ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

stochastic processes and their applications

www.elsevier.com/locate/spa

Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes

Mateusz B. Majka

Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany

Received 27 October 2015; received in revised form 10 February 2017; accepted 24 March 2017

Available online xxxx

Abstract

We present a novel idea for a coupling of solutions of stochastic differential equations driven by Lévy noise, inspired by some results from the optimal transportation theory. Then we use this coupling to obtain exponential contractivity of the semigroups associated with these solutions with respect to an appropriately chosen Kantorovich distance. As a corollary, we obtain exponential convergence rates in the total variation and standard L^1-Wasserstein distances.

MSC: 60G51; 60H10

Keywords: Stochastic differential equations; Lévy processes; Exponential ergodicity; Couplings; Wasserstein distances

1. Introduction

We consider stochastic differential equations of the form

$$dX_t = b(X_t)dt + dL_t, \tag{1.1}$$

where $(L_t)_{t\geq 0}$ is an $\mathbb{R}^d$-valued Lévy process and $b : \mathbb{R}^d \to \mathbb{R}^d$ is a continuous vector field satisfying a one-sided Lipschitz condition, i.e., there exists a constant $C_L > 0$ such that for all x,

E-mail address: majka@uni-bonn.de.

http://dx.doi.org/10.1016/j.spa.2017.03.020
0304-4149/

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

$y \in \mathbb{R}^d$ we have

$$\langle b(x) - b(y), x - y\rangle \leq C_L |x - y|^2. \tag{1.2}$$

These assumptions are sufficient in order for (1.1) to have a unique strong solution (see Theorem 2 in [6]). For any $t \geq 0$, denote the distribution of the random variable L_t by μ_t. Its Fourier transform $\widehat{\mu}_t$ is of the form

$$\widehat{\mu}_t(z) = e^{t\psi(z)}, \quad z \in \mathbb{R}^d,$$

where the *Lévy symbol* (or *Lévy exponent*) $\psi : \mathbb{R}^d \to \mathbb{C}$ is given by the Lévy–Khintchine formula (see e.g. [1] or [20]),

$$\psi(z) = i\langle l, z\rangle - \frac{1}{2}\langle z, Az\rangle + \int_{\mathbb{R}^d} (e^{i\langle z,x\rangle} - 1 - i\langle z, x\rangle \mathbf{1}_{\{|x|\leq 1\}})\nu(dx),$$

for $z \in \mathbb{R}^d$. Here l is a vector in $\mathbb{R}^d$, A is a symmetric nonnegative-definite $d \times d$ matrix and ν is a measure on $\mathbb{R}^d$ satisfying

$$\nu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} (|x|^2 \wedge 1)\nu(dx) < \infty.$$

We call (l, A, ν) the *generating triplet* of the Lévy process $(L_t)_{t\geq 0}$, whereas A and ν are called, respectively, the *Gaussian covariance matrix* and the *Lévy measure* (or *jump measure*) of $(L_t)_{t\geq 0}$.

In this paper we will be working with pure jump Lévy processes. We assume that in the generating triplet of $(L_t)_{t\geq 0}$ we have $l = 0$ and $A = 0$. By the Lévy–Itô decomposition we know that there exists a Poisson random measure N associated with $(L_t)_{t\geq 0}$ in such a way that

$$L_t = \int_0^t \int_{\{|v|>1\}} vN(ds, dv) + \int_0^t \int_{\{|v|\leq 1\}} v\widetilde{N}(ds, dv), \tag{1.3}$$

where

$$\widetilde{N}(ds, dv) = N(ds, dv) - ds\, \nu(dv)$$

is the compensated Poisson random measure.

We will be considering the class of Kantorovich (L^1-Wasserstein) distances. For $p \geq 1$, we can define the *L^p-Wasserstein distance* between two probability measures μ_1 and μ_2 on $\mathbb{R}^d$ by the formula

$$W_p(\mu_1, \mu_2) := \left(\inf_{\pi \in \Pi(\mu_1,\mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \rho(x, y)^p \pi(dx\, dy) \right)^{\frac{1}{p}},$$

where ρ is a metric on $\mathbb{R}^d$ and $\Pi(\mu_1, \mu_2)$ is the family of all couplings of μ_1 and μ_2, i.e., $\pi \in \Pi(\mu_1, \mu_2)$ if and only if π is a measure on $\mathbb{R}^{2d}$ having μ_1 and μ_2 as its marginals. We will be interested in the particular case of $p = 1$ and the distance ρ being given by a concave function $f : [0, \infty) \to [0, \infty)$ with $f(0) = 0$ and $f(x) > 0$ for $x > 0$ as

$$\rho(x, y) := f(|x - y|) \quad \text{for all } x, y \in \mathbb{R}^d.$$

We will denote the L^1-Wasserstein distance associated with a function f by W_f. The most well-known examples are given by $f(x) = \mathbf{1}_{(0,\infty)}(x)$, which leads to the total variation distance (with $W_f(\mu_1, \mu_2) = \frac{1}{2}\|\mu_1 - \mu_2\|_{TV}$) and by $f(x) = x$, which defines the standard L^1-Wasserstein distance (denoted later by W_1). For a detailed exposition of Wasserstein distances, see e.g. Chapter 6 in [27].

For an $\mathbb{R}^d$-valued Markov process $(X_t)_{t\geq 0}$ with transition kernels $(p_t(x,\cdot))_{t\geq 0, x\in\mathbb{R}^d}$ we say that an $\mathbb{R}^{2d}$-valued process $(X'_t, X''_t)_{t\geq 0}$ is a *coupling* of two copies of the Markov process $(X_t)_{t\geq 0}$ if both $(X'_t)_{t\geq 0}$ and $(X''_t)_{t\geq 0}$ are Markov processes with transition kernels p_t but possibly with different initial distributions. We define the *coupling time* T for the marginal processes $(X'_t)_{t\geq 0}$ and $(X''_t)_{t\geq 0}$ by $T := \inf\{t \geq 0 : X'_t = X''_t\}$. The coupling is called *successful* if T is almost surely finite. It is known (see e.g. [13] or [26]) that the condition

$$\|\mu_1 p_t - \mu_2 p_t\|_{TV} \to 0 \quad \text{as } t \to \infty \text{ for any probability measures } \mu_1 \text{ and } \mu_2 \text{ on } \mathbb{R}^d$$

is equivalent to the property that for any two probability measures μ_1 and μ_2 on $\mathbb{R}^d$ there exist marginal processes $(X'_t)_{t\geq 0}$ and $(X''_t)_{t\geq 0}$ with μ_1 and μ_2 as their initial distributions such that the coupling $(X'_t, X''_t)_{t\geq 0}$ is successful. Here $\mu p_t(dy) = \int \mu(dx) p_t(x, dy)$.

Couplings of Lévy processes and related bounds in the total variation distance have recently attracted considerable attention. See e.g. [2,21,22] for couplings of pure jump Lévy processes, [23,28,29] for the case of Lévy-driven Ornstein–Uhlenbeck processes and [12,31,25] for more general Lévy-driven SDEs with non-linear drift. See also [11,19] for general considerations concerning ergodicity of SDEs with jumps. Furthermore, in a recent paper [32], J. Wang investigated the topic of using couplings for obtaining bounds in the L^p-Wasserstein distances.

Previous attempts at constructing couplings of Lévy processes or couplings of solutions to Lévy-driven SDEs include e.g. a coupling of subordinate Brownian motions by making use of the coupling of Brownian motions by reflection (see [2]), a coupling of compound Poisson processes obtained from certain couplings of random walks (see [22] for the original construction and [31] for a related idea applied to Lévy-driven SDEs) and a combination of the coupling by reflection and the synchronous coupling defined via its generator for solutions to SDEs driven by Lévy processes with a symmetric α-stable component (see [32]). In the present paper we use a different idea for a coupling, as well as a different method of construction. Namely, we define a coupling by reflection modified in such a way that it allows for a positive probability of bringing the marginal processes to the same point if the distance between them is small enough. Such a behaviour makes it possible to obtain better convergence rates than a regular coupling by reflection, since it significantly decreases the probability that the marginal processes suddenly jump far apart once they have already been close to each other. We construct our coupling as a solution to an explicitly given SDE, much in the vein of the seminal paper [14] by Lindvall and Rogers, where they constructed a coupling by reflection for diffusions with a drift. The formulas for the SDEs defining the marginal processes in our coupling are given by (2.9) and (2.10) and the way we obtain them is explained in detail in Section 2.2. Then, using this coupling, we construct a carefully chosen Kantorovich distance W_f for an appropriate concave function f such that

$$W_f(\mu_1 p_t, \mu_2 p_t) \leq e^{-ct} W_f(\mu_1, \mu_2)$$

holds for some constant $c > 0$ and all $t \geq 0$, where μ_1 and μ_2 are arbitrary probability measures on $\mathbb{R}^d$ and $(p_t)_{t\geq 0}$ is the transition semigroup associated with $(X_t)_{t\geq 0}$. Here f and c are mutually dependent and are chosen with the aim to make c as large as possible, which leads to bounds that are in some cases close to optimal. A similar approach has been recently taken by Eberle in [5], where he used a specially constructed distance in order to investigate exponential ergodicity of diffusions with a drift. Historically, related ideas have been used e.g. by Chen and Wang in [3] and by Hairer and Mattingly in [7], to investigate spectral gaps for diffusion operators on $\mathbb{R}^d$ and to investigate ergodicity in infinite dimensions, respectively. It is important to point out that the distance function we choose is discontinuous. It is in fact of the form

$$f = f_1 + a\mathbf{1}_{(0,\infty)},$$

where f_1 is a concave, strictly increasing $\mathcal{C}^2$ function with $f_1(0) = 0$, which from some point $R_1 > 0$ is extended in an affine way and a is a positive constant. This choice of the distance (which is directly tied to our choice of the coupling) has an advantage in that it gives us upper bounds in both the total variation and standard L^1-Wasserstein distances (see Corollaries 1.4 and 1.5 and the discussion in Remark 1.6).

Let us now state the assumptions that we will impose on the Lévy measure ν of the process $(L_t)_{t\geq 0}$.

Assumption 1. ν is rotationally invariant, i.e.,

$$\nu(AB) = \nu(B)$$

for every Borel set $B \in \mathcal{B}(\mathbb{R}^d)$ and every $d \times d$ orthogonal matrix A.

Assumption 2. ν is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, with a density q that is almost everywhere continuous on $\mathbb{R}^d$.

Assumption 3. There exist constants $m, \delta > 0$ such that $\delta < 2m$ and

$$\inf_{x\in\mathbb{R}^d:0<|x|\leq\delta} \int_{\{|v|\leq m\}\cap\{|v+x|\leq m\}} q(v) \wedge q(v+x)dv > 0. \tag{1.4}$$

Assumption 4. There exists a constant $\varepsilon > 0$ such that $\varepsilon \leq \delta$ (with δ defined via (1.4)) and

$$\int_{\{|v|\leq\varepsilon/2\}} q(v)dv > 0.$$

Assumptions 1 and 2 are used in the proof of Theorem 1.1 to show that the solution to the SDE that we construct there is actually a coupling. Assumption 1 is quite natural since we want to use reflection of the jumps. It is possible to extend our results to the case where the Lévy measure is only required to have a rotationally invariant component, but we do not do this in the present paper. Assumption 3 is used in our calculations regarding the Wasserstein distances and is basically an assumption about sufficient overlap of the Lévy density q and its translation. A related condition is used e.g. in [22] (see (1.3) in Theorem 1.1 therein) and in [29] to ensure that there is enough jump activity to provide a successful coupling. The restriction in (1.4) to the jumps bounded by m is related to our coupling construction, see the discussion in Section 2.2. Assumption 4 ensures that we have enough small jumps to make use of the reflected jumps in our coupling (cf. the proof of Lemma 3.3). All the assumptions together are satisfied by a large class of rotationally invariant Lévy processes, with symmetric α-stable processes for $\alpha \in (0, 2)$ being one of the most important examples. Note however, that our framework covers also the case of finite Lévy measures and even some cases of Lévy measures with supports separated from zero (see Example 1.7 for further discussion).

We must also impose some conditions on the drift function b. We have already assumed that it satisfies a one-sided Lipschitz condition, which guarantees the existence and uniqueness of a

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

strong solution to (1.1). Now we define the function $\kappa : \mathbb{R}_+ \to \mathbb{R}$ by setting $\kappa(|x-y|)$ to be the largest quantity such that

$$\langle b(x) - b(y), x - y \rangle \leq -\kappa(|x-y|)|x-y|^2 \quad \text{for any } x, y \in \mathbb{R}^d,$$

and therefore it has to be defined as

$$\kappa(r) := \inf \left\{ -\frac{\langle b(x) - b(y), x - y \rangle}{|x-y|^2} : x, y \in \mathbb{R}^d \text{ such that } |x-y| = r \right\}. \tag{1.5}$$

We have the following assumption.

Assumption 5. κ is a continuous function satisfying

$$\liminf_{r \to \infty} \kappa(r) > 0.$$

The above condition means that there exist constants $M > 0$ and $R > 0$ such that for all x, $y \in \mathbb{R}^d$ with $|x-y| \geq R$ we have

$$\langle b(x) - b(y), x - y \rangle \leq -M|x-y|^2. \tag{1.6}$$

In other words, the drift b is dissipative outside some ball of radius R. Note that if the drift is dissipative everywhere, i.e., (1.6) holds for all x, $y \in \mathbb{R}^d$, then the proof of exponential convergence in the L^1-Wasserstein distance is quite straightforward, using just the synchronous coupling for $(L_t)_{t \geq 0}$ and the Gronwall inequality. Thus it is an interesting problem to try to obtain exponential convergence under some weaker assumptions on the drift.

We finally formulate our main results.

Theorem 1.1. *Let us consider a stochastic differential equation*

$$dX_t = b(X_t)dt + dL_t, \tag{1.7}$$

where $(L_t)_{t \geq 0}$ is a pure jump Lévy process with the Lévy measure ν satisfying Assumptions 1 and 2, whereas $b : \mathbb{R}^d \to \mathbb{R}^d$ is a continuous, one-sided Lipschitz vector field. Then a coupling $(X_t, Y_t)_{t \geq 0}$ of solutions to (1.7) can be constructed as a strong solution to the 2d-dimensional SDE given by (2.9) and (2.10), driven by a d-dimensional noise. If we additionally require Assumptions 3–5 to hold, then there exist a concave function f and a constant $c > 0$ such that for any $t \geq 0$ we have

$$\mathbb{E} f(|X_t - Y_t|) \leq e^{-ct} \mathbb{E} f(|X_0 - Y_0|) \tag{1.8}$$

and the coupling $(X_t, Y_t)_{t \geq 0}$ is successful.

Since the inequality (1.8) holds for all couplings of the laws of X_0 and Y_0, directly from the definition of the Wasserstein distance W_f we obtain the following result.

Corollary 1.2. *Let $(X_t)_{t \geq 0}$ be a solution to the SDE (1.7) with $(L_t)_{t \geq 0}$ and b as in Theorem 1.1, satisfying Assumptions 1–5. Then there exist a concave function f and a constant $c > 0$ such that for any $t \geq 0$ and any probability measures μ_1 and μ_2 on $\mathbb{R}^d$ we have*

$$W_f(\mu_1 p_t, \mu_2 p_t) \leq e^{-ct} W_f(\mu_1, \mu_2), \tag{1.9}$$

where $(p_t)_{t \geq 0}$ is the semigroup associated with $(X_t)_{t \geq 0}$.

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

The function f in the theorem and the corollary above is given as $f = a\mathbf{1}_{(0,\infty)} + f_1$, where

$$\begin{aligned} f_1(r) &= \int_0^r \phi(s)g(s)ds \\ \phi(r) &= \exp\left(-\int_0^r \frac{\bar{h}(t)}{C_\varepsilon}dt\right), \qquad \bar{h}(r) = \sup_{t\in(r,r+\varepsilon)} t\kappa^-(t), \\ g(r) &= 1 - \frac{1}{2}\int_0^{r\wedge R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)}dt \left(\int_0^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)}dt\right)^{-1}, \qquad \Phi(r) = \int_0^r \phi(s)ds, \end{aligned} \tag{1.10}$$

while the contractivity constant c is given by $c = \min\{c_1/2K, \widetilde{C}_\delta/4\}$ with

$$c_1 = \frac{C_\varepsilon}{2}\left(\int_0^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)}dt\right)^{-1} \quad \text{and} \quad \widetilde{C}_\delta = \inf_{x\in\mathbb{R}^d:0<|x|\le\delta} \int_{\mathbb{R}^d} q(v)\wedge q(v+x)dv.$$

Here κ is the function defined by (1.5), the constants R_0 and R_1 are defined by

$$\begin{aligned} R_0 &= \inf\{R \ge 0 : \forall r \ge R : \kappa(r) \ge 0\}, \\ R_1 &= \inf\left\{R \ge R_0 + \varepsilon : \forall r \ge R : \kappa(r) \ge \frac{2C_\varepsilon}{(R-R_0)R}\right\}, \end{aligned} \tag{1.11}$$

the constant δ comes from Assumption 3, the constant $\varepsilon \le \delta$ comes from Assumption 4 (see also Remark 3.4) and we have

$$C_\varepsilon = 2\int_{-\varepsilon/4}^0 |y|^2\nu_1(dy), \qquad K = \frac{C_L\delta + \widetilde{C}_\delta f_1(\delta)/2}{\widetilde{C}_\delta f_1(\delta)/2} \quad \text{and} \quad a = Kf_1(\delta), \tag{1.12}$$

where ν_1 is the first marginal of ν and the constant C_L comes from (1.2). Note that due to Assumptions 3 and 4 it is always possible to choose δ and ε in such a way that $\widetilde{C}_\delta > 0$ and $C_\varepsilon > 0$ and due to Assumption 5 the constants R_0 and R_1 are finite.

Remark 1.3. The formulas for the function f and the constant c for which (1.9) holds are quite sophisticated, but they are chosen in such a way as to try to make c as large as possible and their choice is clearly motivated by the calculations in the proof, see Section 3 for details. The contractivity constant c can be seen to be in some sense close to optimal (at least in certain cases). See the discussion in Section 4 for comparison of convergence rates in the L^1-Wasserstein distance in the case where the drift is assumed to be the gradient of a strongly convex potential and the case where convexity is only required to hold outside some ball.

With the above notation and assumptions, we immediately get some important corollaries.

Corollary 1.4. *For any $t \ge 0$ and any probability measures μ_1 and μ_2 on $\mathbb{R}^d$ we have*

$$\|\mu_1 p_t - \mu_2 p_t\|_{TV} \le 2a^{-1}e^{-ct}W_f(\mu_1,\mu_2), \tag{1.13}$$

where $a > 0$ is the constant defined by (1.12).

