

Stability of Stochastic Differential Equations with Jumps by the Coupling Method

DISSERTATION

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Mateusz Bogdan Majka

aus

Jasło, Polen

Bonn 2017

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Andreas Eberle
2. Gutachter: Prof. Dr. Arnaud Guillin
Tag der Promotion: 27.09.2017
Erscheinungsjahr: 2017

Abstract

The topic of this thesis is the study of \mathbb{R}^d -valued stochastic processes defined as solutions to stochastic differential equations (SDEs) driven by a noise with a jump component. Our main focus are SDEs driven by pure jump Lévy processes and, more generally, by Poisson random measures, but our framework includes also cases in which the noise has a diffusion component. We present proofs of results guaranteeing existence of solutions and invariant measures for a broad class of such SDEs. Next we introduce a probabilistic technique known as the coupling method. We present an original construction of a coupling of solutions to SDEs with jumps, which we subsequently apply to study various stability properties of these solutions. We investigate the rates of their convergence to invariant measures, bounds on their Malliavin derivatives (both in the jump and the diffusion case) and transportation inequalities, which characterize concentration of their distributions. In all these cases the use of the coupling method allows us to significantly strengthen results that have been available in the literature so far. We conclude by discussing potential extensions of our techniques to deal with SDEs with jump noise which is inhomogeneous in time and space.

Acknowledgements

I would like to thank my PhD advisor, Prof. Andreas Eberle, for all the help and guidance he provided during the years when I was working on this thesis.

I am also grateful to Prof. Szymon Peszat for introducing me to the topic of Lévy processes and to Prof. Arnaud Guillin for getting me interested in transportation inequalities. Additional thanks go to Profs. Zdzisław Brzeźniak, Jian Wang and Liming Wu for discussions and suggestions regarding my research.

Finally, I would like to thank all my colleagues with whom I spent time discussing mathematics - in Bonn, Kraków and all the other places I visited during my PhD studies.

Contents

1	Introduction	11
2	Stochastic differential equations with jumps	17
2.1	Poisson random measures	17
2.2	Lévy processes	19
2.3	Stochastic integration for processes with jumps	21
2.4	Stochastic differential equations	24
2.4.1	Interlacing	27
2.4.2	Solutions of SDEs as Markov processes	31
2.5	Martingale problems for processes with jumps	34
3	The coupling method	37
3.1	Couplings	37
3.1.1	Coupling by reflection for diffusions	41
3.2	Coupling constructions for SDEs with jumps	45
3.2.1	Coupling by reflection for Lévy-driven SDEs	46
3.2.2	Optimal transport construction	49
3.2.3	Martingale problem approach	55
3.3	Applications of couplings to ergodicity	58
3.4	Applications of couplings to Malliavin calculus	61
3.5	Applications of couplings to transportation inequalities	65
4	Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes	69
5	Transportation inequalities for non-globally dissipative SDEs with jumps via Malliavin calculus and coupling	111
6	A note on existence of global solutions and invariant measures for jump SDEs with locally one-sided Lipschitz drift	153
7	Diffusions with inhomogeneous jumps	169
7.1	Formulation of the problem	170
7.2	Nonlinear flows of probability measures	174
7.3	Stability via coupling	178

1 Introduction

The theory of stochastic differential equations traces back to the paper [Itô51] by Itô and is by now a classical subfield of probability theory. Initially its development was focused on equations driven by Brownian motion. However, from the late 70s there was a surge of interest in SDEs driven by semimartingales with discontinuous paths, see e.g. [DD76] or [Jac79]. There are by now numerous monographs treating the subject of stochastic equations with jump noise, see e.g. [App09] and [Sit05] for SDEs driven by Lévy processes (or, more generally, by Poisson random measures), [Pro05] for SDEs driven by general semimartingales or [PZ07] for SDEs with Lévy noise in infinite dimensional spaces. One of the most important reasons behind the development of this new theory were applications of SDEs in mathematical finance. It was realized that stochastic models with jumps can represent certain kinds of financial markets better than the ones with continuous paths, see e.g. [CT04] and the references therein. However, the theory of SDEs with jumps has found numerous applications also in other fields such as non-linear filtering (see e.g. Section 7 of [Sit05] or [GM11]), self-similar Markov processes ([Dör15]), branching processes ([BLG15b]) and mathematical biology ([PP15]).

The main contribution of this thesis to the literature is an introduction of some novel techniques based on the so-called coupling method, which allow for studying certain stability properties of a broad class of jump SDEs. At the core of the coupling method lies the idea that one can compare two random objects (on two potentially different state spaces) by defining a new object on a product state space in a way which prescribes a specific joint distribution of the given two marginals. It turns out that by considering two copies of the same object and making them have an appropriately chosen joint distribution, we can obtain valuable information on the behaviour of the initially given object. The coupling method, although by now a widespread tool in probability theory (see e.g. [Lin92], [Tho00] or [Vil09]), has not been applied to study continuous-time processes with infinite jump activity to the same extent as to diffusions or Markov chains. Papers dealing with couplings of Lévy processes or, more generally, jump SDEs, started appearing regularly in the past decade. Kulik in [Kul09] applied couplings to study ergodicity of a certain class of SDEs with jump noise. This was followed by the paper [SW11] by Schilling and J. Wang, where they considered couplings of compound Poisson processes based on some couplings of random walks, and subsequently by their joint work with Böttcher [BSW11], where they studied a coupling of subordinate Brownian motions based on the coupling of Brownian motions by reflection. In parallel, F. Y. Wang in [Wan11] considered couplings of Ornstein-Uhlenbeck processes with jumps, whereas Xu in [Xu14] used couplings to study ergodicity of two dimensional SDEs driven by a degenerate Lévy noise. Other examples of papers concerning couplings of Lévy processes or jump SDEs include e.g. [SW12], [SSW12], [LW12], [PSXZ12] and [Son15]. The

1 Introduction

two most recent instances are the paper [JWa16] by J. Wang considering a coupling of solutions to SDEs driven by Lévy processes with a symmetric α -stable component and his joint work [LW16] with Luo, where they constructed a coupling for SDEs with quite a general, not necessarily symmetric jump noise. However, the topic has certainly not been exhausted and there remains a lot of space for further contributions. By employing a novel coupling construction inspired by the optimal transport theory, we show in this thesis how to significantly improve some existing stability results for a broad class of SDEs with jumps.

The most basic type of SDE that we consider is an equation on \mathbb{R}^d of the form

$$dX_t = b(X_t)dt + dL_t, \quad (1.0.1)$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a (possibly non-linear) drift function and $(L_t)_{t \geq 0}$ is a pure jump Lévy process on \mathbb{R}^d (i.e., it does not have a diffusion component). SDEs in which the coefficient near the noise does not depend on X_t are called equations with an additive noise. For such equations we present a novel coupling construction in [Maj15] and then apply it to investigate their ergodic properties. As long as the noise in the SDE has a pure jump additive component, it may be possible to extend the methods from [Maj15] to more general equations of the form

$$dX_t = b(X_t)dt + g(X_{t-})dL_t, \quad (1.0.2)$$

where $g : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ is a sufficiently regular coefficient. For SDEs of the type (1.0.2) we say that the noise is multiplicative. By the Lévy-Itô decomposition (see Theorem 2.2.4), we can write (1.0.2) as

$$dX_t = b(X_t)dt + \int_{\{|v| \leq 1\}} g(X_{t-})v\tilde{N}(dt, dv) + \int_{\{|v| > 1\}} g(X_{t-})vN(dt, dv), \quad (1.0.3)$$

where N is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ and $\tilde{N}(dt, dv) = N(dv, dt) - dt\nu(dv)$ is the compensated Poisson random measure with ν being the Lévy measure of the process $(L_t)_{t \geq 0}$, see Sections 2.1 and 2.2. We can generalize (1.0.3) further by considering any σ -finite measure ν on \mathbb{R}^d and a corresponding Poisson random measure N on $\mathbb{R}_+ \times \mathbb{R}^d$ with intensity $dt\nu(dv)$, two sets $U_0, U_1 \subset \mathbb{R}^d$ such that $\nu(U_0) < \infty$ and $\nu(U_1) = \infty$ and two functions $f : \mathbb{R}^d \times U_0 \rightarrow \mathbb{R}^d$ and $g : \mathbb{R}^d \times U_1 \rightarrow \mathbb{R}^d$. Then we can consider an equation

$$dX_t = b(X_t)dt + \int_{U_1} g(X_{t-}, v)\tilde{N}(dt, dv) + \int_{U_0} f(X_{t-}, v)N(dt, dv).$$

Finally, we can include a diffusion component by considering a Brownian motion $(W_t)_{t \geq 0}$ in \mathbb{R}^m and a coefficient $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$. Then we arrive at

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{U_1} g(X_{t-}, v)\tilde{N}(dt, dv) + \int_{U_0} f(X_{t-}, v)N(dt, dv). \quad (1.0.4)$$

We study such equations in [Maj16] and [Maj16b]. In [Maj16b] we combine methods from [GK80] and [ABW10] to extend some of the results from the latter regarding

existence of solutions and invariant measures for a certain class of such SDEs. The paper [Maj16b] does not make use of the coupling method. However, couplings lie at the core of all the other parts of this thesis. Even though we do not construct a coupling directly for equations with such a general multiplicative noise as in (1.0.4), if there is an additive component of either the Gaussian or the jump noise, we can use the methods from [Ebe16] or [Maj15] to treat also the case of (1.0.4).

The first kind of a stability result that we consider is the problem of quantifying the rate of convergence of solutions of such SDEs to their equilibrium states. Namely, if $(p_t)_{t \geq 0}$ is the transition semigroup for a solution to jump SDE, we show that

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

holds for all $t \geq 0$ and all probability measures μ_1 and μ_2 with some constant $c > 0$, where W_f is a specially constructed Wasserstein distance associated with a concave function f , see Section 3.1. This allows us to quantify the rate of convergence to equilibrium both in the total variation and the standard L^1 -Wasserstein distances under quite mild assumptions on the noise and the drift in the equation. Hence we improve some of the results available in such papers as [Kul09], [PSXZ12], [Son15], [JWa16] or [LW16], see Section 1 in [Maj15] for details. The second stability result concerns obtaining bounds on Malliavin derivatives, which describe the sensitivity of solutions of jump SDEs to perturbations of the driving noise in the equation. Namely, a solution X_t to an SDE with both a Gaussian noise induced by a Brownian motion $(W_t)_{t \geq 0}$ and a jump noise induced by a Poisson random measure N can be considered as a functional of these two noises, i.e., $X_t = F(W, N)$ for a suitably chosen function F . Thus we can consider quantities

$$\lim_{\varepsilon \rightarrow 0} \frac{F(W. + \varepsilon \int_0^\cdot h_s ds) - F(W.)}{\varepsilon} \quad \text{and} \quad F(N + \delta_{(t,u)}) - F(N),$$

which describe changes of F with respect to perturbations of $(W_t)_{t \geq 0}$ in some specific directions and with respect to perturbations of N by adding an additional jump at time t of size u , respectively. The third stability problem that we consider is the problem of obtaining transportation inequalities which characterize the level of concentration of the distributions of these solutions. These inequalities relate the Wasserstein distance between $\delta_x p_t$ which is the distribution at time t of a process with a transition semigroup $(p_t)_{t \geq 0}$ and initial point $x \in \mathbb{R}^d$, and an arbitrary probability measure η , with a functional of their relative entropy, i.e.,

$$W_1(\eta, \delta_x p_t) \leq \alpha_t(H(\eta|\delta_x p_t))$$

for some function $\alpha_t : \mathbb{R} \rightarrow \mathbb{R}$. If this holds with a fixed $t > 0$, $x \in \mathbb{R}^d$ and a function α_t for every probability measure η , then $\delta_x p_t$ is said to satisfy an α_t - $W_1 H$ inequality. The results from [Maj16] on the latter two topics extend the ones obtained in [Wu10], [Ma10] and, to some extent, also in [DGW04], see Section 1 in [Maj16] and Section 3.5 in this thesis.

1 Introduction

In addition, in the last chapter we study a different type of SDEs, where the jump noise is inhomogeneous in both time and space, meaning that the distributions of jump times and the jump vectors depend on the time and the position of the process before the jump. We consider a problem of investigating stability of a specific class of such processes with respect to perturbations of their initial distributions. We present an outline of an attempt to solve this problem by employing the coupling technique.

The biggest and the most important part of this thesis consists of the following three papers.

- [Maj15] Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, *Stochastic Process. Appl.* (2017), in press, DOI: 10.1016/j.spa.2017.03.020, arXiv:1509.08816.
- [Maj16] Transportation inequalities for non-globally dissipative SDEs with jumps via Malliavin calculus and coupling, 2016, submitted, arXiv:1610.06916.
- [Maj16b] A note on existence of global solutions and invariant measures for jump SDEs with locally one-sided Lipschitz drift, 2016, submitted, arXiv:1612.03824.

These papers constitute Chapters 4, 5 and 6, respectively. In addition, there is a large introductory part consisting of Chapters 2 and 3, which serve a twofold purpose. On one hand, they present definitions and results which are important for understanding the material in the papers. This puts the papers in a wider context and increases their readability. Whereas the papers themselves are aimed at an experienced reader who is a researcher in stochastic analysis, the material from the first two chapters should make it possible for the thesis to be understood by an advanced graduate student in probability. Moreover, Chapters 2 and 3 contain some extensions of the results from the papers.

The structure of these two chapters is as follows. In Chapter 2 we introduce all the necessary definitions required to study stochastic differential equations with jumps. Sections 2.1, 2.2 and 2.3 serve a purely introductory purpose. In Section 2.4, in addition to providing background information, we present the results from [Maj16b] regarding existence of solutions and invariant measures to a certain class of jump SDEs. We also introduce the interlacing technique, which allows for extending some of the results presented in [Maj16b]. Section 2.5 presents briefly the theory of martingale problems for processes with jumps and thus lays the groundwork for the next chapter, where it is used to extend some results from [Maj15].

Chapter 3 starts with a general introduction to the coupling method. Next, in Section 3.1.1 we present some results obtained by Eberle in [Ebe16] for diffusions without jumps by using the coupling by reflection, which served as a motivation for the paper [Maj15]. Afterwards, in Section 3.2.1 we construct a coupling by reflection for SDEs driven by rotationally invariant pure jump Lévy processes, by analogy to the coupling used in [Ebe16]. While it turns out that such a coupling is not very useful for obtaining the kind of results that we are interested in, understanding its construction may help the reader to better prepare for what comes next. Namely, in Section 3.2.2 we present in detail the much more sophisticated coupling construction from [Maj15], which lays at the

foundation of the results from both [Maj15] and [Maj16]. Section 3.2.3 is a new material which explains how to improve the results from [Maj15] by employing the theory of martingale problems for jump processes presented earlier in Section 2.5. Afterwards, we present applications of the coupling construction from [Maj15] to investigate ergodicity (Section 3.3), Malliavin derivatives (Section 3.4) and transportation inequalities (Section 3.5).

The thesis is concluded by Chapter 7, where in Sections 7.1 and 7.3 we present an outline of a possible application of the coupling technique to study diffusion processes with jumps that are inhomogeneous in time and space. Such processes appear e.g. in the theory of sequential Monte Carlo methods, see Section 7.2 for a brief discussion about such connections. This designates some future research goals and showcases the power and flexibility of the methods presented earlier in this thesis.

Even though the most important part of the thesis is comprised of the papers [Maj15], [Maj16] and [Maj16b], we would like to stress that the remaining part, in addition to the introductory and explanatory material, contains the following extensions of the contents of the papers.

- A detailed explanation of the interlacing technique for constructing solutions of SDEs with jumps by including an additional jump noise with finite intensity (Section 2.4.1), which is used in Section 2.4 in [Maj15] and can be used to extend Theorem 1.1 from [Maj16b], cf. Theorem 2.4.5.
- A full proof of Theorem 2.4.8, which is a result guaranteeing that a solution to a jump SDE satisfying necessary conditions for uniqueness in law is a Markov process (which is used in [Maj16], see Remark 2.5 therein).
- A discussion of the results and methods from [Ebe16], which motivate our techniques in [Maj15] (Section 3.1.1).
- A description of a coupling by reflection for Lévy-driven SDEs with rotationally invariant jump noise (Section 3.2.1).
- An extended presentation of the construction of the coupling from [Maj15] (Section 3.2.2).
- An alternative approach via the theory of martingale problems to the proof that the process $(X_t, Y_t)_{t \geq 0}$ considered in Section 2.2 in [Maj15] is a coupling of solutions to (1.0.1), which allows for weakening of the assumptions from Theorem 1.1 in [Maj15] (Section 3.2.3).
- An extended presentation of different approaches to Malliavin calculus (Section 3.4).
- An extended discussion of various types of transportation inequalities and their characterization (Section 3.5).

All these additions should help the reader to better understand the contents of [Maj15], [Maj16] and [Maj16b] and to put the results obtained there in a broader context.

2 Stochastic differential equations with jumps

In this chapter we introduce the notion of a stochastic differential equation with noise induced by a stochastic process with jumps. Examples that are the most important for us are the noise induced by a Lévy process (with or without a diffusion component) and, more generally, noise induced by a Poisson random measure.

However, before we can start considering stochastic differential equations, we need a suitable notion of stochastic integration. For stochastic integrals with respect to Brownian motion, we use the classical theory, available nowadays in almost every textbook on stochastic analysis (see [Kuo06], [IW89] or [Pro05], to name but a few). Since the theory of stochastic integration with respect to Poisson random measures, although by now also classical, is considerably less known, we present here a brief account of all the necessary definitions and give more specific references. Our presentation in Sections 2.1, 2.2 and 2.3 is based mainly on the monographs [App09], [IW89], [Sat99] and [PZ07]. We start by defining Poisson random measures.

2.1 Poisson random measures

Let (E, \mathcal{E}) be a measurable space. Consider the space M of all $\mathbb{Z}_+ \cup \{\infty\}$ -valued measures on (E, \mathcal{E}) . Equip M with the smallest σ -field \mathcal{M} with respect to which all the mappings $M \ni \mu \mapsto \mu(B) \in \mathbb{Z}_+ \cup \{\infty\}$ for $B \in \mathcal{E}$ are measurable.

Definition 2.1.1. *Let λ be a σ -finite measure on (E, \mathcal{E}) . An (M, \mathcal{M}) -valued random variable N on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (that is, a \mathcal{F}/\mathcal{M} -measurable mapping $N : \Omega \rightarrow M$) is called a Poisson random measure on E with intensity measure λ if*

1. *for every $B \in \mathcal{E}$ the random variable $N(B)$ has the Poisson distribution with parameter $\lambda(B)$, i.e., $\mathbb{P}(N(B) = n) = \lambda(B)^n \exp(-\lambda(B))/n!$ for $n \in \mathbb{Z}_+$;*
2. *for any disjoint sets $B_1, \dots, B_k \in \mathcal{E}$, the random variables $N(B_1), \dots, N(B_k)$ are independent.*

For any given σ -finite measure λ on (E, \mathcal{E}) , there exists a Poisson random measure N on E which has λ as its intensity, see Theorem I-8.1 in [IW89], Theorem 6.4 in [PZ07] or Proposition 19.4 in [Sat99]. Moreover, we can easily deduce a representation of such a Poisson random measure as a sum of Dirac masses at points randomly distributed according to the measure λ . More specifically, since λ is assumed to be σ -finite, there exist pairwise disjoint sets $E_n \in \mathcal{E}$ such that $\lambda(E_n) < \infty$ for $n \in \mathbb{N}$ and $\bigcup_{n=1}^{\infty} E_n = E$.

2 Stochastic differential equations with jumps

We can consider a doubly indexed sequence of independent random variables ξ_m^n on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for $m, n \in \mathbb{N}$ such that ξ_m^n has values in E_n and $\mathbb{P}(\xi_m^n \in A) = \lambda(A \cap E_n) / \lambda(E_n)$. Then we can consider a sequence of random variables q_n with the Poisson distribution with parameter $\lambda(E_n)$, such that q_n and ξ_m^n are mutually independent for $m, n \in \mathbb{N}$. We set

$$N := \sum_{n=1}^{\infty} \sum_{m=1}^{q_n} \delta_{\xi_m^n}.$$

Then we can show that N is indeed a Poisson random measure on E with intensity λ . In other words, N can be represented as a sum of independent Poisson random measures N_n with finite intensities, where for each $n \in \mathbb{N}$ we have $N_n = \sum_{m=1}^{q_n} \delta_{\xi_m^n}$ and its intensity is the measure λ_n defined for $B \in \mathcal{E}$ as $\lambda_n(B) := \lambda(B \cap E_n)$. As a corollary, we can infer that any Poisson random measure N on (E, \mathcal{E}) can be represented as

$$N(A)(\omega) = \sum_{k=1}^{\infty} \delta_{\xi_k(\omega)}(A), \quad \omega \in \Omega, A \in \mathcal{E},$$

for some sequence $(\xi_k)_{k=1}^{\infty}$ of random elements in E . Hence N can be interpreted as a random distribution of a countable number of points ξ_k in E and for any set $A \in \mathcal{E}$ the quantity $N(A)$ is the number of points in A . Moreover, the expected number of points in A is given by the intensity measure λ , in the sense that $\mathbb{E}N(A) = \lambda(A)$, which follows straight from Definition 2.1.1.

From now on, we will consider Poisson random measures N on $(0, \infty) \times U$ equipped with the product σ -field $\mathcal{B}((0, \infty)) \times \mathcal{U}$, where (U, \mathcal{U}) is a measurable space and $\mathcal{B}((0, \infty))$ denotes the Borel sets in $(0, \infty)$. Moreover, we will focus on the case in which the intensity λ is of the form

$$\lambda(dt dx) = dt \nu(dx),$$

i.e., it is a product of the Lebesgue measure on $(0, \infty)$ and some σ -finite measure ν on (U, \mathcal{U}) . For each $B \in \mathcal{U}$ we can consider a stochastic process $(N_t(B))_{t \geq 0}$ defined by

$$N_t(B) := N((0, t] \times B), \tag{2.1.1}$$

which is the Poisson point process associated with the Poisson random measure N (cf. Section I-9 in [IW89]). In Section 2.3 we will define stochastic integrals with respect to such processes. It is possible to define stochastic integrals with respect to a more general class of point processes, see e.g. Section II-3 in [IW89]. However, here we focus only on integration with respect to Poisson point processes (or, equivalently, with respect to Poisson random measures) defined above.

Natural examples of such Poisson random measures arise as counting measures of jumps of Lévy processes, see Example 2.2.3 in the next section. Before we end the present section, let us briefly discuss the behaviour of Poisson point processes on sets whose intensity measure is finite.

Remark 2.1.2. Let N be a Poisson random measure on $\mathbb{R}_+ \times U$ with intensity $\lambda(dt dv) = dt \nu(dv)$. If we consider a set $A \in \mathcal{U}$ with $\nu(A) < \infty$, then almost surely $N((0, t] \times A) < \infty$ for every $t > 0$ and the process $(N_t(A))_{t \geq 0}$ defined by $N_t(A) := N((0, t] \times A)$ is a Poisson process with intensity $\nu(A)$ (see e.g. Lemma 2.3.4 and Theorem 2.3.5 in [App09], the discussion in Section 6.1 in [PZ07] or the proof of Theorem I-9.1 in [IW89]). Therefore $N_t(A)$ can be written as

$$N_t(A) = \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n^A \leq t\}},$$

where T_n^A are the times of jumps of the process $(N_t(A))_{t \geq 0}$ and for every $n \in \mathbb{N}$ the random variable $T_{n+1}^A - T_n^A$ is exponentially distributed with parameter $\nu(A)$ (i.e., with mean $1/\nu(A)$).

2.2 Lévy processes

Definition 2.2.1. Let $(X_t)_{t \geq 0}$ be a stochastic process on \mathbb{R}^d . We call it a Lévy process if the following conditions are satisfied.

1. $X_0 = 0$ a.s.
2. The increments of $(X_t)_{t \geq 0}$ are independent, i.e., for any $n \geq 1$ and any $0 \leq t_0 < t_1 < \dots < t_n < \infty$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent.
3. The increments of $(X_t)_{t \geq 0}$ are stationary, i.e., for any $t > s \geq 0$ we have $\text{Law}(X_t - X_s) = \text{Law}(X_{t-s})$.
4. $(X_t)_{t \geq 0}$ is stochastically continuous, i.e., for all $a > 0$ and all $s \geq 0$ we have

$$\lim_{t \rightarrow s} \mathbb{P}(|X_t - X_s| > a) = 0.$$

Every Lévy process defined in this way has a càdlàg modification, see e.g. Theorem 2.1.8 in [App09] or Theorem 4.3 in [PZ07]. Hence we can consider a process $(\Delta X_t)_{t \geq 0}$, which is the process of jumps of $(X_t)_{t \geq 0}$, i.e.,

$$\Delta X_t := X_t - X_{t-},$$

where X_{t-} is the left limit of X_t for any $t \geq 0$.

The most important examples of Lévy processes include the Brownian motion and the Poisson process. They are in fact building blocks for all more general Lévy processes, as we shall see in the sequel of this section.

It is easy to show that if $(X_t)_{t \geq 0}$ is a Lévy process, then for each $t \geq 0$ the random variable X_t is *infinitely divisible*, in the sense that for all $n \in \mathbb{N}$ there exist i.i.d. random variables Y_1, \dots, Y_n such that

$$\text{Law}(X_t) = \text{Law}(Y_1 + \dots + Y_n),$$

2 Stochastic differential equations with jumps

cf. Proposition 1.3.1 in [App09] or Example 7.3 in [Sat99]. This connection between the notions of Lévy processes and infinitely divisible distributions allows for a very useful characterization of the former.

There are two ways of characterizing Lévy processes, either via their characteristic functions or via properties of their paths. The first one is the famous Lévy-Khintchine formula (see e.g. Theorem 1.2.14 and (1.19) in [App09] or Theorem 8.1 in [Sat99]).

Theorem 2.2.2. *Let $(L_t)_{t \geq 0}$ be a Lévy process on \mathbb{R}^d . For any $t \geq 0$, denote the law of L_t by μ_t . Then the Fourier transform $\widehat{\mu}_t$ of μ_t is given as*

$$\widehat{\mu}_t(z) := \int \exp(i\langle z, x \rangle) \mu_t(dx) = \exp(t\psi(z)), \quad z \in \mathbb{R}^d,$$

where

$$\psi(z) = i\langle l, z \rangle - \frac{1}{2}\langle z, Az \rangle + \int_{\mathbb{R}^d} (e^{i\langle z, x \rangle} - 1 - i\langle z, x \rangle \mathbf{1}_{\{|x| \leq 1\}}) \nu(dx), \quad (2.2.1)$$

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

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

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 $(L_t)_{t \geq 0}$.

Conversely, if $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$ is a function of the form (2.2.1), then there exists an infinitely divisible distribution μ such that $\widehat{\mu}(z) = \exp(\psi(z))$ for $z \in \mathbb{R}^d$.

Moreover, if μ is an infinitely divisible distribution on \mathbb{R}^d , then there exists a Lévy process $(L_t)_{t \geq 0}$ on \mathbb{R}^d such that $\text{Law}(L_1) = \mu$ (cf. Corollary 11.6 in [Sat99]).

Note that the result above is often stated in the literature for Fourier transforms of infinitely divisible distributions and not for Lévy processes. Obviously the formulation for Lévy processes presented above is then a straightforward corollary if we use the fact that for a Lévy process $(L_t)_{t \geq 0}$ we have

$$\mathbb{E} \exp(i\langle u, L_t \rangle) = \exp(t\psi(u)),$$

where $\psi(u) = \log \mathbb{E} \exp(i\langle u, L_1 \rangle)$, and that L_1 is an infinitely divisible random variable (see Section 1.3 of [App09]).

We can now discuss a very important class of Poisson random measures of the type considered in Section 2.1.

