbb{R}^d} |z|^2 \nu(dz) < \infty$, there is a constant $C_2 > 0$ such that for all $x \in \mathbb{R}^d$,

$$\begin{aligned}
 &\int (f(x+z) - f(x) - \langle \nabla f(x), z \rangle \mathbf{1}_{\{|z| \leq 1\}}) \nu(dz) \\
 &= \int_{\{|z| \leq 1\}} (|x+z|^2 - |x|^2 - 2\langle x, z \rangle) \nu(dz) + \int_{\{|z| > 1\}} (|x+z|^2 - |x|^2) \nu(dz) \\
 &\leq C_2(1 + |x|).
 \end{aligned}$$

Hence, we get that there is a constant $C_3 > 0$ such that

$$\begin{aligned}
 L[\mu]f(x) &\leq (-2K_{2,b_1} + 2K_{b_2})|x|^2 + 2K_{b_2}|x| \int |z| \mu(dz) \\
 &\quad + (2|b_1(0)| + C_2)|x| + C_1 + C_2 \\
 &\leq (-2K_{2,b_1} + 4K_{b_2})|x|^2 + K_{b_2} \int |z|^2 \mu(dz) + C_3,
 \end{aligned}$$

where in the second inequality we used the facts that $2ab \leq a^2 + b^2$ for all $a, b \geq 0$, $(\int |z| \mu(dz))^2 \leq \int |z|^2 \mu(dz)$, and we chose the constant $C_3 > 0$ so that

$$(2|b_1(0)| + C_2)|x| + C_1 + C_2 \leq K_{b_2}|x|^2 + C_3, \quad x \in \mathbb{R}^d.$$

Using the inequality above and following the argument of Lemma 3.5, we can get that for any $0 < K_{b_2} < 2K_{2,b_1}/5$,

$$\int_{\mathbb{R}^d} |z|^2 \mu_{X_t}(dz) = \mathbb{E}|X_t|^2 \leq \mathbb{E}|X_0|^2 + C_4$$

holds with some constant $C_4 > 0$, depending only on the second moment of the Lévy measure ν , $b_1(0)$ and the constants in condition (1.10). Hence, for all $1 \leq i \leq n$,

$$\mathbb{E}A_t^{i,*} \leq \frac{C_5}{\sqrt{n-1}},$$

where $C_5 := K_{b_2}(\sqrt{2} + 1)(\mathbb{E}|X_0|^2 + C_4)^{1/2}$. This, along with (4.3), yields that if $K_{b_2} < \min\{2K_{2,b_1}/5, \lambda_0 c_1/(2(1+c_1))\}$, then

$$\frac{1}{n} \sum_{i=1}^n \frac{d\mathbb{E}\psi(|U_t^{i,\delta}|)}{dt} \leq -\frac{\lambda}{n} \sum_{i=1}^n \mathbb{E}\psi(|U_t^{i,\delta}|) dt + \frac{C_5}{\sqrt{n-1}} + C(\delta).$$

Therefore,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}\psi(|U_t^i|) \leq e^{-\lambda t} \frac{1}{n} \sum_{i=1}^n \mathbb{E}\psi(|U_0^i|) + \frac{C_5}{\lambda\sqrt{n-1}} + \tilde{C}(\delta),$$

where $\tilde{C}(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. Combining this with the facts that $W_\psi(\mu_t, \mu_t^n) \leq \mathbb{E}\psi(|U_t^i|)$ for all $1 \leq i \leq n$ and $c_1 r \psi(r) \leq (1+c_1)r$ yields the desired assertion. \square

Acknowledgements. We thank the two anonymous referees for their helpful comments and valuable corrections that have led to significant improvements of the presentation in the article. The research of Mingjie Liang is supported by the Science Foundation of the Education Department of Fujian Province (No. JAT190701) and the National Science Foundation of Sanming University (No. B201914). The majority of this work was completed when Mateusz B. Majka was affiliated to the University of Warwick and supported by the EPSRC grant no. EP/P003818/1. The research of Jian Wang is supported by the National Natural Science Foundation of China (Nos. 11831014 and 12071076), the Program for Probability and Statistics: Theory and Application (No. IRTL1704) and the Program for Innovative Research Team in Science and Technology in Fujian Province University (IRTSTFJ).

REFERENCES

- [1] Agarwal, A. and Pagliarani, S.: A Fourier-based Picard-iteration approach for a class of McKean-Vlasov SDEs with Lévy jumps, to appear in *Stochastics*, see also arXiv:1812.05026.
- [2] Andreis, L., Dai Pra, P. and Fischer, M.: McKean-Vlasov limit for interacting systems with simultaneous jumps, *Stoch. Anal. Appl.*, **36**, 960–995, 2018.
- [3] Arapostathis, A., Pang, G. and Sandrić, N.: Ergodicity of a Lévy-driven SDE arising from multiclass many-server queues, *Ann. Appl. Probab.*, **29**, 1070–1126, 2019.
- [4] Arapostathis, A., Pang, G. and Sandrić, N.: Subexponential upper and lower bounds in Wasserstein distance for Markov processes, arXiv:1907.05250.