Corollary 1.5. *For any $t \ge 0$ and any probability measures μ_1 and μ_2 on $\mathbb{R}^d$ we have*

$$W_1(\mu_1 p_t, \mu_2 p_t) \le 2\phi(R_0)^{-1}e^{-ct}W_f(\mu_1,\mu_2), \tag{1.14}$$

where the function ϕ and the constant $R_0 > 0$ are defined by (1.10) and (1.11), respectively.

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

Remark 1.6. The corollaries above follow in a straightforward way from (1.9) by comparing the underlying distance function f from below with the $\mathbf{1}_{(0,\infty)}$ function (corresponding to the total variation distance) and the identity function (corresponding to the standard L^1-Wasserstein distance), see Section 4 for explicit proofs. In the paper [5] by Eberle, which treated the diffusion case, a related concave function was constructed, although without a discontinuity at zero (and also extended in an affine way from some point). This leads to bounds of the form

$$W_1(\mu_1 p_t, \mu_2 p_t) \leq L e^{-ct} W_1(\mu_1, \mu_2) \tag{1.15}$$

with some constants $L \geq 1$ and $c > 0$, since such a continuous function f can be compared with the identity function both from above and below. In our case we are not able to produce an inequality like (1.15) due to the discontinuity at zero, but on the other hand we can obtain upper bounds (1.13) in the total variation distance, which is impossible in the framework of [5]. Several months after the submission of the first version of the present manuscript, its author managed to modify the method presented here in order to obtain (1.9) for Lévy-driven SDEs with a continuous function f (which leads to (1.15)) by replacing Assumptions 3 and 4 with an assumption stating that the function $\varepsilon \mapsto \varepsilon / C_\varepsilon$ is bounded in a neighbourhood of zero (with C_ε defined by (1.12)), which is an assumption about sufficient concentration of the Lévy measure ν around zero (sufficient small jump activity, much higher than in the case of Assumptions 3 and 4). This result was presented in [16], where trying to obtain the inequality (1.15) was motivated by showing how it can lead to so-called α-$W_1 H$ transportation inequalities that characterize the concentration of measure phenomenon for solutions of SDEs of the form (1.1). The difference between the approach presented here and the approach in [16] is in the method chosen to deal with the case in which the marginal processes in the coupling are already close to each other and contractivity can be spoilt by having undesirable large jumps. This can be dealt with either by introducing a discontinuity in the distance function and proceeding like in the proof of Lemma 3.7 or by making sure that we have enough small jumps. It is worth mentioning that in the meantime the inequality (1.15) in the Lévy jump case was independently obtained by D. Luo and J. Wang in [15], by using a different coupling and under different assumptions (which are also, however, assumptions about sufficiently high small jump activity). In conclusion, it seems that in order to obtain (1.15) one needs the noise to exhibit a diffusion-like type of behaviour (a lot of small jumps), while estimates of the type (1.13) and (1.14) can be obtained under much milder conditions.

Example 1.7. In order to better understand when Assumptions 3 and 4 are satisfied, let us examine a class of simple examples. We already mentioned that our assumptions hold for symmetric α-stable processes with $\alpha \in (0, 2)$, for which it is sufficient to take arbitrary $m > 0$ and arbitrary $\varepsilon = \delta < 2m$. Now let us consider one-dimensional Lévy measures of the form $\nu(dx) = \left(\mathbf{1}_{[-\theta,-\theta/\beta]}(x) + \mathbf{1}_{[\theta/\beta,\theta]}(x)\right) dx$ for arbitrary $\theta > 0$ and $\beta > 1$. If we would like the quantity appearing in Assumption 3 to be positive, it is then best to take $m = \theta$. Note that if $\beta \leq 3$, then $2\theta/\beta \geq \theta - \theta/\beta$ (the gap in the support of ν is larger than the size of the part of the support contained in $\mathbb{R}_+$) and thus we need to have $\delta < \theta - \theta/\beta$ (taking $\delta = \theta - \theta/\beta$ or larger would result in an overlap of zero mass). This means that $\varepsilon/2 \leq \theta/2 - \theta/2\beta \leq \theta/\beta$ and thus the quantity in Assumption 4 cannot be positive. On the other hand for $\beta > 3$ we can take any $\delta < 2\theta$ in Assumption 3 and thus Assumption 4 can also be satisfied.

Corollary 1.8. *In addition to Assumptions 1–5, suppose that the semigroup $(p_t)_{t\geq 0}$ preserves finite first moments, i.e., if a measure μ has a finite first moment, then for all $t > 0$ the measure*

μp_t also has a finite first moment. Then there exists an invariant measure μ_* for the semigroup $(p_t)_{t\geq 0}$. Moreover, for any $t \geq 0$ and any probability measure η we have

$$W_f(\mu_*, \eta p_t) \leq e^{-ct} W_f(\mu_*, \eta) \tag{1.16}$$

and therefore

$$\|\mu_* - \eta p_t\|_{TV} \leq 2a^{-1} e^{-ct} W_f(\mu_*, \eta) \tag{1.17}$$

and

$$W_1(\mu_*, \eta p_t) \leq 2\phi(R_0)^{-1} e^{-ct} W_f(\mu_*, \eta). \tag{1.18}$$

To illustrate the usefulness of our approach, we can briefly compare our estimates with the ones obtained by other authors, who also investigated exponential convergence rates for semigroups $(p_t)_{t\geq 0}$ associated with solutions of equations like (1.1). In his recent paper [25], Y. Song obtained exponential upper bounds for $\|\delta_x p_t - \delta_y p_t\|_{TV}$ for $x, y \in \mathbb{R}^d$ using Malliavin calculus for jump processes, under some technical assumptions on the Lévy measure (which, however, does not have to be rotationally invariant) and under a global dissipativity condition on the drift. By our Corollary 1.4, we get such bounds under a much weaker assumption on the drift. In [30], J. Wang proved exponential ergodicity in the total variation distance for equations of the form (1.1) driven by α-stable processes, while requiring the drift b to satisfy a condition of the type $\langle b(x), x\rangle \leq -C|x|^2$ when $|x| \geq R$ for some $R > 0$ and $C > 0$. In the proof he used a method involving the notions of T-processes and petite sets. His assumption on the drift is weaker than ours, but our results work for a much larger class of noise. Furthermore, in [19] the authors showed exponential ergodicity, again only in the α-stable case, under some Hölder continuity assumptions on the drift, using two different approaches: by applying the Harris theorem and by a coupling argument. Kulik in [11] also used a coupling argument to give some general conditions for exponential ergodicity, but in practice they can be difficult to verify. However, he gave a simple one-dimensional example of an equation like (1.1), with the drift satisfying a condition similar to the one in [30], whose solution is exponentially ergodic under some relatively mild assumptions on the Lévy measure (see Proposition 0.1 in [11]). It is important to point out that his results, similarly to ours, apply to some cases when the Lévy measure is finite (i.e., Eq. (1.1) is driven by a compound Poisson process). All the papers mentioned above were concerned with bounds only in the total variation distance. On the other hand, J. Wang in [32] has recently obtained exponential convergence rates in the L^p-Wasserstein distances for the case when the noise in (1.1) has an α-stable component and the drift is dissipative outside some ball. By our Corollary 1.5, we get similar results in the L^1-Wasserstein distance for α-stable processes with $\alpha \in (1, 2)$, but also for a much larger class of Lévy processes without α-stable components.

Several months after the previous version of the present manuscript had been submitted, a new paper [15] by D. Luo and J. Wang appeared on arXiv. There the authors introduced yet another idea for a coupling of solutions to equations of the form (1.1) and used it to obtain exponential convergence rates for associated semigroups in both the total variation and the L^1-Wasserstein distances, as well as contractivity in the latter (cf. Remark 1.6). Their construction works under a technical assumption on the Lévy measure, which is essentially an assumption about its sufficient concentration around zero and it does not require the Lévy measure to be symmetric. However, the assumption in [15] is significantly more restrictive than our Assumptions 3 and 4. For example, it does not hold for finite Lévy measures as they do not have enough small jump activity, while our method works even in some cases where the support of the Lévy measure ν is separated from zero (cf. Example 1.7).

The remaining part of this paper is organized as follows: In Section 2 we explain the construction of our coupling and we formally prove that it is actually well defined. In Section 3 we use it to prove the inequality (1.8). In Section 4 we prove Corollaries 1.4, 1.5 and 1.8 and present some further calculations that provide additional insight into optimality of our choice of the contractivity constant c.

2. Construction of the coupling

2.1. Related ideas

The idea for the coupling that we construct in this section comes from the paper [17] by McCann, where he considered the optimal transport problem for concave costs on $\mathbb{R}$. Namely, given two probability measures μ_1 and μ_2 on $\mathbb{R}$, the problem is to find a measure γ on $\mathbb{R}^2$ with marginals μ_1 and μ_2, such that the quantity

$$C(\gamma) := \int_{\mathbb{R}^2} c(x, y) d\gamma(x, y),$$

called the transport cost, is minimized for a given concave function $c : \mathbb{R}^2 \to [0, \infty]$. McCann proved (see the remarks after the proof of Theorem 2.5 in [17] and Proposition 2.12 therein) that the minimizing measure γ (i.e., the optimal coupling of μ_1 and μ_2) is unique and independent of the choice of c, and gave an explicit expression for γ. Intuitively speaking, in the simplest case the idea behind the construction of γ (i.e., of transporting the mass from μ_1 to μ_2) is to keep in place the common mass of μ_1 and μ_2 and to apply reflection to the remaining mass. McCann's paper only treats the one-dimensional case, but since in our setting the jump measure is rotationally invariant, it seems reasonable to try to use a similar idea for a coupling also in the multidimensional case. Note that we do not formally prove in this paper that the constructed coupling is in fact the optimal one. Statements like this are usually difficult to prove, but what we really need is just a good guess of how a coupling close to the optimal one should look. Then usefulness of the constructed coupling is verified by the good convergence rates that we obtain by its application.

A related idea appeared in the paper [8] by Hsu and Sturm, where they dealt with couplings of Brownian motions, but the construction of what they call the mirror coupling can be also applied to other Markov processes. Assume we are given a symmetric transition density $p_t(x, z)$ on $\mathbb{R}$ and that we want to construct a coupling starting from (x_1, x_2) as a joint distribution of an $\mathbb{R}^2$-valued random variable $\zeta = (\zeta_1, \zeta_2)$. We put

$$\mathbb{P}(\zeta_2 = \zeta_1 | \zeta_1 = z_1) = \frac{p_t(x_1, z_1) \wedge p_t(x_2, z_1)}{p_t(x_1, z_1)} \tag{2.1}$$

and

$$\mathbb{P}(\zeta_2 = x_1 + x_2 - \zeta_1 | \zeta_1 = z_1) = 1 - \frac{p_t(x_1, z_1) \wedge p_t(x_2, z_1)}{p_t(x_1, z_1)}$$

so the idea is that if the first marginal process moves from x_1 to z_1, then the second marginal can move either to the same point or to the point reflected with respect to $x_0 = \frac{x_1+x_2}{2}$, with

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

appropriately defined probabilities, taking into account the overlap of transition densities fixed at points x_1 and x_2. Alternatively, we can define this coupling by the joint transition kernel as

$$m_t(x_1, x_2, dy_1, dy_2) := \delta_{y_1}(dy_2)h_0(y_1)dy_1 + \delta_{Ry_1}(dy_2)h_1(y_1)dy_1,$$

where $h_0(z) = p_t(x_1, z) \wedge p_t(x_2, z)$, $h_1(z) = p_t(x_1, z) - h_0(z)$ and $Ry_1 = x_1 + x_2 - y_1$. Hsu and Sturm prove that such a coupling is in fact optimal for concave, strictly increasing cost functions.

Now let us also recall the ideas from [14] by Lindvall and Rogers, where they constructed a coupling $(X_t, Y_t)_{t\geq 0}$ by reflection for diffusions by defining the second marginal process $(Y_t)_{t\geq 0}$ as a solution to an appropriate SDE. If we have a stochastic differential equation

$$dX_t = b(X_t)dt + dB_t \tag{2.2}$$

driven by a d-dimensional Brownian motion $(B_t)_{t\geq 0}$, we can define $(Y_t)_{t\geq 0}$ by setting

$$dY_t = b(Y_t)dt + (I - 2e_t e_t^T)dB_t, \tag{2.3}$$

where

$$e_t := \frac{X_t - Y_t}{|X_t - Y_t|}. \tag{2.4}$$

Of course, Eq. (2.3) only makes sense for $t < T$, where $T := \inf\{t \geq 0 : X_t = Y_t\}$, but we can set $Y_t := X_t$ for $t \geq T$. The proof that Eqs. (2.2) and (2.3) together define a coupling, i.e., the solution $(Y_t)_{t\geq 0}$ to Eq. (2.3) has the same finite dimensional distributions as the solution $(X_t)_{t\geq 0}$ to Eq. (2.2), is quite simple in the Brownian setting. It is sufficient to use the Lévy characterization theorem for Brownian motion, since the process $A_t := I - 2e_t e_t^T$ takes values in orthogonal matrices (and thus the process $(\tilde{B}_t)_{t\geq 0}$ defined by $d\tilde{B}_t := A_t dB_t$ is also a Brownian motion).

Similarly, if we consider an equation like (2.3) but driven by a rotationally invariant Lévy process $(L_t)_{t\geq 0}$ instead of the Brownian motion, it is possible to show that the process $(\tilde{L}_t)_{t\geq 0}$ defined by $d\tilde{L}_t := A_{t-}dL_t$ with $A_{t-} := I - 2e_{t-}e_{t-}^T$ is a Lévy process with the same finite dimensional distributions as $(L_t)_{t\geq 0}$. However, a corresponding coupling by reflection for Lévy processes would not be optimal and we were not able to obtain contractivity in any distance W_f using this coupling. Intuitively, this follows from the fact that such a construction allows for a situation in which two jumping processes, after they have already been close to each other, suddenly jump far apart. We need to somehow restrict such behaviour and therefore we use a more sophisticated construction.

2.2. *Construction of the SDE*

We apply the ideas from [17,8] by coupling the jumps of $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ in an appropriate way. Namely, we would like to use the coupling by reflection modified in such a way that it allows for a positive probability of $(Y_t)_{t\geq 0}$ jumping to the same point as $(X_t)_{t\geq 0}$. In order to employ this additional feature, we need to modify the Poisson random measure N associated with $(L_t)_{t\geq 0}$ via (1.3). Recall that there exists a sequence $(\tau_j)_{j=1}^{\infty}$ of random variables in $\mathbb{R}_+$ encoding the jump times and a sequence $(\xi_j)_{j=1}^{\infty}$ of random variables in $\mathbb{R}^d$ encoding the jump sizes such that

$$N((0,t], A)(\omega) = \sum_{j=1}^{\infty} \delta_{(\tau_j(\omega), \xi_j(\omega))}((0,t] \times A) \quad \text{for all } \omega \in \Omega \text{ and } A \in \mathcal{B}(\mathbb{R}^d)$$

(see e.g. [18], Chapter 6). At the jump time τ_j the process $(X_t)_{t\geq 0}$ jumps from the point $X_{\tau_j -}$ to X_{τ_j} and our goal is to find a way to determine whether the jump of $(Y_t)_{t\geq 0}$ should be reflected or whether $(Y_t)_{t\geq 0}$ should be forced to jump to the same point that $(X_t)_{t\geq 0}$ jumped to. In order to achieve this, let us observe that instead of considering the Poisson random measure N on $\mathbb{R}_+ \times \mathbb{R}^d$, we can extend it to a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d \times [0, 1]$, replacing the d-dimensional random variables ξ_j determining the jump sizes of $(L_t)_{t\geq 0}$, with the $(d+1)$-dimensional random variables (ξ_j, η_j), where each η_j is a uniformly distributed random variable on $[0, 1]$. Thus we have

$$N((0,t], A)(\omega) = \sum_{j=1}^{\infty} \delta_{(\tau_j(\omega), \xi_j(\omega), \eta_j(\omega))}((0,t] \times A \times [0,1])$$
$$\text{for all } \omega \in \Omega \text{ and } A \in \mathcal{B}(\mathbb{R}^d)$$

and by a slight abuse of notation we can write

$$L_t = \int_0^t \int_{\{|v|>1\}\times[0,1]} v N(ds, dv, du) + \int_0^t \int_{\{|v|\leq 1\}\times[0,1]} v \widetilde{N}(ds, dv, du), \tag{2.5}$$

denoting our extended Poisson random measure also by N. With this notation, if there is a jump at time t, then the process $(X_t)_{t\geq 0}$ moves from the point X_{t-} to $X_{t-} + v$ and we draw a random number $u \in [0, 1]$ which is then used to determine whether the process $(Y_t)_{t\geq 0}$ should jump to the same point that $(X_t)_{t\geq 0}$ jumped to, or whether it should be reflected just like in the "pure" reflection coupling. In order to make this work, we introduce a control function ρ with values in $[0, 1]$ that will determine the probability of bringing the processes together. Our idea is based on the formula (2.1) and uses the minimum of the jump density q and its translation by the difference of the positions of the two coupled processes before the jump time, that is, by the vector

$$Z_{t-} := X_{t-} - Y_{t-}.$$

Our first guess would be to define our control function by

$$\rho(v, Z_{t-}) := \min\left\{ \frac{q(v + Z_{t-})}{q(v)}, 1 \right\} = \frac{q(v + Z_{t-}) \wedge q(v)}{q(v)} \tag{2.6}$$

when $q(v) > 0$. We set $\rho(v, Z_{t-}) := 1$ if $q(v) = 0$. Note that we have $q(v + Z_{t-})/q(v) = q(v + X_{t-} - Y_{t-})/q(v + X_{t-} - X_{t-})$, so we can look at this formula as comparing the translations of q by the vectors Y_{t-} and X_{t-}, respectively. The idea here is that "on average" the probability of bringing the processes together should be equal to the ratio of the overlapping mass of the jump density q and its translation and the total mass of q. However, for technical reasons, we will slightly modify this definition.