Example 2.2.3. *Consider a Lévy process $(X_t)_{t \geq 0}$ in \mathbb{R}^d with Lévy measure ν . We can define*

$$N((0, t] \times A) := \sum_{s \in (0, t], \Delta X_s \neq 0} \delta_{(s, \Delta X_s)}((0, t] \times A),$$

2.3 Stochastic integration for processes with jumps

i.e., $N((0, t] \times A)$ counts the number of jumps of $(X_t)_{t \geq 0}$ of size within the set A that happen up to time t . We can then show that N is a Poisson random measure on $(0, \infty) \times \mathbb{R}^d$ with intensity $\lambda(dt dx) = dt \nu(dx)$ (see Theorem 19.2 in [Sat99]). In particular, for any set $A \in \mathcal{B}(\mathbb{R}^d)$ we have $\mathbb{E}N((0, t] \times A) = t\nu(A)$, i.e., the product of the Lebesgue measure on $(0, \infty)$ and the measure ν describes the average number of jumps up to time t of size within the set A (see e.g. Theorem I-8.1 in [IW89] or the proof of Proposition 19.4 in [Sat99]).

The other way of looking at Lévy processes is the Lévy-Itô decomposition of their paths (see e.g. Theorem 2.4.16 in [App09] or Theorem 19.2 in [Sat99]).

Theorem 2.2.4. *If $(X_t)_{t \geq 0}$ is a Lévy process, then there exist $l \in \mathbb{R}^d$, a Brownian motion $(B_t^A)_{t \geq 0}$ with covariance matrix A and an independent Poisson random measure N on $\mathbb{R}_+ \times \mathbb{R}^d$ such that for each $t \geq 0$ we have*

$$X_t = lt + B_t^A + \int_0^t \int_{\{|v| \leq 1\}} v \tilde{N}(ds, dv) + \int_0^t \int_{\{|v| > 1\}} v N(ds, dv).$$

The choice of 1 in the domain of integration above is arbitrary. It can be replaced with any number $m > 0$ by modifying the drift l accordingly (cf. Section 2.2 in [Maj15]). Note that the Poisson random measure appearing in the representation above is the counting measure of jumps of the process $(X_t)_{t \geq 0}$, cf. (19.1) in [Sat99].

There are two alternative ways of approaching the proofs of the results presented above. We can start by proving the Lévy-Khintchine formula in an analytic way (see Section 8 in [Sat99]) and then use it to obtain the Lévy-Itô decomposition of the paths (Section 20 in [Sat99]). The other way is to start by proving the Lévy-Itô decomposition in a probabilistic way and then to obtain the Lévy-Khintchine formula as a corollary (Section 2.4 in [App09]).

2.3 Stochastic integration for processes with jumps

Let us fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, a measure space (U, \mathcal{U}, ν) and consider a Poisson random measure N on $\mathbb{R}_+ \times U$ with intensity $\lambda(dt dv) = dt \nu(dv)$. We can assume that N has the representation

$$N = \sum_{k=1}^{\infty} \delta_{(\tau_k, \xi_k)}, \tag{2.3.1}$$

where $(\tau_k)_{k=1}^{\infty}$ and $(\xi_k)_{k=1}^{\infty}$ are sequences of \mathbb{R}_+ and U -valued random variables, respectively (cf. the discussion in Section 2.1). By

$$\tilde{N}(dt, dv) := N(dt, dv) - dt \nu(dv)$$

we denote the *compensated Poisson random measure*.

2 Stochastic differential equations with jumps

We need to make sense of the following two types of integrals

$$\int_0^t \int_{U_0} f(s, v) N(ds, dv) \quad \text{and} \quad \int_0^t \int_{U_1} g(s, v) \tilde{N}(ds, dv), \quad (2.3.2)$$

where $U_0, U_1 \subset U$, $\nu(U_0) < \infty$, $\nu(U_1) = \infty$ and f and g are random functions satisfying certain assumptions which will be specified in the sequel. By a standard practice, we suppress the dependence on the random parameter $\omega \in \Omega$ in our notation.

Let us first define predictable processes.

Definition 2.3.1. *Let (U, \mathcal{U}) be a measurable space. A real valued stochastic process $f = f(t, x, \omega)$ defined on $[0, \infty) \times U \times \Omega$ is called $(\mathcal{F}_t)_{\geq 0}$ -predictable if it is $\mathcal{S}/\mathcal{B}(\mathbb{R})$ -measurable, where \mathcal{S} is the σ -field on $[0, \infty) \times U \times \Omega$ generated by all the functions g on $[0, \infty) \times X \times \Omega$ such that*

1. *for all $t > 0$ the function $(x, \omega) \mapsto g(t, x, \omega)$ is $\mathcal{U} \times \mathcal{F}_t$ -measurable;*
2. *for all $(x, \omega) \in U \times \Omega$ the function $t \mapsto g(t, x, \omega)$ is left continuous.*

For any predictable f it is possible to define an integral of f with respect to N as a Lebesgue-Stieltjes integral. For any set $A \in \mathcal{U}$ we have

$$\int_0^t \int_A f(s, v) N(ds, dv) = \sum_{k=1}^{\infty} f(s, \xi_k) \mathbf{1}_{\{\tau_k \leq t, \xi_k \in A\}} \quad (2.3.3)$$

(recalling the representation (2.3.1)), whenever the sum is absolutely convergent. We can rigorously define this class of integrands in the following way.

$\mathcal{M} = \left\{ f : [0, \infty) \times U \times \Omega \rightarrow \mathbb{R} \text{ such that } f \text{ is predictable and for each } t > 0 \text{ we have} \right.$

$$\left. \int_0^t \int_U |f(s, x)| N(ds, dx) < \infty \text{ a.s.} \right\}.$$

If we additionally assume that

$$\mathbb{E} \int_0^t \int_U |f(s, x)| ds \nu(dx) < \infty,$$

then we obtain a class of integrands f for which

$$\mathbb{E} \int_0^t \int_U |f(s, x)| N(ds, dx) = \mathbb{E} \int_0^t \int_U |f(s, x)| ds \nu(dx)$$

and we can define

$$\int_0^t \int_U |f(s, x)| \tilde{N}(ds, dx) := \int_0^t \int_U |f(s, x)| N(ds, dx) - \int_0^t \int_U |f(s, x)| ds \nu(dx),$$

2.3 Stochastic integration for processes with jumps

which is then a martingale (cf. Section II-3 in [IW89]). Note that the predictability condition for the integrands is related to the fact that in order for a function f to be Stieltjes integrable with respect to a right continuous integrator, f has to be left continuous (see e.g. Section 6.3 in [Kuo06]). Note also that if $\nu(U_0) < \infty$, then any predictable function f belongs to \mathcal{M} , as the sum appearing in (2.3.3) has in such a case only a finite number of terms, cf. Remark 2.1.2. Moreover, in the case of $\nu(U_0) < \infty$ we can actually drop the predictability assumption (we can integrate any function, since the integral in such a case is just a finite sum), but then for integrals with respect to the compensated Poisson random measure we lose the martingale property of the integrals, cf. e.g. Exercise 4.3.3 in [App09]. Thus we have already achieved our goal of defining the first integral in (2.3.2).

Now we can define

$$\mathcal{M}^2 = \left\{ f : [0, \infty) \times U \times \Omega \rightarrow \mathbb{R} \text{ such that } f \text{ is predictable and for each } t > 0 \text{ we have} \right. \\ \left. \mathbb{E} \int_0^t \int_U |f(s, x)|^2 ds \nu(dx) < \infty \right\}.$$

As usual in the theory of stochastic integration, we can show that every process from \mathcal{M}^2 can be approximated by a sequence of step processes, for which the definition of the stochastic integral

$$\int_0^t \int_U f(s, x) \tilde{N}(ds, dx) \tag{2.3.4}$$

is natural, i.e.,

$$\int_0^t \int_U \left(\sum_{j,k=1}^{m,n} f_k(s_j) \mathbf{1}_{(s_j, s_{j+1}]} \mathbf{1}_{A_k} \right) \tilde{N}(ds, dx) = \sum_{j,k=1}^{m,n} f_k(s_j) \tilde{N}((s_j, s_{j+1}], A_k),$$

for some $0 = s_1 < \dots < s_m = t$, sets $A_1, \dots, A_k \in \mathcal{U}$ and \mathcal{F}_{s_j} -measurable random variables $f_k(s_j)$. Then for integrals of step processes we can show the isometry

$$\mathbb{E} \left(\int_0^t \int_U f(s, x) \tilde{N}(ds, dx) \right)^2 = \mathbb{E} \int_0^t \int_U |f(s, x)|^2 ds \nu(dx),$$

which allows us to extend the definition to $f \in \mathcal{M}^2$. The details of such a construction can be found e.g. in Chapter 4 of [App09].

Having defined the integral (2.3.4) for $f \in \mathcal{M}^2$, we can extend the definition to locally square integrable integrands. Namely, let us define

$$\mathcal{M}_{loc}^2 = \left\{ f : [0, \infty) \times U \times \Omega \rightarrow \mathbb{R} \text{ such that } f \text{ is predictable and there is a sequence} \right. \\ \left. \text{of } (\mathcal{F}_t)_{t \geq 0}\text{-stopping times } \sigma_n \text{ such that } \sigma_n \rightarrow \infty \text{ a.s. and} \right. \\ \left. (t, x, \omega) \mapsto \mathbf{1}_{[0, \sigma_n(\omega)]}(t) f(t, x, \omega) \in \mathcal{M}^2 \text{ for } n \in \mathbb{N} \right\}.$$

2 Stochastic differential equations with jumps

We can show that an equivalent description of the space \mathcal{M}_{loc}^2 is given by

$$\mathcal{M}_{loc}^2 = \left\{ f : [0, \infty) \times U \times \Omega \rightarrow \mathbb{R} \text{ such that } f \text{ is predictable and for all } t > 0 \text{ we have} \right. \\ \left. \mathbb{P} \left(\int_0^t \int_U |f(s, x)|^2 ds \nu(dx) < \infty \right) = 1 \right\}.$$

In order to see that these two definitions are indeed equivalent, it is sufficient to consider a sequence of stopping times defined as

$$\sigma_n(\omega) := \inf \left\{ t \geq 0 : \int_0^t \int_U |f(s, x)|^2 ds \nu(dx) > n \right\}$$

for $\omega \in \Omega$ and $n \in \mathbb{N}$ (see also the remark after Definition 83 in [Sit05]). The integral (2.3.4) defined for $f \in \mathcal{M}_{loc}^2$ is a local martingale and has a càdlàg modification (cf. Theorem 4.2.12 in [App09]). If we assume that $f \in \mathcal{M}^2$, then the integral (2.3.4) is a true, square integrable martingale (see Theorem 4.2.3 in [App09], cf. also the discussion in Section II-3 in [IW89]).

As the last remark in this section, note that if $(X_t)_{t \geq 0}$ is an $(\mathcal{F}_t)_{t \geq 0}$ -adapted càdlàg process, then the process $(X_{t-})_{t \geq 0}$ (the process of left limits) is predictable according to Definition 2.3.1. Thus the framework of stochastic integration presented here covers integrals such as

$$\int_0^t \int_U f(X_{s-}, v) \tilde{N}(ds, dv),$$

for sufficiently regular $f : \mathbb{R}^d \times U \rightarrow \mathbb{R}$, which will play an important role in the next section. Finally, note that even though all the definitions in this section have been formulated for real-valued integrands, extending them to the vector-valued case is straightforward by considering integrals defined in a component-wise way.

2.4 Stochastic differential equations

We consider equations of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{U_1} g(X_{t-}, u) \tilde{N}(dt, du) + \int_{U_0} f(X_{t-}, u) N(dt, du), \quad (2.4.1)$$

where $(W_t)_{t \geq 0}$ is an m -dimensional Brownian motion, N is a Poisson random measure on $\mathbb{R}_+ \times U$ for some σ -finite measure space (U, \mathcal{U}, ν) , $\tilde{N}(dt, dv) = N(dt, dv) - dt \nu(dv)$, $U_1 \subset U$ with $\nu(U_1) = \infty$, $U_0 \subset U$ with $\nu(U_0) < \infty$, $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ and $g, f : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$.

We start this section by providing the classical definitions of two types of solutions to (2.4.1).

Definition 2.4.1. We say that a process $(X_t)_{t \geq 0}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ is a weak solution to (2.4.1) if there exist a Brownian motion $(W_t)_{t \geq 0}$ and a Poisson random measure N adapted to $(\mathcal{F}_t)_{t \geq 0}$ such that almost surely

$$\begin{aligned} X_t = & X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dW_s \\ & + \int_0^t \int_{U_1} g(X_{s-}, u) \tilde{N}(dt, du) + \int_0^t \int_{U_0} f(X_{s-}, u) N(dt, du). \end{aligned} \quad (2.4.2)$$

Equivalently, a weak solution is a tuple $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}, (W_t)_{t \geq 0}, N, (X_t)_{t \geq 0})$ satisfying the conditions above.

For more information on the concept of weak solutions to SDEs see e.g. Section 6.7.3 in [App09], Definition 127 in [Sit05], Definition 3 in [BLG15] or Definition IV-1.2 in [IW89] (the latter only for the Brownian case). Here we implicitly assume that the coefficients in the SDE (2.4.1) are sufficiently regular so that all the integrals appearing in (2.4.2) are well-defined. We also have the concept of a strong solution.

Definition 2.4.2. Suppose we have a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ with an $(\mathcal{F}_t)_{t \geq 0}$ -adapted Brownian motion $(W_t)_{t \geq 0}$, an $(\mathcal{F}_t)_{t \geq 0}$ -adapted Poisson random measure N and a random variable $\xi \in \mathbb{R}^d$ independent of $(W_t)_{t \geq 0}$ and N . Then a strong solution to (2.4.1) is a process $(X_t)_{t \geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$, adapted to $(\mathcal{F}_t^{\xi, W, N})_{t \geq 0}$, which is the augmented filtration generated by ξ , $(W_t)_{t \geq 0}$ and N , such that (2.4.2) holds almost surely.

The reader is encouraged to compare this with Definition IV-1.6 in [IW89], Definition 112 in [Sit05], Definition 11 in [BLG15] or Section 6.2 in [App09]. It is obvious straight from the definition that every strong solution is also a weak solution.

Now we turn our attention to two different concepts of uniqueness of solutions to (2.4.1).

Definition 2.4.3. We say that uniqueness in law holds for solutions of (2.4.1) if for every two weak solutions $(X_t)_{t \geq 0}$ and $(\bar{X}_t)_{t \geq 0}$ of (2.4.1) with the same initial law on \mathbb{R}^d , the laws of the processes $(X_t)_{t \geq 0}$ and $(\bar{X}_t)_{t \geq 0}$ on $D([0, \infty); \mathbb{R}^d)$ (the space of càdlàg functions from $[0, \infty)$ to \mathbb{R}^d) coincide. More precisely, if $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}, (W_t)_{t \geq 0}, N, (X_t)_{t \geq 0})$ and $(\bar{\Omega}, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t \geq 0}, \bar{\mathbb{P}}, (\bar{W}_t)_{t \geq 0}, \bar{N}, (\bar{X}_t)_{t \geq 0})$ are two weak solutions to (2.4.1) such that $\mathbb{P}(X_0 \in B) = \bar{\mathbb{P}}(\bar{X}_0 \in B)$ for all $B \in \mathcal{B}(\mathbb{R}^d)$, then $\mathbb{P}(X \in C) = \bar{\mathbb{P}}(\bar{X} \in C)$ for all $C \in \mathcal{D}([0, \infty); \mathbb{R}^d)$.

For the concept of uniqueness in law, see e.g. Definition IV-1.4 in [IW89] or Definition 9 in [BLG15]. The other notion, which turns out to be stronger, is the pathwise uniqueness.

Definition 2.4.4. We say that pathwise uniqueness holds for solutions of (2.4.1) if for every two weak solutions $(X_t)_{t \geq 0}$ and $(\bar{X}_t)_{t \geq 0}$ of (2.4.1) defined on the same filtered probability space with the same Brownian motion and Poisson random measure and such that $\mathbb{P}(X_0 = \bar{X}_0) = 1$, we have $\mathbb{P}(X_t = \bar{X}_t \text{ for all } t > 0) = 1$. More precisely, if $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}, (W_t)_{t \geq 0}, N, (X_t)_{t \geq 0})$ and $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}, (W_t)_{t \geq 0}, N, (\bar{X}_t)_{t \geq 0})$ are two weak solutions to (2.4.1) such that $\mathbb{P}(X_0 = \bar{X}_0) = 1$, then $\mathbb{P}(X_t = \bar{X}_t \text{ for all } t > 0) = 1$.

2 Stochastic differential equations with jumps

As a reference, see e.g. Definition IV-1.5 in [IW89] or Definition 7 in [BLG15].

It is known that pathwise uniqueness for equations of the type (2.4.1) implies uniqueness in law. In the case of equations driven by Brownian motion, this is a classical result of Yamada and Watanabe. For the case including jumps induced by a Poisson random measure, see e.g. Theorem 137 in [Sit05] or Theorem 1 in [BLG15]. This result can be also inferred from Proposition 2.10 in [Kur07]. Note that this result does not require any explicit assumptions on the coefficients of (2.4.1), but an implicit assumption is that all the integrals appearing in 2.4.2 are well defined.

Many different versions of results guaranteeing existence and uniqueness of various types of solutions to SDEs of the form (2.4.1) can be found in the literature. In the most classical case, existence of a strong, pathwise unique solution is obtained under Lipschitz continuity of the coefficients, i.e.,

$$|b(x) - b(y)|^2 + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq C|x - y|^2.$$

is required to hold for all $x, y \in \mathbb{R}^d$ with some constant $C > 0$, where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm of a matrix (see e.g. Theorem 6.2.3 in [App09] or Theorem IV-9.1 in [IW89]). An additional linear growth condition is also required to hold, see Section 2 of [ABW10] for a detailed discussion on what assumptions are actually needed in Theorem IV-9.1 in [IW89]. Note, however, that the coefficient f is not included in the formulation of the Lipschitz condition above, and in Section 2.4.1 it will become apparent why this is the case.

A stronger result, providing existence of a pathwise unique, strong solution under a relaxed condition of one-sided Lipschitz continuity for the drift, i.e.,

$$\langle b(x) - b(y), x - y \rangle + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq C|x - y|^2,$$

was obtained by Gyöngy and Krylov in [GK80], see Theorem 2 therein. The paper [Maj16b] contains an alternative proof of a specific version of this result (see Theorem 1.1 in [Maj16b]), using methods developed by Albeverio, Brzeźniak and Wu in [ABW10], where they obtained yet another result of similar type (see the discussion in Section 1 in [Maj16b] for details).

Using the interlacing technique, which we will present in detail in Section 2.4.1, we can extend the existence result presented in [Maj16b]. Namely, Theorem 1.1 in [Maj16b] was formulated only for a noise induced by a compensated Poisson random measure with possibly infinite intensity, i.e., it works for an SDE of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du).$$

Its natural extension would be to add noise induced by a (non-compensated) Poisson random measure on a set of finite intensity and consider the equation (2.4.1). Usually, when we consider Poisson random measures on $\mathbb{R}_+ \times \mathbb{R}^d$, this corresponds to adding

large jumps to the equation, i.e., we have SDEs of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{\{|u| \leq c\}} g(X_{t-}, u)\tilde{N}(dt, du) + \int_{\{|u| > c\}} f(X_{t-}, u)N(dt, du)$$

for some $c > 0$. In the more general setting of (2.4.1) we get the following result.

Theorem 2.4.5. *Consider the equation*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{U_1} g(X_{t-}, u)\tilde{N}(dt, du) + \int_{U_0} f(X_{t-}, u)N(dt, du). \quad (2.4.3)$$

Assume that the coefficients b , σ and g in (2.4.3) satisfy the following local one-sided Lipschitz condition, i.e., for every $R > 0$ there exists $C_R > 0$ such that for any $x, y \in \mathbb{R}^d$ with $|x|, |y| \leq R$ we have

$$\langle b(x) - b(y), x - y \rangle + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_{U_1} |g(x, u) - g(y, u)|^2 \nu(du) \leq C_R |x - y|^2. \quad (2.4.4)$$

Moreover, assume a global one-sided linear growth condition, i.e., there exists $C > 0$ such that for any $x \in \mathbb{R}^d$ we have

$$\langle b(x), x \rangle + \|\sigma(x)\|_{HS}^2 + \int_{U_1} |g(x, u)|^2 \nu(du) \leq C(1 + |x|^2). \quad (2.4.5)$$

Furthermore, f is only assumed to be measurable. Under (2.4.4) and (2.4.5) and an additional assumption that $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuous, there exists a pathwise unique strong solution to (2.4.3).

This result follows easily from Theorem 1.1 in [Maj16b] after applying the interlacing procedure presented in detail in the next section.

2.4.1 Interlacing

Suppose we have a solution to the SDE

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t + \int_{U_1} g(Y_{t-}, u)\tilde{N}(dt, du) \quad (2.4.6)$$

and we would like to use it to construct a solution to the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{U_1} g(X_{t-}, u)\tilde{N}(dt, du) + \int_{U_0} f(X_{t-}, u)N(dt, du), \quad (2.4.7)$$

where $\nu(U_0) < \infty$, $\nu(U_1) = \infty$ and the sets U_0 and U_1 are disjoint. We can do so by employing the so-called interlacing technique. The main idea is that since $\nu(U_0) < \infty$, there is almost surely only a finite number of jumps that the Poisson point process $t \mapsto N((0, t] \times U_0)$ makes on any finite time interval (cf. Remark 2.1.2). Hence we can define a sequence of stopping times $\tau_1 < \tau_2 < \dots$ denoting the times of these jumps and

2 Stochastic differential equations with jumps

add the quantity defined by the last integral in (2.4.7) to the solution $(Y_t)_{t \geq 0}$ of (2.4.6) by modifying $(Y_t)_{t \geq 0}$ at times $(\tau_n)_{n=1}^\infty$ accordingly. This method is briefly explained in the proof of Theorem IV-9.1 in [IW89] and is used more extensively throughout the book [App09]. However, the formulas given in Theorem 6.2.9 in [App09] in the context of SDEs of the form (2.4.6) and (2.4.7) are incorrect. More precisely, the formula for the process $Y(t)$ for $\tau_1 < t < \tau_2$ appearing there does not define a solution to the equation it is supposed to, cf. the online errata [AppErr]. Therefore we give here a careful explanation of the interlacing technique which is based on the presentation from Section 4.2 in the paper [BLZ14] where it appears in the context of SDEs driven by Poisson random measures on infinite dimensional spaces.

Consider a stopping time τ such that $\mathbb{P}(\tau < \infty) = 1$. Define

$$W_t^\tau := W_{t+\tau} - W_\tau$$

and

$$N_t^\tau := N_{t+\tau},$$

where $(N_t)_{t \geq 0}$ is the Poisson point process defined in (2.1.1). Then we can prove that $(W_t^\tau)_{t \geq 0}$ is an $(\mathcal{F}_t^\tau)_{t \geq 0}$ -Wiener process and $(N_t^\tau)_{t \geq 0}$ is an $(\mathcal{F}_t^\tau)_{t \geq 0}$ -Poisson point process with intensity measure ν , where $\mathcal{F}_t^\tau := \mathcal{F}_{t+\tau}$ for $t \in [0, T - \tau]$, see e.g. Theorems II-6.4 and II-6.5 in [IW89]. Using this fact, we can quite easily prove the following.

Proposition 2.4.6. *Let τ be a stopping time with values in $[0, T]$ and let X_τ be an \mathcal{F}_τ -measurable random variable. Then, under the assumptions sufficient for the existence of a solution to (2.4.6), there also exists an $(\mathcal{F}_t)_{t \geq 0}$ -adapted process $(Y_t)_{t \geq 0}$ such that*

$$Y_t = Y_\tau + \int_\tau^t b(Y_s) ds + \int_\tau^t \sigma(Y_s) dW_s + \int_\tau^t \int_{U_1} g(Y_{s-}, u) \tilde{N}(ds, du) \quad \text{for } t \in [\tau, T]. \quad (2.4.8)$$

The way to prove the statement above leads first through showing that for any $x \in \mathbb{R}^d$ there exists an $(\mathcal{F}_t^\tau)_{t \geq 0}$ -adapted process $(Y_t^{\tau, x})_{t \geq 0}$ such that

$$Y_t^{\tau, x} = x + \int_0^t b(Y_s^{\tau, x}) ds + \int_0^t \sigma(Y_s^{\tau, x}) dW_s^\tau + \int_0^t \int_{U_1} g(Y_{s-}^{\tau, x}, u) \tilde{N}^\tau(ds, du) \quad \text{for } t \in [0, T - \tau]. \quad (2.4.9)$$

This can be done by following the proof of the existence of a solution to (2.4.6) and replacing all the expectations with conditional expectations with respect to \mathcal{F}_τ . Then we can replace the initial condition $x \in \mathbb{R}^d$ in (2.4.9) by an \mathcal{F}_τ -measurable random variable Y_τ , using the fact that the solution $Y_t^{\tau, x}$ of (2.4.9) is a measurable function of x . This way we obtain a process $(Y_t^\tau)_{t \in [0, T - \tau]}$ satisfying (2.4.9) on $[0, T - \tau]$ with initial condition Y_τ . Finally, setting $Y_t := Y_{t-\tau}^\tau$ for $t \in [\tau, T]$, we obtain a solution to (2.4.8). See the proof of Corollary 4.6 in [BLZ14] for details of a similar reasoning.

Now we are ready to proceed with our construction. We will denote by $Y_{a,t}(\xi)$ for $t \in [a, T]$ the solution to (2.4.6) on $[a, T]$ with initial condition ξ . Due to our reasoning above, we know that we can also replace a number $a \in [0, T]$ with a stopping time τ .

We will denote by $\tau_1 < \tau_2 < \dots$ the stopping times that encode the times of jumps of the Poisson process $N_t(U_0)$ (recall once again that for a Poisson point process evaluated at a set U_0 with $\nu(U_0) < \infty$ there is almost surely only a finite number of jumps on any finite time interval, cf. Remark 2.1.2). Denote by ξ_1, ξ_2, \dots the sizes of respective jumps, i.e., we have

$$N((0, t] \times U_0) = \sum_{k=1}^{\infty} \delta_{(\tau_k, \xi_k)}((0, t] \times U_0).$$

We will first construct a solution to (2.4.7) on the interval $[0, \tau_1]$. We set

$$X_{0,t}(x) := \begin{cases} Y_{0,t}(x) & \text{for } 0 \leq t < \tau_1, \\ Y_{0,\tau_1}(x) + f(Y_{0,\tau_1}(x), \xi_1) & \text{for } t = \tau_1, \end{cases}$$

where ξ_1 is the jump of $N_t(U_0)$ that occurs at time τ_1 . Hence we get

$$\begin{aligned} X_{0,\tau_1}(x) &= Y_{0,\tau_1}(x) + f(Y_{0,\tau_1}(x), \xi_1) \\ &= x + \int_0^{\tau_1} b(Y_{0,s}(x)) ds + \int_0^{\tau_1} \sigma(Y_{0,s}(x)) dW_s \\ &\quad + \int_0^{\tau_1} \int_{U_1} g(Y_{0,s-}(x), u) \tilde{N}(ds, du) + f(Y_{0,\tau_1}(x), \xi_1). \end{aligned}$$

Observe that the process $(Y_{0,t}(x))_{t \in [0, T]}$ has no jumps at time τ_1 , since the sets U_0 and U_1 are disjoint. Thus $Y_{0,\tau_1-}(x) = Y_{0,\tau_1}(x)$ and we have

$$\int_0^t \int_{U_0} f(Y_{0,s-}(x), u) N(ds, du) = \begin{cases} 0 & \text{for } t \in [0, \tau_1), \\ f(Y_{0,\tau_1}(x), \xi_1) & \text{for } t = \tau_1. \end{cases}$$

Hence for $t \in [0, \tau_1]$ we have

$$\begin{aligned} X_{0,t}(x) &= x + \int_0^t b(Y_{0,s}(x)) ds + \int_0^t \sigma(Y_{0,s}(x)) dW_s \\ &\quad + \int_0^t \int_{U_1} g(Y_{0,s-}(x), u) \tilde{N}(ds, du) + \int_0^t \int_{U_0} f(Y_{0,s-}(x), u) N(ds, du) \end{aligned}$$

and in consequence

$$\begin{aligned} X_{0,t}(x) &= x + \int_0^t b(X_{0,s}(x)) ds + \int_0^t \sigma(X_{0,s}(x)) dW_s \\ &\quad + \int_0^t \int_{U_1} g(X_{0,s-}(x), u) \tilde{N}(ds, du) + \int_0^t \int_{U_0} f(X_{0,s-}(x), u) N(ds, du). \end{aligned} \tag{2.4.10}$$

Thus we get a unique solution to (2.4.7) on $[0, \tau_1]$.