- [5] Barbu, V. and Röckner, M.: From nonlinear Fokker-Planck equations to solutions of distribution dependent SDE, *Ann. Probab.*, **48**, 1902–1920, 2020.
- [6] Benazzoli, C., Campi, L. and Di Persio, L.: Mean field games with controlled jump-diffusion dynamics: Existence results and an illiquid interbank market model, to appear in *Stoch. Proc. Appl.*, see also arXiv:1703.01919.
- [7] Butkovsky, O.: On ergodic properties of nonlinear Markov chains and stochastic McKean-Vlasov equations, *Theory Probab. Appl.*, **58**, 661–674, 2014.
- [8] Cai, Y., Huang, J. and Maroulas, V.: Large deviations of mean-field stochastic differential equations with jumps, *Statist. Probab. Lett.*, **96**, 1–9, 2015.
- [9] Carmona, R. and Delarue, F.: *Probabilistic Theory of Mean Field Games with Applications I-II*, Springer, 2017.
- [10] Cattiaux, P. and Guillin, A.: Semi log-concave Markov diffusions, in: *Séminaire de Probabilités XLVI*, Lecture Notes in Mathematics Vol. **2123**, pp. 231–292, Springer Verlag, 2015.
- [11] Cattiaux, P., Guillin, A. and Malrieu, F.: Probabilistic approach for granular media equations in the non uniformly convex case, *Probab. Theory Related Fields*, **140**, 19–40, 2008.
- [12] Chen, M.-F.: *Eigenvalues, Inequalities, and Ergodic Theory*, Springer-Verlag London, Ltd., London, 2005.
- [13] Chen, Z.-Q., Song, R. and Zhang, X.: Stochastic flows for Lévy processes with Hölder drifts, *Revista Matemática Iberoamericana*, **34**, 1755–1788, 2018.
- [14] dos Reis, G., Salkeld, W. and Tugaut, J.: Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law, *Ann. Appl. Probab.*, **29**, 1487–1540, 2019.
- [15] Durmus, A., Eberle, A., Guillin, A. and Zimmer, R.: An elementary approach to uniform in time propagation of chaos, to appear in *Proceedings AMS*, see also arXiv:1805.11387.
- [16] Eberle, A., Guillin, A. and Zimmer, R.: Quantitative Harris type theorems for diffusions and McKean-Vlasov processes, *Trans. Amer. Math. Soc.*, **371**, 7135–7173, 2019.
- [17] Graham, C.: McKean-Vlasov Ito-Skorohod equations, and nonlinear diffusions with discrete jump sets, *Stoch. Proc. Appl.*, **40**, 69–82, 1992.
- [18] Gyöngy, I. and Krylov, N.V.: On stochastic equations with respect to semimartingales. I., *Stochastics*, **4**, 1–21, 1980/81.
- [19] Hairer, M., Mattingly, J.C. and Scheutzow, M.: Asymptotic coupling and a general form of Harris’ theorem with applications to stochastic delay equations, *Probab. Theory Related Fields*, **149**, 223–259, 2011.
- [20] Hairer, M. and Mattingly, J.C.: Yet another look at Harris’s ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI, Vol. 63 of Progr. Probab.*, 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.
- [21] Hammersley, W., Siska, D. and Szpruch, L.: McKean-Vlasov SDEs under measure dependent Lyapunov conditions, arXiv:1802.03974.
- [22] Huang, L.-J., Majka, M.B. and Wang, J.: Approximation of heavy-tailed distributions via stable-driven SDEs, arXiv:2007.02212.
- [23] Huang, X., Röckner, M. and Wang, F.-Y.: Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs, *Discrete and Continuous Dynamical Systems*, **39**, 3017–3035, 2019.
- [24] Hao, T. and Li, J.: Mean-field SDEs with jumps and nonlocal integral-PDEs, *NoDEA Nonlinear Differential Equations Appl.*, **23**, Art. 17, 51, 2016.
- [25] Jourdain, B., Méléard, S. and Woyczynski, W.A.: Nonlinear SDEs driven by Lévy processes and related PDEs, *ALEA*, **4**, 1–29, 2008.
- [26] Kac, M.: *Foundations of kinetic theory*, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, vol. III, pp. 171–197, University of California Press, Berkeley and Los Angeles, 1956.
- [27] Kolokoltsov, V.N.: *Nonlinear Markov Processes and Kinetic Equations*, Cambridge University Press, Cambridge, 2010.
- [28] Kulik, A.M.: Exponential ergodicity of the solutions to SDE’s with a jump noise, *Stoch. Proc. Appl.*, **119**, 602–632, 2009.
- [29] Kühn, F. and Schilling, R.L.: Strong convergence of the Euler–Maruyama approximation for a class of Lévy-driven SDEs, *Stoch. Proc. Appl.* **129**, 2654–2680, 2019.