Namely, we will only apply our coupling construction presented above to the jumps of size bounded by a constant $m > 0$ satisfying Assumption 3. For the larger jumps we will apply the synchronous coupling, i.e., whenever $(X_t)_{t\geq 0}$ makes a jump of size greater than m, we will let $(Y_t)_{t\geq 0}$ make exactly the same jump. The rationale behind this is the following. First, this modification allows us to control the size of jumps of the difference process $Z_t := X_t - Y_t$. If $(X_t)_{t\geq 0}$ makes a large jump v, then instead of reflecting the jump for $(Y_t)_{t\geq 0}$ and having a large change in the value of Z_t, we make the same jump v with $(Y_t)_{t\geq 0}$ and the value of Z_t does not change at all. Secondly, by doing this we do not in any way spoil the contractivity in W_f that we want to show. As will be evident in the proof, what is crucial for the contractivity is on one hand the reflection applied to small jumps only (see Lemmas 3.3 and 3.6) and on the other the quantity (1.4) from Assumption 3 (see Lemma 3.7). If the latter, however, holds for some $m_0 > 0$ then it also holds for all $m \geq m_0$ and in our calculations we can always choose m large enough if needed (see the inequality (3.16) in the proof of Lemma 3.3 and (3.39) after the proof of Lemma 3.7). Therefore choosing a large but finite m is a better solution than constructing a coupling with $m = \infty$ (i.e., applying our "mirror" construction to jumps of all sizes), which would require us to impose an additional assumption on the size of jumps of the noise $(L_t)_{t\geq 0}$.

Now that we have justified making such an adjustment, note that for any fixed $m > 1$ we can always write (2.5) as

$$L_t = \int_0^t \int_{\{|v|>m\}\times[0,1]} vN(ds, dv, du) + \int_0^t \int_{\{|v|\leq m\}\times[0,1]} v\widetilde{N}(ds, dv, du) + \int_0^t \int_{\{m\geq|v|>1\}\times[0,1]} v\nu(dv)duds.$$

Then we can include the last term appearing above in the drift b in Eq. (1.1) describing $(X_t)_{t\geq 0}$. Obviously such a change of the drift does not influence its dissipativity properties. Thus, once we have fixed a large enough m (see the discussion above), we can for notational convenience redefine $(L_t)_{t\geq 0}$ and b by setting

$$L_t := \int_0^t \int_{\{|v|>m\}\times[0,1]} vN(ds, dv, du) + \int_0^t \int_{\{|v|\leq m\}\times[0,1]} v\widetilde{N}(ds, dv, du) \tag{2.7}$$

and modifying b accordingly.

Since we want to apply different couplings for the compensated and uncompensated parts of $(L_t)_{t\geq 0}$, we actually need to modify the definition (2.6) of the control function ρ by putting

$$\rho(v, Z_{t-}) := \frac{q(v) \wedge q(v + Z_{t-})\mathbf{1}_{\{|v+Z_{t-}|\leq m\}}}{q(v)}.$$

Observe that with our new definition for any integrable function f and any $z \in \mathbb{R}^d$ we have

$$\int_{\{|v|\leq m\}} f(v)\rho(v, z)\nu(dv) = \int_{\{|v|\leq m\}} f(v)\frac{q(v) \wedge q(v + z)\mathbf{1}_{\{|v+z|\leq m\}}}{q(v)} q(v)dv = \int_{\{|v|\leq m\}\cap\{|v+z|\leq m\}} f(v)\,(q(v) \wedge q(v + z))\,dv,$$

while with (2.6) we would just have

$$\int_{\{|v|\leq m\}} f(v)\rho(v, z)\nu(dv) = \int_{\{|v|\leq m\}} f(v)\,(q(v) \wedge q(v + z))\,dv.$$

We will use this fact later in the proof of Lemma 2.5. On an intuitive level, if the distance Z_{t-} between the processes before the jump is big (much larger than m), and we are only considering the jumps bounded by m (and thus $|v + Z_{t-}|$ is still big), then the probability of bringing the processes together should be zero, while the quantity (2.6) can still be positive in such a situation. The restriction we introduce in the definition of ρ eliminates this problem.

To summarize, in our construction once we have the number $u \in [0, 1]$, if the jump vector of $(X_t)_{t\geq 0}$ at time t is v and $|v| \leq m$, then the jump vector of $(Y_t)_{t\geq 0}$ should be $X_{t-} - Y_{t-} + v$ (so

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

that $(Y_t)_{t\geq 0}$ jumps from Y_{t-} to $X_{t-}+v$) when

$$u < \rho(v, Z_{t-}). \tag{2.8}$$

Otherwise the jump of $(Y_t)_{t\geq 0}$ should be v reflected with respect to the hyperplane spanned by the vector $e_{t-} = (X_{t-} - Y_{t-})/|X_{t-} - Y_{t-}|$. If $|v| > m$, then the jump of $(Y_t)_{t\geq 0}$ is the same as the one of $(X_t)_{t\geq 0}$, i.e., it is also given by the vector v.

We are now ready to define our coupling by choosing an appropriate SDE for the process $(Y_t)_{t\geq 0}$. Recall that $(X_t)_{t\geq 0}$ is given by (1.1) and thus

$$dX_t = b(X_t)dt + \int_{\{|v|>m\}\times[0,1]} vN(dt,dv,du) + \int_{\{|v|\leq m\}\times[0,1]} v\widetilde{N}(dt,dv,du). \tag{2.9}$$

Now, in view of the above discussion, we consider the SDE

$$\begin{aligned} dY_t &= b(Y_t)dt + \int_{\{|v|>m\}\times[0,1]} vN(dt,dv,du) \\ &\quad + \int_{\{|v|\leq m\}\times[0,1]} (X_{t-} - Y_{t-} + v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}}\widetilde{N}(dt,dv,du) \\ &\quad + \int_{\{|v|\leq m\}\times[0,1]} R(X_{t-},Y_{t-})v\mathbf{1}_{\{u\geq\rho(v,Z_{t-})\}}\widetilde{N}(dt,dv,du), \end{aligned} \tag{2.10}$$

where

$$R(X_{t-},Y_{t-}) := I - 2\frac{(X_{t-}-Y_{t-})(X_{t-}-Y_{t-})^T}{|X_{t-}-Y_{t-}|^2} = I - 2e_{t-}e_{t-}^T$$

is the reflection operator like in (2.3) with e_t defined by (2.4). Observe that if $Z_{t-} = 0$, then $\rho(v, Z_{t-}) = 1$ and the condition (2.8) is satisfied almost surely, so after Z_t hits zero once, it stays there forever. Thus, if we denote

$$T := \inf\{t \geq 0 : X_t = Y_t\}, \tag{2.11}$$

then $X_t = Y_t$ for any $t \geq T$.

We can equivalently write (2.10) in a more convenient way as

$$\begin{aligned} dY_t &= b(Y_t)dt + \int_{\{|v|>m\}\times[0,1]} vN(dt,dv,du) \\ &\quad + \int_{\{|v|\leq m\}\times[0,1]} R(X_{t-},Y_{t-})v\widetilde{N}(dt,dv,du) \\ &\quad + \int_{\{|v|\leq m\}\times[0,1]} (X_{t-} - Y_{t-} + v - R(X_{t-},Y_{t-})v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}}\widetilde{N}(dt,dv,du). \end{aligned} \tag{2.12}$$

2.3. Auxiliary estimates

At first glance, it is not clear whether the above equation even has a solution or if $(X_t, Y_t)_{t\geq 0}$ indeed is a coupling. Before we answer these questions, we will first show some estimates of the coefficients of (2.12), which will be useful in the sequel (see Lemmas 2.5 and 3.2).

Lemma 2.1 (*Linear Growth*). *There exists a constant $C = C(m) > 0$ such that for any x, $y \in \mathbb{R}^d$ we have*

$$\int_{\{|v|\le m\}\times[0,1]} |x-y+v-R(x,y)v|^2 \mathbf{1}_{\{u<\rho(v,x-y)\}}\nu(dv)du \le C(1+|x-y|^2).$$

Proof. We will keep using the notation $z = x - y$. We have

$$\int_{\{|v|\le m\}} |z+v-R(x,y)v|^2\rho(v,z)\nu(dv) \le 2\int_{\{|v|\le m\}} |z+v|^2\rho(v,z)\nu(dv) + 2\int_{\{|v|\le m\}} |R(x,y)v|^2\rho(v,z)\nu(dv) \quad (2.13)$$

and, since R is an isometry, we can estimate

$$\begin{aligned} 2\int_{\{|v|\le m\}} |R(x,y)v|^2\rho(v,z)\nu(dv) &= 2\int_{\{|v|\le m\}} |v|^2\rho(v,z)\nu(dv) \\ &\le 2\int_{\{|v|\le m\}} |v|^2 q(v+z)\wedge q(v)dv \le 2\int_{\{|v|\le m\}} |v|^2 q(v)dv = 2\int_{\{|v|\le m\}} |v|^2\nu(dv). \end{aligned}$$

The last integral is of course finite, since ν is a Lévy measure. We still have to bound the first integral on the right hand side of (2.13). We have

$$\begin{aligned} 2\int_{\{|v|\le m\}} |z+v|^2\rho(v,z)\nu(dv) &\le 2\int_{\{|v|\le m\}} |z+v|^2 q(v+z)\wedge q(v)dv \\ &= 2\int_{\{|v-z|\le m\}} |v|^2 q(v)\wedge q(v-z)dv. \end{aligned}$$

Now let us consider two cases. First assume that $|z| \le 2m$ (instead of 2 we can also take any positive number strictly greater than 1). Then

$$2\int_{\{|v-z|\le m\}} |v|^2 q(v)\wedge q(v-z)dv \le 2\int_{\{|v-z|\le m\}} |v|^2\nu(dv) \le 2\int_{\{|v|\le 3m\}} |v|^2\nu(dv) < \infty.$$

On the other hand, when $|z| > 2m$, we have

$$\{v\in\mathbb{R}^d : |v-z|\le m\} \subset \{v\in\mathbb{R}^d : |v|\le m\}^c =: B(m)^c,$$

and $\nu(B(m)^c) < \infty$, which allows us to estimate

$$\begin{aligned} &2\int_{\{|v-z|\le m\}} |v|^2 q(v)\wedge q(v-z)dv \\ &\le 4\int_{\{|v-z|\le m\}} |v-z|^2 q(v)\wedge q(v-z)dv + 4\int_{\{|v-z|\le m\}} |z|^2 q(v)\wedge q(v-z)dv \\ &\le 4\int_{\{|v-z|\le m\}} |v-z|^2 q(v-z)dv + 4\int_{\{|v-z|\le m\}} |z|^2 q(v)dv \\ &\le 4\int_{\{|v|\le m\}} |v|^2\nu(dv) + 4|z|^2\nu(B(m)^c). \end{aligned}$$

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

Hence, by choosing

$$C := \max\left\{2\int_{\{|v|\le 3m\}}|v|^2\nu(dv) + 2\int_{\{|v|\le m\}}|v|^2\nu(dv),\, 6\int_{\{|v|\le m\}}|v|^2\nu(dv),\, 4\nu(B(m)^c)\right\}$$

we get the desired result. □

Here we should remark that by the above lemma we have

$$\mathbb{P}\left(\int_0^t\int_{\{|v|\le m\}\times[0,1]}|Z_{s-} + v - R(X_{s-}, Y_{s-})v|^2\mathbf{1}_{\{u<\rho(v,Z_{s-})\}}\nu(dv)duds < \infty\right) = 1.$$

We will use this fact later on.

The next thing we need to show is that the (integrated) coefficients are continuous in the solution variable. Note that obviously

$$\int_{\{|v|\le m\}\times[0,1]}|R(x+h, y)v - R(x, y)v|^2\nu(dv)du \to 0, \quad \text{as } h \to 0,$$

so we just need to take care of the part involving $\rho(v, z)$. Before we proceed though, let us make note of the following fact.

Remark 2.2. For a fixed value of $z \neq 0$, the measure

$$\rho(v, z)\nu(dv)$$

is a finite measure on $\mathbb{R}^d$. Indeed, if $z \neq 0$, we can choose a neighbourhood U of z such that $0 \notin \bar{U}$. Then $U - z$ is a neighbourhood of 0 and we have

$$\begin{aligned}\int_{\mathbb{R}^d}\rho(v, z)\nu(dv) &= \int_U \rho(v, z)\nu(dv) + \int_{U^c}\rho(v, z)\nu(dv)\\ &\le \int_U q(v)dv + \int_{U^c} q(v+z)dv\\ &= \int_U q(v)dv + \int_{(U-z)^c} q(v)dv < \infty,\end{aligned}$$

since ν is a Lévy measure.

Lemma 2.3 (*Continuity Condition*)**.** *For any $x, y \in \mathbb{R}^d$ and $z = x - y$ we have*

$$\begin{aligned}&\int_{\{|v|\le m\}\times[0,1]}|(x+h-y+v-R(x+h, y)v)\mathbf{1}_{\{u<\rho(v,z+h)\}}\\ &\quad - (x-y+v-R(x, y)v)\mathbf{1}_{\{u<\rho(v,z)\}}|^2\nu(dv)du \to 0, \quad \textit{as } h \to 0.\end{aligned}$$

Proof. We have

$$\begin{aligned}&\int_{\{|v|\le m\}\times[0,1]}|(x+h-y+v-R(x+h, y)v)\mathbf{1}_{\{u<\rho(v,z+h)\}}\\ &\qquad - (x-y+v-R(x, y)v)\mathbf{1}_{\{u<\rho(v,z)\}}|^2\nu(dv)du\\ &\quad = \int_{\{|v|\le m\}\times[0,1]}|(x+h-y+v-R(x+h, y)v)\mathbf{1}_{\{u<\rho(v,z+h)\}}\\ &\qquad - (x-y+v-R(x, y)v)\mathbf{1}_{\{u<\rho(v,z+h)\}} + (x-y+v-R(x, y)v)\mathbf{1}_{\{u<\rho(v,z+h)\}}\end{aligned}$$

$$
\begin{aligned}
&\quad - (x - y + v - R(x,y)v)\mathbf{1}_{\{u<\rho(v,z)\}}|^2 \nu(dv)du \\
&\le 2\int_{\{|v|\le m\}} |h - R(x+h,y)v + R(x,y)v|^2 \rho(v,z+h)\nu(dv) \\
&\quad + 2\int_{\{|v|\le m\}} |x-y+v-R(x,y)v|^2 |\rho(v,z+h)-\rho(v,z)|\nu(dv) \\
&=: I_1 + I_2.
\end{aligned}
$$

Taking into account Remark 2.2 and using the dominated convergence theorem, we can easily show that I_1 converges to zero when $h \to 0$. As for I_2, observe that

$$
\begin{aligned}
&|\rho(v,z+h)-\rho(v,z)|\mathbf{1}_{\{|v|\le m\}} \\
&= \frac{|q(v+z+h)\mathbf{1}_{\{|v+z+h|\le m\}} \wedge q(v) - q(v+z)\mathbf{1}_{\{|v+z|\le m\}} \wedge q(v)|}{|q(v)|} \mathbf{1}_{\{|v|\le m\}}.
\end{aligned}
$$

Recall that by Assumption 2, the density q is continuous almost everywhere on $\mathbb{R}^d$. Moreover, for a fixed $z \in \mathbb{R}^d$ the function $\mathbf{1}_{\{|v+z|\le m\}}$ is continuous outside of the set $\{v \in \mathbb{R}^d : |v+z| = m\}$, which is of measure zero. Therefore, using the dominated convergence theorem once again, we show that $I_2 \to 0$ when $h \to 0$. $\square$

2.4. *Existence of a solution*

Note that having the above estimates, it would be possible to prove existence of a weak solution to the $2d$-dimensional system given by (2.9) and (2.10), using Theorem 175 in [24]. However, there is a simpler method allowing to prove even more, namely, existence of a unique strong solution. To this end, we will use the so-called interlacing technique. This technique of modifying the paths of a process by adding jumps defined by a Poisson random measure of finite intensity is well known, cf. e.g. Theorem IV-9.1 in [9] or Theorem 6.2.9 in [1]. We first notice that without loss of generality it allows us to focus on the small jumps of size bounded by m, as we can always add the big jumps later, both to $(X_t)_{t\ge 0}$ and $(Y_t)_{t\ge 0}$. Hence we can consider the equation for $(Y_t)_{t\ge 0}$ written as

$$
\begin{aligned}
dY_t &= b(Y_t)dt + \int_{\{|v|\le m\}\times[0,1]} R(X_{t-},Y_{t-})v\widetilde{N}(dt,dv,du) \\
&\quad + \int_{\{|v|\le m\}\times[0,1]} (X_{t-} - Y_{t-} + v - R(X_{t-},Y_{t-})v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}}\widetilde{N}(dt,dv,du).
\end{aligned}
\tag{2.14}
$$

Now observe that if we only consider the equation

$$
dY_t^1 = b(Y_t^1)dt + \int_{\{|v|\le m\}\times[0,1]} R(X_{t-},Y_{t-}^1)v\widetilde{N}(dt,dv,du), \tag{2.15}
$$

it is easy to see that it has a unique strong solution since the process $(X_t, Y_t^1)_{t\ge 0}$ up to its coupling time T takes values in the region of $\mathbb{R}^{2d}$ in which the function R is locally Lipschitz and has linear growth. Then note that the second integral appearing in (2.14) represents a sum of jumps of which (almost surely) there is only a finite number on any finite time interval, since

$$
\int_{\mathbb{R}^d\times[0,1]} \mathbf{1}_{\{u<\rho(v,Z_{t-})\}}\nu(dv)du = \int_{\mathbb{R}^d} \rho(v,Z_{t-})\nu(dv) < \infty,
$$

as long as $Z_{t-} \neq 0$ (see Remark 2.2). Then in principle in such situations it is possible to use the interlacing technique to modify the paths of the process $(Y_t^1)_{t\geq 0}$ by adding the jumps defined by the second integral in (2.14), see e.g. the proof of Proposition 2.2 in [15] for a similar construction. Here, however, our particular case is even simpler. Namely, let us consider a uniformly distributed random variable $\xi \in [0,1]$ and define

$$\tau_1 := \inf\{t > 0 : \xi < \rho(\Delta L_t, Z_{t-}^1)\},$$

where $Z_t^1 := X_t - Y_t^1$ and $(L_t)_{t\geq 0}$ is the Lévy process associated with N. Then if we define a process $(Y_t^2)_{t\geq 0}$ by adding the jump of size $X_{\tau_1 -} - Y_{\tau_1 -}^1 + \Delta L_{\tau_1} - R(X_{\tau_1 -}, Y_{\tau_1 -}^1)\Delta L_{\tau_1}$ to the path of $(Y_t^1)_{t\geq 0}$ at time τ_1, we see that $Y_{\tau_1}^2 = X_{\tau_1}$. Moreover, since $\rho(v,0) = 1$ for any $v \in \mathbb{R}^d$, we have $Y_t^2 = X_t$ for all $t \geq \tau_1$. Thus we only need to add one jump to the solution of (2.15) in order to obtain a process which behaves like a solution to (2.14) up to the coupling time, and like the process $(X_t)_{t\geq 0}$ later on. In consequence we obtain a solution $(X_t, Y_t)_{t\geq 0}$ to the system defined by (2.9) and (2.10).