2 Stochastic differential equations with jumps

Due to Proposition 2.4.6, there exists a process $Y_{\tau_1,t}(X_{0,\tau_1}(x))$ for $t \in [\tau_1, T]$, which is a solution to (2.4.6) on $[\tau_1, T]$ with initial condition $X_{0,\tau_1}(x)$ at time τ_1 . Hence we have

$$\begin{aligned} Y_{\tau_1,t}(X_{0,\tau_1}(x)) &= X_{0,\tau_1}(x) + \int_{\tau_1}^t b(Y_{\tau_1,s}(X_{0,\tau_1}(x)))ds \\ &\quad + \int_{\tau_1}^t \sigma(Y_{\tau_1,s}(X_{0,\tau_1}(x)))dW_s + \int_{\tau_1}^t \int_{U_1} g(Y_{\tau_1,s-}(X_{0,\tau_1}(x)), u)\tilde{N}(ds, du) \end{aligned} \quad (2.4.11)$$

for $t \in [\tau_1, T]$ and we can define

$$X_{0,t}(x) := \begin{cases} X_{0,t}(x) & \text{for } 0 \leq t \leq \tau_1, \\ Y_{\tau_1,t}(X_{0,\tau_1}(x)) & \text{for } \tau_1 < t < \tau_2, \\ Y_{\tau_1,\tau_2}(X_{0,\tau_1}(x)) + f(Y_{\tau_1,\tau_2}(X_{0,\tau_1}(x)), \xi_2) & \text{for } t = \tau_2. \end{cases}$$

Now it is easy to see that so defined $X_{0,t}(x)$ satisfies the equation (2.4.10) for $t \in (\tau_1, \tau_2)$. Namely, it is sufficient to split all the integrals appearing in (2.4.10) by writing, as shown on the example of the drift component,

$$\begin{aligned} \int_0^t b(X_{0,s}(x))ds &= \int_0^{\tau_1} b(X_{0,s}(x))ds + \int_{\tau_1}^t b(X_{0,s}(x))ds \\ &= \int_0^{\tau_1} b(X_{0,s}(x))ds + \int_{\tau_1}^t b(Y_{\tau_1,s}(X_{0,\tau_1}(x)))ds \end{aligned}$$

and then combining (2.4.10) for $t \in [0, \tau_1]$ with (2.4.11) for $t \in (\tau_1, \tau_2)$. Moreover, since $X_{0,\tau_2-}(x) = Y_{\tau_1,\tau_2-}(X_{0,\tau_1}(x)) = Y_{\tau_1,\tau_2}(X_{0,\tau_1}(x))$, we have

$$\begin{aligned} \int_0^{\tau_2} \int_{U_0} f(X_{0,s-}(x), u)N(ds, du) &= f(X_{0,\tau_1-}(x), \xi_1) + f(Y_{\tau_1,\tau_2}(X_{0,\tau_1}(x)), \xi_2) \\ &= f(X_{0,\tau_1-}(x), \xi_1) + f(X_{0,\tau_2-}(x), \xi_2) \end{aligned}$$

and hence we get

$$\begin{aligned} X_{0,\tau_2}(x) &= x + \int_0^{\tau_2} b(X_{0,s}(x))ds + \int_0^{\tau_2} \sigma(X_{0,s}(x))dW_s \\ &\quad + \int_0^{\tau_2} \int_{U_1} g(X_{0,s-}(x), u)\tilde{N}(ds, du) + \int_0^{\tau_2} \int_{U_0} f(X_{0,s-}(x), u)N(ds, du). \end{aligned}$$

Thus we showed how to construct a solution to (2.4.7) on the interval $[0, \tau_2]$. By iterating this procedure, we obtain a solution on every interval $[0, \tau_n]$ for $n \in \mathbb{N}$ and hence on the entire $[0, T]$. Note that, by construction, if the solution $(Y_t)_{t \geq 0}$ to (2.4.6) is pathwise unique, then the solution $(X_t)_{t \geq 0}$ to (2.4.7) built from $(Y_t)_{t \geq 0}$ is also pathwise unique.

2.4.2 Solutions of SDEs as Markov processes

Here we keep considering SDEs of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t + \int_{U_1} g(X_{t-}, u)\tilde{N}(dt, du) + \int_{U_0} f(X_{t-}, u)N(dt, du). \quad (2.4.12)$$

Recall that $X_{s,t}(\zeta)$ denotes the value at time t of the solution to (2.4.12) started at time s with initial condition ζ .

For a solution $(X_t)_{t \geq 0}$ to (2.4.12), let us define the transition semigroup

$$p_{s,t}f(x) := \mathbb{E}f(X_{s,t}(x))$$

for any bounded measurable $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and any $x \in \mathbb{R}^d$. If $p_{s,t}$ is time-homogeneous, i.e., if we have $p_{s,t} = p_{0,t-s}$ for all $0 \leq s \leq t$, then we use the notation $p_t := p_{0,t}$.

In the present section we would like to show that in our setting, solutions to (2.4.12) are Markov processes. In other words, we want to show that

$$\mathbb{P}(X_{s,t}(\zeta) \in A | \mathcal{F}_u) = \mathbb{P}(X_{s,t}(\zeta) \in A | X_{s,u}). \quad (2.4.13)$$

for any $0 \leq s \leq u \leq t$ and any \mathcal{F}_s -measurable random variable ζ . We will in fact prove something stronger, i.e., we will show that

$$\mathbb{E}(\varphi(X_{s,t}(\zeta)) | \mathcal{F}_u) = p_{u,t}\varphi(X_{s,u}(\zeta)) \quad (2.4.14)$$

for any $\varphi \in B_b(\mathbb{R}^d)$. It is trivial to see that (2.4.14) implies (2.4.13), since

$$\begin{aligned} \mathbb{E}(\varphi(X_{s,t}(\zeta)) | X_{s,u}(\zeta)) &= \mathbb{E}(\mathbb{E}(\varphi(X_{s,t}(\zeta)) | \mathcal{F}_u) | X_{s,u}(\zeta)) = \mathbb{E}(p_{u,t}\varphi(X_{s,u}(\zeta)) | X_{s,u}(\zeta)) \\ &= p_{u,t}\varphi(X_{s,u}(\zeta)) = \mathbb{E}(\varphi(X_{s,t}(\zeta)) | \mathcal{F}_u) \end{aligned}$$

and then it is sufficient to take $\varphi = \mathbf{1}_A$.

Before we proceed with the proof of (2.4.14), let us cite a useful theorem, which is an original result from [Maj16b].

Theorem 2.4.7. *Consider an SDE*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du) \quad (2.4.15)$$

with coefficients satisfying (2.4.4) and (2.4.5). Additionally assume that there exists a constant $L > 0$ such that for all $x \in \mathbb{R}^d$ we have

$$\|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq L(1 + |x|^2) \quad (2.4.16)$$

and that b is continuous. Then the solution $(X_t)_{t \geq 0}$ to (2.4.15) depends on its initial condition in a continuous way, i.e., if $x_n \rightarrow x$ in \mathbb{R}^d , then for any $t > 0$ we have $X_{0,t}(x_n) \rightarrow X_{0,t}(x)$ in probability.

2 Stochastic differential equations with jumps

This result follows from the proof of Lemma 2.5 in [Maj16b]. It implies that if we show that $(X_t)_{t \geq 0}$ is a time-homogeneous Markov process, we will automatically know that its transition semigroup $(p_t)_{t \geq 0}$ is Feller, i.e., we have $p_t f \in C_b(\mathbb{R}^d)$ for every $f \in C_b(\mathbb{R}^d)$.

It is worth noting that if $(X_t)_{t \geq 0}$ is a Feller process (i.e., it has a Feller transition semigroup) and has càdlàg paths, then the Markov property automatically implies strong Markov property (see e.g. Theorem 16.21 in [Bre68], which is formulated for Feller processes with continuous paths, but the proof also works with only right continuity). This obviously applies to solutions of (2.4.12), which are càdlàg (cf. e.g. Theorem 6.2.3 in [App09]). Hence, if we show the Markov property for solutions of (2.4.15), we will know that under assumptions of Theorem 2.4.7 they are strong Markov, Feller processes.

Theorem 2.4.8. *If the solution to (2.4.12) is unique (either pathwise or in law) and if it depends continuously on the initial data, then it is a Markov process.*

Proof. Our proof is based on Theorem 9.30 in [PZ07] and Proposition 4.2 in [ABW10], see also Theorem 9.14 in [DPZ14] and Theorem 5.1.5 in [SV79] for similar reasonings. Let $0 \leq s \leq u \leq t$ and fix an \mathcal{F}_s -measurable random variable ζ . We consider the process $(X_{s,t}(\zeta))_{t \geq 0}$, which is a solution to (2.4.12) started at time s with initial condition ζ . We will show that

$$\mathbb{E}(\varphi(X_{s,t}(\zeta)) | \mathcal{F}_u) = p_{u,t} \varphi(X_{s,u}(\zeta)) \quad (2.4.17)$$

for any $\varphi \in B_b(\mathbb{R}^d)$. First, let us observe that $X_{s,t}(\zeta) = X_{u,t}(X_{s,u}(\zeta))$. Indeed,

$$\begin{aligned} X_{s,t}(\zeta) &= \zeta + \int_s^t \int_{U_1} g(X_{s,r-}(\zeta), v) \tilde{N}(dr, dv) \\ &= \zeta + \int_s^u \int_{U_1} g(X_{s,r-}(\zeta), v) \tilde{N}(dr, dv) \\ &\quad + \int_u^t \int_{U_1} g(X_{s,r-}(\zeta), v) \tilde{N}(dr, dv) \\ &= X_{s,u}(\zeta) + \int_u^t \int_{U_1} g(X_{s,r-}(\zeta), v) \tilde{N}(dr, dv). \end{aligned}$$

On the other hand,

$$X_{u,t}(X_{s,u}(\zeta)) = X_{s,u}(\zeta) + \int_u^t \int_{U_1} g(X_{u,r-}(X_{s,u}(\zeta)), v) \tilde{N}(dr, dv),$$

but the solution to (2.4.12) is unique, and hence $\text{Law}(X_{s,t}(\zeta)) = \text{Law}(X_{u,t}(X_{s,u}(\zeta)))$. Thus

$$\mathbb{E}(\varphi(X_{s,t}(\zeta)) | \mathcal{F}_u) = \mathbb{E}(\varphi(X_{u,t}(X_{s,u}(\zeta))) | \mathcal{F}_u)$$

and we see that in order to prove (2.4.17) it is sufficient to show that

$$\mathbb{E}(\varphi(X_{u,t}(\eta)) | \mathcal{F}_u) = p_{u,t} \varphi(\eta) \quad (2.4.18)$$

for any \mathcal{F}_u -measurable random variable η , and then to take $\eta = X_{s,u}(\zeta)$.

We can use approximation arguments in order to show that it is actually sufficient to prove (2.4.18) for continuous φ and for simple η , i.e., η of the form

$$\eta = \sum_{j=1}^N x_j \mathbf{1}_{A_j}, \quad (2.4.19)$$

where $x_j \in \mathbb{R}^d$, $A_j \subset \mathcal{F}_u$, $\bigcup A_j = \Omega$, A_j pairwise disjoint (see Theorem 9.30 in [PZ07] or Theorem 9.14 in [DPZ14] for details of these approximation arguments and note that in order to consider a simplified version of the initial condition η we need to use the continuous dependence of $X_{u,t}$ on the initial data).

For η given by (2.4.19) we have

$$X_{u,t}(\eta) = \sum_{j=1}^N X_{u,t}(x_j) \mathbf{1}_{A_j}$$

and

$$\begin{aligned} \mathbb{E}(\varphi(X_{u,t}(\eta)) | \mathcal{F}_u) &= \sum_{j=1}^N \mathbb{E}(\varphi(X_{u,t}(x_j)) \mathbf{1}_{A_j} | \mathcal{F}_u) \\ &= \sum_{j=1}^N (\mathbb{E}\varphi(X_{u,t}(x_j))) \mathbf{1}_{A_j} \\ &= \sum_{j=1}^N p_{u,t}\varphi(x_j) \mathbf{1}_{A_j} = p_{u,t}\varphi(\eta), \end{aligned}$$

where we use the fact that $X_{u,t}(x_j)$ is independent of \mathcal{F}_u and $\mathbf{1}_{A_j}$ are \mathcal{F}_u -measurable.

Now we show that the laws of $X_{t,t+h}(\zeta)$ and $X_{0,h}(\zeta)$ are the same based on Proposition 4.2 in [ABW10]. This shows that the Markov process is time-homogeneous.

We have

$$\begin{aligned} X_{t,t+h}(\zeta) &= \zeta + \int_t^{t+h} b(X_{t,r-})(\zeta) dr + \int_t^{t+h} \sigma(X_{t,r-})(\zeta) dW_s \\ &\quad + \int_t^{t+h} \int_{U_1} g(X_{t,r-}(\zeta), v) \tilde{N}(dr, dv) + \int_t^{t+h} \int_{U_0} f(X_{t,r-}(\zeta), v) N(dr, dv) \\ &= \zeta + \int_0^h b(X_{t,(t+u)-})(\zeta) du + \int_0^h \sigma(X_{t,(t+u)-})(\zeta) dW_u^t \\ &\quad + \int_0^h \int_{U_1} g(X_{t,(t+u)-}(\zeta), v) \tilde{N}^t(du, dv) + \int_0^h \int_{U_0} f(X_{t,(t+u)-}(\zeta), v) N^t(du, dv), \end{aligned}$$

where $W_u^t := W_{t+u} - W_t$ and $N_u^t := N_{t+u} - N_t$ are a Wiener process and a Poisson point process with the same laws as W and N , respectively (cf. Theorems II-6.4 and II-6.5 in [IW89]). Now it is easy to see that the process $(X_{0,h}(\zeta))_{h \geq 0}$ satisfies the same SDE as $(X_{t,t+h}(\zeta))_{h \geq 0}$. Since its solution is unique in law, our assertion follows. \square

2 Stochastic differential equations with jumps

Corollary 2.4.9. *A solution $(X_t)_{t \geq 0}$ to (2.4.15) under the assumptions of Lemma 2.5 in [Maj16b] is a time homogeneous Markov process. Moreover, it is strong Markov and Feller.*

For a Markov process $(X_t)_{t \geq 0}$ with transition kernels $p_t(x, dy)$ we denote its distribution at time t , provided its initial distribution is μ , by μp_t . More precisely,

$$\mu p_t(dy) := \int \mu(dx) p_t(x, dy).$$

We say that a measure μ_0 is invariant for $(X_t)_{t \geq 0}$ (or, equivalently, for the transition semigroup $(p_t)_{t \geq 0}$) if

$$\mu_0 p_t = \mu_0$$

for every $t \geq 0$. This means that $\int \mu_0(dx) p_t(x, A) = \mu_0(A)$ for every $t \geq 0$ and for every set $A \in \mathcal{B}(\mathbb{R}^d)$. This is equivalent to the fact that

$$\int p_t f(x) \mu_0(dx) = \int f(x) \mu_0(dx)$$

for every $f \in B_b(\mathbb{R}^d)$.

The following theorem is an original result from [Maj16b].

Theorem 2.4.10. *Assume that the coefficients in (2.4.15) satisfy the local one-sided Lipschitz condition (2.4.4) and the linear growth condition (2.4.16) for σ and g . Moreover, assume that there exist constants $K, M > 0$ such that for all $x \in \mathbb{R}^d$ we have*

$$\langle b(x), x \rangle + \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq -K|x|^2 + M. \quad (2.4.20)$$

Finally, let the drift coefficient b in (2.4.15) be continuous. Then there exists an invariant measure for the solution of (2.4.15).

The result above is proved based on the Krylov-Bogoliubov method, see Section 2 in [Maj16b] and the references therein. While it guarantees existence of an invariant measure, it does not tell us anything about quantitative behaviour of the process. In Chapter 3 we will explain how to use the coupling method to obtain explicit convergence rates of the distributions of solutions to (2.4.15) to the invariant measure, while replacing the condition (2.4.20) with a stronger assumption of dissipativity at infinity.

2.5 Martingale problems for processes with jumps

In 1975, Stroock in [Str75] considered operators of the form

$$\mathcal{L}_t = L_t + K_t \quad (2.5.1)$$

where

$$L_t f(x) = \frac{1}{2} \sum_{i,j=1}^d a_{ij}(t, x) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_{i=1}^d b_i(t, x) \frac{\partial}{\partial x_i} f(x)$$

2.5 Martingale problems for processes with jumps

with some sufficiently regular coefficients a_{ij} , b_i , and

$$K_t f(x) = \int (f(x+y) - f(x) - \langle y, \nabla f(x) \rangle \mathbf{1}_{\{|y| \leq 1\}}) \nu(t, x, dy).$$

Here $\nu : [0, \infty) \times \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, \infty]$ is a kernel such that for each (t, x) the measure $\nu(t, x, \cdot)$ is a Lévy measure, i.e.,

$$\int_{\mathbb{R}^d} (1 \wedge |y|^2) \nu(t, x, dy) < \infty.$$

We say that there exists a solution to the martingale problem for \mathcal{L}_t if for each $(s, x) \in [0, t) \times \mathbb{R}^d$ there exists a probability measure \mathbb{P} on $D([0, \infty); \mathbb{R}^d)$ (the space of \mathbb{R}^d -valued right continuous functions with left limits) such that $\mathbb{P}(X_s = x) = 1$ and the process

$$f(X_t) - \int_s^t \mathcal{L}_u f(X_u) du$$

is a \mathbb{P} -martingale for all $f \in C_0^\infty(\mathbb{R}^d)$, where $(X_t)_{t \geq 0}$ is the canonical process on the space $D([0, \infty); \mathbb{R}^d)$. If there is at most one such measure, we say that the solution to the martingale problem for \mathcal{L}_t is unique. A martingale problem having exactly one solution is said to be well-posed. In [Str75], there are conditions guaranteeing existence and uniqueness of solutions to martingale problems for generators of the type (2.5.1), see also Theorems 8.3.3 and 8.3.4 in [EK86]. However, in this thesis we focus on processes defined as solutions to stochastic differential equations and hence we are interested in martingale problems mainly through their connection to SDEs.

Such a connection was provided in 2010 by Kurtz in [Kur11], where he considered generators of the form

$$\begin{aligned} Af(x) &= \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \langle b(x), \nabla f(x) \rangle \\ &\quad + \int_S \lambda(x, u) (f(x + \gamma(x, u)) - f(x) - \mathbf{1}_{S_1}(u) \langle \gamma(x, u), \nabla f(x) \rangle) \nu(du), \end{aligned}$$

where ν is a σ -finite measure on some measurable space S , $S_1 \subset S$, whereas $\lambda : \mathbb{R}^d \times S \rightarrow [0, 1]$ and $\gamma : \mathbb{R}^d \times S \rightarrow \mathbb{R}^d$ are such that

$$\int_S \lambda(x, u) (\mathbf{1}_{S_1}(u) |\gamma(x, u)|^2 + \mathbf{1}_{S \setminus S_1}(u)) \nu(du) < \infty.$$

Obviously, by taking $\lambda \equiv 1$ and $\gamma(x, u) = u$ we obtain an operator of the form (2.5.1). Kurtz proved in [Kur11] (see Theorem 2.3 therein) that if a process $(X_t)_{t \geq 0}$ is a solution of the $D([0, \infty); \mathbb{R}^d)$ -martingale problem for the operator A , then it is also a weak solution

2 Stochastic differential equations with jumps

to

$$\begin{aligned}
X_t &= X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s \\
&+ \int_0^t \int_{[0,1] \times S_1} \mathbf{1}_{[0,\lambda(X_{s-},u)]} \gamma(X_{s-}, u) \tilde{N}(ds, dv, du) \\
&+ \int_0^t \int_{[0,1] \times (S \setminus S_1)} \mathbf{1}_{[0,\lambda(X_{s-},u)]} \gamma(X_{s-}, u) N(ds, dv, du).
\end{aligned} \tag{2.5.2}$$

In other words, if a process $(X_t)_{t \geq 0}$ is such that for any $f \in C_0^\infty(\mathbb{R}^d)$ the process

$$f(X_t) - f(X_0) - \int_0^t Af(X_s)ds$$

is a martingale, then we can construct a Brownian motion $(W_t)_{t \geq 0}$ and a Poisson random measure N such that $(X_t)_{t \geq 0}$ solves the equation (2.5.2), i.e., $(X_t)_{t \geq 0}$ is a weak solution to (2.5.2). Note that the result in [Kur11] is proved under an assumption that the generator A is such that it maps $C_c^2(\mathbb{R}^d)$ into $C_b(\mathbb{R}^d)$. However, if $\lambda \equiv 1$ and $\gamma(x, u) = u$ (and if the drift and the diffusion coefficients are sufficiently regular) then this is indeed the case, which can be inferred e.g. from Lemma 2.1 in [Küh17]. By applying the Itô formula we can also show that every weak solution to (2.5.2) solves the martingale problem for A . Hence we can infer that uniqueness in law of weak solutions for (2.5.2) is equivalent to uniqueness of solutions to the martingale problem for A (cf. Corollary 2.5 in [Kur11]). This will prove to be useful in Section 3.2.3.

3 The coupling method

In this chapter we introduce the coupling method. It is a widespread tool in probability theory, with numerous applications in stochastic analysis, ergodic theory of stochastic processes, stochastic inequalities, etc. There are by now many monographs treating this subject in detail and the reader is encouraged to consult positions such as [Lin92], [Tho00] or [Vil09] and the references therein for a wider perspective. Here we focus on couplings of stochastic processes which are given as solutions to stochastic differential equations with jumps and we show multiple applications of the coupling method for investigating various properties of such processes.

3.1 Couplings

We start with a measure theoretic definition of a coupling.

Definition 3.1.1. *Suppose we have two probability measures μ and ν on measurable spaces (E_1, \mathcal{E}_1) and (E_2, \mathcal{E}_2) , respectively. Then a probability measure π on the product space $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ is called a coupling of μ and ν if it has marginals μ and ν . In other words, if we consider projections $p_1 : E_1 \times E_2 \rightarrow E_1$ and $p_2 : E_1 \times E_2 \rightarrow E_2$ given as $p_1(x_1, x_2) = x_1$ and $p_2(x_1, x_2) = x_2$, then we require the measure π to satisfy $\mu = \pi \circ p_1^{-1}$ and $\nu = \pi \circ p_2^{-1}$ (or, using the push-forward notation, $(p_1)_\# \pi = \mu$ and $(p_2)_\# \pi = \nu$). Equivalently, for all sets $A \in \mathcal{E}_1$, $B \in \mathcal{E}_2$ we have $\pi(A \times E_2) = \mu(A)$ and $\pi(E_1 \times B) = \nu(B)$.*

The definition above will be used in the definition of the Wasserstein distances later in this section. Meanwhile, we introduce a probabilistic definition of a coupling, which we will use extensively in the context of couplings of stochastic processes.

Definition 3.1.2. *Let $(E_1, \mathcal{E}_1, \mu)$ and $(E_2, \mathcal{E}_2, \nu)$ be two probability spaces. Consider random elements $X : \Omega_X \rightarrow E_1$ and $Y : \Omega_Y \rightarrow E_2$ defined on some probability spaces $(\Omega_X, \mathcal{F}_X, \mathbb{P}_X)$ and $(\Omega_Y, \mathcal{F}_Y, \mathbb{P}_Y)$, respectively, such that $\text{Law}(X) = \mu$ and $\text{Law}(Y) = \nu$. A coupling is a random element (X', Y') in $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\text{Law}(X') = \text{Law}(X) = \mu$ and $\text{Law}(Y') = \text{Law}(Y) = \nu$.*

Simply put, if we have two random objects (defined on possibly different probability spaces) with some given distributions (laws), constructing their coupling amounts to constructing two new random objects on a common probability space with these given respective distributions. The important point here is that the original random objects do not need to have any specific joint distribution, while constructing a coupling involves choosing a right way in which the new random objects are jointly distributed.

3 The coupling method

The second definition can be seen as a special case of the first one, if we consider canonical random variables X and Y on $(E_1, \mathcal{E}_1, \mu)$ and $(E_2, \mathcal{E}_2, \nu)$, respectively (i.e., X and Y are identity functions on E_1 and E_2 with distributions μ and ν , respectively). Then constructing a coupling of μ and ν in the sense of Definition 3.1.1 corresponds to constructing a random vector (X', Y') on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ with respective marginal distributions (the law of (X', Y') is a coupling of μ and ν). These definitions extend naturally to any finite number of measures or random elements (see Section 2 in Chapter 3 in [Tho00]).

The simplest example of a coupling of two random objects is their independent coupling, i.e., in Definition 3.1.2 we can choose X' and Y' as independent copies of X and Y , respectively. Then their joint distribution is just the product measure $\mu \otimes \nu$. A more interesting type of a coupling is what is called in [Vil09] a deterministic coupling, i.e., a coupling in which the second random object is a deterministic function of the first one. In other words, (X', Y') is deterministic if there exists a measurable function $T : E_1 \rightarrow E_2$ such that $Y' = T(X')$. Using the push-forward notation, we have $T_{\#}\mu = \nu$.

Example 3.1.3. *Let μ and ν be two atomless probability measures on \mathbb{R} and denote by F and G , respectively, their cumulative distribution functions. We can define their right continuous inverses as*

$$F^{-1}(t) := \inf\{x \in \mathbb{R} : F(x) > t\}$$

and in an analogous way for G . Then we can define the map $T = G^{-1} \circ F$ and we have $T_{\#}\mu = \nu$. This coupling can be called a coupling by monotone rearrangement (or increasing rearrangement - cf. page 19 of [Vil09]), as it maps the left-most quantile of μ onto the left-most quantile of ν .

In this thesis we will be mainly interested in constructing couplings of two copies of the same random object, i.e., we will want the marginal distributions to be the same. The challenge will be in choosing an appropriate joint distribution. Performing such constructions turns out to be a powerful tool in studying stochastic processes and we turn our attention now to this type of couplings.

If we consider a time index set I (usually in our case it will be the interval $[0, \infty)$ or $[0, T]$ for some $T > 0$), then we can interpret (E, \mathcal{E}) -valued stochastic processes indexed by I as random elements in (E^I, \mathcal{E}^I) . Since the distribution of a stochastic process as a random element in (E^I, \mathcal{E}^I) is uniquely determined by its finite dimensional distributions, we arrive at the following definition, which is just a special case of Definition 3.1.2.

Definition 3.1.4. *Let $(X_t)_{t \geq 0}$ be an (E, \mathcal{E}) -valued stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then an $(E \times E, \mathcal{E} \otimes \mathcal{E})$ -valued stochastic process $(Y_t^1, Y_t^2)_{t \geq 0}$ on a (possibly different) probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ is a coupling of two copies of $(X_t)_{t \geq 0}$ if both the marginal processes $(Y_t^1)_{t \geq 0}$ and $(Y_t^2)_{t \geq 0}$ have the same finite dimensional distributions as $(X_t)_{t \geq 0}$. Namely, for any $n \geq 1$, any $t_1, \dots, t_n \geq 0$ and any sets $A_1, \dots, A_n \in \mathcal{E}$ we have*

$$\mathbb{P}'(Y_{t_1}^1 \in A_1, \dots, Y_{t_n}^1 \in A_n) = \mathbb{P}'(Y_{t_1}^2 \in A_1, \dots, Y_{t_n}^2 \in A_n) = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n).$$

See Chapter 4 of [Tho00] for this point of view on couplings of stochastic processes.