- [30] Li, J.: Mean-field forward and backward SDEs with jumps and associated nonlocal quasi-linear integral-PDEs, *Stoch. Proc. Appl.*, **128**, 3118–3180, 2018.
- [31] Liang, M., Schilling, R.L. and Wang, J.: A unified approach to coupling SDEs driven by Lévy noise and some applications, *Bernoulli*, **26**, 664–693, 2020.
- [32] Luo, D. and Wang, J.: Coupling by reflection and Hölder regularity for non-local operators of variable order, *Trans. Amer. Math. Soc.*, **371**, 431–459, 2019.
- [33] Luo, D. and Wang, J.: Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises, *Stoch. Proc. Appl.* **129**, 3129–3173, 2019.
- [34] Majka, M.B.: Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, *Stoch. Proc. Appl.*, **127**, 4083–4125, 2017.
- [35] Majka, M.B.: Transportation inequalities for non-globally dissipative SDEs with jumps via Malliavin calculus and coupling, *Ann. Inst. Henri Poincaré Probab. Stat.*, **55**, 2019–2057, 2019.
- [36] Majka, M.B.: A note on existence of global solutions and invariant measures for jump SDEs with locally one-sided Lipschitz drift, *Probab. Math. Statist.*, **40**, 37–55, 2020.
- [37] Masuda, H.: Ergodicity and exponential β -mixing bounds for multidimensional diffusions with jumps, *Stoch. Proc. Appl.*, **117**, 35–56, 2007.
- [38] Mehri, S., Scheutzwow, M., Stannat, W. and Zangeneh, B.Z.: Propagation of chaos for stochastic spatially structured neuronal networks with fully path dependent delays and monotone coefficients driven by jump diffusion noise, *Ann. Appl. Probab.*, **30**, 175–207, 2020.
- [39] Méléard, S.: *Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models*, in : Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), Lecture Notes in Math., vol. **1642**, pp. 42–95, Springer, Berlin, 1996.
- [40] Meyn, S.P. and Tweedie, R.T.: Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes, *Adv. Appl. Probab.*, **25**, 518–548, 1993.
- [41] Priola, E.: Pathwise uniqueness for singular SDEs driven by stable processes, *Osaka J. Math.* **49**, 421–447, 2012.
- [42] Song, Y.: Gradient estimates and exponential ergodicity for mean-field SDEs with jumps, *J. Theoret. Probab.*, **33**, 201–238, 2020.
- [43] Schilling, R.L. and Wang, J.: On the coupling property and the Liouville theorem for Ornstein-Uhlenbeck processes, *J. Evol. Equ.*, **12**, 119–140, 2012.
- [44] Sznitman, A.-S.: *Topics in propagation of chaos*, in: École d’Été de Probabilités de Saint-Flour XIX–1989, Lecture Notes in Math., vol **1464**, pp. 165–251, Springer, Berlin, 1991.
- [45] Tanaka, H., Tsuchiya, M. and Watanabe, S.: Perturbation of drift-type for Lévy processes, *J. Math. Kyoto Univ.* **14**, 73–92, 1974.
- [46] Villani, C.: *Optimal Transport: Old and New*, Springer-Verlag, Berlin, 2009.
- [47] Wang, F.-Y.: Distribution dependent SDEs for Landau type equations, *Stoch. Proc. Appl.*, **128**, 595–621, 2018.
- [48] Wang, J.: Criteria for ergodicity of Lévy type operators in dimension one, *Stoch. Proc. Appl.*, **118**, 1909–1928, 2008.
- [49] Wang, J.: On the exponential ergodicity for Lévy driven Ornstein–Uhlenbeck processes, *J. Appl. Probab.*, **49**, 990–1104, 2012.
- [50] Wang, J.: L^p -Wasserstein distance for stochastic differential equations driven by Lévy processes, *Bernoulli*, **22**, 1598–1616, 2016.
- [51] Zhang, X.: Stochastic differential equations with Sobolev drifts and driven by α -stable processes, *Ann. Inst. Henri Poincaré Probab. Stat.*, **49**, 1057–1079, 2013.
- [52] Zimmer, R.: *Kantorovich contractions with explicit constants for diffusion semigroups*, PhD Thesis, University of Bonn, 2017.