2.5. *Proof that* $(X_t, Y_t)_{t\geq 0}$ *is a coupling*

By the previous subsection, we already have the existence of the process $(X_t, Y_t)_{t\geq 0}$ defined as a solution to (2.9) and (2.10). However, we still need to show that $(X_t, Y_t)_{t\geq 0}$ is indeed a coupling. If we denote

$$B(X_{t-}, Y_{t-}, v, u) := R(X_{t-}, Y_{t-})v + (Z_{t-} + v - R(X_{t-}, Y_{t-})v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}} \tag{2.16}$$

and

$$\begin{aligned}\widetilde{L}_t := &\int_0^t \int_{\{|v|>m\}\times[0,1]} vN(ds,dv,du)\\ &+ \int_0^t \int_{\{|v|\leq m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds,dv,du),\end{aligned} \tag{2.17}$$

then we can write Eq. (2.12) for $(Y_t)_{t\geq 0}$ as

$$dY_t = b(Y_t)dt + d\widetilde{L}_t.$$

Then, if we show that $(\widetilde{L}_t)_{t\geq 0}$ is a Lévy process with the same finite dimensional distributions as $(L_t)_{t\geq 0}$ defined by (2.7), our assertion follows from the uniqueness in law of solutions to Eq. (1.1). An analogous fact in the Brownian case was proved using the Lévy characterization theorem for Brownian motion. Here the proof is more involved, although the idea is very similar. It is sufficient to show two things. First we need to prove that for any $z \in \mathbb{R}^d$ and any $t \geq 0$ we have

$$\mathbb{E}\exp(i\langle z, \widetilde{L}_t\rangle) = \mathbb{E}\exp(i\langle z, L_t\rangle). \tag{2.18}$$

Then we must also show that for any $t > s \geq 0$ the increment

$$\widetilde{L}_t - \widetilde{L}_s$$

is independent of $\mathcal{F}_s$, where $(\mathcal{F}_t)_{t\geq 0}$ is the filtration generated by $(L_t)_{t\geq 0}$. We will need the following lemma.

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

Lemma 2.4. *Let $f(v,u)$ be a random function on $\{|v| \leq m\} \times [0,1]$, measurable with respect to $\mathcal{F}_{t_1}$. If*

$$\mathbb{P}\left(\int_{\{|v|\leq m\}\times[0,1]} |f(v,u)|^2 \nu(dv)du < \infty\right) = 1, \tag{2.19}$$

then

$$\mathbb{E}\left[\exp\left(i\left\langle z, \int_{t_1}^{t_2}\int_{\{|v|\leq m\}\times[0,1]} f(v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\Big|\mathcal{F}_{t_1}\right]$$
$$= \exp\left((t_2-t_1)\int_{\{|v|\leq m\}\times[0,1]}\left(e^{i\langle z, f(v,u)\rangle} - 1 - i\langle z, f(v,u)\rangle\right)\nu(dv)du\right). \tag{2.20}$$

Proof. By a standard argument, if the condition (2.19) is satisfied, we can approximate $\int_{t_1}^{t_2}\int_{\{|v|\leq m\}\times[0,1]} f(v,u)\widetilde{N}(ds,dv,du)$ in probability by integrals of step functions f^n of the form

$$f^n(v,u) = \sum_{j=1}^{l_n} c_j \mathbf{1}_{A_j}$$

where A_j are pairwise disjoint subsets of $\{|v| \leq m\} \times [0,1]$ such that $(\nu \times \lambda)(A_j) < \infty$ for all j, where λ is the Lebesgue measure on $[0,1]$ and c_j are $\mathcal{F}_{t_1}$-measurable random variables. Thus it is sufficient to show (2.20) for the step functions f^n and then pass to the limit using the dominated convergence theorem for conditional expectations. Indeed, for every f^n we can show that

$$\mathbb{E}\left[\exp\left(i\left\langle z, \int_{t_1}^{t_2}\int_{\{|v|\leq m\}\times[0,1]} f^n(v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\Big|\mathcal{F}_{t_1}\right]$$
$$= \mathbb{E}\left[\prod_{j=1}^{l_n}\exp\left(i\left\langle z, \int_{t_1}^{t_2}\int_{\{|v|\leq m\}\times[0,1]} c_j\mathbf{1}_{A_j}\widetilde{N}(ds,dv,du)\right\rangle\right)\Big|\mathcal{F}_{t_1}\right]$$
$$= \mathbb{E}\left[\prod_{j=1}^{l_n}\exp\left(i\left\langle z, c_j\widetilde{N}((t_1,t_2],A_j)\right\rangle\right)\Big|\mathcal{F}_{t_1}\right].$$

The random variables $\widetilde{N}((t_1,t_2],A_j)$ are mutually independent and they are all independent of $\mathcal{F}_{t_1}$ and the random variables c_j are $\mathcal{F}_{t_1}$-measurable so we know that we can calculate the above conditional expectation as just an expectation with c_j constant and then plug the random c_j back in. Thus we get

$$\mathbb{E}\prod_{j=1}^{l_n}\exp\left(i\left\langle z, c_j\widetilde{N}((t_1,t_2],A_j)\right\rangle\right) = \prod_{j=1}^{l_n}\mathbb{E}\exp\left(i\left\langle z, c_j\widetilde{N}((t_1,t_2],A_j)\right\rangle\right)$$
$$= \prod_{j=1}^{l_n}\exp\left((t_2-t_1)\left(e^{i\langle z,c_j\rangle}(\nu\times\lambda)(A_j) - 1 - i\langle z, c_j\rangle(\nu\times\lambda)(A_j)\right)\right)$$
$$= \exp\left((t_2-t_1)\int_{\{|v|\leq m\}\times[0,1]}\left(e^{i\langle z, f^n(v,u)\rangle} - 1 - i\langle z, f^n(v,u)\rangle\right)\nu(dv)du\right),$$

where in the second step we just used the formula for the characteristic function of the Poisson distribution. □

Now we will prove (2.18) in the special case where

$$\widetilde{L}_t = \int_0^t \int_{\{|v|\leq m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds, dv, du)$$

and the process $(L_t)_{t\geq 0}$ is also considered without the large jumps. Once we have this, it is easy to extend the result to the general case where $(\widetilde{L}_t)_{t\geq 0}$ is given by (2.17).

Lemma 2.5. *For every $t > 0$ and every $z \in \mathbb{R}^d$ we have*

$$\begin{aligned}&\mathbb{E}\exp\left(i\left\langle z, \int_0^t \int_{\{|v|\leq m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds, dv, du)\right\rangle\right)\\ &\quad = \mathbb{E}\exp\left(i\left\langle z, \int_0^t \int_{\{|v|\leq m\}\times[0,1]} v\widetilde{N}(ds, dv, du)\right\rangle\right).\end{aligned}$$

Proof. First recall that we have

$$\mathbb{P}\left(\int_{\{|v|\leq m\}\times[0,1]} |B(X_{t-}, Y_{t-}, v, u)|^2\nu(dv)du < \infty\right) = 1$$

(see the remark after the proof of Lemma 2.1). Then observe that by Lemma 2.3 we know that the square integrated process B, i.e., the process

$$\int_{\{|v|\leq m\}\times[0,1]} |B(X_{t-}, Y_{t-}, v, u)|^2\nu(dv)du$$

has left-continuous trajectories. This means that (almost surely) we can approximate $B(X_{t-}, Y_{t-}, v, u)$ in $L^2([0,t]\times(\{|v|\leq m\};\nu)\times[0,1])$ by Riemann sums of the form

$$B^n(s, v, u) := \sum_{k=0}^{m_n-1} B(X_{t_k^n}, Y_{t_k^n}, v, u)\mathbf{1}_{(t_k^n, t_{k+1}^n]}(s) \tag{2.21}$$

for some sequence of partitions $0 = t_0^n < t_1^n < \cdots < t_{m_n}^n = t$ of the interval $[0,t]$ with the mesh size going to zero as $n \to \infty$. From the general theory of stochastic integration with respect to Poisson random measures (see e.g. [1], Section 4.2) it follows that the sequence of integrals $\int_0^t \int_{\{|v|\leq m\}\times[0,1]} B^n(s, v, u)\widetilde{N}(ds, dv, du)$ converges in probability to the integral $\int_0^t \int_{\{|v|\leq m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds, dv, du)$. Thus we have

$$\begin{aligned}&\mathbb{E}\exp\left(i\left\langle z, \int_0^t \int_{\{|v|\leq m\}\times[0,1]} B^n(s, v, u)\widetilde{N}(ds, dv, du)\right\rangle\right)\\ &\quad \to \mathbb{E}\exp\left(i\left\langle z, \int_0^t \int_{\{|v|\leq m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds, dv, du)\right\rangle\right)\end{aligned}$$

for any $z \in \mathbb{R}^d$ and $t > 0$, as $n \to \infty$. We will show now that in fact for all $n \in \mathbb{N}$ we have

$$\mathbb{E}\exp\left(i\left\langle z,\int_0^t\int_{\{|v|\leq m\}\times[0,1]}B^n(s,v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)$$
$$=\mathbb{E}\exp\left(i\left\langle z,\int_0^t\int_{\{|v|\leq m\}\times[0,1]}v\widetilde{N}(ds,dv,du)\right\rangle\right),\tag{2.22}$$

which will prove the desired assertion. To this end, let us calculate

$$\begin{aligned}
&\mathbb{E}\exp\left(i\left\langle z,\int_0^t\int_{\{|v|\leq m\}\times[0,1]}B^n(s,v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\\
&=\mathbb{E}\exp\left(i\left\langle z,\sum_{k=0}^{m_n-1}\int_{t_k^n}^{t_{k+1}^n}\int_{\{|v|\leq m\}\times[0,1]}B(X_{t_k^n},Y_{t_k^n},v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\\
&=\mathbb{E}\Bigg(\mathbb{E}\Bigg[\prod_{k=0}^{m_n-2}\exp\left(i\left\langle z,\int_{t_k^n}^{t_{k+1}^n}\int_{\{|v|\leq m\}\times[0,1]}B(X_{t_k^n},Y_{t_k^n},v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\\
&\quad\times\exp\left(i\left\langle z,\int_{t_{m_n-1}^n}^{t_{m_n}^n}\int_{\{|v|\leq m\}\times[0,1]}B(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n},v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\Bigg|\mathcal{F}_{m_n-1}^n\Bigg]\Bigg)\\
&=\mathbb{E}\Bigg(\prod_{k=0}^{m_n-2}\exp\left(i\left\langle z,\int_{t_k^n}^{t_{k+1}^n}\int_{\{|v|\leq m\}\times[0,1]}B(X_{t_k^n},Y_{t_k^n},v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\\
&\quad\times\mathbb{E}\left[\exp\left(i\left\langle z,\int_{t_{m_n-1}^n}^{t_{m_n}^n}\int_{\{|v|\leq m\}\times[0,1]}B(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n},v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\Bigg|\mathcal{F}_{m_n-1}^n\right]\Bigg).
\end{aligned}\tag{2.23}$$

Now we can use Lemma 2.4 to evaluate the conditional expectation appearing above as

$$\exp\Bigg((t_{m_n-1}^n-t_{m_n}^n)$$
$$\times\int_{\{|v|\leq m\}\times[0,1]}\left(e^{i\langle z,B(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n},v,u)\rangle}-1-i\langle z,B(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n},v,u)\rangle\right)\nu(dv)du\Bigg).$$

Here comes the crucial part of our proof. We will show that

$$\int_{\{|v|\leq m\}\times[0,1]}\left(e^{i\langle z,B(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n},v,u)\rangle}-1-i\langle z,B(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n},v,u)\rangle\right)\nu(dv)du$$
$$=\int_{\{|v|\leq m\}\times[0,1]}\left(e^{i\langle z,v\rangle}-1-i\langle z,v\rangle\right)\nu(dv)du.\tag{2.24}$$

Let us fix the values of $X_{t_{m_n-1}^n}$ and $Y_{t_{m_n-1}^n}$ for the moment and denote

$$R:=R(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n})\quad\text{and}\quad c:=X_{t_{m_n-1}^n}-Y_{t_{m_n-1}^n}=Z_{t_{m_n-1}^n}.\tag{2.25}$$

Then, using the formula (2.16) we can write

$$B(X_{t_{m_n-1}^n},Y_{t_{m_n-1}^n},v,u)=Rv+(c+v-Rv)\mathbf{1}_{\{u<\rho(v,c)\}}.$$

Next, integrating over [0, 1] with respect to u, we get

$$\int_{\{|v|\le m\}\times[0,1]}\left(e^{i\langle z,B(X_{t^n_{m_n-1}},Y_{t^n_{m_n-1}},v,u)\rangle}-1-i\langle z,B(X_{t^n_{m_n-1}},Y_{t^n_{m_n-1}},v,u)\rangle\right)\nu(dv)du$$
$$=\int_{\{|v|\le m\}}\Big(e^{i\langle z,Rv\rangle}\left(e^{i\langle z,c+v-Rv\rangle}\rho(v,c)+(1-\rho(v,c))\right)-1$$
$$-i\langle z,Rv\rangle-i\langle z,c+v-Rv\rangle\rho(v,c)\Big)\nu(dv).$$

Since $|B(X_{t^n_{m_n-1}},Y_{t^n_{m_n-1}},v,u)|^2$ is integrable with respect to $\nu\times\lambda$ over $\{|v|\le m\}\times[0,1]$,

$$e^{i\langle z,Rv\rangle}\left(e^{i\langle z,c+v-Rv\rangle}\rho(v,c)+(1-\rho(v,c))\right)-1-i\langle z,Rv\rangle-i\langle z,c+v-Rv\rangle\rho(v,c)$$

is integrable with respect to ν over $\{|v|\le m\}$. Moreover, $e^{i\langle z,Rv\rangle}-1-i\langle z,Rv\rangle$ is also integrable over $\{|v|\le m\}$. In fact, since ν is assumed to be rotationally invariant and R is an orthogonal matrix, we easily see that

$$\int_{\{|v|\le m\}}\left(e^{i\langle z,Rv\rangle}-1-i\langle z,Rv\rangle\right)\nu(dv)=\int_{\{|v|\le m\}}\left(e^{i\langle z,v\rangle}-1-i\langle z,v\rangle\right)\nu(dv).$$

We infer that $\left(e^{i\langle z,Rv\rangle}(e^{i\langle z,c+v-Rv\rangle}-1)-i\langle z,c+v-Rv\rangle\right)\rho(v,c)$ is also integrable with respect to ν over $\{|v|\le m\}$. Now we will show that the integral of this function actually vanishes. Note that we have $R=I-2cc^T/|c|^2$ and since q is the density of a rotationally invariant measure ν, we have $q(Rv)=q(v)$ and $q(Rv-c)=q(v+c)$ for any $v\in\mathbb{R}^d$. Now

$$\int_{\{|v|\le m\}}\left(e^{i\langle z,Rv\rangle}(e^{i\langle z,c+v-Rv\rangle}-1)-i\langle z,c+v-Rv\rangle\right)\rho(v,c)\nu(dv)$$
$$=\int_{\{|v|\le m\}}\left(e^{i\langle z,c+v\rangle}-e^{i\langle z,Rv\rangle}-i\langle z,c+v\rangle+i\langle z,Rv\rangle\right)q(v)\wedge q(v+c)\mathbf{1}_{\{|v+c|\le m\}}dv$$
$$=\int_{\{|v-c|\le m\}\cap\{|v|\le m\}}\left(e^{i\langle z,v\rangle}-e^{i\langle z,R(v-c)\rangle}-i\langle z,v\rangle+i\langle z,R(v-c)\rangle\right)q(v-c)\wedge q(v)dv$$
$$=\int_{\{|v-c|\le m\}\cap\{|v|\le m\}}\left(e^{i\langle z,v\rangle}-e^{i\langle z,Rv+c\rangle}-i\langle z,v\rangle+i\langle z,Rv+c\rangle\right)q(v-c)\wedge q(v)dv$$
$$=\int_{\{|Rv-c|\le m\}\cap\{|Rv|\le m\}}\left(e^{i\langle z,Rv\rangle}-e^{i\langle z,v+c\rangle}-i\langle z,Rv\rangle+i\langle z,v+c\rangle\right)q(Rv-c)\wedge q(Rv)dv$$
$$=\int_{\{|v+c|\le m\}\cap\{|v|\le m\}}\left(e^{i\langle z,Rv\rangle}-e^{i\langle z,v+c\rangle}-i\langle z,Rv\rangle+i\langle z,v+c\rangle\right)q(v+c)\wedge q(v)dv$$
$$=\int_{\{|v|\le m\}}\left(e^{i\langle z,Rv\rangle}-e^{i\langle z,v+c\rangle}-i\langle z,Rv\rangle+i\langle z,v+c\rangle\right)\rho(v,c)\nu(dv)$$
$$=-\int_{\{|v|\le m\}}\left(e^{i\langle z,Rv\rangle}(e^{i\langle z,c+v-Rv\rangle}-1)-i\langle z,c+v-Rv\rangle\right)\rho(v,c)\nu(dv),$$

where in the second step we use a change of variables from v to $v-c$, in the third step we use the fact that $Rc=-c$, in the fourth step we change the variables from v to Rv and in the fifth step we use the symmetry properties $|Rv-c|=|v+c|$ and $|Rv|=|v|$. Hence we have shown (2.24). Now we return to our calculations in (2.23) and compute

$$\mathbb{E}\left(\prod_{k=0}^{m_n-2}\exp\left(i\left\langle z,\int_{t^n_k}^{t^n_{k+1}}\int_{\{|v|\le m\}\times[0,1]}B(X_{t^n_k},Y_{t^n_k},v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)\right.$$

$$\times \mathbb{E}\left[\exp\left(i\left\langle z, \int_{t^n_{m_n-1}}^{t^n_{m_n}} \int_{\{|v|\le m\}\times[0,1]} B(X_{t^n_{m_n-1}}, Y_{t^n_{m_n-1}}, v, u)\widetilde{N}(ds, dv, du)\right\rangle\right)\Big| \mathcal{F}^n_{m_n-1}\right]\right)$$
$$= \exp\left((t^n_{m_n} - t^n_{m_n-1}) \int_{\{|v|\le m\}\times[0,1]} \left(e^{i\langle z,v\rangle} - 1 - i\langle z, v\rangle\right) \nu(dv)du\right)$$
$$\times \mathbb{E}\left(\prod_{k=0}^{m_n-2} \exp\left(i\left\langle z, \int_{t^n_k}^{t^n_{k+1}} \int_{\{|v|\le m\}\times[0,1]} B(X_{t^n_k}, Y_{t^n_k}, v, u)\widetilde{N}(ds, dv, du)\right\rangle\right)\right). \quad (2.26)$$

Then we can just repeat all the steps from (2.23) to (2.26), this time conditioning on $\mathcal{F}^n_{m_n-2}$, and after repeating this procedure $m_n - 1$ times, we get (2.22). □

It remains now to show the independence of the increments of $(\widetilde{L}_t)_{t\ge0}$.