In the sequel we will focus on couplings of \mathbb{R}^d -valued Markov processes. It is a simple observation that for a Markov process, its transition kernels uniquely determine its finite dimensional distributions (see e.g. Section 1 in Chapter 4 of [EK86]). Hence, given an \mathbb{R}^d -valued Markov process $(X_t)_{t \geq 0}$ with transition kernels $(p_t(x, \cdot))_{x \in \mathbb{R}^d, t \geq 0}$, we can determine whether an \mathbb{R}^{2d} -valued process $(X'_t, X''_t)_{t \geq 0}$ is a coupling of two copies of $(X_t)_{t \geq 0}$ by comparing the transition kernels of $(X'_t)_{t \geq 0}$ and $(X''_t)_{t \geq 0}$ with $(p_t(x, \cdot))_{x \in \mathbb{R}^d, t \geq 0}$. This formulation of the definition of a coupling is used both in [Maj15] and [Maj16], as well as in papers by other authors such as [SW11] or [Wan11]. Observe that in this definition we do not require $(X'_t, X''_t)_{t \geq 0}$ to be a Markov process. In fact, if $(X'_t, X''_t)_{t \geq 0}$ is a Markov process, such a coupling is called Markovian, but there exist also non-Markovian couplings, see e.g. [HS13] and the references therein for examples.

A famous example of a coupling of two copies of a Markov process is the coupling of Brownian motions by reflection (see Figure 3.1). If we have a Brownian motion $(B_t^1)_{t \geq 0}$ with initial point $x \in \mathbb{R}^d$, we can construct a new Brownian motion $(B_t^2)_{t \geq 0}$ with initial point $y \in \mathbb{R}^d$ (where $y \neq x$) by reflecting the path of $(B_t^1)_{t \geq 0}$ with respect to the hyperplane orthogonal to the vector $x - y$, as shown in the picture below. In the case of couplings of Markov processes we are usually interested only in the behaviour of the path of the second process up to the first time when it meets with the path of the first process (the *coupling time* τ). More precisely, for two processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$, we have $\tau = \inf\{t > 0 : X_t = Y_t\}$. For $t > \tau$, the path of $(Y_t)_{t \geq 0}$ is required to just follow the path of $(X_t)_{t \geq 0}$ (cf. the picture below).

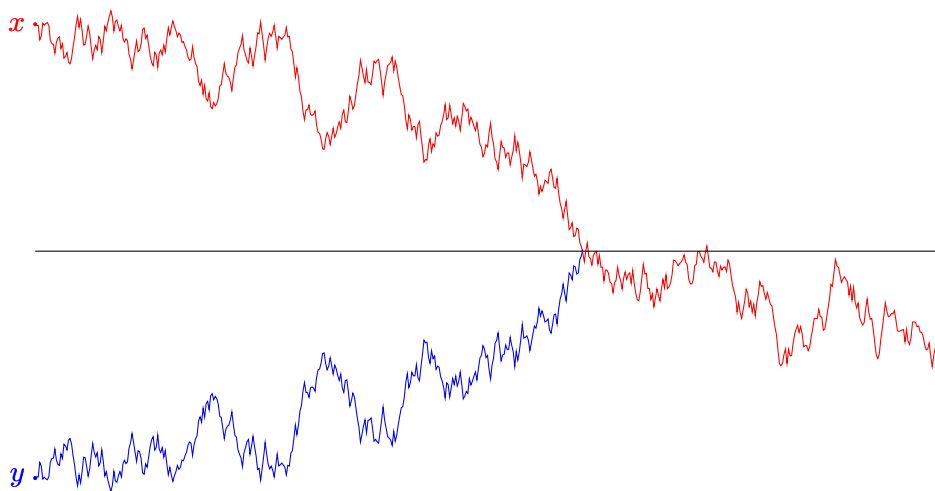


Figure 3.1: Coupling by reflection of Brownian motions.

To make this rigorous, given a coupling $(X_t, Y_t)_{t \geq 0}$ we can construct a new coupling

3 The coupling method

$(X_t, Y'_t)_{t \geq 0}$ by setting

$$Y'_t = \begin{cases} Y_t & \text{for } t < \tau \\ X_t & \text{for } t \geq \tau. \end{cases}$$

It can be easily shown that $(Y'_t)_{t \geq 0}$ has the same finite dimensional distributions as $(Y_t)_{t \geq 0}$ (and hence also $(X_t)_{t \geq 0}$), e.g. under the assumption that $(X_t)_{t \geq 0}$ is a solution to a well-posed martingale problem, cf. Section 2.2 in [JWa16] or Section 3.1 in [PW06].

Usually in the coupling method the main challenge is to construct couplings which satisfy some specific optimality criteria. Consider two copies of a Markov process $(X_t)_{t \geq 0}$ with transition semigroup $(p_t)_{t \geq 0}$ and let their initial distributions be μ and ν , respectively. It is well-known (see e.g. (2.12) in [Lin92] or Theorem 5.1 in Chapter 4 in [Tho00]) that every coupling satisfies the so-called coupling inequality

$$\|\mu p_t - \nu p_t\|_{TV} \leq 2\mathbb{P}(\tau > t),$$

for any $t > 0$, where τ is the coupling time. It is natural to ask whether one can construct a coupling for which this inequality becomes an equality. Such a coupling is usually called the maximal coupling, see Theorem 6.1 in Chapter 4 in [Tho00] for a detailed discussion about its existence. Another notion of optimality appears in [BK00], where Burdzy and Kendall studied *efficient* Markovian couplings, i.e., couplings whose coupling time gives a sharp estimate on the spectral gap λ of the process' generator, in the sense that $\mathbb{P}(\tau > t) \asymp \exp(-\lambda t)$. Yet another optimality criterion appears in [JMS14], where the authors try to minimize or maximize certain functionals of $\mathbb{P}(\tau > t)$.

However, here we are interested in an optimality criterion motivated by the optimal transport theory, where we consider a problem of transporting the mass of one probability measure onto the mass of another in a way which minimizes a given cost function (the Kantorovich problem). Rigorously, given two probability measures μ on E_1 and ν on E_2 , the task is to construct a measure γ on $E_1 \times E_2$ with μ and ν as its marginals (i.e., a coupling of μ and ν) such that for a given function $c : E_1 \times E_2 \rightarrow \mathbb{R}$ we minimize the quantity

$$\int_{E_1 \times E_2} c(x, y) d\gamma(x, y). \quad (3.1.1)$$

If we are only interested in deterministic couplings (in the sense defined above), i.e., if we want to minimize

$$\int_{E_1 \times E_2} c(x, T(x)) \mu(dx)$$

by finding the right $T : E_1 \rightarrow E_2$ such that $\nu = \mu \circ T^{-1}$, we call it the Monge problem, cf. Chapter 1 in [Vil09].

A related notion which is widely used in many areas of probability theory is that of the Wasserstein distance between two given measures. Assume we have two probability measures μ and ν on E , choose $p \in [1, \infty)$ and let ρ be a metric on E . Then we define the L^p -Wasserstein distance between μ and ν by

$$W_{p,\rho}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{E \times E} \rho(x, y)^p d\pi(x, y) \right)^{1/p},$$

where $\Pi(\mu, \nu)$ is the family of all couplings of μ and ν . Equivalently,

$$W_{p,\rho}(\mu, \nu) = \inf_{X,Y} \left\{ [\mathbb{E}\rho(X, Y)^p]^{1/p} : \text{Law}(X) = \mu, \text{Law}(Y) = \nu \right\} .$$

If $E = \mathbb{R}^d$ and the metric ρ is chosen to be the Euclidean metric, we denote $W_{p,\rho}$ simply by W_p .

In the sequel we will be interested mainly in the case when $p = 1$ and the underlying metric ρ is specified by some concave function $f : [0, \infty) \rightarrow [0, \infty)$ in the sense that

$$\rho(x, y) := f(|x - y|) . \quad (3.1.2)$$

It is easy to see that if f is increasing, concave, $f(0) = 0$ and $f(x) > 0$ for $x > 0$, then ρ defined by (3.1.2) is indeed a metric on E . In such a case, we will denote the L^1 -Wasserstein (Kantorovich) distance associated with f via ρ by W_f , i.e., we have

$$W_f(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{E \times E} f(|x - y|) d\pi(x, y) . \quad (3.1.3)$$

Consider now a Markov process $(X_t)_{t \geq 0}$ in \mathbb{R}^d with associated transition kernels $(p_t(x, \cdot))_{t \geq 0, x \in \mathbb{R}^d}$. Recall that if $\text{Law}(X_0) = \mu$ for some probability measure μ on \mathbb{R}^d , then the distribution of the random variable X_t for any $t > 0$ can be denoted by μp_t , where $\mu p_t(dy) = \int \mu(dx) p_t(x, dy)$. We can then be interested in finding a coupling $(X_t, Y_t)_{t \geq 0}$ of two copies of this process, with initial distributions, say, μ and ν , such that for any $t > 0$ the Wasserstein distance between μp_t and νp_t is as small as possible. Note here that finding a coupling of stochastic processes in the sense of Definition 3.1.4, gives us a coupling of the laws of these processes at any time $t > 0$ in the sense of Definition 3.1.1. Note also that any coupling $(X_t, Y_t)_{t \geq 0}$ gives us an upper bound on $W_{p,\rho}(\mu p_t, \nu p_t)$ for any $t > 0$, i.e.,

$$W_{p,\rho}(\mu p_t, \nu p_t) \leq (\mathbb{E}\rho(X_t, Y_t)^p)^{1/p} .$$

Thus, if we want to find sharp upper bounds on Wasserstein distances between laws of a Markov process, we need to find a coupling $(X_t, Y_t)_{t \geq 0}$ of two copies of this process, for which the expressions $(\mathbb{E}\rho(X_t, Y_t)^p)^{1/p}$ are as close to the infimum over all couplings as possible. This will be our task in the sequel.

3.1.1 Coupling by reflection for diffusions

We start by recalling a classical construction of a coupling by reflection for diffusions with non-linear drift, which is due to Lindvall and Rogers [LR86]. Namely, consider a stochastic differential equation of the form

$$dX_t = b(X_t)dt + dB_t , \quad (3.1.4)$$

where $(B_t)_{t \geq 0}$ is a Brownian motion in \mathbb{R}^d and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a drift function. Then we can consider another equation

$$dY_t = b(Y_t)dt + R(X_t, Y_t)dB_t , \quad (3.1.5)$$

3 The coupling method

where

$$R(X_t, Y_t) := I - 2e_t e_t^T \quad (3.1.6)$$

with I being the $d \times d$ identity matrix and

$$e_t := \frac{X_t - Y_t}{|X_t - Y_t|}. \quad (3.1.7)$$

Note that (3.1.5) makes sense only for $t < \tau$ where $\tau = \inf\{t > 0 : X_t = Y_t\}$. For $t \geq \tau$ we can set $Y_t = X_t$. Under assumptions guaranteeing existence of a unique strong solution to the system of equations given by (3.1.4) and (3.1.5), we can easily show that the process $(X_t, Y_t)_{t \geq 0}$ obtained this way is indeed a coupling. To this end, let us notice that the random operator $R(X_t, Y_t)$ defined by (3.1.6) takes values in orthogonal matrices, i.e.,

$$R(X_t, Y_t)R(X_t, Y_t)^T = (I - 2e_t e_t^T)(I - 2e_t e_t^T)^T = I \quad (3.1.8)$$

and hence the process $(\tilde{B}_t)_{t \geq 0}$ defined as

$$\tilde{B}_t := \int_0^t R(X_s, Y_s) dB_s$$

is a Brownian motion in \mathbb{R}^d due to the Lévy characterization theorem (cf. Theorem II-6.1 in [IW89]). To see this, observe that its i -th component is given as

$$\tilde{B}_t^i = \int_0^t R^{ik}(X_s, Y_s) dB_s^k$$

and thus

$$\begin{aligned} [\tilde{B}_t^i, \tilde{B}_t^j] &= \sum_{k,l=1}^d \int_0^t R^{ik}(X_s, Y_s) R^{jl}(X_s, Y_s) d[B^k, B^l]_s \\ &= \int_0^t \sum_{k=1}^d R^{ik}(X_s, Y_s) R^{jk}(X_s, Y_s) ds \\ &= \int_0^t \delta_{ij} ds, \end{aligned}$$

where δ_{ij} is the Kronecker delta and the last equality follows from (3.1.8). Then we use uniqueness in law of solutions to (3.1.4) to conclude that the solution $(Y_t)_{t \geq 0}$ to (3.1.5) has the same finite dimensional distributions as the solution $(X_t)_{t \geq 0}$ to (3.1.4). We can easily replace (3.1.4) by an SDE with a slightly more general additive noise of the form

$$dX_t = b(X_t)dt + \sigma dB_t, \quad (3.1.9)$$

where σ is a constant matrix with $\det \sigma > 0$. Then the construction presented above still works after replacing the vector e_t defined in (3.1.7) with $e_t = \sigma^{-1}(X_t - Y_t)/|\sigma^{-1}(X_t - Y_t)|$

and the noise in (3.1.5) with $\sigma R(X_t, Y_t)dB_t$ (see [LR86] or [Ebe16]). It is possible to consider also the case of a multiplicative noise, i.e., an equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

and to modify the construction accordingly, but the reflection coupling in this case does not have as good properties as in the additive noise case, unless we impose some strict assumptions on σ . See Section 3 in [LR86] for a detailed discussion.

The coupling by reflection introduced above can be used in order to obtain convergence rates to equilibrium for solutions to equations of the form (3.1.9), under assumptions on the drift which will be discussed in detail in the remaining part of this section. This has recently been done by Eberle in [Ebe11] and [Ebe16].

Let us denote by $p_t(x, dy)$ the transition kernels associated with the solution to (3.1.9). We use the notation $\mu p_t(dy) = \int \mu(dx)p_t(x, dy)$ to denote the distribution of X_t at time $t \geq 0$ provided that X_0 is distributed according to a measure μ . We are then interested in properties of the mapping $\mu \mapsto \mu p_t$. Now consider the following assumption on the drift

$$\langle b(x) - b(y), x - y \rangle \leq -K|x - y|^2 \quad (3.1.10)$$

for some constant $K > 0$. If there exists a $K > 0$ such that (3.1.10) holds for all $x, y \in \mathbb{R}^d$, then we say that b satisfies a *global dissipativity* condition. If there exists a $K > 0$ and an $R > 0$ such that (3.1.10) holds for x and $y \in \mathbb{R}^d$ such that $|x - y| > R$, then we say that b satisfies a *dissipativity at infinity* condition.

An example which often appears in the literature is when the drift b in the equation (3.1.4) is given in terms of a gradient of some potential function U . If $b(x) = -\nabla U(x)$ and U is strongly convex, then b is globally dissipative. On the other hand, the dissipativity at infinity assumption allows us to cover a much wider spectrum of examples, including the double-well potential (also triple-well etc.) and other potentials which behave well outside a compact set of arbitrary size (roughly speaking, the drift is supposed to “point towards the origin” from large distances).

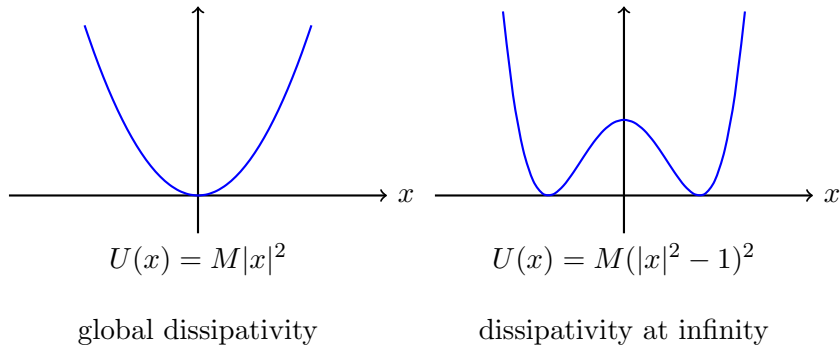


Figure 3.2: Comparison of the global dissipativity and the dissipativity at infinity conditions.

3 The coupling method

It can be easily shown that if the drift in (3.1.9) is globally dissipative, then the semigroup $(p_t)_{t \geq 0}$ associated with the solution $(X_t)_{t \geq 0}$ to (3.1.9) is exponentially contractive with respect to Wasserstein distances W_p for all $p \in [1, \infty)$. Namely, let us use the synchronous coupling $(X_t, Y_t)_{t \geq 0}$ for solutions to (3.1.9) by defining the second marginal process in the coupling via the equation

$$dY_t = b(Y_t)dt + \sigma dB_t. \quad (3.1.11)$$

This means that we apply exactly the same noise as in (3.1.9). Then obviously

$$X_t - Y_t = X_0 - Y_0 + \int_0^t (b(X_s) - b(Y_s)) ds.$$

Hence, by the Itô formula we get

$$|X_t - Y_t|^p = |X_0 - Y_0|^p + \int_0^t \langle b(X_s) - b(Y_s), (X_s - Y_s)p |X_s - Y_s|^{p-2} \rangle ds$$

and using (3.1.10) we obtain

$$|X_t - Y_t|^p \leq |X_0 - Y_0|^p - Kp \int_0^t |X_s - Y_s|^p ds.$$

Since this holds on any time interval $[r, u] \subset [0, t]$, by the differential version of the Gronwall inequality we have

$$|X_t - Y_t|^p \leq |X_0 - Y_0|^p e^{-Kpt}$$

and hence

$$W_p(\mu p_t, \nu p_t) \leq (\mathbb{E}|X_0 - Y_0|^p)^{1/p} e^{-Kt}.$$

However, the inequality above holds for all couplings of X_0 and Y_0 and thus

$$W_p(\mu p_t, \nu p_t) \leq e^{-Kt} W_p(\mu, \nu). \quad (3.1.12)$$

It is worth pointing out that for equations of the form (3.1.9), the global dissipativity of the drift not only implies (3.1.12) for all $p \in [1, \infty)$ but the two conditions are actually equivalent (this can be inferred from Corollary 1.4 in [vRS05]). However, it turns out that at least for the W_1 distance we can prove a similar inequality to (3.1.12) under a much weaker dissipativity at infinity condition on the drift. Namely, Eberle in [Ebe16] used a class of L^1 -Wasserstein (Kantorovich) distances W_f associated with concave functions (recall (3.1.3)) to show that under the dissipativity at infinity condition we have

$$W_1(\mu p_t, \nu p_t) \leq C e^{-ct} W_1(\mu, \nu) \quad (3.1.13)$$

with some constants $c > 0$ and $C > 1$. This can be done by showing first that for an appropriately chosen distance W_f we have

$$W_f(\mu p_t, \nu p_t) \leq e^{-ct} W_f(\mu, \nu). \quad (3.1.14)$$

If W_f is chosen in such a way that there is a constant $C > 0$ such that $C^{-1}x \leq f(x) \leq x$ for all $x \in \mathbb{R}^d$, then

$$C^{-1}W_1(\mu p_t, \nu p_t) \leq W_f(\mu p_t, \nu p_t) \leq e^{-ct}W_f(\mu, \nu) \leq e^{-ct}W_1(\mu, \nu)$$

and we arrive at (3.1.13). To this end, f needs to be chosen as a non-decreasing, concave function, which is extended in an affine way from some point $R_1 > 0$ (cf. Figure 3.3).

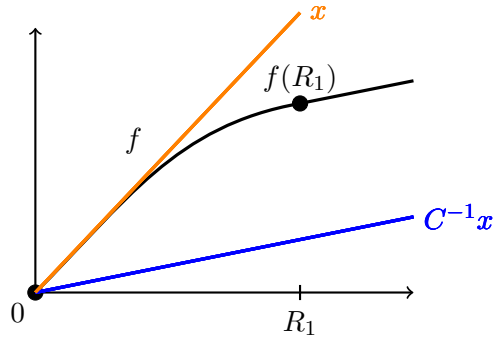


Figure 3.3: The choice of f in [Ebe16].

The details of this construction can be found in [Ebe16], see also Section 4 of [Maj16]. It is worth pointing out that there exist some extensions of the technique from [Ebe16], which allow us to cover some cases of equations of the form (3.1.4) in which the drift is not dissipative even at infinity. While we cannot expect to get results of the form (3.1.13) in such a case, it is still possible to obtain some upper bounds on the W_1 distance by using Lyapunov functions, cf. [EGZ16]. Here, however, we focus on extending the results from [Ebe16] in a different direction.

3.2 Coupling constructions for SDEs with jumps

One of the main goals of this PhD dissertation is to present analogous results in the case of SDEs driven by Lévy processes with jumps. We first focus on equations of the form

$$dX_t = b(X_t)dt + dL_t, \tag{3.2.1}$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a continuous, one-sided Lipschitz drift, whereas $(L_t)_{t \geq 0}$ is a pure jump Lévy process on \mathbb{R}^d . Existence of a unique strong solution to (3.2.1) follows in such a case from Theorem 2.4.5. We consider rotationally invariant Lévy processes (i.e., processes with rotationally invariant Lévy measure ν), in hope of obtaining a construction similar to the one due to Lindvall and Rogers [LR86], i.e., we would like to use some kind of a reflection technique to provide a coupling which would allow us to obtain inequalities of the type (3.1.14) for appropriately chosen Wasserstein distances.

3.2.1 Coupling by reflection for Lévy-driven SDEs

When $(L_t)_{t \geq 0}$ is rotationally invariant, in principle it is possible to define a coupling by reflection in an analogous manner to [LR86], i.e., we can set

$$dY_t = b(Y_t)dt + R(X_{t-}, Y_{t-})dL_t \quad (3.2.2)$$

with the reflection operator R defined by (3.1.6). Such a coupling turns out not to be very useful for our purposes but we believe that studying it can be instructive and serve as a prelude to the much more sophisticated construction presented in [Maj15]. The proof that it is indeed a coupling of two copies of the solution to (3.2.1) is more involved than in the Brownian case, since there is no straightforward analogue of the Lévy characterization theorem in the case of jump processes. However, we can still show that a process defined by

$$\tilde{L}_t := \int_0^t R(X_{s-}, Y_{s-})dL_s \quad (3.2.3)$$

is a Lévy process with the same finite dimensional distributions as $(L_t)_{t \geq 0}$, which allows us to use uniqueness in law of solutions to (3.2.1) to conclude that the process $(X_t, Y_t)_{t \geq 0}$ defined as a solution to the system given by (3.2.1) and (3.2.2) is indeed a coupling of two copies of the solution to (3.2.1). Note that uniqueness in law for (3.2.1) follows from Theorem 2.4.5 combined with the Yamada-Watanabe result from Section 2.4. One way to handle the process $(\tilde{L}_t)_{t \geq 0}$ is to use the Lévy-Itô decomposition for $(L_t)_{t \geq 0}$ to rewrite (3.2.3) as

$$\tilde{L}_t = \int_0^t \int_{\{|v|>1\}} R(X_{s-}, Y_{s-})vN(dv, ds) + \int_0^t \int_{\{|v|\leq 1\}} R(X_{s-}, Y_{s-})v\tilde{N}(dv, ds)$$

and then to show that

$$\begin{aligned} \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v|\leq 1\}} R(X_{s-}, Y_{s-})v\tilde{N}(dv, ds) \right\rangle \right) \\ = \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v|\leq 1\}} v\tilde{N}(dv, ds) \right\rangle \right) \end{aligned} \quad (3.2.4)$$

and

$$\begin{aligned} \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v|>1\}} R(X_{s-}, Y_{s-})vN(dv, ds) \right\rangle \right) \\ = \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v|>1\}} vN(dv, ds) \right\rangle \right) \end{aligned}$$

for all $z \in \mathbb{R}^d$. If we then additionally show that the integrals

$$\int_s^t \int_{\{|v|\leq 1\}} R(X_{r-}, Y_{r-})v\tilde{N}(dv, dr)$$

and

$$\int_s^t \int_{\{|v|>1\}} R(X_{r-}, Y_{r-}) v N(dv, dr)$$

are both independent of \mathcal{F}_s , where $(\mathcal{F}_t)_{t \geq 0}$ is the filtration generated by $(L_t)_{t \geq 0}$, then we arrive at our conclusion (cf. e.g. the proof of Theorem II-6.1 in [IW89]). We present here a detailed proof of how to obtain (3.2.4), since it can be seen as a simplified version of a reasoning presented in Section 2.5 in [Maj15] in a more sophisticated setting, and thus it may help the reader to better understand the rationale behind the proof in [Maj15]. In order to prove (3.2.4), we can approximate the integral

$$\int_0^t \int_{\{|v| \leq 1\}} R(X_{s-}, Y_{s-}) v \tilde{N}(dv, ds)$$

by integrals of Riemann sums of the form

$$\int_0^t \int_{\{|v| \leq 1\}} \left(\sum_{k=0}^{m_n-1} R(X_{t_k^n}, Y_{t_k^n}) v \mathbf{1}_{(t_k^n, t_{k+1}^n]}(s) \right) \tilde{N}(dv, ds),$$

where $0 = t_0^n < t_1^n < \dots < t_{m_n}^n = t$ is a sequence of partitions of the interval $[0, t]$ with the mesh size $\delta_n := \max_{k \in \{0, \dots, m_n-1\}} |t_{k+1}^n - t_k^n| \rightarrow 0$ as $n \rightarrow \infty$. Then of course

$$\begin{aligned} & \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v| \leq 1\}} \left(\sum_{k=0}^{m_n-1} R(X_{t_k^n}, Y_{t_k^n}) v \mathbf{1}_{(t_k^n, t_{k+1}^n]}(s) \right) \tilde{N}(dv, ds) \right\rangle \right) \\ & \rightarrow \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v| \leq 1\}} R(X_{s-}, Y_{s-}) v \tilde{N}(dv, ds) \right\rangle \right) \end{aligned}$$

for any $z \in \mathbb{R}^d$, as $n \rightarrow \infty$. But we can actually show that for all $n \in \mathbb{N}$ we have

$$\begin{aligned} & \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v| \leq 1\}} \left(\sum_{k=0}^{m_n-1} R(X_{t_k^n}, Y_{t_k^n}) v \mathbf{1}_{(t_k^n, t_{k+1}^n]}(s) \right) \tilde{N}(dv, ds) \right\rangle \right) \\ & = \mathbb{E} \exp \left(i \left\langle z, \int_0^t \int_{\{|v| \leq 1\}} v \tilde{N}(dv, ds) \right\rangle \right), \end{aligned} \tag{3.2.5}$$

which then finishes the proof. This last claim is based on the fact that for any $n \in \mathbb{N}$ and any $k \in \{0, \dots, m_n\}$ we have

$$\begin{aligned} & \mathbb{E} \left[\exp \left(i \left\langle z, \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq 1\}} R(X_{t_k^n}, Y_{t_k^n}) v \tilde{N}(ds, dv) \right\rangle \right) \middle| \mathcal{F}_{t_k^n} \right] \\ & = \exp \left((t_{k+1}^n - t_k^n) \int_{\{|v| \leq 1\}} \left(e^{i \langle z, R(X_{t_k^n}, Y_{t_k^n}) v \rangle} - 1 - i \langle z, R(X_{t_k^n}, Y_{t_k^n}) v \rangle \right) \nu(dv) \right), \end{aligned} \tag{3.2.6}$$

3 The coupling method

cf. Lemma 2.4 in [Maj15]. Thus if we compute

$$\begin{aligned}
& \mathbb{E} \exp \left(i \left\langle z, \sum_{k=0}^{m_n-1} \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq 1\}} R(X_{t_k^n}, Y_{t_k^n}) v \tilde{N}(ds, dv) \right\rangle \right) \\
&= \mathbb{E} \left(\mathbb{E} \left[\prod_{k=0}^{m_n-2} \exp \left(i \left\langle z, \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq 1\}} R(X_{t_k^n}, Y_{t_k^n}) v \tilde{N}(ds, dv) \right\rangle \right) \right. \right. \\
&\quad \left. \left. \times \exp \left(i \left\langle z, \int_{t_{m_n-1}^n}^{t_{m_n}^n} \int_{\{|v| \leq 1\}} R(X_{t_{m_n-1}^n}, Y_{t_{m_n-1}^n}) v \tilde{N}(ds, dv) \right\rangle \right) \middle| \mathcal{F}_{t_{m_n-1}^n} \right] \right) \\
&= \mathbb{E} \left(\prod_{k=0}^{m_n-2} \exp \left(i \left\langle z, \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq 1\}} R(X_{t_k^n}, Y_{t_k^n}) v \tilde{N}(ds, dv) \right\rangle \right) \right) \\
&\quad \times \mathbb{E} \left[\exp \left(i \left\langle z, \int_{t_{m_n-1}^n}^{t_{m_n}^n} \int_{\{|v| \leq 1\}} R(X_{t_{m_n-1}^n}, Y_{t_{m_n-1}^n}) v \tilde{N}(ds, dv) \right\rangle \right) \middle| \mathcal{F}_{t_{m_n-1}^n} \right],
\end{aligned}$$

we can use (3.2.6) to evaluate the conditional expectation appearing in our calculations. Moreover, due to rotational invariance of the Lévy measure ν , we easily see that for any $n \in \mathbb{N}$ and any $k \in \{0, \dots, m_n\}$ we have

$$\begin{aligned}
& \exp \left((t_{k+1}^n - t_k^n) \int_{\{|v| \leq 1\}} \left(e^{i \langle z, R(X_{t_k^n}, Y_{t_k^n}) v \rangle} - 1 - i \langle z, R(X_{t_k^n}, Y_{t_k^n}) v \rangle \right) \nu(dv) \right) \\
&= \exp \left((t_{k+1}^n - t_k^n) \int_{\{|v| \leq 1\}} \left(e^{i \langle z, v \rangle} - 1 - i \langle z, v \rangle \right) \nu(dv) \right).
\end{aligned} \tag{3.2.7}$$

This gives us

$$\begin{aligned}
& \mathbb{E} \exp \left(i \left\langle z, \sum_{k=0}^{m_n-1} \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq 1\}} R(X_{t_k^n}, Y_{t_k^n}) v \tilde{N}(ds, dv) \right\rangle \right) \\
&= \exp \left((t_{m_n}^n - t_{m_n-1}^n) \int_{\{|v| \leq 1\}} \left(e^{i \langle z, v \rangle} - 1 - i \langle z, v \rangle \right) \nu(dv) \right) \\
&\quad \times \mathbb{E} \left(\prod_{k=0}^{m_n-2} \exp \left(i \left\langle z, \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq 1\}} R(X_{t_k^n}, Y_{t_k^n}) v \tilde{N}(ds, dv) \right\rangle \right) \right).
\end{aligned}$$

By iterating this procedure, we arrive at (3.2.5). A similar argument is used in Section 2.5 in [Maj15]. The crucial difference is that in the setting from [Maj15] it is much more difficult to prove an analogue of (3.2.7).