Lemma 2.6. *Under the above assumptions, for any $t_2 > t_1 \ge 0$ the random variable $\widetilde{L}_{t_2} - \widetilde{L}_{t_1}$ is independent of $\mathcal{F}_{t_1}$.*

Proof. We will show that for an arbitrary $\mathcal{F}_{t_1}$-measurable random variable ξ and for any $z_1, z_2 \in \mathbb{R}^d$ we have

$$\mathbb{E}\exp\left(i\left\langle z_1, \int_{t_1}^{t_2} \int_{\{|v|\le m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds, dv, du)\right\rangle + i\langle z_2, \xi\rangle\right)$$
$$= \mathbb{E}\exp\left(i\left\langle z_1, \int_{t_1}^{t_2} \int_{\{|v|\le m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds, dv, du)\right\rangle\right) \cdot \mathbb{E}\exp(i\langle z_2, \xi\rangle).$$

As in the proof of Lemma 2.5, the integral $\int_{t_1}^{t_2} \int_{\{|v|\le m\}\times[0,1]} B(X_{s-}, Y_{s-}, v, u)\widetilde{N}(ds, dv, du)$ can be approximated by integrals of Riemann sums $B^n(s, v, u)$ that have been defined by (2.21) for some sequence of partitions $t_1 = t^n_0 < t^n_1 < \cdots < t^n_{m_n} = t_2$ such that $\delta_n := \max_{k\in\{0,\ldots,m_n-1\}} |t^n_{k+1} - t^n_k| \to 0$ as $n \to \infty$. Denote

$$I^n_k := \int_{t^n_k}^{t^n_{k+1}} \int_{\{|v|\le m\}\times[0,1]} B(X_{t^n_k}, Y_{t^n_k}, v, u)\widetilde{N}(ds, dv, du), \qquad I^n := \sum_{k=0}^{m_n-1} I^n_k.$$

Then we have

$$\mathbb{E}\exp\left(i\langle z_1, I^n\rangle + i\langle z_2, \xi\rangle\right) = \mathbb{E}\left(\exp(i\langle z_2, \xi\rangle) \prod_{k=0}^{m_n-1} \exp(i\langle z_1, I^n_k\rangle)\right)$$
$$= \mathbb{E}\left(\mathbb{E}\left[\exp(i\langle z_2, \xi\rangle) \prod_{k=0}^{m_n-1} \exp(i\langle z_1, I^n_k\rangle)\Big| \mathcal{F}_{t^n_{m_n-1}}\right]\right)$$
$$= \mathbb{E}\left(\exp(i\langle z_2, \xi\rangle) \prod_{k=0}^{m_n-2} \exp(i\langle z_1, I^n_k\rangle)\mathbb{E}\left[\exp(i\langle z_1, I^n_{m_n-1}\rangle)\Big| \mathcal{F}_{t^n_{m_n-1}}\right]\right), \quad (2.27)$$

where in the last step we used the fact that for every $k \in \{0, \ldots, m_n - 1\}$ the random variable ξ is $\mathcal{F}_{t_1} \subset \mathcal{F}_{t^n_k}$-measurable. Now, using Lemma 2.4 and our calculations from the proof of Lemma 2.5, we can show that

$$\mathbb{E}\left[\exp(i\langle z_1, I^n_{m_n-1}\rangle)\Big| \mathcal{F}_{t^n_{m_n-1}}\right] = \mathbb{E}\exp\left(i\left\langle z_1, \int_{t^n_{m_n-1}}^{t^n_{m_n}} \int_{\{|v|\le m\}\times[0,1]} v\widetilde{N}(ds, dv, du)\right\rangle\right)$$

and thus we see that the expression on the right hand side of (2.27) is equal to

$$\mathbb{E}\exp\left(i\left\langle z_1, \int_{t^n_{m_n-1}}^{t^n_{m_n}}\int_{\{|v|\leq m\}\times[0,1]} v\widetilde{N}(ds,dv,du)\right\rangle\right)$$
$$\mathbb{E}\left(\exp(i\langle z_2,\xi\rangle)\prod_{k=0}^{m_n-2}\exp(i\langle z_1, I^n_k\rangle)\right).$$

Thus, by repeating the above procedure $m_n - 1$ times (conditioning on the consecutive σ-fields $\mathcal{F}_{t^n_k}$), we get

$$\mathbb{E}\exp\left(i\langle z_1, I^n\rangle + i\langle z_2,\xi\rangle\right) = \mathbb{E}\exp(i\langle z_2,\xi\rangle)$$
$$\times\prod_{k=0}^{m_n-1}\mathbb{E}\exp\left(i\left\langle z_1, \int_{t^n_k}^{t^n_{k+1}}\int_{\{|v|\leq m\}\times[0,1]} v\widetilde{N}(ds,dv,du)\right\rangle\right). \tag{2.28}$$

However, by the same argument as above we can show that

$$\prod_{k=0}^{m_n-1}\mathbb{E}\exp\left(i\left\langle z_1, \int_{t^n_k}^{t^n_{k+1}}\int_{\{|v|\leq m\}\times[0,1]} v\widetilde{N}(ds,dv,du)\right\rangle\right) = \mathbb{E}\exp(i\langle z_1, I^n\rangle).$$

Since I^n converges in probability to $\int_{t_1}^{t_2}\int_{\{|v|\leq m\}\times[0,1]} B(X_{s-},Y_{s-},v,u)\widetilde{N}(ds,dv,du)$, we get

$$\mathbb{E}\exp(i\langle z_1, I^n\rangle) \to \mathbb{E}\exp\left(i\left\langle z_1, \int_{t_1}^{t_2}\int_{\{|v|\leq m\}\times[0,1]} B(X_{s-},Y_{s-},v,u)\widetilde{N}(ds,dv,du)\right\rangle\right)$$

and, by passing to a subsequence for which almost sure convergence holds and using the dominated convergence theorem, we get

$$\mathbb{E}\exp(i\langle z_1, I^n\rangle + i\langle z_2,\xi\rangle)$$
$$\to \mathbb{E}\exp\left(i\left\langle z_1, \int_{t_1}^{t_2}\int_{\{|v|\leq m\}\times[0,1]} B(X_{s-},Y_{s-},v,u)\widetilde{N}(ds,dv,du)\right\rangle + i\langle z_2,\xi\rangle\right),$$

which proves the desired assertion. □

3. Proof of the inequality (1.8)

In this section we want to apply the coupling that we constructed in Section 2 to prove Corollary 1.2, which follows easily from the inequality (1.8). Namely, in order to obtain

$$W_f(\mu p_t, \nu p_t) \leq e^{-ct} W_f(\mu,\nu), \tag{3.1}$$

we will prove that

$$\mathbb{E}f(|X_t - Y_t|) \leq e^{-ct}\mathbb{E}f(|X_0 - Y_0|), \tag{3.2}$$

where $(X_t, Y_t)_{t\geq 0}$ is the coupling defined by (2.9) and (2.10) and the laws of the random variables X_0 and Y_0 are μ and ν, respectively. Obviously, straight from the definition of the distance W_f we see that for any coupling $(X_t, Y_t)_{t\geq 0}$ the expression $\mathbb{E}f(|X_t - Y_t|)$ gives an upper bound for $W_f(\mu p_t, \nu p_t)$ and since we can prove (3.2) for any coupling of the initial conditions X_0 and Y_0, it is easy to see that (3.2) indeed implies (3.1). Note that without loss of generality we can assume that $\mathbb{P}(X_0 \neq Y_0) = 1$. Indeed, given any probability measures μ and ν we can decompose them by writing

$$\mu = \mu \wedge \nu + \widetilde{\mu} \quad \text{and} \quad \nu = \mu \wedge \nu + \widetilde{\nu} \tag{3.3}$$

for some finite measures $\widetilde{\mu}$ and $\widetilde{\nu}$ on $\mathbb{R}^d$. Then, if $\alpha := (\mu \wedge \nu)(\mathbb{R}^d) \in (0,1)$, we can define probability measures $\overline{\mu} := \widetilde{\mu}/\widetilde{\mu}(\mathbb{R}^d)$ and $\overline{\nu} := \widetilde{\nu}/\widetilde{\nu}(\mathbb{R}^d)$ and we can easily show that $W_f(\mu,\nu) = (1-\alpha)W_f(\overline{\mu},\overline{\nu})$. Obviously, the decomposition (3.3) is preserved by the semigroup $(p_t)_{t\geq 0}$ and thus we see that in order to show (3.1) it is sufficient to show that $W_f(\overline{\mu}p_t, \overline{\nu}p_t) \leq e^{-ct} W_f(\overline{\mu},\overline{\nu})$.

In our proof we will aim to obtain estimates of the form

$$\mathbb{E}f(|Z_t|) - \mathbb{E}f(|Z_0|) \leq \mathbb{E}\int_0^t -cf(|Z_s|)ds, \tag{3.4}$$

for some constant $c > 0$, where $Z_t = X_t - Y_t$, which by the Gronwall inequality will give us (3.2). We assume that f is of the form

$$f = f_1 + f_2,$$

where $f_1 \in \mathcal{C}^2$, $f_1' \geq 0$, $f_1'' \leq 0$ and $f_1(0) = 0$ and $f_2 = a\mathbf{1}_{(0,\infty)}$ for some constant $a > 0$ to be chosen later. We also choose f_1 in such a way that $f_1'(0) = 1$ and thus $f_1' \leq 1$ since f_1' is decreasing. Recall that our coupling is defined in such a way that the equation for the difference process $Z_t = X_t - Y_t$ is given by

$$\begin{aligned} dZ_t &= (b(X_t) - b(Y_t))dt + \int_{\{|v|\leq m\}\times[0,1]} (I - R(X_{t-},Y_{t-}))v\widetilde{N}(dt,dv,du) \\ &\quad - \int_{\{|v|\leq m\}\times[0,1]} (Z_{t-} + v - R(X_{t-},Y_{t-})v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}}\widetilde{N}(dt,dv,du). \end{aligned} \tag{3.5}$$

Note that the jumps of size greater than m cancel out, since we apply synchronous coupling for $|v| > m$ in our construction of the process $(Y_t)_{t\geq 0}$. In order to simplify the notation, let us denote

$$A(X_{t-},Y_{t-},v,u) := -(Z_{t-} + v - R(X_{t-},Y_{t-})v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}}. \tag{3.6}$$

Then we can write

$$\begin{aligned} dZ_t &= (b(X_t) - b(Y_t))dt + \int_{\{|v|\leq m\}\times[0,1]} (I - R(X_{t-},Y_{t-}))v\widetilde{N}(dt,dv,du) \\ &\quad + \int_{\{|v|\leq m\}\times[0,1]} A(X_{t-},Y_{t-},v,u)\widetilde{N}(dt,dv,du). \end{aligned} \tag{3.7}$$

Let us split our computations into two parts by writing

$$\mathbb{E}f(|Z_t|) - \mathbb{E}f(|Z_0|) = \mathbb{E}f_1(|Z_t|) - \mathbb{E}f_1(|Z_0|) + a\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_t|) - a\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_0|). \tag{3.8}$$

We will first deal with finding an appropriate formula for f_1 by bounding the difference $\mathbb{E}f_1(|Z_t|) - \mathbb{E}f_1(|Z_0|)$ from above. This way we will obtain some estimates that are valid only under the assumption that $|Z_s| > \delta$ for some $\delta > 0$ and all $s \in [0,t]$. We will then use the discontinuous part f_2 of our distance function f to improve these results and obtain bounds that

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

hold regardless of the value of $|Z_s|$. We will start the proof by applying the Itô formula for Lévy processes (see e.g. [1], Theorem 4.4.10) to Eq. (3.7) and the function $g(x) := f_1(|x|)$. We have

$$\partial_i g(x) = f_1'(|x|)\frac{x_i}{|x|} \quad \text{and} \quad \partial_j \partial_i g(x) = f_1''(|x|)\frac{x_j x_i}{|x|^2} + f_1'(|x|)\left(\delta_{ij}\frac{1}{|x|} - \frac{x_j x_i}{|x|^3}\right), \tag{3.9}$$

where δ_{ij} is the Kronecker delta. By the Itô formula we have

$$\begin{aligned} g(Z_t) - g(Z_0) &= \sum_{i=1}^{d} \int_0^t \partial_i g(Z_{s-}) dZ_s^i \\ &\quad + \sum_{s\in(0,t]} \left(g(Z_s) - g(Z_{s-}) - \sum_{i=1}^{d} \partial_i g(Z_{s-}) \Delta Z_s^i \right), \end{aligned} \tag{3.10}$$

where $Z_t = (Z_t^1, \dots, Z_t^d)$ and $\Delta Z_t = Z_t - Z_{t-}$. Using the Taylor formula we can write

$$\begin{aligned} & g(Z_s) - g(Z_{s-}) - \sum_{i=1}^{d} \partial_i g(Z_{s-}) \Delta Z_s^i \\ & = \sum_{i,j=1}^{d} \int_0^1 (1-u) \partial_j \partial_i g(Z_{s-} + u\Delta Z_s) du \Delta Z_s^i \Delta Z_s^j . \end{aligned}$$

Denoting $W_{s,u} := Z_{s-} + u\Delta Z_s$ and using (3.9), we can further evaluate the above expression as

$$\begin{aligned} & \sum_{i,j=1}^{d} \int_0^1 (1-u) \Bigg[f_1''(|W_{s,u}|) \frac{W_{s,u}^j W_{s,u}^i}{|W_{s,u}|^2} \\ & \quad + f_1'(|W_{s,u}|) \frac{1}{|W_{s,u}|} \left(\delta_{ij} - \frac{W_{s,u}^j W_{s,u}^i}{|W_{s,u}|^2} \right) \Bigg] du \Delta Z_s^i \Delta Z_s^j . \end{aligned} \tag{3.11}$$

Observe now that for every $s \in (0, t]$ and every $u \in (0, 1)$ the vectors ΔZ_s and $W_{s,u}$ are parallel. This follows from the fact that if $\Delta Z_s \neq 0$ (i.e., there is a jump at s) then Y_s is equal either to X_s or to $R(X_{s-}, Y_{s-})X_s$ and hence Z_s is equal either to zero or to $2e_{s-}e_{s-}^T X_s$, which is obviously parallel to Z_{s-}. Thus we always have

$$\sum_{i=1}^{d} W_{s,u}^i \Delta Z_s^i = \langle W_{s,u}, \Delta Z_s \rangle = \pm |W_{s,u}| \cdot |\Delta Z_s|$$

and in consequence (3.11) is equal to

$$\begin{aligned} & \int_0^1 (1-u) \Bigg[f_1''(|W_{s,u}|) \frac{|W_{s,u}|^2 |\Delta Z_s|^2}{|W_{s,u}|^2} \\ & \quad + f_1'(|W_{s,u}|) \frac{1}{|W_{s,u}|} \left(|\Delta Z_s|^2 - \frac{|W_{s,u}|^2 |\Delta Z_s|^2}{|W_{s,u}|^2} \right) \Bigg] du \\ & = \int_0^1 (1-u) f_1''(|W_{s,u}|) |\Delta Z_s|^2 du, \end{aligned}$$

so we see that the second sum in (3.10) is of the form

$$\sum_{s\in(0,t]}\left(|\Delta Z_s|^2\int_0^1(1-u)f_1''(|Z_{s-}+u\Delta Z_s|)du\right).$$

Hence we can write (3.10) as

$$\begin{aligned}f_1(|Z_t|)-f_1(|Z_0|)&=\int_0^t f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-},b(X_{s-})-b(Y_{s-})\rangle ds\\&+\int_0^t\int_{\{|v|\le m\}\times[0,1]}f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-},(I-R(X_{s-},Y_{s-}))v\rangle\widetilde{N}(ds,dv,du)\\&+\int_0^t\int_{\{|v|\le m\}\times[0,1]}f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-},A(X_{s-},Y_{s-},v,u)\rangle\widetilde{N}(ds,dv,du)\\&+\sum_{s\in(0,t]}\left(|\Delta Z_s|^2\int_0^1(1-u)f_1''(|Z_{s-}+u\Delta Z_s|)du\right).\end{aligned}\tag{3.12}$$

Note that the above formula holds only for $t<T$, where T is the coupling time defined by (2.11). However, for $t\ge T$ we have $Z_t=0$ so if we want to obtain (3.2), it is sufficient to bound $\mathbb{E}f(|Z_{t\wedge T}|)$. In order to calculate the expectations of the above terms we will use a sequence of stopping times $(\tau_n)_{n=1}^\infty$ defined by

$$\tau_n:=\inf\{t\ge0:|Z_t|\notin(1/n,n)\}.$$

Note that we have $\tau_n\to T$ as $n\to\infty$, which follows from non-explosiveness of $(Z_t)_{t\ge0}$, which in turn is a consequence of non-explosiveness of the solution to (1.1). Now we will split our computations into several lemmas.

Lemma 3.1. *We have*

$$\mathbb{E}\int_0^{t\wedge\tau_n}\int_{\{|v|\le m\}\times[0,1]}f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-},(I-R(X_{s-},Y_{s-}))v\rangle\widetilde{N}(ds,dv,du)=0.$$

Proof. Observe that

$$\begin{aligned}\langle Z_{s-},(I-R(X_{s-},Y_{s-}))v\rangle&=\langle Z_{s-},2e_{s-}e_{s-}^Tv\rangle=2\langle e_{s-},v\rangle\left\langle Z_{s-},\frac{Z_{s-}}{|Z_{s-}|}\right\rangle\\&=2\langle e_{s-},v\rangle|Z_{s-}|\end{aligned}$$

and therefore

$$\begin{aligned}&\int_0^{t\wedge\tau_n}\int_{\{|v|\le m\}\times[0,1]}f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-},(I-R(X_{s-},Y_{s-}))v\rangle\widetilde{N}(ds,dv,du)\\&=2\int_0^{t\wedge\tau_n}\int_{\{|v|\le m\}\times[0,1]}f_1'(|Z_{s-}|)\langle e_{s-},v\rangle\widetilde{N}(ds,dv,du).\end{aligned}$$

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

By the Cauchy–Schwarz inequality and the fact that $f_1' \leq 1$, for any $t \geq 0$ we have

$$\int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} |f_1'(|Z_{s-}|)|^2 |\langle e_{s-}, v\rangle|^2 \nu(dv)duds$$
$$\leq \int_0^t \int_{\{|v|\leq m\}\times[0,1]} |v|^2 \nu(dv)duds < \infty,$$

which implies that

$$\int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} f_1'(|Z_{s-}|)\langle e_{s-}, v\rangle \widetilde{N}(ds, dv, du)$$

is a martingale, from which we immediately obtain our assertion. □

Lemma 3.2. *We have*

$$\mathbb{E}\int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u)\rangle \widetilde{N}(ds, dv, du) = 0.$$

Proof. By the Cauchy–Schwarz inequality and the fact that $f_1' \leq 1$, we have

$$\int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} \left| f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u)\rangle \right|^2 \nu(dv)duds$$
$$\leq \int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} |A(X_{s-}, Y_{s-}, v, u)|^2 \nu(dv)duds.$$

Using the bounds obtained in Lemma 2.1 and the fact that $|Z_s| \leq n$ for $s \leq \tau_n$, we can bound the integral above by a constant. Thus we see that the process

$$\int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u)\rangle \widetilde{N}(ds, dv, du)$$

is a martingale. □

Lemma 3.3. *For any $t > 0$, we have*

$$\mathbb{E}\sum_{s\in(0,t]} \left(|\Delta Z_s|^2 \int_0^1 (1-u) f_1''(|Z_{s-} + u\Delta Z_s|)du \right) \leq C_\varepsilon \mathbb{E}\int_0^t \bar{f}_\varepsilon(|Z_{s-}|)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds,$$

where $\delta > 0$, $\varepsilon \leq \delta \wedge 2m$, the constant C_ε is defined by

$$C_\varepsilon := 2\int_{-\varepsilon/4}^0 |y|^2 \nu_1(dy),$$

where ν_1 is the first marginal of ν and the function $\bar{f}_\varepsilon$ is defined by

$$\bar{f}_\varepsilon(y) := \sup_{x\in(y-\varepsilon,y)} f_1''(x).$$

Remark 3.4. Note that the above estimate holds for any $\delta > 0$ and $\varepsilon \leq \delta \wedge 2m$ as long as ε satisfies Assumption 4 and m is sufficiently large (see (3.16)). Even though our calculations from the proof of Lemma 3.7 indicate that later on we should choose δ and m to be the constants from

Assumption 3, here in Lemma 3.3 we do not use the condition (1.4). Note that if the condition (1.4) from Assumption 3 is satisfied by more than one value of δ (which is the case for most typical examples), there appears a question of the optimal choice of δ and ε that would maximize the contractivity constant c defined by (3.44) via (3.29) and (3.39). The answer to this depends on the particular choice of the noise $(L_t)_{t\geq 0}$. It is non-trivial though, even in simple cases, since c depends on δ and ε in a convoluted way (see the discussion in Example 4.2).

Remark 3.5. In the proof of the inequality (1.8), if we want to obtain an inequality of the form (3.4) from (3.12), we need to bound the sum appearing in (3.12) by a strictly negative term. For technical reasons that will become apparent in the proof of Lemma 3.6 (see the remarks after (3.22)), we will use the supremum of the second derivative of f_1 over "small" jumps that decrease the distance between X_t and Y_t.

Proof. Observe that for every $u \in (0,1)$ we have

$$\begin{aligned} f_1''(|Z_{s-} + u\Delta Z_s|) &= f_1''(|Z_{s-} + u\Delta Z_s|)(\mathbf{1}_{\{|Z_s| \in (|Z_{s-}|-\varepsilon, |Z_{s-}|)\}} + \mathbf{1}_{\{|Z_s| \notin (|Z_{s-}|-\varepsilon, |Z_{s-}|)\}}) \\ &\leq \sup_{x \in (|Z_{s-}|-\varepsilon, |Z_{s-}|)} f_1''(x)\mathbf{1}_{\{|Z_s| \in (|Z_{s-}|-\varepsilon, |Z_{s-}|)\}}. \end{aligned} \tag{3.13}$$

Indeed, f_1 is assumed to be concave, and thus f_1'' is negative, so

$$f_1''(|Z_{s-} + u\Delta Z_s|)\mathbf{1}_{\{|Z_s| \notin (|Z_{s-}|-\varepsilon, |Z_{s-}|)\}} \leq 0.$$

We also know that the vectors Z_{s-} and ΔZ_s are parallel, hence if $|Z_s| \in (|Z_{s-}| - \varepsilon, |Z_{s-}|)$, then $|Z_{s-} + u\Delta Z_s| = |Z_{s-}| - u|\Delta Z_s|$ for all $u \in (0,1)$. In particular, we have $|\Delta Z_s| \in (0, \varepsilon)$ and $|Z_{s-} + u\Delta Z_s| \in (|Z_{s-}| - \varepsilon, |Z_{s-}|)$ for all $u \in (0,1)$ and hence we have (3.13).

Now let $\delta > 0$ be a positive constant (as mentioned in Remark 3.4, it can be the constant from Assumption 3). Here we introduce an additional factor involving δ in order for the integral in (3.15) to be bounded from below by a positive constant. We have

$$\sup_{x \in (y-\varepsilon, y)} f_1''(x) \cdot \mathbf{1}_{\{|y|>\delta\}} \geq \sup_{x \in (y-\varepsilon, y)} f_1''(x),$$

so we can write

$$\begin{aligned} &\sum_{s\in(0,t]} \left(|\Delta Z_s|^2 \int_0^1 (1-u) f_1''(|Z_{s-} + u\Delta Z_s|) du \right) \\ &\quad\leq \sum_{s\in(0,t]} \left(\frac{1}{2} |\Delta Z_s|^2 \bar{f}_\varepsilon(|Z_{s-}|) \right) \mathbf{1}_{\{|Z_s| \in (|Z_{s-}|-\varepsilon, |Z_{s-}|)\}} \mathbf{1}_{\{|Z_{s-}|>\delta\}}. \end{aligned} \tag{3.14}$$

Now observe that

$$\{|Z_s| \in (|Z_{s-}| - \varepsilon, |Z_{s-}|)\} = \{|Z_s| < |Z_{s-}|\} \cap \{|\Delta Z_s| < \varepsilon\},$$

and the condition $|Z_s| < |Z_{s-}|$ is equivalent to $\langle \Delta Z_s, 2Z_{s-} + \Delta Z_s \rangle < 0$, so we have

$$\mathbf{1}_{\{|Z_s| \in (|Z_{s-}|-\varepsilon, |Z_{s-}|)\}} = \mathbf{1}_{\{|\Delta Z_s| < \varepsilon\}} \mathbf{1}_{\{\langle \Delta Z_s, 2Z_{s-} + \Delta Z_s \rangle < 0\}}.$$

Now we can use Eq. (3.5) describing the dynamics of the jumps of the process $(Z_t)_{t\geq 0}$ and express the sum on the right hand side of (3.14) as an integral with respect to the Poisson random measure N associated with $(L_t)_{t\geq 0}$. However, since all the terms in this sum are negative, we can additionally bound it from above by a sum taking into account only the jumps for which $u \geq \rho(v, Z_{s-})$, i.e., only the reflected jumps. After doing all this, we get

$$\begin{aligned}
&\mathbb{E}\sum_{s\in(0,t]}\left(\frac{1}{2}|\Delta Z_s|^2\bar{f}_\varepsilon(|Z_{s-}|)\right)\mathbf{1}_{\{|Z_s|\in(|Z_{s-}|-\varepsilon,|Z_{s-}|)\}}\mathbf{1}_{\{|Z_{s-}|>\delta\}}\\
&\leq \frac{1}{2}\mathbb{E}\int_0^t\int_{\{|v|\leq m\}\times[0,1]}|2e_{s-}e_{s-}^T v|^2\bar{f}_\varepsilon(|Z_{s-}|)\mathbf{1}_{\{|2e_{s-}e_{s-}^T v|<\varepsilon\}}\\
&\quad\times\mathbf{1}_{\{\langle 2e_{s-}e_{s-}^T v, 2Z_{s-}+2e_{s-}e_{s-}^T v\rangle<0\}}\mathbf{1}_{\{|Z_{s-}|>\delta\}}N(ds,dv,du).
\end{aligned}$$

Note that

$$\begin{aligned}
\langle e_{s-}e_{s-}^T v, Z_{s-}+e_{s-}e_{s-}^T v\rangle &= \langle\, \langle e_{s-}, v\rangle e_{s-}, |Z_{s-}|e_{s-}+\langle e_{s-}, v\rangle e_{s-}\,\rangle\\
&= \langle e_{s-}, v\rangle\,(|Z_{s-}|+\langle e_{s-}, v\rangle)
\end{aligned}$$

and thus we can express the expectation above as

$$\begin{aligned}
&2\mathbb{E}\int_0^t\bar{f}_\varepsilon(|Z_{s-}|)\int_{\{|v|\leq m\}\times[0,1]}|\langle e_{s-}, v\rangle|^2\mathbf{1}_{\{|\langle e_{s-},v\rangle|<\varepsilon/2\}}\\
&\quad\times\mathbf{1}_{\{\langle e_{s-},v\rangle(|Z_{s-}|+\langle e_{s-},v\rangle)<0\}}\mathbf{1}_{\{|Z_{s-}|>\delta\}}\nu(dv)duds.
\end{aligned}$$

Now denote $\nu^m(dv) := \mathbf{1}_{\{|v|\leq m\}}\nu(dv)$ and observe that if we consider the image $\nu^m\circ h_w^{-1}$ of the measure ν^m by the mapping $h_w:\mathbb{R}^d\to\mathbb{R}$ defined by $h_w(v)=\langle w, v\rangle$ for a unit vector $w\in\mathbb{R}^d$, then due to the rotational invariance of ν^m, the measure $\nu^m\circ h_v^{-1}$ is independent of the choice of w, i.e.,

$$\nu^m\circ h_w^{-1}=\nu_1^m \quad \text{for all unit vectors } w\in\mathbb{R}^d,$$

where ν_1^m is the first marginal of ν^m (and therefore it is the jump measure of a one-dimensional Lévy process being a projection of $(L_t)_{t\geq 0}$ with truncated jumps, see e.g. [20], Proposition 11.10). Hence we can calculate the above integral with respect to ν^m as an integral with respect to ν_1^m and write the expression we are investigating as

$$2\mathbb{E}\int_0^t\bar{f}_\varepsilon(|Z_{s-}|)\left(\int_{\mathbb{R}}|y|^2\mathbf{1}_{\{|y|<\varepsilon/2\}}\mathbf{1}_{\{y(|Z_{s-}|+y)<0\}}\nu_1^m(dy)\right)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds.$$

The condition $y(|Z_{s-}|+y)<0$ holds if and only if $y<0$ and $y\geq -|Z_{s-}|$, so for those $s\in[0,t]$ for which $|Z_{s-}|\geq\delta$ holds, we can bound the above integral with respect to ν_1^m from below, i.e.,

$$\begin{aligned}
\int_{\mathbb{R}}|y|^2\mathbf{1}_{\{|y|<\varepsilon/2\}}\mathbf{1}_{\{y(|Z_{s-}|+y)<0\}}\nu_1^m(dy) &\geq \int_{\mathbb{R}}|y|^2\mathbf{1}_{\{|y|<\varepsilon/2\}}\mathbf{1}_{\{y<0\wedge y>-\delta\}}\nu_1^m(dy)\\
&\geq \int_{\max\{-\delta.-\varepsilon/2\}}^0|y|^2\nu_1^m(dy)>0. \qquad (3.15)
\end{aligned}$$

Obviously, since $\varepsilon\leq\delta$, we have $\max\{-\delta.-\varepsilon/2\}=-\varepsilon/2$. Moreover, we can take m in our construction large enough so that

$$\int_{-\varepsilon/2}^0|y|^2\nu_1^m(dy)\geq\int_{-\varepsilon/4}^0|y|^2\nu_1(dy), \qquad (3.16)$$

where ν_1 is the first marginal of ν (note that if the dimension is greater than one, the measures $\mathbf{1}_{\{|y|\leq m\}}\nu_1(dy)$ and $\nu_1^m(dy)$ do not coincide and hence we need to change the integration limit on

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

the right hand side above). Thus we can estimate

$$2\mathbb{E}\int_0^t \bar{f}_\varepsilon(|Z_{s-}|)\left(\int_{\mathbb{R}} |y|^2 \mathbf{1}_{\{|y|<\varepsilon/2\}}\mathbf{1}_{\{y(|Z_{s-}|+y)<0\}}\nu_1(dy)\right)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds$$
$$\leq C_\varepsilon \mathbb{E}\int_0^t \bar{f}_\varepsilon(|Z_{s-}|)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds. \quad \square$$

The calculations in the above lemma still hold if we replace the time t with $t \wedge \tau_n$. Hence, after writing down the formula (3.12) for the stopped process $(Z_{t\wedge\tau_n})_{t\geq 0}$, taking the expectation and using Lemmas 3.1–3.3, we obtain

$$\mathbb{E}f_1(|Z_{t\wedge\tau_n}|) - \mathbb{E}f_1(|Z_0|) \leq \mathbb{E}\int_0^{t\wedge\tau_n} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, b(X_{s-}) - b(Y_{s-})\rangle ds$$
$$+ C_\varepsilon \mathbb{E}\int_0^{t\wedge\tau_n} \bar{f}_\varepsilon(|Z_{s-}|)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds. \tag{3.17}$$

We have managed to use the second derivative of f_1 to obtain a negative term that works only when $|Z_{s-}| > \delta$. Recall that it was necessary to bound $|Z_{s-}|$ from below since we needed to bound the integral in (3.15) from below. In order to obtain a negative term for $|Z_{s-}| \leq \delta$ we will later use the discontinuous part f_2 of our distance function f. Now we focus on finding a continuous function f_1 that will give us proper estimates for $|Z_{s-}| > \delta$. The argument we use here is based on arguments used by Eberle for diffusions in his papers [4,5].

Lemma 3.6. *There exist a concave, strictly increasing $\mathcal{C}^2$ function f_1 and a constant $c_1 > 0$ defined by (3.25) and (3.29) respectively, such that*

$$-f_1'(r)\kappa(r)r + C_\varepsilon \bar{f}_\varepsilon(r) \leq -c_1 f_1(r) \tag{3.18}$$

holds for all $r > \delta$, where κ is the function defined by (1.5).

Proof. Our assertion (3.18) is equivalent to

$$C_\varepsilon \bar{f}_\varepsilon(r) \leq -c_1 f_1(r) + f_1'(r)\kappa(r)r \quad \text{for all } r > \delta$$

or, explicitly,

$$\sup_{x\in(r-\varepsilon,r)} f_1''(x) \leq -\frac{c_1}{C_\varepsilon}f_1(r) + f_1'(r)\frac{r\kappa(r)}{C_\varepsilon} \quad \text{for all } r > \delta. \tag{3.19}$$

Observe that for this to make sense, we should have $\delta \geq \varepsilon$. Define

$$h(r) := r\kappa(r).$$

If we use the fact that $-h^- \leq h$, where h^- is the negative part of h, then we see that in order to show (3.19), it is sufficient to show

$$\sup_{x\in(r-\varepsilon,r)} f_1''(x) \leq -\frac{c_1}{C_\varepsilon}f_1(r) - f_1'(r)\frac{h^-(r)}{C_\varepsilon} \quad \text{for all } r > \delta,$$

which is equivalent to

$$f_1''(r-a) \leq -\frac{c_1}{C_\varepsilon}f_1(r) - f_1'(r)\frac{h^-(r)}{C_\varepsilon} \quad \text{for all } a \in (0,\varepsilon) \text{ and } r > \delta.$$

We will look for f_1 such that

$$f_1'(r) = \phi(r)g(r)$$

for some appropriately chosen functions ϕ and g. Then of course

$$f_1''(r-a) = \phi'(r-a)g(r-a) + \phi(r-a)g'(r-a).$$

We will choose ϕ and g in such a way that

$$\phi(r-a)g'(r-a) \leq -\frac{c_1}{C_\varepsilon} f_1(r) \tag{3.20}$$

and

$$\phi'(r-a)g(r-a) \leq -f_1'(r)\frac{h^-(r)}{C_\varepsilon}. \tag{3.21}$$

Since we assume that $f_1'' \leq 0$, which means f_1' is decreasing, we have $f_1'(r) \leq f_1'(r-a)$ and (3.21) is implied by

$$\phi'(r-a)g(r-a) \leq -f_1'(r-a)\frac{h^-(r)}{C_\varepsilon}. \tag{3.22}$$

Note that our ability to replace (3.21) with the above condition is a consequence of our choice to consider only the jumps that decrease the distance between X_t and Y_t (see Remark 3.5), which is equivalent to considering the supremum of f_1'' over a non-symmetric interval. In order to obtain (3.22), we need ϕ such that

$$\phi'(r-a) \leq -\frac{h^-(r)}{C_\varepsilon}\phi(r-a) \quad \text{for all } a \in (0,\varepsilon) \text{ and } r > \delta,$$

which is implied by

$$\phi'(r) \leq -\frac{h^-(r+a)}{C_\varepsilon}\phi(r) \quad \text{for all } a \in (0,\varepsilon) \text{ and } r > 0. \tag{3.23}$$

Define

$$\bar{h}(r) := \sup_{t\in(r,r+\varepsilon)} h^-(t) = \sup_{t\in(r,r+\varepsilon)} t\kappa^-(t).$$

Then of course

$$-\bar{h}(r) \leq -h^-(r+a) \quad \text{for all } a \in (0,\varepsilon)$$

and thus the condition

$$\phi'(r) \leq -\frac{\bar{h}(r)}{C_\varepsilon}\phi(r) \quad \text{for all } r > 0$$

implies (3.23). In view of the above considerations, we can choose ϕ by setting

$$\phi(r) := \exp\left(-\int_0^r \frac{\bar{h}(t)}{C_\varepsilon} dt\right) \tag{3.24}$$

and this ensures that (3.21) holds.