The coupling by reflection constructed this way is, however, not efficient for obtaining good convergence rates in Wasserstein distances. An intuitive reason for this is that for jump processes, unlike for diffusions, it can easily happen that two coupled (reflected) processes, after already coming close to each other, will rapidly spread apart after a jump in a wrong direction (see Figure 3.4).

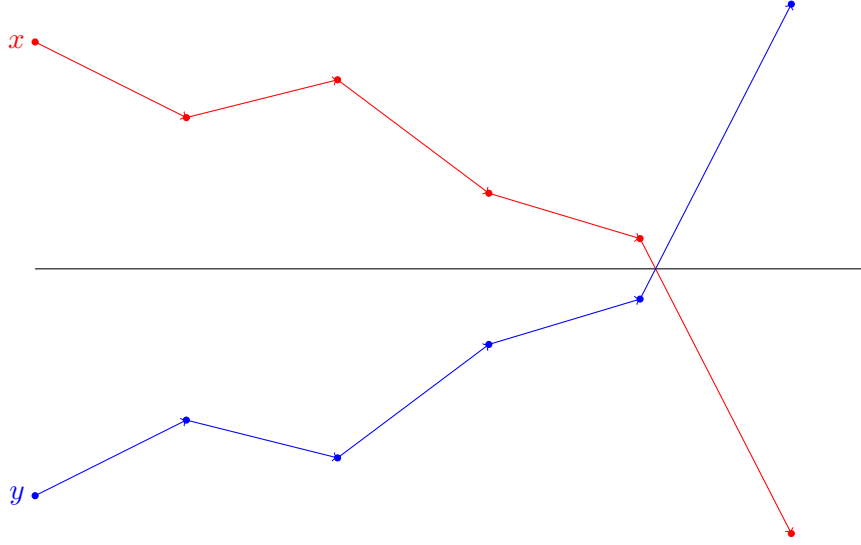


Figure 3.4: Coupling by reflection for jump processes.

Such a behaviour can obviously significantly disturb convergence to equilibrium and therefore, in order to obtain good convergence rates, we need to introduce an alternative construction that will significantly restrict the probability of something like this happening.

3.2.2 Optimal transport construction

The idea we use comes from the optimal transport theory, which considers a problem of coupling two probability measures in a way which minimizes the transport cost, defined as an integral of a given cost function with respect to that coupling (cf. (3.1.1) and the discussion in the first part of Section 3.1 or see the book by Villani [Vil09] for a comprehensive treatment).

In a paper by McCann [McC99], it was proved that for two probability measures μ and ν on \mathbb{R} such that the signed measure $\mu - \nu$ changes its sign at most twice along the real line, there exists a coupling γ which minimizes the transport cost

$$\int c(x, y) \gamma(dx dy)$$

for all concave cost functions c and that optimal coupling is given by an explicit formula. Roughly speaking, the idea for the coupling is to keep in place the common mass of μ and ν and to apply antimonotone rearrangement to the remaining mass. See the example in Figure 3.5, where the left-most quantile of the measure μ is mapped onto the right-most quantile of the measure ν . Compare this with Example 3.1.3, where we considered a coupling by monotone rearrangement. For details, see Section 2 in [McC99], in particular the comments after Theorem 2.5 and Propositions 2.11 and 2.12.

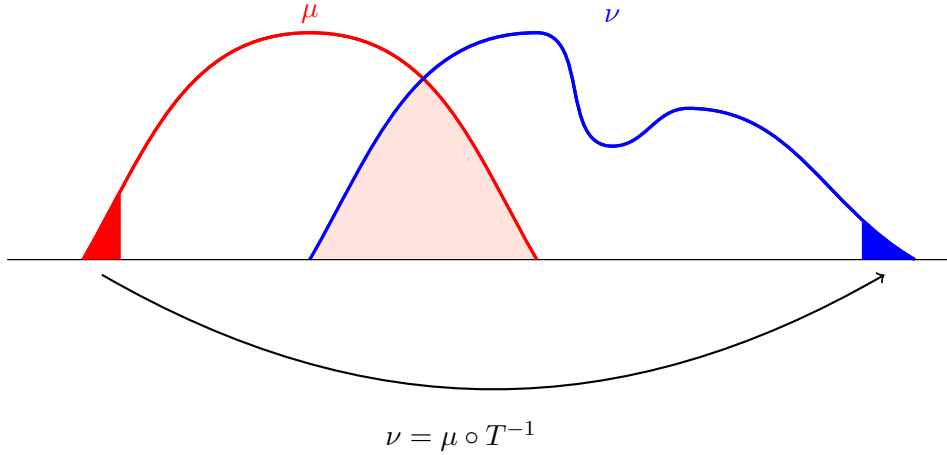


Figure 3.5: Coupling by antimonotone rearrangement.

The important question for us is how to apply this idea in order to construct a coupling of Lévy-driven SDEs. A helpful hint can be found by analyzing a related construction presented in the paper [HS13] by Hsu and Sturm. They consider a family of symmetric densities $(p(x, \cdot))_{x \in \mathbb{R}}$ on \mathbb{R} and they construct a coupling of $p(x_1, \cdot)$ and $p(x_2, \cdot)$ for $x_1 \neq x_2$ in a way which minimizes transport cost for all concave cost functions. Namely, given $p(x_1, \cdot)$ and $p(x_2, \cdot)$ they construct a measure $m(x_1, x_2, \cdot)$ on \mathbb{R}^2 with $p(x_1, \cdot)$ and $p(x_2, \cdot)$ as its marginals such that for all concave costs $\phi : \mathbb{R} \rightarrow \mathbb{R}$ and for any other coupling $\gamma(x_1, x_2, \cdot)$ of $p(x_1, \cdot)$ and $p(x_2, \cdot)$ we have

$$\int_{\mathbb{R}^2} \phi(|x - y|) \gamma(dx dy) \geq \int_{\mathbb{R}^2} \phi(|x - y|) m(dx dy). \quad (3.2.8)$$

The way they present their construction is by taking a random variable ζ_1 with the density $p(x_1, \cdot)$ and defining a new random variable ζ_2 whose values depend on those of ζ_1 . The coupling $m(x_1, x_2, \cdot)$ is then obtained as the joint density of (ζ_1, ζ_2) . It turns out that in order for m to satisfy (3.2.8), when ζ_1 takes a value z_1 , we should assign to ζ_2 either the same value or the value reflected with respect to the point $\frac{x_1 + x_2}{2}$. The crucial problem is obviously specifying the probabilities of these events in an appropriate way. These should be

$$\mathbb{P}(\zeta_2 = \zeta_1 | \zeta_1 = z_1) = \frac{p(x_1, z_1) \wedge p(x_2, z_1)}{p(x_1, z_1)} \quad (3.2.9)$$

and

$$\mathbb{P}(\zeta_2 = x_1 + x_2 - \zeta_1 | \zeta_1 = z_1) = 1 - \frac{p(x_1, z_1) \wedge p(x_2, z_1)}{p(x_1, z_1)}. \quad (3.2.10)$$

The way in which these quantities are chosen is illustrated in Figure 3.6.

Rigorously, this means that $m(x_1, x_2, \cdot)$ is defined by

$$\begin{aligned} m(x_1, x_2, dy_1, dy_2) = & \delta_{y_1}(dy_2) (p(x_1, y_1) \wedge p(x_2, y_1)) dy_1 \\ & + \delta_{Ry_1}(dy_2) (p(x_1, y_1) - p(x_1, y_1) \wedge p(x_2, y_1)) dy_1, \end{aligned}$$

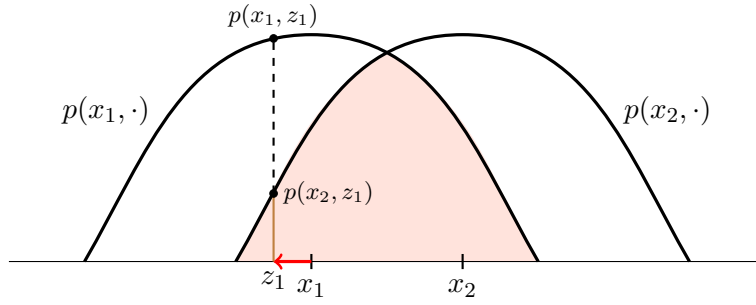


Figure 3.6: Coupling of $p(x_1, \cdot)$ and $p(x_2, \cdot)$.

where

$$Ry_1 = x_1 + x_2 - y_1.$$

In order to check that $m(x_1, x_2, \cdot)$ is indeed a coupling of $p(x_1, \cdot)$ and $p(x_2, \cdot)$, we need to use symmetry of the densities $(p(x, \cdot))_{x \in \mathbb{R}}$, i.e., we use the fact that for any x_1, x_2 and $y_1 \in \mathbb{R}$ we have

$$p(x_1, y_1) = p(x_2, Ry_1).$$

Note that with the formulas (3.2.9) and (3.2.10), when the point z_1 is chosen somewhere near x_2 , we may have $p(x_2, z_1) > p(x_1, z_1)$ and then with probability 1 we have to choose $\zeta_2 = \zeta_1 = z_1$. On the other hand, if z_1 is chosen such that $p(x_2, z_1) = 0$, we may be forced to choose $\zeta_2 = x_1 + x_2 - z_1$ almost surely.

If we interpret $p(x_1, \cdot)$ and $p(x_2, \cdot)$ as jump densities of two processes which are, before the jump, at the positions x_1 and x_2 , we may get some intuition regarding the appropriate way to couple two jump processes.

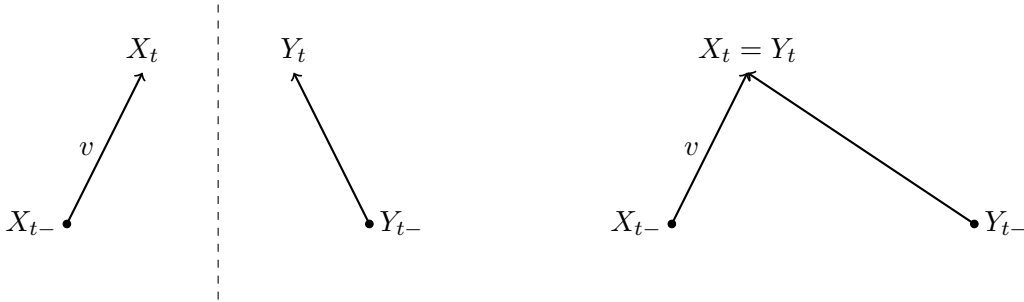


Figure 3.7: The way to couple jumps of two stochastic processes.

Suppose we have two processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ that both jump at some time $t > 0$. Before the jump they are at positions X_{t-} and Y_{t-} , respectively. Assume that the size of the jump of $(X_t)_{t \geq 0}$ at time t is described by a vector $v \in \mathbb{R}^d$ and that its distribution is rotationally invariant. How do we choose the size of the jump of $(Y_t)_{t \geq 0}$?

3 The coupling method

The construction presented above suggests that the jump of $(Y_t)_{t \geq 0}$ should be either the jump of $(X_t)_{t \geq 0}$ reflected with respect to the hyperplane orthogonal to the vector $X_{t-} - Y_{t-}$ or that we should force $(Y_t)_{t \geq 0}$ to jump to the same point that $(X_t)_{t \geq 0}$ jumped to (see Figure 3.7).

We will now try to construct a coupling of solutions to Lévy-driven SDEs which will work exactly in such a way. We do not formally verify that our coupling is indeed optimal in the sense of minimizing Wasserstein distances W_f based on concave functions f (whether this is true remains an open problem potentially worth investigating), but in [Maj15] this coupling was successfully applied to obtain good convergence rates in such distances under quite mild assumptions on the drift and the noise in (3.2.1).

In order for our construction to work, in addition to the assumption that the Lévy measure ν of the process $(L_t)_{t \geq 0}$ is rotationally invariant, we also require it to be absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d with some density q , i.e., we have

$$\nu(dv) = q(v)dv.$$

We start by rewriting the noise $(L_t)_{t \geq 0}$ in (3.2.1) using the Lévy-Itô decomposition as

$$L_t = \int_0^t \int_{\{|v| > 1\}} v N(dv, ds) + \int_0^t \int_{\{|v| \leq 1\}} v \tilde{N}(dv, ds).$$

Note that we can always represent the Poisson random measure N on $\mathbb{R}_+ \times \mathbb{R}^d$ as a sum of Dirac measures, i.e., we have

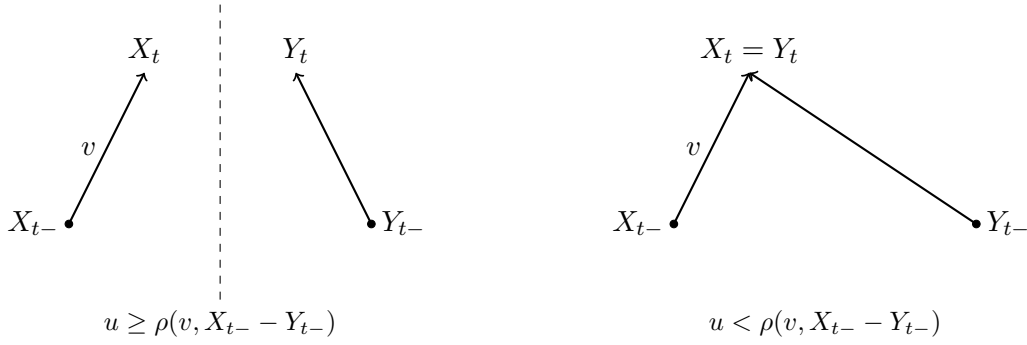
$$N([0, t], A)(\omega) = \sum_{j=1}^{\infty} \delta_{(\tau_j(\omega), \xi_j(\omega))}([0, t] \times A) \text{ for all } \omega \in \Omega \text{ and } A \in \mathcal{B}(\mathbb{R}^d)$$

where τ_j are random variables in \mathbb{R}_+ representing times of jumps and ξ_j are random variables in \mathbb{R}^d representing sizes of jumps, cf. the discussion in Section 2.1. We can embed N in $\mathbb{R}_+ \times \mathbb{R}^d \times [0, 1]$ by replacing each ξ_j with (ξ_j, η_j) , where η_j is a uniformly distributed random variable on $[0, 1]$. Then we can rewrite

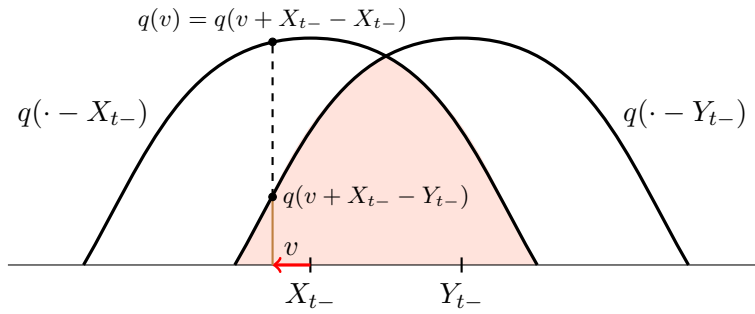
$$L_t = \int_0^t \int_{\{|v| > 1\} \times [0, 1]} v N(dv, du, ds) + \int_0^t \int_{\{|v| \leq 1\} \times [0, 1]} v \tilde{N}(dv, du, ds), \quad (3.2.11)$$

where, by a slight abuse of notation, we still denote our new, extended Poisson random measure by N . This operation obviously does not change the behaviour of the process $(L_t)_{t \geq 0}$, but now, with its every jump, we additionally draw a random number $u \in [0, 1]$, which will serve as a control number helping us to decide what we should do with the jump of the second marginal process $(Y_t)_{t \geq 0}$ in our coupling.

Suppose $(X_t)_{t \geq 0}$ makes a jump of size $v \in \mathbb{R}^d$ at time $t > 0$. We associate with this jump a number $u \in [0, 1]$. We now need to define a control function $\rho = \rho(v, X_{t-} - Y_{t-})$ whose value depends on the size of the jump and the distance between the two coupled processes before the jump. We will then compare the values of u and ρ , and, based on the result of this comparison, we will decide whether we should reflect the jump of $(Y_t)_{t \geq 0}$ or force it to jump to the same place as $(X_t)_{t \geq 0}$.


 Figure 3.8: Choosing where to jump based on comparison between u and ρ .

In order to get an idea on how to define ρ , recall how we defined the probabilities (3.2.9) and (3.2.10) when we coupled one-dimensional densities $p(x_1, \cdot)$ and $p(x_2, \cdot)$. Now we use the same idea to couple two copies of the jump density q , corresponding to the jump of $(X_t)_{t \geq 0}$ from the point X_{t-} and the jump of $(Y_t)_{t \geq 0}$ from the point Y_{t-} , cf. Figure 3.9.


 Figure 3.9: Coupling of two copies of the jump density q .

We thus arrive at the formula

$$\rho(v, X_{t-} - Y_{t-}) := \frac{q(v) \wedge q(v + X_{t-} - Y_{t-})}{q(v)}. \quad (3.2.12)$$

Now we are ready to finally write down the formula for a coupling of solutions to (3.2.1), expressed in terms of a system of two SDEs, just as the classical coupling by reflection for diffusions from [LR86] given by (3.1.4) and (3.1.5). Before we proceed though, let us remark that for technical reasons that are explained in Section 2.2 in [Maj15], we will apply our construction presented above only to the jumps of size smaller than m (where $m > 0$ is chosen accordingly, based on properties of the Lévy measure ν of the process $(L_t)_{t \geq 0}$). To the bigger jumps we just apply the synchronous coupling,

3 The coupling method

i.e., we do not change the noise from the original equation, cf. (3.1.11). This forces us to modify the definition of ρ given in (3.2.12) to

$$\rho(v, X_{t-} - Y_{t-}) := \frac{q(v) \wedge q(v + X_{t-} - Y_{t-}) \mathbf{1}_{\{|v+X_{t-}-Y_{t-}|\leq m\}}}{q(v)}. \quad (3.2.13)$$

We can always choose m large enough if necessary, hence we assume that $m > 1$. Note that then we can rewrite the Lévy-Itô decomposition of $(L_t)_{t \geq 0}$ in (3.2.11) as

$$\begin{aligned} L_t &= \int_0^t \int_{\{|v|>m\} \times [0,1]} v N(ds, dv, du) + \int_0^t \int_{\{|v|\leq m\} \times [0,1]} v \tilde{N}(ds, dv, du) \\ &\quad + \int_0^t \int_{\{m \geq |v| > 1\} \times [0,1]} v \nu(dv) du ds. \end{aligned}$$

For convenience, we can include the last term above in the drift of our SDE. Since changing the drift function by a constant does not influence its properties such as continuity or dissipativity, we will slightly abuse the notation and keep denoting the new drift by b . Equivalently, we may just as well assume that the process $(L_t)_{t \geq 0}$ is given from the start by

$$L_t = \int_0^t \int_{\{|v|>m\} \times [0,1]} v N(ds, dv, du) + \int_0^t \int_{\{|v|\leq m\} \times [0,1]} v \tilde{N}(ds, dv, du). \quad (3.2.14)$$

Now we rewrite the equation (3.2.1) using the representation (3.2.14) of $(L_t)_{t \geq 0}$ and we have

$$dX_t = b(X_t)dt + \int_{\{|v|>m\} \times [0,1]} v N(dt, dv, du) + \int_{\{|v|\leq m\} \times [0,1]} v \tilde{N}(dt, dv, du). \quad (3.2.15)$$

We write the equation for $(Y_t)_{t \geq 0}$ as

$$\begin{aligned} dY_t &= b(Y_t)dt + \int_{\{|v|>m\} \times [0,1]} v N(dt, dv, du) \\ &\quad + \int_{\{|v|\leq m\} \times [0,1]} (X_{t-} - Y_{t-} + v) \mathbf{1}_{\{u < \rho(v, Z_{t-})\}} \tilde{N}(dt, dv, du) \\ &\quad + \int_{\{|v|\leq m\} \times [0,1]} R(X_{t-}, Y_{t-}) v \mathbf{1}_{\{u \geq \rho(v, Z_{t-})\}} \tilde{N}(dt, dv, du), \end{aligned} \quad (3.2.16)$$

where

$$Z_t := X_t - Y_t.$$

The first integral in the formula above corresponds to the synchronous coupling for jumps larger than m , the second integral corresponds to the jump bringing $(Y_t)_{t \geq 0}$ to the position of $(X_t)_{t \geq 0}$ when $u < \rho(v, Z_{t-})$ and the last integral corresponds to the reflected jump which happens whenever $u \geq \rho(v, Z_{t-})$. In Section 2 in [Maj15] it is rigorously proved that the system of equations defined by (3.2.15) and (3.2.16) has a

unique strong solution $(X_t, Y_t)_{t \geq 0}$, which is a coupling of two copies of the solution to (3.2.15). The existence is shown using the interlacing technique (Section 2.4 in [Maj15]), while the fact that the process obtained this way is indeed a coupling of solutions to (3.2.15) is shown by using a modification of the technique presented in Section 3.2.1 in the context of the coupling by reflection for Lévy-driven SDEs (Section 2.5 in [Maj15]).

Before we proceed to discuss multiple applications of this coupling, let us have a look at an alternative approach to proving that $(Y_t)_{t \geq 0}$ defined by (3.2.16) has the same finite dimensional distributions as $(X_t)_{t \geq 0}$ defined by (3.2.15).

3.2.3 Martingale problem approach

For an SDE of the form

$$dX_t = b(X_t)dt + dL_t \tag{3.2.17}$$

with

$$L_t = \int_0^t \int_{\{|v|>m\}} vN(dv, ds) + \int_0^t \int_{\{|v|\leq m\}} v\tilde{N}(dv, ds),$$

by Theorem 6.7.4 in [App09] the generator A associated with its solution is given by

$$\begin{aligned} Af(x) &= \langle b(x), \nabla f(x) \rangle + \int_{\{|z|>m\}} (f(x+z) - f(x)) \nu(dz) \\ &\quad + \int_{\{|z|\leq m\}} (f(x+z) - f(x) - \langle z, \nabla f(x) \rangle) \nu(dz). \end{aligned} \tag{3.2.18}$$

By the Yamada-Watanabe result for SDEs with jumps (see the discussion in Section 2.4 and e.g. Corollary 140 in [Sit05] or Theorem 1 in [BLG15]) it is known that pathwise uniqueness of solutions to (3.2.17) implies uniqueness in law of weak solutions. Moreover, by Corollary 2.5 in [Kur11] we know that uniqueness in law for (3.2.17) is equivalent to uniqueness of solutions to the martingale problem associated with the generator A defined by (3.2.18) for $f \in C_c^2(\mathbb{R}^d)$. Hence, if we define a process $(X_t, Y_t)_{t \geq 0}$ on \mathbb{R}^{2d} via its generator \mathcal{L} and we show that the marginal generators of \mathcal{L} on \mathbb{R}^d coincide with A given by (3.2.18), we will know that $(X_t, Y_t)_{t \geq 0}$ is indeed a coupling of two copies of the solution to (3.2.17).

The process $(X_t, Y_t)_{t \geq 0}$ given as a strong solution to the system of equations (3.2.15)

3 The coupling method

and (3.2.16) is given for $h \in C_c^2(\mathbb{R}^{2d})$ by

$$\begin{aligned}
\mathcal{L}h(x, y) &= \langle b(x), \nabla_x h(x, y) \rangle + \langle b(y), \nabla_y h(x, y) \rangle \\
&+ \int_{\{|z|>m\}} (h(x+z, y+z) - h(x, y)) \nu(dz) \\
&+ \int_{\{|z|\leq m\}} \left(h(x+z, x+z) - h(x, y) - \langle \nabla_x h(x, y), z \rangle \right. \\
&\quad \left. - \langle \nabla_y h(x, y), z+x-y \rangle \right) \rho(z, x-y) \nu(dz) \\
&+ \int_{\{|z|\leq m\}} \left(h(x+z, y+R(x, y)z) - h(x, y) - \langle \nabla_x h(x, y), z \rangle \right. \\
&\quad \left. - \langle \nabla_y h(x, y), R(x, y)z \rangle \right) (\nu(dz) - \rho(z, x-y) \nu(dz)) .
\end{aligned}$$

Now fix a function $g \in C_c^2(\mathbb{R}^d)$. If we take $h(x, y) = g(x)$, then we immediately see that $\mathcal{L}h(x, y) = Ag(x)$. Now take $h(x, y) = g(y)$. Then we have

$$\begin{aligned}
\mathcal{L}g(y) &= \langle b(y), \nabla_y g(y) \rangle + \int_{\{|z|>m\}} (g(y+z) - g(y)) \nu(dz) \\
&+ \int_{\{|z|\leq m\}} \left(g(x+z) - g(y) - \langle \nabla_y g(y), z+x-y \rangle \right) \\
&\quad \times q(z) \wedge q(z+x-y) \mathbf{1}_{\{|z+x-y|\leq m\}} dz \\
&+ \int_{\{|z|\leq m\}} \left(g(y+R(x, y)z) - g(y) - \langle \nabla_y g(y), R(x, y)z \rangle \right) \\
&\quad \times (\nu(dz) - q(z) \wedge q(z+x-y) \mathbf{1}_{\{|z+x-y|\leq m\}} dz) ,
\end{aligned} \tag{3.2.19}$$

where we used the fact that $\rho(z, x-y) \nu(dz) = q(z) \wedge q(z+x-y) \mathbf{1}_{\{|z+x-y|\leq m\}} dz$, cf. (3.2.13). Now we make the substitution $u = z+x-y$ in order to write the second integral in (3.2.19) as

$$\int_{\{|u|\leq m\} \cap \{|u-(x-y)|\leq m\}} \left(g(y+u) - g(y) - \langle \nabla_y g(y), u \rangle \right) q(u-(x-y)) \wedge q(u) du .$$

Then observe that due to rotational invariance of ν we have

$$\begin{aligned}
&\int_{\{|z|\leq m\}} \left(g(y+R(x, y)z) - g(y) - \langle \nabla_y g(y), R(x, y)z \rangle \right) \nu(dz) \\
&= \int_{\{|z|\leq m\}} \left(g(y+z) - g(y) - \langle \nabla_y g(y), z \rangle \right) \nu(dz) .
\end{aligned}$$

Moreover, from the definition of R we have $|z+x-y| = |R(x, y)z - x+y|$ and hence,

from the rotational invariance of q , we obtain

$$\begin{aligned} & \int_{\{|z|\leq m\}} \left(g(y + R(x, y)z) - g(y) - \langle \nabla_y g(y), R(x, y)z \rangle \right) \\ & \times q(z) \wedge q(z + x - y) \mathbf{1}_{\{|z+x-y|\leq m\}} dz \\ & = \int_{\{|z|\leq m\}} \left(g(y + R(x, y)z) - g(y) - \langle \nabla_y g(y), R(x, y)z \rangle \right) \\ & \times q(R(x, y)z) \wedge q(R(x, y)z - x + y) \mathbf{1}_{\{|R(x, y)z-x+y|\leq m\}} dz. \end{aligned}$$

Now we can substitute $u = R(x, y)z$ and we see that the last integral above equals

$$\int_{\{|u|\leq m\} \cap \{|u-(x-y)|\leq m\}} \left(g(y + u) - g(y) - \langle \nabla_y g(y), u \rangle \right) q(u) \wedge q(u - x + y) du.$$

Combining all our calculations, we conclude that

$$\mathcal{L}g(y) = Ag(y)$$

for any $g \in C_c^2(\mathbb{R}^d)$, with A given by (3.2.18), which finishes the proof. Note that the method presented above allows us to omit the assumption about absolute continuity of the Lévy measure ν by replacing the function $\rho(z, x - y)$ defined via (3.2.13) with the truncated Radon-Nikodym derivative

$$\rho(z, x - y) = \frac{\nu \wedge (\delta_{x-y} * \nu)(dz)}{\nu(dz)} \mathbf{1}_{\{|z+x-y|\leq m\}}. \quad (3.2.20)$$

For applications of this coupling presented in [Maj15], under the absolute continuity assumption and with ρ defined by (3.2.13), we need an additional assumption guaranteeing that after translating q by some vector, the overlap of the translated q and the q at its original position will have a positive mass. More precisely, we will require that there exist constants $m > 0$ and $\delta > 0$ such that $\delta < 2m$ and

$$\inf_{x \in \mathbb{R}^d: 0 < |x| \leq \delta} \int_{\{|v|\leq m\} \cap \{|v+x|\leq m\}} q(v) \wedge q(v + x) dv > 0. \quad (3.2.21)$$

This not only guarantees that our coupling does not just reduce to the coupling by reflection presented in Section 3.2.1, but also allows us to study convergence rates in appropriately chosen Kantorovich distances, see the discussion in Section 3.3.

With ρ redefined by (3.2.20), the assumption (3.2.21) about sufficient overlap of the density q and its translation becomes an assumption about non-triviality of the measure $\nu \wedge (\delta_{x-y} * \nu)$ under some conditions on x and y . Bearing in mind that such an extension is possible, the author claims that the approach presented in [Maj15] has an advantage in being more straightforward and intuitive, as it does not require any tools from the martingale problem theory and it does not rely on the results from [Kur11]. The assumption about absolute continuity of ν is also not restrictive, as it is satisfied by most cases encountered in applications. Very singular jump measures such as combinations of Dirac masses are excluded anyway on the account of assumption (3.2.21) (i.e., the measure $\nu \wedge (\delta_{x-y} * \nu)(dz)$ becomes trivial).

3.3 Applications of couplings to ergodicity

The coupling we introduced in Section 3.2.2 allows us to obtain convergence rates to equilibrium for solutions to SDEs of the form (3.2.1) in both the standard L^1 -Wasserstein and total variation distances.

First we obtain exponential contractivity for the transition semigroup $(p_t)_{t \geq 0}$ of the solution $(X_t)_{t \geq 0}$ to (3.2.1) in a Kantorovich distance W_f associated with some specially constructed concave function f , i.e., we have

$$W_f(\mu_1 p_t, \mu_2 p_t) \leq e^{-ct} W_f(\mu_1, \mu_2) \quad (3.3.1)$$

for all $t > 0$ and all probability measures μ_1 and μ_2 on \mathbb{R}^d . In the Lévy jump case we can work with two types of distance functions f . We can either choose a continuous, increasing, concave function f_1 similar to the one used in the diffusion case by Eberle in [Ebe16] (see Section 3.1.1) or a discontinuous function of the form

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

with some constant $a > 0$.

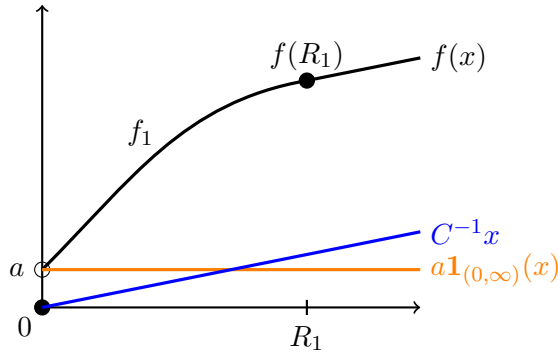


Figure 3.10: The choice of f in [Maj15].

The choice of the approach depends on the assumptions satisfied by the noise $(L_t)_{t \geq 0}$ in our equation. In principle, obtaining (3.3.1) for a continuous f in the Lévy jump case requires stronger assumptions on the Lévy measure ν than getting the same results with a discontinuous f . Both approaches have their advantages and disadvantages. Since a discontinuous function f can be compared from below with both $\mathbf{1}_{(0, \infty)}$ and the identity function, the corresponding Kantorovich distance gives us an upper bound on both the standard L^1 -Wasserstein and the total variation distances. Hence we get

$$\begin{aligned} W_1(\mu_1 p_t, \mu_2 p_t) &\leq C(\mu_1, \mu_2) e^{-ct}, \\ \|\mu_1 p_t - \mu_2 p_t\|_{TV} &\leq C(\mu_1, \mu_2) e^{-ct} \end{aligned} \quad (3.3.2)$$

for all $t > 0$, with some constants $c, C = C(\mu_1, \mu_2) > 0$, for any probability measures μ_1, μ_2 on \mathbb{R}^d . However, in this case we cannot get upper bounds in the L^1 -Wasserstein

3.3 Applications of couplings to ergodicity

distance, since f cannot be bounded from above by a rescaled identity function. On the other hand, if we obtain (3.3.1) with a continuous function f , we can then compare the corresponding Kantorovich distance W_f with W_1 both from above and below, and thus we get

$$W_1(\mu_1 p_t, \mu_2 p_t) \leq C e^{-ct} W_1(\mu_1, \mu_2), \quad (3.3.3)$$

similarly to the (3.1.13) in the diffusion case. In this case, however, we obviously do not get upper bounds for the total variation distance. See the discussion in Remark 1.6 in [Maj15] for more details.

For the construction of our coupling, as presented in Section 3.2.2, we need the following assumptions.

Assumption 1. *The Lévy measure ν is rotationally invariant, i.e.,*

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

for any Borel set $B \in \mathcal{B}(\mathbb{R}^d)$ and any $d \times d$ 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 continuous almost everywhere on \mathbb{R}^d .*

This is sufficient to prove that the system of equations defined by (3.2.15) and (3.2.16) has a solution and that this solution is a coupling of two copies of the solution to (3.2.15), see Theorem 1.1 in [Maj15]. Actually, as we indicated in Section 3.2.3, it is possible to remove the assumption about absolute continuity of ν by modifying the coupling construction accordingly.

If we want to obtain bounds (3.3.2), we additionally need the following two assumptions on the Lévy measure.

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

$$\inf_{x \in \mathbb{R}^d: 0 < |x| \leq \delta} \int_{\{|v| \leq m\} \cap \{|v+x| \leq m\}} q(v) \wedge q(v+x) dv > 0. \quad (3.3.4)$$

Assumption 4. *There exists a constant $\varepsilon > 0$ such that $\varepsilon \leq \delta$ (with δ defined by (3.3.4) above) and*

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

Assumptions 1-4 are satisfied by a large class of rotationally invariant Lévy processes, such as all symmetric α -stable processes for $\alpha \in (0, 2)$, many compound Poisson processes and even some processes with Lévy measures with supports separated from zero (which shows the real strength of our coupling, as usually in the literature similar results are obtained under assumptions of high concentration of the Lévy measure around zero, cf. the discussion near the end of Section 1 in [Maj15] and Example 1.7 therein).

We also need the dissipativity at infinity assumption on the drift, as discussed in Section 3.1.1.

On the other hand, if we would like to obtain (3.3.3), we should replace Assumptions 3 and 4 with the following condition.

3 The coupling method

Assumption 5.

$$\varepsilon / \left(\int_0^\varepsilon |y|^2 \nu_1(dy) \right) \text{ is bounded as } \varepsilon \rightarrow 0,$$

where ν_1 is the first marginal of the rotationally invariant measure ν , i.e., $\nu_1(A) := \nu(A \times \mathbb{R}^{d-1})$ for $A \in \mathcal{B}(\mathbb{R})$.

This assumption, on the other hand, is satisfied (together with Assumptions 1 and 2) by symmetric α -stable processes only when $\alpha \in [1, 2)$. Generally speaking, in order to obtain estimates of the form (3.3.3) we need a noise in our SDE which exhibits diffusion-like behaviour, see also Remark 1.6 in [Maj15]. One additional technical assumption which is required for (3.3.3) is that if we put

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

then we have

$$\lim_{r \rightarrow 0} r \kappa(r) = 0.$$

This is, however, satisfied whenever the coefficients satisfy a one-sided Lipschitz condition and the drift b is continuous.

The discontinuous case was treated in [Maj15] (Theorem 1.1 and Corollary 1.2 therein), while the continuous case was proved in Theorem 3.1 in [Maj16]. It is worth pointing out that in both cases as a corollary we can prove existence of a unique invariant measure for the equation (3.2.1) and thus we obtain exponential rate of convergence of the distributions of its solution to equilibrium. In the continuous case this is an almost immediate consequence of completeness of the space of probability measures with finite first moments equipped with the L^1 -Wasserstein metric (see Theorem 6.18 in [Vil09] for this result and Corollary 3 in [Ebe16] or the beginning of Section 3 in [KW12] for how it implies existence of a unique invariant measure by an application of the Banach fixed point theorem). In the discontinuous case the matter is slightly more complicated, but we still obtain an invariant measure if we additionally assume that the transition semigroup $(p_t)_{t \geq 0}$ associated with the solution to our SDE preserves finite first moments of measures, i.e., if μ has a finite first moment, then μp_t also does (cf. Corollary 1.8 in [Maj15]).

In [Maj16] some extensions of these results were presented in the context of equations of the form

$$dX_t = b(X_t)dt + dB_t^1 + \sigma(X_t)dB_t^2 + dL_t + \int_U g(X_{t-}, u) \tilde{N}(dt, du),$$

where $(L_t)_{t \geq 0}$ is a Lévy process as above, $(B_t^1)_{t \geq 0}$ and $(B_t^2)_{t \geq 0}$ are Brownian motions, $\tilde{N}(dt, du) = N(dt, du) - dt \nu(du)$ is a compensated Poisson random measure on $\mathbb{R}_+ \times U$, all the sources of noise are independent and the coefficients $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ and $g : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$ satisfy a dissipativity at infinity condition, i.e., there exist constants

$K, R > 0$ such that

$$\begin{aligned} \langle b(x) - b(y), x - y \rangle + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \\ \leq -K|x - y|^2, \end{aligned} \quad (3.3.5)$$

for all $x, y \in \mathbb{R}^d$ with $|x - y| > R$. In such a case we can apply the coupling from [Maj15] to $(L_t)_{t \geq 0}$ (let us denote it by an operator $M(\cdot, \cdot)$), the reflection coupling from [LR86] to $(B_t^1)_{t \geq 0}$ and the synchronous coupling to the other two noises. Hence we have

$$\begin{aligned} dY_t = b(Y_t)dt + R(X_t, Y_t)dB_t^1 + \sigma(Y_t)dB_t^2 \\ + M(X_{t-}, Y_{t-})dL_t + \int_U g(Y_{t-}, u)\tilde{N}(dt, du). \end{aligned}$$

This allows us to get

$$W_f(\mu_{pt}, \nu_{pt}) \leq e^{-ct}W_f(\mu, \nu)$$

for either a continuous or a discontinuous function f , depending on which component of the noise we use in our calculations, see Section 4 in [Maj16] for details.

3.4 Applications of couplings to Malliavin calculus

Consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ with an m -dimensional Brownian motion $(W_t)_{t \geq 0}$ and a Poisson random measure N on $\mathbb{R}_+ \times U$ (where (U, \mathcal{U}) is a measure space), both adapted to $(\mathcal{F}_t)_{t \geq 0}$. We will now define the Malliavin derivative for a certain class of measurable functionals F with respect to the process $(W_t)_{t \geq 0}$, as well as the Malliavin derivative of F with respect to N . Consider the family \mathcal{S} of smooth functionals of $(W_t)_{t \geq 0}$ of the form

$$F = f(W(h_1), \dots, W(h_n)) \text{ for } n \geq 1, \quad (3.4.1)$$

where $W(h) = \int_0^T h(s)dW_s$ for $h \in H = L^2([0, T]; \mathbb{R}^m)$ and $f \in C^\infty(\mathbb{R}^n)$. We define the Malliavin derivative of F with respect to $(W_t)_{t \geq 0}$ as the unique element ∇F in $L^2(\Omega; H) \simeq L^2(\Omega \times [0, T]; \mathbb{R}^m)$ such that for any $h \in H$ we have

$$\langle \nabla F, h \rangle_{L^2([0, T]; \mathbb{R}^m)} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left(F(W. + \int_0^\cdot h_s ds) - F(W.) \right), \quad (3.4.2)$$

where the convergence holds in $L^2(\Omega)$ (see e.g. Definition A.10 in [DØP09]). This can be thought of as a Fréchet derivative of F in directions from the so-called Cameron-Martin space H_{CM} , i.e.,

$$H_{CM} = \left\{ \bar{h} \in H : \exists h \in H \text{ s.t. } \bar{h}(t) = \int_0^t h(s)ds \right\},$$

3 The coupling method

cf. Appendix A in [DØP09]. An equivalent definition states that for $F \in \mathcal{S}$ of the form (3.4.1) we can put

$$\nabla_t F = \sum_{i=1}^n \partial_i f(W(h_1), \dots, W(h_n)) h_i(t). \quad (3.4.3)$$

Then the definition can be extended to all random variables F in the space $\mathbb{D}^{1,2}$ which is the completion of \mathcal{S} in $L^2(\Omega)$ with respect to the norm

$$\|F\|_{\mathbb{D}^{1,2}}^2 := \|F\|_{L^2(\Omega)}^2 + \|\nabla F\|_{L^2(\Omega; H)}^2.$$

The other approach leads through so-called chaos expansions. It turns out that every square integrable, \mathcal{F}_T -measurable (where $(W_t)_{t \geq 0}$ is adapted to $(\mathcal{F}_t)_{t \geq 0}$) random variable F can be represented as

$$F = \sum_{n=0}^{\infty} I_n(f_n),$$

where I_n are n -fold iterated Itô integrals of symmetric square integrable deterministic functions $f_n \in L^2([0, T]^n)$ with respect to $(W_t)_{t \geq 0}$, see e.g. Chapter 1 in [DØP09]. Then it turns out that $F \in \mathbb{D}^{1,2}$ if and only if

$$\sum_{n=1}^{\infty} n n! \|f_n\|_{L^2([0, T]^n)}^2 < \infty,$$

see e.g. Theorem A.22 in [DØP09], and for such F we can define

$$\nabla_t F = \sum_{n=1}^{\infty} n I_{n-1}(f_n(\cdot, t)),$$

where $I_{n-1}(f_n(\cdot, t))$ means that we integrate f_n treating it as a function of $n-1$ variables with a parameter t . It turns out that for $F \in \mathbb{D}^{1,2}$ this approach gives us the same operator as the one defined by (3.4.2) or (3.4.3), cf. e.g. Theorem A.22 in [DØP09].

However, for the Poisson case, the approach through the chaos expansion and the one using a differential operator are non-equivalent.

The original approach to the Malliavin calculus for jump processes involved introducing a differential operator similar to (3.4.2) for Poisson random measures. Namely, consider a predictable random field

$$V : [0, \infty) \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^d.$$

The Poisson random measure N perturbed by V is denoted by N^V and is formally defined by setting

$$\int_0^t \int_{\mathbb{R}^d} \psi(s, z) N^V(ds, dz) = \int_0^t \int_{\mathbb{R}^d} \psi(s, z + V(s, z)) N(ds, dz)$$

for an appropriate class of test functions ψ . In other words, if

$$N = \sum \delta_{(\tau_j, \xi_j)},$$

then

$$N^V = \sum \delta_{(\tau_j, \xi_j + V(\tau_j, \xi_j))}.$$

Let us now fix $p \in [1, \infty)$. We say that a functional $F = F(N)$ has an L^p -derivative in the direction of V if there is an L^p -integrable random variable $D_V F$ such that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left| \frac{F(N^{\varepsilon V}) - F(N)}{\varepsilon} - D_V F \right|^p = 0.$$

This approach is useful for proving existence and regularity of densities of some functionals of Lévy processes, as well as obtaining some gradient estimates for their transition semigroups, see e.g. [Bis83], [BC86], [BGJ87] or [Nor88].

However, here we are interested in another approach, in which the Malliavin derivative with respect to a Poisson random measure is not defined as a differential operator, but as a difference operator.

Let $N = \sum_{j=1}^{\infty} \delta_{(\tau_j, \xi_j)}$. Then for any random variable F we define

$$D_{t,u} F := F(N + \delta_{(t,u)}) - F(N),$$

which we call the Malliavin derivative of F with respect to N .

It turns out that in an analogous way to the Brownian case, if N is adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$, we can represent square integrable, \mathcal{F}_T -measurable random variables F as

$$F = \sum_{n=0}^{\infty} I_n(f_n), \tag{3.4.4}$$

where this time I_n are n -fold iterated integrals with respect to the compensated Poisson random measure \tilde{N} , of symmetric functions f_n on $[0, T] \times U$, which are square integrable with respect to the product of the Lebesgue measure on $[0, T]$ and the jump measure ν on U , see e.g. Chapter 10 in [DØP09] or Theorem 1.3 in [LP11]. Similarly to the Brownian case we can define

$$D_{t,u} F := \sum_{n=1}^{\infty} n I_{n-1}(f_n(\cdot, t, u))$$

for $F \in \mathbb{D}^{1,2}$, with $\mathbb{D}^{1,2}$ defined as the set of random variables F having representation (3.4.4) such that

$$\sum_{n=1}^{\infty} n n! \|f_n\|_{L^2([0,T] \times (U;\nu))^n}^2 < \infty.$$

It turns out that for such random variables the two approaches to defining $D_{t,u}$ are equivalent, see e.g. Section 4 of [Løk04] or Theorem 3.3 in [LP11]. The importance of the approach via the chaos expansion comes from the fact that it allows us to prove a Clark-Ocone formula, see e.g. Theorem 7 in [Løk04] for the pure jump case or Section 12.5 in [DØP09] and the references therein for how to combine the Gaussian and the jump case. Based on Theorem 12.20 in [DØP09], we have

3 The coupling method

Theorem 3.4.1. *Let $F \in \mathbb{D}^{1,2}$. Then*

$$F = \mathbb{E}F + \int_0^T \mathbb{E} [\nabla_t F | \mathcal{F}_t] dW_t + \int_0^T \int_U \mathbb{E} [D_{t,u} F | \mathcal{F}_t] \tilde{N}(dt, du). \quad (3.4.5)$$

See Section 12.5 in [DØP09] for details of how to rigorously define the space $\mathbb{D}^{1,2}$ in the combined Wiener-Poisson setting. What is important for us in the sequel is that solutions to SDEs of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du) \quad (3.4.6)$$

belong to $\mathbb{D}^{1,2}$ under standard Lipschitz and linear growth conditions on the coefficients, see Theorem 17.4 in [DØP09].

Note that the approach to the Malliavin calculus for jump processes via a difference operator, as opposed to the differential operator introduced in [Bis83], traces back to the paper [NV90] by Nualart and Vives and was later extended by Picard in [Pic96] and [Pic96b]. For a detailed discussion on differences between these two approaches, see the book [Ish13] by Ishikawa, specifically Sections 2.1.1 and 2.1.2 where they are called Bismut's method and Picard's method, respectively.

Under a global dissipativity condition on the coefficients of (3.4.6), i.e., the condition (3.3.5) holding with some constant $K > 0$ for all $x, y \in \mathbb{R}^d$, the following bounds for the Malliavin derivatives D and ∇ of solutions to (3.4.6) were obtained in [Wu10] and [Ma10], respectively. Namely, we have

$$\mathbb{E} [|D_{t,u} f(X_T)| | \mathcal{F}_t] \leq e^{-K(T-t)} |g(X_{t-}, u)|$$

and

$$\mathbb{E} [\|\nabla_t X_T\|_{HS}^2 | \mathcal{F}_t] \leq e^{-2K(T-t)} \|\sigma(X_t)\|_{HS}^2,$$

where $K > 0$ is the constant with which the global dissipativity condition holds, $T > t > 0$ and $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is a Lipschitz function with $\|f\|_{\text{Lip}} \leq 1$, see page 476 in [Wu10] and Lemma 3.4 in [Ma10].

Using the coupling method, similar estimates were obtained in [Maj16] under the dissipativity at infinity condition. The estimates in [Maj16] are of exactly the same form as in [Wu10] for $D_{t,u}$ and of slightly weaker form for ∇_t , see Section 5.2 and Corollary 2.15 in [Maj16], respectively.

If we can apply the Clark-Ocone formula (3.4.5) to a certain class of functionals of solutions to (3.4.6), estimates of such type allow us to obtain some information on the behaviour of these solutions. Following this idea, in [Wu10] and [Ma10] these estimates were used to obtain transportation inequalities for such SDEs under the global dissipativity condition, while in [Maj16] these results were improved to the case of dissipativity at infinity. This will be the topic of the next section. However, the author believes that the technique of using couplings in order to obtain such kind of estimates on Malliavin derivatives may also find other applications in the future.

3.5 Applications of couplings to transportation inequalities

Transportation cost-information inequalities (often in the literature shortly called transportation inequalities) compare the transportation cost between two measures (understood as the Wasserstein distance between them) with their Kullback-Leibler information (relative entropy). For probability measures μ_1 and μ_2 on a metric space (E, ρ) , the latter is defined as

$$H(\mu_1|\mu_2) := \begin{cases} \int \log \frac{d\mu_1}{d\mu_2} d\mu_1 & \text{if } \mu_1 \ll \mu_2, \\ +\infty & \text{otherwise.} \end{cases}$$

We say that a probability measure μ satisfies an L^p -transportation cost-information inequality on (E, ρ) if there is a constant $C > 0$ such that for any probability measure η we have

$$W_{p,\rho}(\eta, \mu) \leq \sqrt{2CH(\eta|\mu)}.$$

Then we write $\mu \in T_p(C)$.

The most important cases are $p = 1$ and $p = 2$. Since $W_{1,\rho} \leq W_{2,\rho}$, we see that the L^2 -transportation inequality (the T_2 inequality) implies T_1 , and it is well known that in fact T_2 is much stronger (see e.g. the discussion in Section 1 of [GL10]). The T_1 inequality is related to the phenomenon of measure concentration (see the discussion below in the context of so called α - W_1H inequalities), whereas T_2 has many interesting connections with other functional inequalities.

Consider the log-Sobolev inequality, which holds for the measure μ with some constant $K > 0$ if for any strictly positive function $f \in C_b^2(\mathbb{R}^d)$ we have

$$\mu(f \log f) - \mu(f) \log \mu(f) \leq \frac{1}{2K} \mu \left(\frac{|\nabla f|^2}{f} \right). \quad (3.5.1)$$

Due to Otto and Villani [OV00], we know that the log-Sobolev inequality implies the following T_2 inequality (also known as the Talagrand inequality)

$$W_2(\eta, \mu)^2 \leq \frac{2}{K} H(\eta|\mu) \quad (3.5.2)$$

for any probability measure η on \mathbb{R}^d . Furthermore, they showed that the Talagrand inequality implies the Poincaré inequality

$$\text{Var}_\mu(f) \leq \frac{1}{K} \mu(|\nabla f|^2),$$

where $\text{Var}_\mu(f) := \mu(f^2) - \mu(f)^2$.

As for T_1 inequalities, let us consider their generalization known as α - W_1H inequalities. Namely, let α be a non-decreasing, left continuous function on \mathbb{R}_+ with $\alpha(0) = 0$. We say that a probability measure μ satisfies a W_1H -inequality with deviation function α (or simply α - W_1H inequality) if for any probability measure η we have

$$\alpha(W_{1,\rho}(\eta, \mu)) \leq H(\eta|\mu). \quad (3.5.3)$$

3 The coupling method

In order to better understand the meaning of (3.5.3), let us have a look at a result due to Gozlan and Léonard (see Theorem 2 in [GL07] for the original result and Lemma 2.1 in [Wu10] for the reformulation we use below, see also Theorem 22.10 in [Vil09]) which generalizes a result by Bobkov and Götze (Theorem 3.1 in [BG99]).

Fix a probability measure μ on (E, ρ) and a convex deviation function α . Then the following properties are equivalent:

1. the α - W_1H inequality for the measure μ holds, i.e., for any probability measure η on (E, ρ) we have

$$\alpha(W_{1,\rho}(\eta, \mu)) \leq H(\eta|\mu),$$

2. for every $f : E \rightarrow \mathbb{R}$ bounded and Lipschitz with $\|f\|_{\text{Lip}} \leq 1$ we have

$$\int e^{\lambda(f-\mu(f))} d\mu \leq e^{\alpha^*(\lambda)} \text{ for any } \lambda > 0, \quad (3.5.4)$$

where $\alpha^*(\lambda) := \sup_{r \geq 0} (r\lambda - \alpha(r))$ is the convex conjugate of α ,

3. if $(\xi_k)_{k \geq 1}$ is a sequence of i.i.d random variables with common law μ , then for every $f : E \rightarrow \mathbb{R}$ bounded and Lipschitz with $\|f\|_{\text{Lip}} \leq 1$ we have

$$\mathbb{P} \left(\frac{1}{n} \sum_{k=1}^n f(\xi_k) - \mu(f) > r \right) \leq e^{-n\alpha(r)} \text{ for any } r > 0, n \geq 1. \quad (3.5.5)$$

This gives an intuitive interpretation of α - W_1H in terms of the concentration of measure property (3.5.5), while the second characterization (3.5.4) is very useful for proving such inequalities, see e.g. [Wu10], [Ma10] and [Maj16]. The third characterization (3.5.5) can be interpreted as a bound on the error of Monte Carlo estimation of the integral $\mu(f)$.

It is worth pointing out that the third characterization above, i.e., the formula (3.5.5) is indeed supposed to hold without taking absolute value of the difference $\frac{1}{n} \sum_{k=1}^n f(\xi_k) - \mu(f)$. The proof of (3.5.5) follows from (3.5.4) via an application of the exponential Chebyshev inequality, i.e., a simple result which states that for any real-valued random variable X (not necessarily non-negative) we have

$$\mathbb{P}(X > r) \leq e^{-tr} \mathbb{E} e^{tX}$$

for any $t > 0$ and any $r > 0$. See the proof of Theorem 22.10 in [Vil09] for details. However, we can always replace the function f appearing in (3.5.5) with $-f$ and since the right hand side of (3.5.5) does not depend on f , we see that (3.5.5) implies

$$\begin{aligned} \mathbb{P} \left(\left| \frac{1}{n} \sum_{k=1}^n f(\xi_k) - \mu(f) \right| > r \right) &= \mathbb{P} \left(\frac{1}{n} \sum_{k=1}^n f(\xi_k) - \mu(f) > r \right) \\ &\quad + \mathbb{P} \left(-\frac{1}{n} \sum_{k=1}^n f(\xi_k) + \mu(f) > r \right) \\ &\leq 2e^{-n\alpha(r)}. \end{aligned}$$

3.5 Applications of couplings to transportation inequalities

For a general survey of transportation inequalities the reader may consult [GL10] or Chapter 22 of [Vil09].