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

If we assume $f_1(0) = 0$, then

$$f_1(r) = \int_0^r \phi(s)g(s)ds. \tag{3.25}$$

We will choose g such that $1/2 \leq g \leq 1$, which will give us both a lower and an upper bound on f_1'. We would also like g to be constant for large arguments in order to make $f_1'(r)$ constant for sufficiently large r. This is necessary to get an upper bound for the W_1 distance (see the proof of Corollary 1.5). Hence, we will now proceed to find a formula for g for which (3.20) holds and then we will extend g as a constant function equal to $1/2$ beginning from some point R_1. Next we will show that if R_1 is chosen to be sufficiently large, then (3.19) holds for $r \geq R_1$ and $g = 1/2$. Note that if we set

$$\Phi(r) := \int_0^r \phi(s)ds,$$

then we have $f_1(r) \leq \Phi(r)$ and in order to get (3.20) it is sufficient to choose g in such a way that

$$\phi(r-a)g'(r-a) \leq -\frac{c_1}{C_\varepsilon}\Phi(r) \quad \text{for all } a \in (0,\varepsilon) \text{ and } r > \delta, \tag{3.26}$$

which is implied by

$$\phi(r)g'(r) \leq -\frac{c_1}{C_\varepsilon}\Phi(r+a) \quad \text{for all } a \in (0,\varepsilon) \text{ and } r > 0.$$

Since Φ is increasing, the condition

$$\phi(r)g'(r) \leq -\frac{c_1}{C_\varepsilon}\Phi(r+\varepsilon) \quad \text{for all } a \in (0,\varepsilon) \text{ and } r > 0$$

implies (3.26). This means that we can choose g by setting

$$g(r) := 1 - \frac{c_1}{C_\varepsilon}\int_0^r \frac{\Phi(t+\varepsilon)}{\phi(t)}dt.$$

Then obviously we have $g \leq 1$ and if we want to have $g \geq 1/2$, we must choose the constant c_1 in such a way that

$$1 - \frac{c_1}{C_\varepsilon}\int_0^r \frac{\Phi(t+\varepsilon)}{\phi(t)}dt \geq \frac{1}{2}$$

or equivalently

$$c_1 \leq \frac{C_\varepsilon}{2}\left(\int_0^r \frac{\Phi(t+\varepsilon)}{\phi(t)}dt\right)^{-1}. \tag{3.27}$$

Now define

$$R_0 := \inf\{R \geq 0 : \forall r \geq R : \kappa(r) \geq 0\}. \tag{3.28}$$

Note that R_0 is finite since $\lim_{r\to\infty}\kappa(r) > 0$. For all $r \geq R_0$ we have

$$h^-(r) = 0 \quad \text{and} \quad \phi(r) = \phi(R_0).$$

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

Now we would like to define $R_1 \geq R_0 + \varepsilon$ in such a way that

$$g(r) = \begin{cases} 1 - \dfrac{c_1}{C_\varepsilon} \displaystyle\int_0^r \dfrac{\Phi(t+\varepsilon)}{\phi(t)} dt & r \leq R_1 \\ \dfrac{1}{2} & r \geq R_1 \end{cases}$$

and (3.19) holds for $r \geq R_1$. By setting

$$c_1 := \frac{C_\varepsilon}{2} \left(\int_0^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)} dt \right)^{-1} \tag{3.29}$$

we ensure that g defined above is continuous and that (3.27) and, in consequence, (3.20) holds for $r \leq R_1$.

We will now explain how to find R_1. Since $f_1'(r) = \frac{1}{2}\phi(R_0)$ for $r \geq R_1$, we have

$$\sup_{x \in (r-\varepsilon, r)} f_1''(x) = 0 \quad \text{for all } r \geq R_1$$

and therefore (3.19) for $r \geq R_1$ holds if and only if

$$-f_1'(r) \frac{r\kappa(r)}{C_\varepsilon} \leq -\frac{c_1}{C_\varepsilon} f_1(r) \quad \text{for all } r \geq R_1,$$

which is equivalent to

$$-r\kappa(r) \frac{\phi(R_0)}{2} \leq -c_1 f_1(r) \quad \text{for all } r \geq R_1.$$

Using once again the fact that $f_1 \leq \Phi$, we see that it is sufficient to have

$$-r\kappa(r) \frac{\phi(R_0)}{2} \leq -c_1 \Phi(r) \quad \text{for all } r \geq R_1.$$

By the definition of c_1, the right hand side of the above inequality is equal to

$$-C_\varepsilon \Phi(r) \left(2 \int_0^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)} dt \right)^{-1}.$$

In order to make our computations easier, we will use the inequality

$$\int_{R_0}^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)} dt \leq \int_0^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)} dt$$

and we will look for R_1 such that

$$-r\kappa(r) \frac{\phi(R_0)}{2} \leq -C_\varepsilon \Phi(r) \left(2 \int_{R_0}^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)} dt \right)^{-1} \quad \text{for all } r \geq R_1. \tag{3.30}$$

We can compute

$$\begin{aligned} \int_{R_0}^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)} dt &= \int_{R_0}^{R_1} \frac{\Phi(R_0) + \phi(R_0)(t + \varepsilon - R_0)}{\phi(R_0)} dt \\ &= (R_1 - R_0) \frac{\Phi(R_0)}{\phi(R_0)} + \frac{1}{2}(R_1 + \varepsilon - R_0)^2 - \frac{1}{2}\varepsilon^2 \\ &\geq (R_1 - R_0) \frac{\Phi(R_0)}{\phi(R_0)} + \frac{1}{2}(R_1 - R_0)^2 \end{aligned}$$

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

$$\geq \frac{1}{2}(R_1 - R_0)\frac{\Phi(R_0)}{\phi(R_0)} + \frac{1}{2}(R_1 - R_0)^2$$
$$= \frac{(R_1 - R_0)\Phi(R_1)}{2\phi(R_0)}.$$

Therefore if we find R_1 such that

$$-r\kappa(r)\frac{\phi(R_0)}{2} \leq \frac{-C_\varepsilon \Phi(r)\phi(R_0)}{(R_1 - R_0)\Phi(R_1)} \quad \text{for all } r \geq R_1, \tag{3.31}$$

it will imply (3.30). Observe now that we have

$$\frac{\Phi(r)}{\Phi(R_1)} \leq \frac{r}{R_1} \quad \text{for all } r \geq R_1. \tag{3.32}$$

This follows from the fact that ϕ is decreasing, which implies that $\Phi(R_1) \geq \phi(R_0)R_1$ and thus

$$\frac{\phi(R_0)}{\Phi(R_1)}(r - R_1) \leq \frac{1}{R_1}(r - R_1)$$

and

$$\frac{\phi(R_0)(r - R_1) + \Phi(R_1)}{\Phi(R_1)} \leq \frac{r}{R_1}$$

hold for $r \geq R_1$. If we divide both sides of (3.31) by $\phi(R_0)$ and use (3.32), we see that we need to have

$$\frac{-r\kappa(r)}{2} \leq \frac{-C_\varepsilon r}{(R_1 - R_0)R_1} \quad \text{for all } r \geq R_1$$

or, equivalently,

$$\frac{2C_\varepsilon}{(R_1 - R_0)R_1} \leq \kappa(r) \quad \text{for all } r \geq R_1.$$

This shows that we can define R_1 by

$$R_1 := \inf\left\{R \geq R_0 + \varepsilon : \forall r \geq R : \kappa(r) \geq \frac{2C_\varepsilon}{(R - R_0)R}\right\}, \tag{3.33}$$

which is finite since we assume that $\lim_{r\to\infty} \kappa(r) > 0$. □

Our choice of f_1 and c_1 made above (see (3.25) and (3.29), respectively) allows us to estimate

$$\mathbb{E}\int_0^{t\wedge\tau_n} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, b(X_{s-}) - b(Y_{s-})\rangle \mathbf{1}_{\{|Z_{s-}|>\delta\}}ds$$
$$+ C_\varepsilon \mathbb{E}\int_0^{t\wedge\tau_n} \bar{f}_\varepsilon(|Z_{s-}|)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds \leq \mathbb{E}\int_0^{t\wedge\tau_n} -c_1 f_1(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds. \tag{3.34}$$

If we are to obtain (3.4), then on the right hand side of (3.34) we would like to have the function f instead of f_1, but we can achieve this by assuming

$$a \leq K \inf_{x>\delta} f_1(x) \tag{3.35}$$

or, more explicitly, $a \leq K f_1(\delta)$ (since f_1 is increasing), for some constant $K \geq 1$ to be chosen later. Then we have

$$\begin{aligned} -c_1 f_1(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}} &= -c_1 \left[\frac{1}{2} f_1(|Z_s|) + \frac{1}{2} f_1(|Z_s|)\right] \mathbf{1}_{\{|Z_{s-}|>\delta\}} \\ &\leq -\frac{c_1}{2} f_1(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}} - \frac{c_1 a}{2K}\mathbf{1}_{\{|Z_{s-}|>\delta\}} \\ &\leq -\frac{c_1}{2K}(f_1 + a)(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}} = -\frac{c_1}{2K} f(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}} \end{aligned}$$

and hence

$$\mathbb{E}\int_0^{t\wedge\tau_n} -c_1 f_1(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}} ds \leq \mathbb{E}\int_0^{t\wedge\tau_n} -\frac{c_1}{2K} f(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}} ds. \tag{3.36}$$

Now if we write (3.17) as

$$\begin{aligned} &\mathbb{E} f_1(|Z_{t\wedge\tau_n}|) - \mathbb{E} f_1(|Z_0|) \\ &\leq \mathbb{E}\int_0^{t\wedge\tau_n} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, b(X_{s-}) - b(Y_{s-})\rangle(\mathbf{1}_{\{|Z_{s-}|>\delta\}} + \mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}})ds \\ &\quad + C_\varepsilon \mathbb{E}\int_0^{t\wedge\tau_n} \bar{f}_\varepsilon(|Z_{s-}|)\mathbf{1}_{\{|Z_{s-}|>\delta\}} ds, \end{aligned} \tag{3.37}$$

we see that by (3.34) and (3.36) we already have a good bound for the terms involving $\mathbf{1}_{\{|Z_{s-}|>\delta\}}$. Now we need to obtain estimates for the case when $|Z_{s-}| \leq \delta$. To this end, we should come back to Eq. (3.8) and focus on the expression

$$a\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_t|) - a\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_0|).$$

We have the following lemma.

Lemma 3.7. *For any $t \geq 0$ we have*

$$\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_t|) - \mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_0|) \leq -\mathbb{E}\int_0^t \widetilde{C}_\delta(m)\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}} ds,$$

where

$$\widetilde{C}_\delta(m) := \inf_{x\in\mathbb{R}^d:0<|x|\leq\delta} \int_{\{|v|\leq m\}\cap\{|v+x|\leq m\}} q(v)\wedge q(v+x)dv > 0. \tag{3.38}$$

Note that $\widetilde{C}_\delta(m)$ is positive by Assumption 3 about the sufficient overlap of q and translated q (see the condition (1.4)).

Proof. Observe that almost surely we have

$$\mathbf{1}_{(0,\infty)}(|Z_t|) = 1 - \int_0^t \int_{\{|v|\leq m\}\times[0,1]} \mathbf{1}_{\{u<\rho(v,Z_{s-})\}}\mathbf{1}_{\{|Z_{s-}|\neq 0\}} N(ds, dv, du).$$

The integral with respect to the Poisson random measure N appearing above counts exactly the one jump that brings the processes X_t and Y_t to the same point. Note that if we skipped the condition $\{|Z_{s-}| \neq 0\}$, it would also count all the jumps that happen after the coupling time and it would be possibly infinite. Since we obviously have

$$\int_0^t \int_{\{|v|\leq m\}\times[0,1]} \mathbf{1}_{\{u<\rho(v,Z_{s-})\}}\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}N(ds,dv,du)$$
$$\leq \int_0^t \int_{\{|v|\leq m\}\times[0,1]} \mathbf{1}_{\{u<\rho(v,Z_{s-})\}}\mathbf{1}_{\{|Z_{s-}|\neq 0\}}N(ds,dv,du),$$

we can estimate

$$\mathbf{1}_{(0,\infty)}(|Z_t|) \leq 1 - \int_0^t \int_{\{|v|\leq m\}\times[0,1]} \mathbf{1}_{\{u<\rho(v,Z_{s-})\}}\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}N(ds,dv,du),$$

and therefore we get

$$a\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_t|) - a\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_0|)$$
$$\leq -a\mathbb{E}\int_0^t \int_{\{|v|\leq m\}\times[0,1]} \mathbf{1}_{\{u<\rho(v,Z_{s-})\}}\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}\nu(dv)duds,$$

where we used the assumption that $\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_0|) = \mathbb{P}(|Z_0| \neq 0) = 1$ (see the remarks at the beginning of this section). We also have

$$\mathbb{E}\int_0^t \int_{\{|v|\leq m\}\times[0,1]} \mathbf{1}_{\{u<\rho(v,Z_{s-})\}}\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}\nu(dv)duds$$
$$= \mathbb{E}\int_0^t \int_{\{|v|\leq m\}} \rho(v, Z_{s-})\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}\nu(dv)ds$$
$$= \mathbb{E}\int_0^t \int_{\{|v|\leq m\}\cap\{|v+Z_{s-}|\leq m\}} (q(v+Z_{s-})\wedge q(v))\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}dvds$$
$$\geq \mathbb{E}\int_0^t \widetilde{C}_\delta(m)\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}ds$$

and the assertion follows. □

Note that we can always choose m large enough so that

$$\inf_{x\in\mathbb{R}^d:0<|x|\leq\delta} \int_{\{|v|\leq m\}\cap\{|v+x|\leq m\}} q(v)\wedge q(v+x)dv$$
$$\geq \frac{1}{2}\inf_{x\in\mathbb{R}^d:0<|x|\leq\delta} \int_{\mathbb{R}^d} q(v)\wedge q(v+x)dv =: \frac{1}{2}\widetilde{C}_\delta \tag{3.39}$$

and hence we have

$$\mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_t|) - \mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_0|) \leq -\mathbb{E}\int_0^t \frac{1}{2}\widetilde{C}_\delta\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}ds.$$

Combining the estimate above with (3.8) and (3.37), we obtain

$$\mathbb{E}f(|Z_{t\wedge\tau_n}|) - \mathbb{E}f(|Z_0|)$$
$$\leq \mathbb{E}\int_0^{t\wedge\tau_n} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, b(X_{s-}) - b(Y_{s-})\rangle(\mathbf{1}_{\{|Z_{s-}|>\delta\}} + \mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}})ds$$
$$+ C_\varepsilon\mathbb{E}\int_0^{t\wedge\tau_n} \bar{f}_\varepsilon(|Z_{s-}|)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds - a\mathbb{E}\int_0^{t\wedge\tau_n} \frac{1}{2}\widetilde{C}_\delta\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}ds.$$

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

In order to deal with the expressions involving $\{|Z_{s-}| \leq \delta\}$, we will use the fact that b satisfies the one-sided Lipschitz condition (1.2) with some constant $C_L > 0$ and that $f_1'(r) \leq f_1'(0) = 1$ for all $r \geq 0$ to get

$$\begin{aligned}
&\mathbb{E} \int_0^{t\wedge\tau_n} f_1'(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, b(X_{s-}) - b(Y_{s-})\rangle \mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}} ds \\
&- a\mathbb{E} \int_0^{t\wedge\tau_n} \frac{1}{2} \widetilde{C}_\delta \mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}} ds \\
&\leq \left(C_L\delta - \frac{1}{2} a\widetilde{C}_\delta \right) \mathbb{E} \int_0^{t\wedge\tau_n} \mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}} ds. \qquad (3.40)
\end{aligned}$$

We would like to bound this expression by

$$\mathbb{E} \int_0^{t\wedge\tau_n} -Cf(|Z_{s-}|) \mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}} ds$$

for some positive constant C, but since the function f is bounded on the interval $[0, \delta]$ by $f_1(\delta) + a$, we have

$$-Cf_1(\delta) - Ca \leq -Cf(|Z_{s-}|) \quad \text{if } 0 < |Z_{s-}| \leq \delta$$

and thus it is sufficient if we have

$$C_L\delta + Cf_1(\delta) \leq (\widetilde{C}_\delta/2 - C)a.$$

Of course the right hand side has to be positive, so we can choose e.g. $C := \widetilde{C}_\delta/4$. Then we must have

$$\frac{C_L\delta + \widetilde{C}_\delta f_1(\delta)/4}{\widetilde{C}_\delta/4} \leq a, \qquad (3.41)$$

but on the other hand, by (3.35), we must also have $a \leq Kf_1(\delta)$. Hence we can define

$$K := \frac{C_L\delta + \widetilde{C}_\delta f_1(\delta)/4}{\widetilde{C}_\delta f_1(\delta)/4} \qquad (3.42)$$

and

$$a := Kf_1(\delta). \qquad (3.43)$$

Then obviously both (3.35) and (3.41) hold and we get the required estimate for the right hand side of (3.40). Using all our estimates together, we get

$$\begin{aligned}
\mathbb{E}f(|Z_{t\wedge\tau_n}|) - \mathbb{E}f(|Z_0|) \leq\ &\mathbb{E} \int_0^{t\wedge\tau_n} -\frac{c_1}{2K} f(|Z_s|) \mathbf{1}_{\{|Z_{s-}|>\delta\}} ds \\
&+ \mathbb{E} \int_0^{t\wedge\tau_n} -\frac{1}{4} \widetilde{C}_\delta f(|Z_s|) \mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}} ds.
\end{aligned}$$

Denote

$$c := \min\left\{ \frac{c_1}{2K}, \frac{1}{4}\widetilde{C}_\delta \right\}. \qquad (3.44)$$

Then of course

$$\begin{aligned}\mathbb{E}f(|Z_{t\wedge\tau_n}|)-\mathbb{E}f(|Z_0|) &\leq \mathbb{E}\int_0^{t\wedge\tau_n} -cf(|Z_s|)\mathbf{1}_{\{|Z_{s-}|>\delta\}}ds \\ &\quad + \mathbb{E}\int_0^{t\wedge\tau_n} -cf(|Z_s|)\mathbf{1}_{\{0<|Z_{s-}|\leq\delta\}}ds \\ &= \mathbb{E}\int_0^{t\wedge\tau_n} -cf(|Z_s|)ds. \end{aligned} \tag{3.45}$$

Note that we can perform the same calculations not only on the interval $[0, t\wedge\tau_n]$, but also on any interval $[s\wedge\tau_n, t\wedge\tau_n]$ for arbitrary $0 \leq s < t$. Indeed, by our assumption (see the beginning of this section) we have $\mathbb{P}(|Z_0| \neq 0) = 1$ and hence for any $0 \leq s < T$ we have $\mathbb{P}(|Z_{s\wedge\tau_n}| \neq 0) = 1$. Thus Lemma 3.7 still holds on $[s\wedge\tau_n, t\wedge\tau_n]$. It is easy to see that the other calculations are valid too and we obtain

$$\mathbb{E}f(|Z_{t\wedge\tau_n}|)-\mathbb{E}f(|Z_{s\wedge\tau_n}|) \leq \mathbb{E}\int_s^t -cf(|Z_{r\wedge\tau_n}|)dr. \tag{3.46}$$

Since this holds for any $0 \leq s < t$, by the differential version of the Gronwall inequality we obtain

$$\mathbb{E}f(|Z_{t\wedge\tau_n}|) \leq \mathbb{E}f(|Z_0|)e^{-ct}.$$

Note that we cannot use the integral version of the Gronwall inequality for (3.45) since the right hand side is negative and that is why we need (3.46) to hold for any $s < t$. By the Fatou lemma and the fact that $Z_t = 0$ for $t \geq T$ (see the remarks after (3.12)) we get

$$\mathbb{E}f(|Z_t|) \leq e^{-ct}\mathbb{E}f(|Z_0|) \quad \text{for all } t \geq 0,$$

which finishes the proof of (1.8).

Proof of Theorem 1.1. By everything we proved in Sections 2.4 and 2.5 and the entire Section 3, we obtain a coupling $(X_t, Y_t)_{t\geq 0}$ satisfying the inequality (1.8). The only thing that remains to be shown is the fact that the coupling $(X_t, Y_t)_{t\geq 0}$ is successful. This follows easily from the inequality (1.8) and the form of the function f. Indeed, recalling that $Z_t = X_t - Y_t$ and that T denotes the coupling time for $(X_t, Y_t)_{t\geq 0}$, for a fixed $t > 0$ we have

$$\begin{aligned}\mathbb{P}(T > t) &= \mathbb{P}(|Z_t| > 0) = \mathbb{E}\mathbf{1}_{(0,\infty)}(|Z_t|) \leq \frac{1}{a}\mathbb{E}\left(f_1(|Z_t|) + a\mathbf{1}_{(0,\infty)}(|Z_t|)\right) \\ &= \frac{1}{a}\mathbb{E}f(|Z_t|) \leq \frac{1}{a}e^{-ct}\mathbb{E}f(|Z_0|).\end{aligned}$$

Hence we get

$$\mathbb{P}(T = \infty) = \mathbb{P}\left(\bigcap_{t>0}\{T > t\}\right) = \lim_{t\to\infty}\mathbb{P}(T > t) = 0. \quad \square$$

4. Additional proofs and examples

Proof of Corollary 1.4. We have

$$\mathbf{1}_{(0,\infty)} = a^{-1}a\mathbf{1}_{(0,\infty)} \leq a^{-1}(f_1 + a\mathbf{1}_{(0,\infty)}) = a^{-1}f,$$

hence we get

$$\frac{1}{2}\|\mu_1 p_t - \mu_2 p_t\|_{TV} = W_{\mathbf{1}_{(0,\infty)}}(\mu_1 p_t, \mu_2 p_t)$$
$$\leq a^{-1} W_f(\mu_1 p_t, \mu_2 p_t) \leq a^{-1} e^{-ct} W_f(\mu_1, \mu_2). \quad \square$$

Proof of Corollary 1.5. We have

$$f_1'(r) = \phi(r) g(r) \geq \frac{\phi(r)}{2} \geq \frac{\phi(R_0)}{2}$$

for all $r \geq 0$. But $f_1(0) = 0$, so we get

$$f_1(r) \geq \frac{\phi(R_0)}{2} r$$

for all $r \geq 0$ and in consequence

$$r \leq \frac{2 f_1(r)}{\phi(R_0)} \leq \frac{2 f(r)}{\phi(R_0)},$$

which proves that

$$W_1(\mu_1 p_t, \mu_2 p_t) \leq 2\phi(R_0)^{-1} e^{-ct} W_f(\mu_1, \mu_2). \quad \square$$

Proof of Corollary 1.8. Let us first comment on the assumption we make on the semigroup $(p_t)_{t \geq 0}$ stating that if a measure μ has a finite first moment, then for all $t > 0$ the measure μp_t also has a finite first moment. This assumption seems quite natural for proving existence of invariant measures for Markov processes by using methods based on Wasserstein distances, cf. assumption $(H1)$ in [10]. In our setup, it holds e.g. if we assume that the noise $(L_t)_{t \geq 0}$ has a finite first moment and the drift b satisfies a linear growth condition, i.e., there exists a constant $C > 0$ such that $|b(x)|^2 \leq C(1 + |x^2|)$ for all $x \in \mathbb{R}^d$. By Corollary 1.5, we have

$$W_1(\mu p_t, \eta p_t) \leq L e^{-ct} W_f(\mu, \eta) \tag{4.1}$$

for some constants c, $L > 0$ and any probability measures μ and η. Now let μ be a fixed, arbitrarily chosen probability measure and consider a sequence of measures $(\mu p_n)_{n=0}^{\infty}$. Apply (4.1) to μ and $\eta = \mu p_1$ with $t = n$. We get

$$W_1(\mu p_n, \mu p_{n+1}) \leq L e^{-cn} W_f(\mu, \mu p_1).$$

Similarly, using the triangle inequality for W_1, we get that for any $k \geq 1$

$$W_1(\mu p_n, \mu p_{n+k}) \leq L \sum_{j=0}^{k-1} e^{-c(n+j)} W_f(\mu, \mu p_1) \leq L \frac{e^{-cn}}{1 - e^{-c}} W_f(\mu, \mu p_1).$$

It is now easy to see that $(\mu p_n)_{n=0}^{\infty}$ is a Cauchy sequence with respect to the W_1 distance. Since the space of probability measures with finite first moments equipped with the W_1 distance is complete (see e.g. Theorem 6.18 in [27]), we infer that $(\mu p_n)_{n=0}^{\infty}$ has a limit μ_0. Note that here we use the assumption about the semigroup $(p_t)_{t \geq 0}$ preserving finite first moments. We also know that W_1 actually metrizes the weak convergence of measures and thus

$$\int \varphi \mu p_n \to \int \varphi \mu_0$$

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

as $n \to \infty$ for all continuous bounded ($\mathcal{C}_b$) functions φ. It is easy to check that since the drift in (1.1) is one-sided Lipschitz, the semigroup $(p_t)_{t\geq 0}$ is Feller, in particular for any $\varphi \in \mathcal{C}_b$ we have $p_1\varphi \in \mathcal{C}_b$ and thus

$$\int \varphi(x)\mu p_{n+1}(dx) = \int p_1\varphi(x)\mu p_n(dx) \to \int p_1\varphi(x)\mu_0(dx) = \int \varphi(x)\mu_0 p_1(dx).$$

Hence we infer that

$$\mu_0 = \mu_0 p_1.$$

Now if we define

$$\mu_* := \int_0^1 \mu_0 p_s ds,$$

we can easily show (see e.g. [10], the beginning of Section 3 for details) that for any $t \geq 0$ we have

$$\mu_* p_t = \mu_*,$$

i.e., μ_* is actually an invariant measure for $(p_t)_{t\geq 0}$. Now the inequality (1.16) follows easily from (1.9) applied to μ_* and η. Indeed, we have

$$W_f(\mu_*, \eta p_t) = W_f(\mu_* p_t, \eta p_t) \leq e^{-ct} W_f(\mu_*, \eta).$$

Similarly, the inequalities (1.17) and (1.18) follow easily from (1.13) and (1.14), respectively. □

We would like now to investigate optimality of the contraction constant we obtained in Corollary 1.2. First, let us recall a well-known result. Let $(X_t)_{t\geq 0}$ be the solution to (1.1) and $(p_t)_{t\geq 0}$ its associated semigroup. If there exists a constant $M > 0$ such that for all $x, y \in \mathbb{R}^d$ we have

$$\langle b(x) - b(y), x - y\rangle \leq -M|x - y|^2, \tag{4.2}$$

then for all $t > 0$ and any probability measures μ_1, μ_2 we have

$$W_1(\mu_1 p_t, \mu_2 p_t) \leq e^{-Mt} W_1(\mu_1, \mu_2).$$

Example 4.1. A typical example illustrating the above result is the case when the drift b is given as the gradient of a convex potential, i.e., $b = -\nabla U$ with e.g. $U(x) = M|x^2|/2$ for some constant $M > 0$. Then we obviously have

$$\langle b(x) - b(y), x - y\rangle = -M|x - y|^2$$

and, by the above result, exponential convergence with the rate e^{-Mt} holds for Eq. (1.1) in the standard L^1-Wasserstein distance.

Example 4.2. We will now try to examine the case in which we drop the convexity assumption. Assume

$$\kappa(r) \geq 0 \quad \text{for all } r \geq 0 \quad \text{and} \quad \kappa(r) \geq M \quad \text{for all } r \geq R \tag{4.3}$$

for some constants $M > 0$ and $R > 0$. This means that we have

$$\langle b(x) - b(y), x - y\rangle \leq 0$$

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

everywhere, but the dissipativity condition (4.2) holds only outside some fixed ball of radius R. Then, using the notation from Section 3, we can easily check that the function ϕ is constant and equal to 1. We have

$$f_1(r) = \int_0^r g(s)ds \quad \text{and} \quad g(r) = 1 - \frac{1}{R_1^2 + 2\varepsilon R_1}\left(\frac{1}{2}r^2 + \varepsilon r\right)$$

and therefore

$$f_1(r) = r - \frac{1}{R_1^2 + 2\varepsilon R_1}\left(\frac{1}{6}r^3 + \frac{1}{2}\varepsilon r^2\right).$$

We also have $R_0 = 0$ and it can be shown that

$$R_1 \leq \max(R, W),$$

where W is the positive solution to the equation $M = 2C_\varepsilon/W^2$, i.e., $W = \sqrt{2C_\varepsilon/M}$. Indeed, if $R > W$, then $2C_\varepsilon/R^2 \leq 2C_\varepsilon/W^2 = M$ and thus, by (4.3), for all $r \geq R$ we have $\kappa(r) \geq 2C_\varepsilon/R^2$, which implies that R belongs to the set of which R_1 is the infimum (see (3.33)) and hence $R_1 \leq R$. On the other hand, if $R \leq W$, then for all $r \geq W$ we have $\kappa(r) \geq M = 2C_\varepsilon/W^2$ and thus $R_1 \leq W$. Observe that

$$c_1 = \frac{C_\varepsilon}{R_1^2 + 2\varepsilon R_1} \geq \frac{C_\varepsilon}{\max(R, W)^2 + 2\varepsilon \max(R, W)}.$$

Moreover, $K = 1$ when $C_L = 0$ (see (3.42)). Thus we have

$$\frac{c_1}{2K} \geq \frac{C_\varepsilon}{2\max(R, W)^2 + 4\varepsilon \max(R, W)},$$

which means that the lower bound for $c_1/2K$ is of order $\min(R^{-2}, M)$. This means that the convergence rates in the W_1 distance are not substantially affected by dropping the global dissipativity assumption, as long as the ball in which the dissipativity does not hold is not too large. This behaviour is similar to the diffusion case (see Remark 5 in [5]).

As an example, consider a one-dimensional Lévy process with the jump density given by $q(v) = (1/|v|^{1+\alpha})$ for $\alpha \in (0, 2)$. Then we can easily show that

$$C_\varepsilon = \frac{2}{2-\alpha}\left(\frac{\varepsilon}{4}\right)^{2-\alpha} \quad \text{and} \quad \widetilde{C}_\delta = \frac{2}{\alpha}\left(\frac{2}{\delta}\right)^{\alpha}.$$

Let us focus on the case of $\alpha \in (1, 2)$. If we denote

$$c_1(\varepsilon) := \frac{C_\varepsilon}{2R^2 + 4\varepsilon R},$$

then as a function of ε it obtains its maximum for $\varepsilon_0 := (2-\alpha)R(2\alpha - 2)^{-1}$. Thus if $c_1(\varepsilon_0) \leq c_2(\varepsilon_0)$, where $c_2(\delta) := \widetilde{C}_\delta/4$ (which, as we can check numerically, is true e.g. for any R if $\alpha > 11/10$), then we see that the optimal choice of parameters that maximizes the lower bound for $c = \min\{c_1/2K, \widetilde{C}_\delta/4\}$ is to take $\varepsilon = \delta = \varepsilon_0$, at least as long as $R \geq \sqrt{2C_{\varepsilon_0}/M}$, since only then $c_1(\varepsilon_0)$ is actually a lower bound for $c_1/2K$. But for this to be true, once R and α are fixed, it is sufficient to consider a large enough M (to give specific values, e.g. for $R = 1$ and $\alpha = 3/2$ we have $\varepsilon_0 = 1/2$, $C_{\varepsilon_0} = \sqrt{2}$ and $c_1(\varepsilon_0) = \sqrt{2}/4$, hence when we consider $M \geq 2\sqrt{2}$, it is optimal to take $\varepsilon = \delta = 1/2$ and we obtain $c \geq \sqrt{2}/8$). Note that for fixed values of R and

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

M, when α increases to 2, the values of C_{ε_0}, $c_1(\varepsilon_0)$ and $c_2(\varepsilon_0)$ increase to ∞. However, in such a case $c_1(\varepsilon_0)$ is no longer a lower bound for $c_1/2K$, since $R < \sqrt{2C_{\varepsilon_0}/M}$. Instead we have

$$\frac{c_1}{2K} \geq \frac{C_{\varepsilon_0}}{4C_{\varepsilon_0}M^{-1} + 4\varepsilon_0\sqrt{2C_{\varepsilon_0}M^{-1}}}$$

and the right hand side converges to $M/4$ when $\alpha \to 2$, hence in the limit we get $c \geq M/4$, which is exactly the same bound that can be obtained in the diffusion case (see [5] once again).

Acknowledgements

I would like to thank my Ph.D. advisor, Andreas Eberle, for numerous discussions and valuable comments regarding the contents of this paper. I am also grateful to Zdzisław Brzeźniak for some useful remarks and to the anonymous referee for constructive suggestions that helped me to improve the paper. This work was financially supported by the Bonn International Graduate School of Mathematics.

References

[1] D. Applebaum, Lévy Processes and Stochastic Calculus, second ed., Cambridge University Press, 2009.
[2] B. Böttcher, R.L. Schilling, J. Wang, Construction of coupling processes for Lévy processes, Stochastic Process. Appl. 121 (6) (2011) 1201–1216.
[3] M.F. Chen, F.Y. Wang, Estimation of the first eigenvalue of second order elliptic operators, J. Funct. Anal. 131 (2) (1995) 345–363.
[4] A. Eberle, Reflection coupling and Wasserstein contractivity without convexity, C. R. Math. Acad. Sci. Paris 349 (19–20) (2011) 1101–1104.
[5] A. Eberle, Reflection couplings and contraction rates for diffusions, Probab. Theory Related Fields 166 (3–4) (2016) 851–886.
[6] I. Gyöngy, N.V. Krylov, On stochastic equations with respect to semimartingales. I, Stochastics 4 (1) (1980/1981) 1–21.
[7] M. Hairer, J.C. Mattingly, Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations, Ann. Probab. 36 (6) (2008) 2050–2091.
[8] E.P. Hsu, K.T. Sturm, Maximal coupling of Euclidean Brownian motions, Commun. Math. Stat. 1 (1) (2013) 93–104.
[9] N. Ikeda, S. Watanabe, Stochastic Differential Equations and Diffusion Processes, second ed., North-Holland, Kodansha, 1989.
[10] T. Komorowski, A. Walczuk, Central limit theorem for Markov processes with spectral gap in the Wasserstein metric, Stochastic Process. Appl. 122 (5) (2012) 2155–2184.
[11] A.M. Kulik, Exponential ergodicity of the solutions to SDE's with a jump noise, Stochastic Process. Appl. 119 (2) (2009) 602–632.
[12] H. Lin, J. Wang, Successful couplings for a class of stochastic differential equations driven by Lévy processes, Sci. China Math. 55 (8) (2012) 1735–1748.
[13] T. Lindvall, Lectures on the Coupling Method, Wiley, New York, 1992.
[14] T. Lindvall, L.C.G. Rogers, Coupling of multidimensional diffusions by reflection, Ann. Probab. 14 (3) (1986) 860–872.
[15] D. Luo, J. Wang, Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises, preprint, arXiv:1604.07206.
[16] M.B. Majka, Transportation inequalities for non-globally dissipative SDEs with jumps via Malliavin calculus and coupling, preprint, arXiv:1610.06916.
[17] R. McCann, Exact solutions to the transportation problem on the line, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 455 (1984) (1999) 1341–1380.
[18] S. Peszat, J. Zabczyk, Stochastic Partial Differential Equations with Lévy Noise, Cambridge University Press, 2007.
[19] E. Priola, A. Shirikyan, L. Xu, J. Zabczyk, Exponential ergodicity and regularity for equations with Lévy noise, Stochastic Process. Appl. 122 (1) (2012) 106–133.
[20] K. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, 1999.

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020

[21] R.L. Schilling, P. Sztonyk, J. Wang, Coupling property and gradient estimates of Lévy processes via the symbol, Bernoulli 18 (4) (2012) 1128–1149.

[22] R.L. Schilling, J. Wang, On the coupling property of Lévy processes, Ann. Inst. H. Poincaré Probab. Statist. 47 (4) (2011) 1147–1159.

[23] R.L. Schilling, J. Wang, On the coupling property and the Liouville theorem for Ornstein–Uhlenbeck processes, J. Evol. Equ. 12 (1) (2012) 119–140.

[24] R. Situ, Theory of Stochastic Differential Equations with Jumps and Applications, Springer, New York, 2005.

[25] Y. Song, Gradient estimates and coupling property for semilinear SDEs driven by jump processes, Sci. China Math. 58 (2) (2015) 447–458.

[26] H. Thorisson, Shift-coupling in continuous time, Probab. Theory Related Fields 99 (4) (1994) 477–483.

[27] C. Villani, Optimal Transport. Old and New, Springer-Verlag, Berlin, 2009.

[28] F.Y. Wang, Coupling for Ornstein–Uhlenbeck processes with jumps, Bernoulli 17 (4) (2011) 1136–1158.

[29] J. Wang, On the exponential ergodicity of Lévy-driven Ornstein–Uhlenbeck processes, J. Appl. Probab. 49 (4) (2012) 990–1004.

[30] J. Wang, Exponential ergodicity and strong ergodicity for SDEs driven by symmetric α-stable processes, Appl. Math. Lett. 26 (6) (2013) 654–658.

[31] J. Wang, On the existence and explicit estimates for the coupling property of Lévy processes with drift, J. Theoret. Probab. 27 (3) (2014) 1021–1044.

[32] J. Wang, L^p-Wasserstein distance for stochastic differential equations driven by Lévy processes, Bernoulli 22 (3) (2016) 1598–1616.

Please cite this article in press as: M.B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, Stochastic Processes and their Applications (2017), http://dx.doi.org/10.1016/j.spa.2017.03.020