Our goal in this section is to discuss application of the coupling method to obtain α - W_1H inequalities for solutions of equations of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du). \quad (3.5.6)$$

As an example of a simple equation of the type (3.5.6), consider

$$dX_t = b(X_t)dt + \sqrt{2}dW_t$$

with a d -dimensional Brownian motion $(W_t)_{t \geq 0}$. If the global dissipativity assumption is satisfied, i.e., if there exists $K > 0$ such that

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

then $(X_t)_{t \geq 0}$ has an invariant measure μ and by a result of Bakry and Émery [BE85], μ satisfies the log-Sobolev inequality (3.5.1) with the constant $K > 0$ and thus (by Otto and Villani [OV00]) also the Talagrand $T_2(1/K)$ inequality (3.5.2). More generally, for equations of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

under the global dissipativity assumption, Djellout, Guillin and Wu in [DGW04] showed $T_2(C_1)$ with some modified constant $C_1 > 0$.

For equations of the form

$$dX_t = b(X_t)dt + \int_U g(X_{t-}, u)\tilde{N}(dt, du) \quad (3.5.7)$$

in general the Poincaré inequality does not hold (see Example 1.1 in [Wu10]) and thus we cannot prove the T_2 inequality. However, under the global dissipativity assumption, Wu in [Wu10] showed some α - W_1H inequalities.

The method of proof is based on the bounds on Malliavin derivatives that we presented in Section 3.4. Roughly speaking, using the Clark-Ocone formula for an \mathcal{F}_T -measurable random variable F we can show that if there exists a deterministic function $h : [0, T] \times U \rightarrow \mathbb{R}$ such that $\int_0^T \int_U h(t, u)^2 \nu(du) dt < \infty$ and

$$\mathbb{E}[D_{t,u}F | \mathcal{F}_t] \leq h(t, u)$$

for all $t \in [0, T]$, then for any C^2 convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ such that ϕ' is also convex, we have

$$\mathbb{E}\phi(F - \mathbb{E}F) \leq \mathbb{E}\phi\left(\int_0^T \int_U h(t, u)\tilde{N}(dt, du)\right).$$

Applying this to $\phi(x) = \exp(\lambda x)$ for some $\lambda > 0$ and using the Gozlan-Léonard characterization (3.5.4) we are able to show an α - W_1H inequality for the distribution of F . This result can be in particular applied to $F = X_T$ for any $T > 0$, with $(X_t)_{t \geq 0}$ being

3 The coupling method

a solution to (3.5.7). The method from [Wu10] was subsequently extended by Ma in [Ma10] to include equations of the jump diffusion type (3.5.6). Note that both [Wu10] and [Ma10] require the coefficients in the equation to satisfy a global dissipativity assumption. In [Maj16] it was shown how to use the coupling constructed in [Maj15] for the jump part and the coupling by reflection from [Ebe16] for the Brownian part in order to extend the results from [Wu10] and [Ma10] to the dissipativity at infinity case (see Corollary 2.9 in [Maj16]). However, [Maj16] also presented a more general framework that allows for proving α - W_1H inequalities in other cases in which there exist couplings $(X_t, Y_t)_{t \geq 0}$ of solutions to equations

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du)$$

such that

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq c(t)|x - y|$$

for all $t > 0$, some function $c : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and all initial conditions $x, y \in \mathbb{R}^d$, see Theorem 2.1 in [Maj16]. The point is that we do not necessarily need to work within the framework of [Maj15] and [Ebe16] and we may make use of other coupling constructions available in the literature to obtain transportation inequalities under different sets of assumptions, cf. Remark 2.13 in [Maj16].

Note that T_2 inequalities for equations

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{3.5.8}$$

under a global dissipativity condition were obtained by Djellout, Guillin and Wu in [DGW04] using a method based on the Girsanov theorem, see Theorem 5.6 in [DGW04] and condition (4.5) therein. From this a T_1 inequality follows immediately, as we indicated earlier in this section. The method of proof from [DGW04] does not seem to work for equations of the type

$$dX_t = b(X_t)dt + \int_U g(X_{t-}, u)\tilde{N}(dt, du)$$

and that is why Wu in [Wu10] developed a new method based on the Malliavin calculus to get α - W_1H inequalities in such a case. He covered only the pure jump case, but Ma in [Ma10] showed that similar approach still works in the jump diffusion case, but of course by taking $g = 0$ her result can be also applied to a diffusion of the form (3.5.8). This way we get back to the setting of [DGW04], getting a new method of obtaining a kind of W_1 transportation inequalities for diffusions (but not necessarily with the quadratic deviation function as in [DGW04]). Recall however again that all the papers [DGW04], [Wu10] and [Ma10] treat only the global dissipativity case. Thus the extension of the results from [Ma10] presented in [Maj16] is the first (as far as we know) successful attempt at obtaining W_1 transportation inequalities for diffusions in the non-globally dissipative case and hence can also be seen as an extension of [DGW04].

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

COUPLING AND EXPONENTIAL ERGODICITY FOR STOCHASTIC DIFFERENTIAL EQUATIONS DRIVEN BY LÉVY PROCESSES

MATEUSZ B. MAJKA

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.

1. INTRODUCTION

We consider stochastic differential equations of the form

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

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

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

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*) $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$ 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

2010 *Mathematics Subject Classification.* 60G51, 60H10.

Key words and phrases. Stochastic differential equations, Lévy processes, exponential ergodicity, couplings, Wasserstein distances.

we know that there exists a Poisson random measure N associated with $(L_t)_{t \geq 0}$ in such a way that

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

where

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

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 $f : [0, \infty) \rightarrow [0, \infty)$ with $f(0) = 0$ and $f(x) > 0$ for $x > 0$ as

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

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

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

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] and [22] for couplings of pure jump Lévy processes, [23], [28] and [29] for the case of Lévy-driven Ornstein-Uhlenbeck processes and [12], [31] and [25] for more general Lévy-driven SDEs with non-linear drift. See also [11] and [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 Subsection 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

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

Assumption 4. There exists a constant $\varepsilon > 0$ such that $\varepsilon \leq \delta$ (with δ defined via (1.4) above) 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 strong solution to (1.1). Now we define the function $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}$ by setting $\kappa(|x - y|)$ to be the largest quantity such that

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

and therefore it has to be defined as

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

We have the following assumption.

Assumption 5. κ is a continuous function satisfying

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

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

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

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*

$$(1.7) \quad dX_t = b(X_t)dt + dL_t,$$

where $(L_t)_{t \geq 0}$ is a pure jump Lévy process with the Lévy measure ν satisfying Assumptions 1 and 2, whereas $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ 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

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

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*

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

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

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

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

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

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

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

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

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

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

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 $\tilde{C}_\delta > 0$ 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 $t \geq 0$ and any probability measures μ_1 and μ_2 on \mathbb{R}^d we have*

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

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

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

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

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

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

$$(1.15) \quad W_1(\mu_1 p_t, \mu_2 p_t) \leq Le^{-ct}W_1(\mu_1, \mu_2)$$

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_1H 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 spoiled 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 below 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 $\nu(dx) = (\mathbf{1}_{[-\theta, -\theta/\beta]}(x) + \mathbf{1}_{[\theta/\beta, \theta]}(x)) dx$ 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*

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

and therefore

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

and

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

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., the equation (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 $c : \mathbb{R}^2 \rightarrow [0, \infty]$. 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

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

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 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

$$(2.2) \quad dX_t = b(X_t)dt + dB_t$$

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

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

where

$$(2.4) \quad e_t := \frac{X_t - Y_t}{|X_t - Y_t|}.$$

Of course, the equation (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 the equations (2.2) and (2.3) together define a coupling, i.e., the solution $(Y_t)_{t \geq 0}$ to the equation (2.3) has the same finite dimensional distributions as the solution $(X_t)_{t \geq 0}$ to the equation (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] and [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

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

(see e.g. [18], Chapter 6). At the jump time τ_j the process $(X_t)_{t \geq 0}$ jumps from the point X_{τ_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

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

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

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

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 Lemma 3.3 and Lemma 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

$$\begin{aligned} L_t &= \int_0^t \int_{\{|v| > m\} \times [0,1]} v N(ds, dv, du) + \int_0^t \int_{\{|v| \leq m\} \times [0,1]} v \tilde{N}(ds, dv, du) \\ &\quad + \int_0^t \int_{\{m \geq |v| > 1\} \times [0,1]} v \nu(dv) dud s. \end{aligned}$$

Then we can include the last term appearing above in the drift b in the equation (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

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

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

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

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

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

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

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

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

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

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

$$(2.11) \quad T := \inf\{t \geq 0 : X_t = Y_t\},$$

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

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

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

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*

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

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

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

and, since R is an isometry, we can estimate

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

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

$$2 \int_{\{|v| \leq m\}} |z + v|^2 \rho(v, z) \nu(dv) \leq 2 \int_{\{|v| \leq m\}} |z + v|^2 q(v + z) \wedge q(v) dv \\ = 2 \int_{\{|v-z| \leq m\}} |v|^2 q(v) \wedge q(v - z) dv.$$

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

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

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

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

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

$$2 \int_{\{|v-z| \leq m\}} |v|^2 q(v) \wedge q(v - z) dv \\ \leq 4 \int_{\{|v-z| \leq m\}} |v - z|^2 q(v) \wedge q(v - z) dv + 4 \int_{\{|v-z| \leq m\}} |z|^2 q(v) \wedge q(v - z) dv \\ \leq 4 \int_{\{|v-z| \leq m\}} |v - z|^2 q(v - z) dv + 4 \int_{\{|v-z| \leq m\}} |z|^2 q(v) dv \\ \leq 4 \int_{\{|v| \leq m\}} |v|^2 \nu(dv) + 4|z|^2 \nu(B(m)^c).$$

Hence, by choosing

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

we get the desired result. \square

Here we should remark that by the above lemma we have

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

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

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

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

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

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

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

Proof. We have

$$\begin{aligned} &\int_{\{|v| \leq m\} \times [0,1]} |(x+h-y+v - R(x+h, y)v) \mathbf{1}_{\{u < \rho(v, z+h)\}} \\ &\quad - (x-y+v - R(x, y)v) \mathbf{1}_{\{u < \rho(v, z)\}}|^2 \nu(dv) du \\ &= \int_{\{|v| \leq m\} \times [0,1]} |(x+h-y+v - R(x+h, y)v) \mathbf{1}_{\{u < \rho(v, z+h)\}} \\ &\quad - (x-y+v - R(x, y)v) \mathbf{1}_{\{u < \rho(v, z+h)\}} \\ &\quad + (x-y+v - R(x, y)v) \mathbf{1}_{\{u < \rho(v, z+h)\}} \\ &\quad - (x-y+v - R(x, y)v) \mathbf{1}_{\{u < \rho(v, z)\}}|^2 \nu(dv) du \\ &\leq 2 \int_{\{|v| \leq m\}} |h - R(x+h, y)v + R(x, y)v|^2 \rho(v, z+h) \nu(dv) \\ &\quad + 2 \int_{\{|v| \leq m\}} |x-y+v - R(x, y)v|^2 |\rho(v, z+h) - \rho(v, z)| \nu(dv) \\ &=: I_1 + I_2. \end{aligned}$$

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

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

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 $\mathbf{1}_{\{|v+z| \leq m\}}$ 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 $I_2 \rightarrow 0$ when $h \rightarrow 0$. \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 $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$. Hence we can consider the equation for $(Y_t)_{t \geq 0}$ written as

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

Now observe that if we only consider the equation

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

it is easy to see that it has a unique strong solution since the process $(X_t, Y_t^1)_{t \geq 0}$ 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 above). 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

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

and

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

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

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

Then, if we show that $(\tilde{L}_t)_{t \geq 0}$ 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 the equation (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

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

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

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

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.

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*

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

then

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

Proof. By a standard argument, if the condition (2.19) is satisfied, we can approximate $\int_{t_1}^{t_2} \int_{\{|v| \leq m\} \times [0,1]} f(v, u) \tilde{N}(ds, dv, du)$ 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

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

The random variables $\tilde{N}((t_1, t_2], A_j)$ 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

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

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

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

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

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 $(\tilde{L}_t)_{t \geq 0}$ is given by (2.17).

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

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

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

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

for some sequence of partitions $0 = t_0^n < t_1^n < \dots < t_{m_n}^n = t$ of the interval $[0, t]$ with the mesh size going to zero as $n \rightarrow \infty$. 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 $\int_0^t \int_{\{|v| \leq m\} \times [0, 1]} B^n(s, v, u) \tilde{N}(ds, dv, du)$ converges in probability to the integral $\int_0^t \int_{\{|v| \leq m\} \times [0, 1]} B(X_{s-}, Y_{s-}, v, u) \tilde{N}(ds, dv, du)$. Thus we have

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

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

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

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

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

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

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

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

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

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

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

Then, using the formula (2.16) we can write

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

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

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

Since $|B(X_{t_{m_{n-1}}^n}^n, Y_{t_{m_{n-1}}^n}^n, v, u)|^2$ is integrable with respect to $\nu \times \lambda$ over $\{|v| \leq m\} \times [0, 1]$,

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

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

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

We infer that $(e^{i\langle z, Rv \rangle} (e^{i\langle z, c+v-Rv \rangle} - 1) - i\langle z, c + v - Rv \rangle) \rho(v, c)$ is also integrable with respect to ν over $\{|v| \leq m\}$. 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

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

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

$$\begin{aligned}
& (2.26) \\
& \mathbb{E} \left(\prod_{k=0}^{m_n-2} \exp \left(i \left\langle z, \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq m\} \times [0,1]} B(X_{t_k^n}, Y_{t_k^n}, v, u) \tilde{N}(ds, dv, du) \right\rangle \right) \right) \\
& \times \mathbb{E} \left[\exp \left(i \left\langle z, \int_{t_{m_n-1}^n}^{t_{m_n}^n} \int_{\{|v| \leq m\} \times [0,1]} B(X_{t_{m_n-1}^n}, Y_{t_{m_n-1}^n}, v, u) \tilde{N}(ds, dv, du) \right\rangle \right) \middle| \mathcal{F}_{t_{m_n-1}^n} \right] \\
& = \exp \left((t_{m_n}^n - t_{m_n-1}^n) \int_{\{|v| \leq m\} \times [0,1]} (e^{i\langle z, v \rangle} - 1 - i\langle z, v \rangle) \nu(dv) du \right) \\
& \times \mathbb{E} \left(\prod_{k=0}^{m_n-2} \exp \left(i \left\langle z, \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq m\} \times [0,1]} B(X_{t_k^n}, Y_{t_k^n}, v, u) \tilde{N}(ds, dv, du) \right\rangle \right) \right).
\end{aligned}$$

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

It remains now to show the independence of the increments of $(\tilde{L}_t)_{t \geq 0}$.

Lemma 2.6. *Under the above assumptions, for any $t_2 > t_1 \geq 0$ the random variable $\tilde{L}_{t_2} - \tilde{L}_{t_1}$ 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

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

As in the proof of Lemma 2.5, the integral $\int_{t_1}^{t_2} \int_{\{|v| \leq m\} \times [0,1]} B(X_{s-}, Y_{s-}, v, u) \tilde{N}(ds, dv, du)$ 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 $t_1 = t_0^n < t_1^n < \dots < t_{m_n}^n = t_2$ such that $\delta_n := \max_{k \in \{0, \dots, m_n-1\}} |t_{k+1}^n - t_k^n| \rightarrow 0$ as $n \rightarrow \infty$. Denote

$$I_k^n := \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq m\} \times [0,1]} B(X_{t_k^n}, Y_{t_k^n}, v, u) \tilde{N}(ds, dv, du), \quad I^n := \sum_{k=0}^{m_n-1} I_k^n.$$

Then we have

$$\begin{aligned} (2.27) \quad & \mathbb{E} \exp(i \langle z_1, I^n \rangle + i \langle z_2, \xi \rangle) = \mathbb{E} \left(\exp(i \langle z_2, \xi \rangle) \prod_{k=0}^{m_n-1} \exp(i \langle z_1, I_k^n \rangle) \right) \\ &= \mathbb{E} \left(\mathbb{E} \left[\exp(i \langle z_2, \xi \rangle) \prod_{k=0}^{m_n-1} \exp(i \langle z_1, I_k^n \rangle) \middle| \mathcal{F}_{t_{m_n-1}^n} \right] \right) \\ &= \mathbb{E} \left(\exp(i \langle z_2, \xi \rangle) \prod_{k=0}^{m_n-2} \exp(i \langle z_1, I_k^n \rangle) \mathbb{E} \left[\exp(i \langle z_1, I_{m_n-1}^n \rangle) \middle| \mathcal{F}_{t_{m_n-1}^n} \right] \right), \end{aligned}$$

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

$$\mathbb{E} \left[\exp(i \langle z_1, I_{m_n-1}^n \rangle) \middle| \mathcal{F}_{t_{m_n-1}^n} \right] = \mathbb{E} \exp \left(i \left\langle z_1, \int_{t_{m_n-1}^n}^{t_{m_n}^n} \int_{\{|v| \leq m\} \times [0,1]} v \tilde{N}(ds, dv, du) \right\rangle \right)$$

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

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

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

$$\begin{aligned} (2.28) \quad & \mathbb{E} \exp(i \langle z_1, I^n \rangle + i \langle z_2, \xi \rangle) = \mathbb{E} \exp(i \langle z_2, \xi \rangle) \\ & \quad \times \prod_{k=0}^{m_n-1} \mathbb{E} \exp \left(i \left\langle z_1, \int_{t_k^n}^{t_{k+1}^n} \int_{\{|v| \leq m\} \times [0,1]} v \tilde{N}(ds, dv, du) \right\rangle \right). \end{aligned}$$

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

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

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

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

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

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

which proves the desired assertion. \square

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

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

we will prove that

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

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

$$(3.3) \quad \mu = \mu \wedge \nu + \tilde{\mu} \text{ and } \nu = \mu \wedge \nu + \tilde{\nu}$$

for some finite measures $\tilde{\mu}$ and $\tilde{\nu}$ on \mathbb{R}^d . Then, if $\alpha := (\mu \wedge \nu)(\mathbb{R}^d) \in (0, 1)$, we can define probability measures $\bar{\mu} := \tilde{\mu}/\tilde{\mu}(\mathbb{R}^d)$ and $\bar{\nu} := \tilde{\nu}/\tilde{\nu}(\mathbb{R}^d)$ and we can easily show that $W_f(\mu, \nu) = (1 - \alpha)W_f(\bar{\mu}, \bar{\nu})$. 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 $W_f(\bar{\mu} p_t, \bar{\nu} p_t) \leq e^{-ct} W_f(\bar{\mu}, \bar{\nu})$.

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

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

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

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

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

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

Then we can write

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

Let us split our computations into two parts by writing

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

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 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 the equation (3.7) and the function $g(x) := f_1(|x|)$. We have

$$(3.9) \quad \partial_i g(x) = f'_1(|x|) \frac{x_i}{|x|} \text{ and } \partial_j \partial_i g(x) = f''_1(|x|) \frac{x_j x_i}{|x|^2} + f'_1(|x|) \left(\delta_{ij} \frac{1}{|x|} - \frac{x_j x_i}{|x|^3} \right),$$

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

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

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

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

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

$$(3.11) \quad \sum_{i,j=1}^d \int_0^1 (1-u) \left[f''_1(|W_{s,u}|) \frac{W_{s,u}^j W_{s,u}^i}{|W_{s,u}|^2} + f'_1(|W_{s,u}|) \frac{1}{|W_{s,u}|} \left(\delta_{ij} - \frac{W_{s,u}^j W_{s,u}^i}{|W_{s,u}|^2} \right) \right] du \Delta Z_s^i \Delta Z_s^j.$$

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

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

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

(3.12)

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

Note that the above formula holds only for $t < T$, where T is the coupling time defined by (2.11). However, for $t \geq T$ 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

$$\tau_n := \inf\{t \geq 0 : |Z_t| \notin (1/n, n)\}.$$

Note that we have $\tau_n \rightarrow T$ as $n \rightarrow \infty$, which follows from non-explosiveness of $(Z_t)_{t \geq 0}$, 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*

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

Proof. Observe that

$$\langle Z_{s-}, (I - R(X_{s-}, Y_{s-}))v \rangle = \langle Z_{s-}, 2e_{s-}e_{s-}^T v \rangle = 2\langle e_{s-}, v \rangle \langle Z_{s-}, \frac{Z_{s-}}{|Z_{s-}|} \rangle = 2\langle e_{s-}, v \rangle |Z_{s-}|$$

and therefore

$$\begin{aligned} \int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, (I - R(X_{s-}, Y_{s-}))v \rangle \tilde{N}(ds, dv, du) \\ = 2 \int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} f'_1(|Z_{s-}|) \langle e_{s-}, v \rangle \tilde{N}(ds, dv, du). \end{aligned}$$

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

$$\int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} f'_1(|Z_{s-}|) \langle e_{s-}, v \rangle \tilde{N}(ds, dv, du)$$

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

Lemma 3.2. *We have*

$$\mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u) \rangle \tilde{N}(ds, dv, du) = 0.$$

Proof. By the Cauchy-Schwarz inequality and the fact that $f'_1 \leq 1$, we have

$$\begin{aligned} \int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} |f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u) \rangle|^2 \nu(dv) duds \\ \leq \int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} |A(X_{s-}, Y_{s-}, v, u)|^2 \nu(dv) duds. \end{aligned}$$

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

$$\int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u) \rangle \tilde{N}(ds, dv, du)$$

is a martingale. \square

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}_ε 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) below). 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

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

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

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

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 the equation (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

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

and thus we can express the expectation above as

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

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 $h_w : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $\nu^m \circ h_w^{-1}$ is independent of the choice of w , i.e.,

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

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.,

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

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

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

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 the right hand side above). Thus we can estimate

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

□

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

$$(3.17) \quad \begin{aligned} \mathbb{E} f_1(|Z_{t \wedge \tau_n}|) - \mathbb{E} f_1(|Z_0|) &\leq \mathbb{E} \int_0^{t \wedge \tau_n} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, b(X_{s-}) - b(Y_{s-}) \rangle ds \\ &+ C_\varepsilon \mathbb{E} \int_0^{t \wedge \tau_n} \bar{f}_\varepsilon(|Z_{s-}|) \mathbf{1}_{\{|Z_{s-}| > \delta\}} ds. \end{aligned}$$

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] and [5].

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

$$(3.18) \quad -f'_1(r)\kappa(r)r + C_\varepsilon \bar{f}_\varepsilon(r) \leq -c_1 f_1(r)$$

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

Proof. Our assertion (3.18) is equivalent to

$$C_\varepsilon \bar{f}_\varepsilon(r) \leq -c_1 f_1(r) + f'_1(r)\kappa(r)r \text{ for all } r > \delta$$

or, explicitly,

$$(3.19) \quad \sup_{x \in (r-\varepsilon, r)} f''_1(x) \leq -\frac{c_1}{C_\varepsilon} f_1(r) + f'_1(r) \frac{r\kappa(r)}{C_\varepsilon} \text{ for all } r > \delta.$$

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

$$\sup_{x \in (r-\varepsilon, r)} f''_1(x) \leq -\frac{c_1}{C_\varepsilon} f_1(r) - f'_1(r) \frac{h^-(r)}{C_\varepsilon} \text{ for all } r > \delta,$$

which is equivalent to

$$f''_1(r-a) \leq -\frac{c_1}{C_\varepsilon} f_1(r) - f'_1(r) \frac{h^-(r)}{C_\varepsilon} \text{ for all } a \in (0, \varepsilon) \text{ and } r > \delta.$$

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

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

and

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

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

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

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

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

which is implied by

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

Define

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

Then of course

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

and thus the condition

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

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

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

and this ensures that (3.21) holds.

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

$$(3.25) \quad f_1(r) = \int_0^r \phi(s)g(s)ds.$$

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

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

which is implied by

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

Since Φ is increasing, the condition

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

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

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

Now define

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

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

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

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

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

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

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

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

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

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

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

which is equivalent to

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

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

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

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

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

We can compute

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

Therefore if we find R_1 such that

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

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

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

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

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

or, equivalently,

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

This shows that we can define R_1 by

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

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

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

$$(3.34) \quad \begin{aligned} & \mathbb{E} \int_0^{t \wedge \tau_n} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, b(X_{s-}) - b(Y_{s-}) \rangle \mathbf{1}_{\{|Z_{s-}| > \delta\}} ds \\ & + C_\varepsilon \mathbb{E} \int_0^{t \wedge \tau_n} \bar{f}_\varepsilon(|Z_{s-}|) \mathbf{1}_{\{|Z_{s-}| > \delta\}} ds \leq \mathbb{E} \int_0^{t \wedge \tau_n} -c_1 f_1(|Z_s|) \mathbf{1}_{\{|Z_{s-}| > \delta\}} ds. \end{aligned}$$

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

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

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

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

Now if we write (3.17) as

$$(3.37) \quad \begin{aligned} & \mathbb{E} f_1(|Z_{t \wedge \tau_n}|) - \mathbb{E} f_1(|Z_0|) \\ & \leq \mathbb{E} \int_0^{t \wedge \tau_n} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, b(X_{s-}) - b(Y_{s-}) \rangle (\mathbf{1}_{\{|Z_{s-}| > \delta\}} + \mathbf{1}_{\{0 < |Z_{s-}| \leq \delta\}}) ds \\ & + C_\varepsilon \mathbb{E} \int_0^{t \wedge \tau_n} \bar{f}_\varepsilon(|Z_{s-}|) \mathbf{1}_{\{|Z_{s-}| > \delta\}} ds, \end{aligned}$$

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 the equation (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

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

where

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

Note that $\tilde{C}_\delta(m)$ 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

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

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

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

and the assertion follows. \square

Note that we can always choose m large enough so that

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

and hence we have

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

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

$$\begin{aligned} & \mathbb{E}f(|Z_{t \wedge \tau_n}|) - \mathbb{E}f(|Z_0|) \\ & \leq \mathbb{E} \int_0^{t \wedge \tau_n} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, b(X_{s-}) - b(Y_{s-}) \rangle (\mathbf{1}_{\{|Z_{s-}| > \delta\}} + \mathbf{1}_{\{0 < |Z_{s-}| \leq \delta\}}) ds \\ & \quad + C_\varepsilon \mathbb{E} \int_0^{t \wedge \tau_n} \bar{f}_\varepsilon(|Z_{s-}|) \mathbf{1}_{\{|Z_{s-}| > \delta\}} ds - a \mathbb{E} \int_0^{t \wedge \tau_n} \frac{1}{2} \tilde{C}_\delta \mathbf{1}_{\{0 < |Z_{s-}| \leq \delta\}} ds. \end{aligned}$$

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

$$\begin{aligned} (3.40) \quad & \mathbb{E} \int_0^{t \wedge \tau_n} f'_1(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, b(X_{s-}) - b(Y_{s-}) \rangle \mathbf{1}_{\{0 < |Z_{s-}| \leq \delta\}} ds - a \mathbb{E} \int_0^{t \wedge \tau_n} \frac{1}{2} \tilde{C}_\delta \mathbf{1}_{\{0 < |Z_{s-}| \leq \delta\}} ds \\ & \leq (C_L \delta - \frac{1}{2} a \tilde{C}_\delta) \mathbb{E} \int_0^{t \wedge \tau_n} \mathbf{1}_{\{0 < |Z_{s-}| \leq \delta\}} ds. \end{aligned}$$

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

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

and thus it is sufficient if we have

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

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

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

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

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

and

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

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

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

Denote

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

Then of course

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

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

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

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

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

which finishes the proof of (1.8).

Proof of Theorem 1.1. By everything we proved in Sections 2.4, 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

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

Hence we get

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

□

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

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

□

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

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

which proves that

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

□

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

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

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

$$\int \varphi \mu p_n \rightarrow \int \varphi \mu_0$$

as $n \rightarrow \infty$ 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

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

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. \square

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

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

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 the equation (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

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

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

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

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

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

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

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

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 $c_2(\delta) := \tilde{C}_\delta/4$ (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 $c = \min\{c_1/2K, \tilde{C}_\delta/4\}$ 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 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

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

and the right hand side converges to $M/4$ when $\alpha \rightarrow 2$, 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).

ACKNOWLEDGEMENT

I would like to thank my PhD 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 (2011), no. 6, 1201-1216.
- [3] M. F. Chen, F. Y. Wang, Estimation of the first eigenvalue of second order elliptic operators, J. Funct. Anal. 131 (1995), no. 2, 345-363.
- [4] A. Eberle, Reflection coupling and Wasserstein contractivity without convexity, C. R. Math. Acad. Sci. Paris 349 (2011), no. 19-20, 1101-1104.
- [5] A. Eberle, Reflection couplings and contraction rates for diffusions, Probab. Theory Related Fields 166 (2016), no. 3-4, 851-886.
- [6] I. Gyöngy, N. V. Krylov, On stochastic equations with respect to semimartingales. I., Stochastics 4 (1980/81), no. 1, 1-21.
- [7] M. Hairer, J. C. Mattingly, Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations, Ann. Probab. 36 (2008), no. 6, 2050-2091.
- [8] E. P. Hsu, K. T. Sturm, Maximal coupling of Euclidean Brownian motions, Commun. Math. Stat. 1 (2013), no. 1, 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 (2012), no. 5, 2155-2184.
- [11] A. M. Kulik, Exponential ergodicity of the solutions to SDE's with a jump noise, Stochastic Process. Appl. 119 (2009), no. 2, 602-632.
- [12] H. Lin, J. Wang, Successful couplings for a class of stochastic differential equations driven by Lévy processes, Sci. China Math. 55 (2012), no. 8, 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 (1986), no. 3, 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, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. 455 (1999), no. 1984, 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 (2012), no. 1, 106-133.

- [20] K. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, 1999.
- [21] R. L. Schilling, P. Sztonyk, J. Wang, Coupling property and gradient estimates of Lévy processes via the symbol, *Bernoulli* 18 (2012), no. 4, 1128-1149.
- [22] R. L. Schilling, J. Wang, On the coupling property of Lévy processes, *Ann. Inst. Henri Poincaré Probab. Stat.* 47 (2011), no. 4, 1147-1159.
- [23] R. L. Schilling, J. Wang, On the coupling property and the Liouville theorem for Ornstein-Uhlenbeck processes, *J. Evol. Equ.* 12 (2012), no. 1, 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 (2015), no. 2, 447-458.
- [26] H. Thorisson, Shift-coupling in continuous time, *Probab. Theory Related Fields* 99 (1994), no. 4, 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 (2011), no. 4, 1136-1158.
- [29] J. Wang, On the exponential ergodicity of Lévy-driven Ornstein-Uhlenbeck processes, *J. Appl. Probab.* 49 (2012), no. 4, 990-1004.
- [30] J. Wang, Exponential ergodicity and strong ergodicity for SDEs driven by symmetric α -stable processes, *Appl. Math. Lett.* 26 (2013), no. 6, 654-658.
- [31] J. Wang, On the existence and explicit estimates for the coupling property of Lévy processes with drift, *J. Theoret. Probab.* 27 (2014), no. 3, 1021-1044.
- [32] J. Wang, L^p -Wasserstein distance for stochastic differential equations driven by Lévy processes, *Bernoulli* 22 (2016), no. 3, 1598-1616.

INSTITUTE FOR APPLIED MATHEMATICS, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY

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

5 Transportation inequalities for non-globally dissipative SDEs with jumps via Malliavin calculus and coupling

TRANSPORTATION INEQUALITIES FOR NON-GLOBALLY DISSIPATIVE SDES WITH JUMPS VIA MALLIAVIN CALCULUS AND COUPLING

MATEUSZ B. MAJKA

ABSTRACT. By using the mirror coupling for solutions of SDEs driven by pure jump Lévy processes, we extend some transportation and concentration inequalities, which were previously known only in the case where the coefficients in the equation satisfy a global dissipativity condition. Furthermore, by using the mirror coupling for the jump part and the coupling by reflection for the Brownian part, we extend analogous results for jump diffusions. To this end, we improve some previous results concerning such couplings and show how to combine the jump and the Brownian case. As a crucial step in our proof, we develop a novel method of bounding Malliavin derivatives of solutions of SDEs with both jump and Gaussian noise, which involves the coupling technique and which might be of independent interest. The bounds we obtain are new even in the case of diffusions without jumps.

1. INTRODUCTION

We consider stochastic differential equations in \mathbb{R}^d of the form

$$(1.1) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du),$$

where $(W_t)_{t \geq 0}$ is an m -dimensional Brownian motion and $\tilde{N}(dt, du) = N(dt, du) - dt \nu(du)$ is a compensated Poisson random measure on $\mathbb{R}_+ \times U$, where (U, \mathcal{U}, ν) is a σ -finite measure space. Moreover, the coefficients $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ and $g : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$ are such that for any $x \in \mathbb{R}^d$ we have

$$\int_U |g(x, u)|^2 \nu(du) < \infty$$

and there exists a continuous function $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}$ such that for all $x, y \in \mathbb{R}^d$ we have

$$(1.2) \quad \langle b(x) - b(y), x - y \rangle + \frac{1}{2} \int_U |g(x, u) - g(y, u)|^2 \nu(du) + \|\sigma(x) - \sigma(y)\|_{HS}^2 \leq -\kappa(|x - y|)|x - y|^2,$$

where $\|\sigma\|_{HS} = \sqrt{\text{tr} \sigma \sigma^T}$ is the Hilbert-Schmidt norm. Note that κ is allowed to take negative values.

If the condition (1.2) holds with a constant function $\kappa \equiv K$ for some $K \in \mathbb{R}$, we call (1.2) a one-sided Lipschitz condition. If $K > 0$, we call it a (global) dissipativity condition. If a one-sided Lipschitz condition is satisfied and we additionally assume that the drift b is continuous and that σ and g satisfy a linear growth condition, we can prove that (1.1) has a unique non-explosive strong solution, even if the one-sided Lipschitz condition is satisfied only locally (see e.g. Theorem 2 in [17]).

2010 *Mathematics Subject Classification.* 60G51, 60H10, 60H07, 60E15.

Key words and phrases. Stochastic differential equations, Lévy processes, transportation inequalities, couplings, Wasserstein distances, Malliavin calculus.

For $p \geq 1$, the L^p -Wasserstein distance (or the L^p -transportation cost) between two probability measures μ_1, μ_2 on a metric space (E, ρ) is defined by

$$W_{p,\rho}(\mu_1, \mu_2) := \inf_{\pi \in \Pi(\mu_1, \mu_2)} \left(\int \int \rho(x, y)^p \pi(dx dy) \right)^{1/p},$$

where $\Pi(\mu_1, \mu_2)$ 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 $E \times E$ with marginals μ_1 and μ_2 . If the metric space (E, ρ) is chosen to be \mathbb{R}^d with the Euclidean metric $\rho(x, y) = |x - y|$, then we denote $W_{p,\rho}$ just by W_p .

If the equation (1.1) is globally dissipative with some constant $K > 0$, then it is well known that the solution $(X_t)_{t \geq 0}$ to (1.1) has an invariant measure and that the transition semigroup $(p_t)_{t \geq 0}$ associated with $(X_t)_{t \geq 0}$ is exponentially contractive with respect to W_p for any $p \in [1, 2]$, i.e.,

$$W_p(\mu_1 p_t, \mu_2 p_t) \leq e^{-Kt} W_p(\mu_1, \mu_2)$$

for any probability measures μ_1 and μ_2 on \mathbb{R}^d and any $t > 0$ (see e.g. the proof of Theorem 2.2 in [25]). However, we will show that for $p = 1$ a related result still holds (under some additional assumptions, see Corollary 2.7) if we replace the global dissipativity condition with the following one.

Assumption D1. (Dissipativity at infinity)

$$\limsup_{r \rightarrow \infty} \kappa(r) > 0.$$

In other words, Assumption D1 states that there exist constants $R > 0$ and $K > 0$ such that for all $x, y \in \mathbb{R}^d$ with $|x - y| > R$ we have

$$\langle b(x) - b(y), x - y \rangle + \frac{1}{2} \int_U |g(x, u) - g(y, u)|^2 \nu(du) + \|\sigma(x) - \sigma(y)\|_{HS}^2 \leq -K|x - y|^2,$$

which justifies calling it a dissipativity at infinity condition. Moreover, in some cases we will also need another condition on the function κ , namely

Assumption D2. (Regularity of the drift at zero)

$$\lim_{r \rightarrow 0} r\kappa(r) = 0.$$

This is obviously satisfied if, e.g., there is a constant $L > 0$ such that we have $\kappa(r) \geq -L$ for all $r \geq 0$ (which is the case whenever the coefficients in (1.1) satisfy a one-sided Lipschitz condition) and if b is continuous. Such an assumption is quite natural in order to ensure existence of a solution to (1.1).

For probability measures μ_1 and μ_2 on (E, ρ) , we define the *relative entropy* (Kullback-Leibler information) of μ_1 with respect to μ_2 by

$$H(\mu_1 | \mu_2) := \begin{cases} \int \log \frac{d\mu_1}{d\mu_2} d\mu_1 & \text{if } \mu_1 \ll \mu_2, \\ +\infty & \text{otherwise.} \end{cases}$$

We say that a probability measure μ satisfies an L^p -transportation cost-information inequality on (E, ρ) if there is a constant $C > 0$ such that for any probability measure η we have

$$W_{p,\rho}(\eta, \mu) \leq \sqrt{2CH(\eta | \mu)}.$$

Then we write $\mu \in T_p(C)$.

The most important cases are $p = 1$ and $p = 2$. Since $W_{1,\rho} \leq W_{2,\rho}$, we see that the L^2 -transportation inequality (the T_2 inequality, also known as the Talagrand inequality) implies T_1 , and it is well known that in fact T_2 is much stronger. The T_2 inequality

has some interesting connections with other well-known functional inequalities. Due to Otto and Villani [29], we know that the log-Sobolev inequality implies T_2 , whereas T_2 implies the Poincaré inequality. On the other hand, the T_1 inequality is related to the phenomenon of measure concentration. Indeed, consider a generalization of T_1 known as the α - W_1H inequality. Namely, let α be a non-decreasing, left continuous function on \mathbb{R}_+ with $\alpha(0) = 0$. We say that a probability measure μ satisfies a W_1H -inequality with deviation function α (or simply α - W_1H inequality) if for any probability measure η we have

$$(1.3) \quad \alpha(W_{1,\rho}(\eta, \mu)) \leq H(\eta|\mu).$$

We have the following result which is due to Gozlan and Léonard (see Theorem 2 in [14] for the original result, cf. also Lemma 2.1 in [39]). It is a generalization of a result by Bobkov and Götze (Theorem 3.1 in [8]), which held only for the quadratic deviation function.

Fix a probability measure μ on (E, ρ) and a convex deviation function α . Then the following properties are equivalent:

- (1) the α - W_1H inequality for the measure μ holds, i.e., for any probability measure η on (E, ρ) we have

$$\alpha(W_{1,\rho}(\eta, \mu)) \leq H(\eta|\mu),$$

- (2) for every $f : E \rightarrow \mathbb{R}$ bounded and Lipschitz with $\|f\|_{\text{Lip}} \leq 1$ we have

$$(1.4) \quad \int e^{\lambda(f-\mu(f))} d\mu \leq e^{\alpha^*(\lambda)} \text{ for any } \lambda > 0,$$

where $\alpha^*(\lambda) := \sup_{r \geq 0} (r\lambda - \alpha(r))$ is the convex conjugate of α ,

- (3) if $(\xi_k)_{k \geq 1}$ is a sequence of i.i.d random variables with common law μ , then for every $f : E \rightarrow \mathbb{R}$ bounded and Lipschitz with $\|f\|_{\text{Lip}} \leq 1$ we have

$$(1.5) \quad \mathbb{P} \left(\frac{1}{n} \sum_{k=1}^n f(\xi_k) - \mu(f) > r \right) \leq e^{-n\alpha(r)} \text{ for any } r > 0, n \geq 1.$$

This gives an intuitive interpretation of α - W_1H in terms of a concentration of measure property (1.5), while the second characterization (1.4) is very useful for proving such inequalities, as we shall see in the sequel. For a general survey of transportation inequalities the reader might consult [15] or Chapter 22 of [37].

As an example of a simple equation of the type (1.1) consider

$$dX_t = b(X_t)dt + \sqrt{2}dW_t$$

with a d -dimensional Brownian motion $(W_t)_{t \geq 0}$. If the global dissipativity assumption is satisfied, then $(X_t)_{t \geq 0}$ has an invariant measure μ and by a result of Bakry and Émery [3], μ satisfies the log-Sobolev inequality and thus (by Otto and Villani [29]) also the Talagrand inequality. More generally, for equations of the form

$$(1.6) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

also under the global dissipativity assumption, Djellout, Guillin and Wu in [11] showed that T_2 holds for the invariant measure, as well as on the path space. As far as we are aware, there are currently no results in the literature concerning transportation inequalities for equations like (1.6) without assuming global dissipativity. Hence, even though in the present paper we focus on SDEs with jumps, our results may be also new in the purely Gaussian case.

For equations of the form

$$(1.7) \quad dX_t = b(X_t)dt + \int_U g(X_{t-}, u)\tilde{N}(dt, du),$$

the Poincaré inequality does not always hold (see Example 1.1 in [39]) and thus in general we cannot have T_2 . However, under the global dissipativity assumption, Wu in [39] showed some α - W_1H inequalities.

Suppose there is a real measurable function g_∞ on U such that $|g(x, u)| \leq g_\infty(u)$ for every $x \in \mathbb{R}^d$ and $u \in U$. We make the following assumption.

Assumption E. (Exponential integrability of the intensity measure)

There exists a constant $\lambda > 0$ such that

$$\beta(\lambda) := \int_U (e^{\lambda g_\infty(u)} - \lambda g_\infty(u) - 1)\nu(du) < \infty,$$

where ν is the intensity measure associated with \tilde{N} .

Remark 1.1. Assumption E is quite restrictive. In particular, let us consider the case where $U \subset \mathbb{R}^d$ and $g(x, u) = \tilde{g}(x)u$ for some $\mathbb{R}^{d \times d}$ -valued function \tilde{g} and hence the equation (1.7) is driven by a d -dimensional Lévy process $(L_t)_{t \geq 0}$ (i.e., we have $dX_t = b(X_t)dt + \tilde{g}(X_{t-})dL_t$). Then Assumption E implies finiteness of an exponential moment of $(L_t)_{t \geq 0}$ (cf. Theorem 25.3 and Corollary 25.8 in [34]). However, there are examples of equations of such type for which the α - W_1H inequality implies Assumption E, and hence in general we cannot prove such inequalities without it (see Remark 2.5 in [39]). Nevertheless, without this assumption it is still possible to obtain some concentration inequalities (see Remark 5.2 in [39] or Theorem 2.2 below).

Fix $T > 0$ and define a deviation function

$$\alpha_T(r) := \sup_{\lambda \geq 0} \left\{ r\lambda - \int_0^T \beta(e^{-Kt}\lambda)dt \right\},$$

where the constants $\lambda > 0$ and $K > 0$ are such that Assumption E is satisfied with λ and that (1.7) is globally dissipative with the dissipativity constant K . Then for any $T > 0$ and any $x \in \mathbb{R}^d$, by Theorem 2.2 in [39] we have the W_1H transportation inequality with deviation function α_T for the measure $\delta_x p_T$, which is the law of the random variable $X_T(x)$, where $(X_t(x))_{t \geq 0}$ is a solution to (1.7) starting from $x \in \mathbb{R}^d$, i.e., we have

$$\alpha_T(W_1(\eta, \delta_x p_T)) \leq H(\eta|\delta_x p_T)$$

for any probability measure η on \mathbb{R}^d , where $W_1 = W_{1,\rho}$ with ρ being the Euclidean metric on \mathbb{R}^d . Analogous results have been proved by a very similar approach in [25] for equations of the form (1.1), i.e., including also the Gaussian noise.

In the sequel we will explain how to modify the proofs in [39] and [25] to replace the global dissipativity assumption with our Assumption D1. We will show that we can obtain α - W_1H inequalities by using couplings to control perturbations of solutions to (1.1), see Theorem 2.1. We will also prove that the construction of the required couplings is possible for a certain class of equations satisfying Assumption D1 (Theorems 2.3 and 2.8). All these results together will imply our extension of the main theorems from [39] and [25], which is stated as Corollary 2.9.

The method of the proof is based on the Malliavin calculus. On any filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ equipped with an m -dimensional Brownian motion $(W_t)_{t \geq 0}$ and a Poisson random measure N on $\mathbb{R}_+ \times U$, we can define the Malliavin derivatives for a certain class of measurable functionals F with respect to the process $(W_t)_{t \geq 0}$ (the

classic Malliavin differential operator ∇), as well as a Malliavin derivative of F with respect to N (the difference operator D). Namely, if we consider the family \mathcal{S} of smooth functionals of $(W_t)_{t \geq 0}$ of the form

$$F = f(W(h_1), \dots, W(h_n)) \text{ for } n \geq 1,$$

where $W(h) = \int_0^T h(s) dW_s$ for $h \in H = L^2([0, T]; \mathbb{R}^m)$ and $f \in C^\infty(\mathbb{R}^n)$, we can define the Malliavin derivative with respect to $(W_t)_{t \geq 0}$ as the unique element ∇F in $L^2(\Omega; H) \simeq L^2(\Omega \times [0, T]; \mathbb{R}^m)$ such that for any $h \in H$ we have

$$\langle \nabla F, h \rangle_{L^2([0, T]; \mathbb{R}^m)} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left(F(W. + \int_0^\cdot h_s ds) - F(W.) \right),$$

where the convergence holds in $L^2(\Omega)$ (see e.g. Definition A.10 in [10]). Then the definition can be extended to all random variables F in the space $\mathbb{D}^{1,2}$ which is the completion of \mathcal{S} in $L^2(\Omega)$ with respect to the norm

$$\|F\|_{\mathbb{D}^{1,2}}^2 := \|F\|_{L^2(\Omega)}^2 + \|\nabla F\|_{L^2(\Omega; H)}^2.$$

For a brief introduction to the Malliavin calculus with respect to Brownian motion see Appendix A in [10] or Chapter VIII in [5] and for a comprehensive treatment the monograph [28]. On the other hand, the definition of the Malliavin derivative with respect to N that we need is much less technical, since it is just a difference operator. Namely, if our Poisson random measure N on $\mathbb{R}_+ \times U$ has the form

$$N = \sum_{j=1}^{\infty} \delta_{(\tau_j, \xi_j)}$$

with \mathbb{R}_+ -valued random variables τ_j and U -valued ξ_j , then for any measurable functional f of N and for any $(t, u) \in \mathbb{R}_+ \times U$ we put

$$(1.8) \quad D_{t,u} f(N) := f(N + \delta_{(t,u)}) - f(N).$$

There is also an alternative approach to the Malliavin calculus for jump processes, where the Malliavin derivative is defined as an actual differential operator, which was in fact the original approach and which traces back to Bismut [7], see also [4] and [6]. However, for our purposes we prefer the definition (1.8), which was introduced by Picard in [30] and [31], and which is suitable for proving the Clark-Ocone formula. Namely, we will need to use the result stating that for any F being a functional of $(W_t)_{t \geq 0}$ and N such that

$$(1.9) \quad \mathbb{E} \int_0^T |\nabla_t F|^2 dt + \mathbb{E} \int_0^T \int_U |D_{t,u} F|^2 \nu(du) dt < \infty,$$

we have

$$F = \mathbb{E} F + \int_0^T \mathbb{E}[\nabla_t F | \mathcal{F}_t] dW_t + \int_0^T \int_U \mathbb{E}[D_{t,u} F | \mathcal{F}_t] \tilde{N}(dt, du).$$

It is proved in [24] that the definition (1.8) is actually equivalent to the definition of the Malliavin derivative for jump processes via the chaos expansion and this approach is used to obtain the Clark-Ocone formula for the pure jump case. For the jump diffusion case, see Theorem 12.20 in [10]. For more general recent extensions of this result, see [21]. Once we apply the Clark-Ocone formula to the solution of (1.1), we can obtain some information on its behaviour by controlling its Malliavin derivatives. Therefore one of the crucial components of the proof of our results in this paper is Theorem 2.14, presenting a novel method of bounding such derivatives, which, contrary to the method used in Lemma 3.4 in [25], works also without the global dissipativity assumption and without any explicit regularity conditions on the coefficients of (1.1), except some sufficient ones

to guarantee Malliavin differentiability of the solution (it is enough if the coefficients are Lipschitz, see e.g. Theorem 17.4 in [10]).

The last notion that we need to introduce before we will be able to formulate our main results is that of a coupling. 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. The construction of appropriate couplings of solutions to equations like (1.1) plays the key role in the proofs of Theorems 2.3 and 2.8. For more information about couplings, see e.g. [22], [12], [27] and the references therein.

The only papers that we are aware of which deal with transportation inequalities directly in the context of SDEs with jumps are [39], [26], [25] and [36]. The latter two actually extend the method developed by Wu in [39], but in both these papers a kind of global dissipativity assumption is required (see Remark 2.12 for a discussion about [36]). In the present paper we explain how to drop this assumption (by imposing some additional conditions) and further extend the method of Wu. Since our extension lies at the very core of the method, it allows us to improve on essentially all the main results and corollaries obtained in [39] and [25] (and it might be also applicable to the results in [36], cf. once again Remark 2.12), replacing the global dissipativity assumption with a weaker condition.

On the other hand, in [26] some convex concentration inequalities of the type (2.6) have been shown for a certain class of additive functionals $S_T = \int_0^T g(X_t) dt$ of solutions $(X_t)_{t \geq 0}$ to equations like (1.1). These are later used to obtain some α - $W_1 I$ inequalities, which are analogous to α - $W_1 H$ inequalities (1.3) but with the Kullback-Leibler information H replaced with the Fisher-Donsker-Varadhan information, see e.g. [16] for more details. The proof in [26], similarly to [39], is based on the forward-backward martingale method from [19], but unlike [39] it does not use the Malliavin calculus. In the framework of Wu from [39] that we use here, it is possible to obtain related α - $W_1 J$ inequalities with J being the modified Donsker-Varadhan information. Once we have transportation inequalities like the ones in our Theorem 2.1, we can use the methods from Corollary 2.15 in [39] and Corollary 2.7 in [25]. This is, however, beyond the scope of the present paper and in the sequel we focus on extending the main results from [39] and [25].

2. MAIN RESULTS

We start with a general theorem, which shows that a key tool to obtain transportation inequalities for a solution $(X_t)_{t \geq 0}$ to

$$(2.1) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du)$$

is to be able to control perturbations of $(X_t)_{t \geq 0}$ via a coupling, with respect to changes in initial conditions (see (2.2) below) as well as changes of the drift (see (2.3)). In the next two theorems we assume that the coefficients in (2.1) satisfy some sufficient conditions for existence of a solution and its Malliavin differentiability (e.g. they are Lipschitz, cf. Theorem 17.4 in [10]). From now on, $(\mathcal{F}_t)_{t \geq 0}$ will always denote the filtration generated by all the sources of noise in the equations that we consider, while $(p_t)_{t \geq 0}$ will be the transition semigroup associated with the solution to the equation. Moreover, for a process $(h_t)_{t \geq 0}$ adapted to $(\mathcal{F}_t)_{t \geq 0}$, we will denote by $(\tilde{X}_t)_{t \geq 0}$ a solution to

$$d\tilde{X}_t = b(\tilde{X}_t)dt + \sigma(\tilde{X}_t)h_t dt + \sigma(\tilde{X}_t)dW_t + \int_U g(\tilde{X}_{t-}, u)\tilde{N}(dt, du).$$

Then we have the following result.

Theorem 2.1. *Assume there exists a constant σ_∞ such that for any $x \in \mathbb{R}^d$ we have $\|\sigma(x)\| \leq \sigma_\infty$, where $\|\cdot\|$ is the operator norm, and there exists a measurable function $g_\infty : U \rightarrow \mathbb{R}$ such that $|g(x, u)| \leq g_\infty(u)$ for any $x \in \mathbb{R}^d$ and $u \in U$. Assume further that there exists some $\lambda > 0$ such that Assumption E is satisfied. Moreover, suppose that there exists a coupling $(X_t, Y_t)_{t \geq 0}$ of solutions to (2.1) and a function $c_1 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that for any $0 \leq s \leq t$ we have*

$$(2.2) \quad \mathbb{E}[|X_t - Y_t|/\mathcal{F}_s] \leq c_1(t-s)|X_s - Y_s|.$$

Furthermore, assume that there exists a coupling $(X_t, Y'_t)_{t \geq 0}$ of solutions to (2.1) and functions $c_2, c_3 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that for any $0 \leq s \leq t$ we have

$$(2.3) \quad \mathbb{E}[|\tilde{X}_t - Y'_t|/\mathcal{F}_s] \leq c_2(t-s)\mathbb{E} \int_s^t c_3(r)|\sigma(\tilde{X}_r)h_r|dr.$$

Then the following assertions hold.

- (1) For any $T > 0$ and for any $x \in \mathbb{R}^d$ the measure $\delta_x p_T$ satisfies

$$(2.4) \quad \alpha_T(W_1(\eta, \delta_x p_T)) \leq H(\eta|\delta_x p_T)$$

for any probability measure η on \mathbb{R}^d . Here $W_1 = W_{1,\rho}$ with ρ being the Euclidean metric on \mathbb{R}^d and

$$\alpha_T(r) := \sup_{\lambda \geq 0} \left\{ r\lambda - \int_0^T \beta(c_1(T-t)\lambda)dt - \frac{\sigma_\infty^2 c_2^2(T)\lambda^2}{2} \int_0^T c_3^2(t)dt \right\}.$$

- (2) For any $T > 0$ and for any $x \in \mathbb{R}^d$ the law $\mathbb{P}_{x,[0,T]}$ of $(X_t(x))_{t \in [0,T]}$ as a measure on the space $\mathbb{D}([0,T]; \mathbb{R}^d)$ of càdlàg \mathbb{R}^d -valued functions on $[0,T]$ satisfies

$$(2.5) \quad \alpha_T^P(W_{1,d_{L^1}}(Q, \mathbb{P}_{x,[0,T]})) \leq H(Q|\mathbb{P}_{x,[0,T]})$$

for any probability measure Q on $\mathbb{D}([0,T]; \mathbb{R}^d)$. Here we take $d_{L^1}(\gamma_1, \gamma_2) := \int_0^T |\gamma_1(t) - \gamma_2(t)|dt$ as the L^1 metric on the path space and

$$\alpha_T^P(r) := \sup_{\lambda \geq 0} \left\{ r\lambda - \int_0^T \beta \left(\lambda \int_t^T c_1(s-t)ds \right) dt - \frac{\sigma_\infty^2 \lambda^2}{2} \int_0^T c_3^2(t) \left(\int_t^T c_2(r)dr \right)^2 dt \right\}.$$

Even without Assumption E, it is still possible to recover some concentration inequalities.

Theorem 2.2. *Assume that all the assumptions of Theorem 2.1 are satisfied except for Assumption E. Instead, suppose that $g_\infty(u)$ is just square integrable with respect to ν . Fix any $T > 0$ and any $x \in \mathbb{R}^d$. Then for any C^2 convex function ϕ such that ϕ' is also convex and for any Lipschitz function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, we have*

$$(2.6) \quad \begin{aligned} & \mathbb{E}\phi\left(f(X_T(x)) - p_T f(x)\right) \\ & \leq \mathbb{E}\phi\left(\|f\|_{\text{Lip}} \left(\int_0^T \int_U c_1(T-t)g_\infty(u)\tilde{N}(dt, du) + c_2(T) \int_0^T c_3(t)j(t)dW_t \right)\right), \end{aligned}$$

where j is any deterministic \mathbb{R}^m -valued function such that for all $t > 0$ we have $|j(t)| = \sigma_\infty$. Moreover, for any Lipschitz function $F : \mathbb{D}([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}$ we have

$$(2.7) \quad \begin{aligned} & \mathbb{E}\phi\left(F(X_{[0, T]}(x)) - \mathbb{E}F(X_{[0, T]}(x))\right) \\ & \leq \mathbb{E}\phi\left(\|F\|_{\text{Lip}}\left(\int_0^T \int_U \left(\int_t^T c_1(r-t)dr\right) g_\infty(u) \tilde{N}(dt, du) \right. \right. \\ & \quad \left. \left. + \int_0^T c_3(t) \left(\int_t^T c_2(r)dr\right) j(t)dW_t\right)\right). \end{aligned}$$

The crucial step in proving the above theorems is to find appropriate bounds on Malliavin derivatives of the solution to (2.1). We will show that we can obtain such bounds on D and ∇ using conditions (2.2) and (2.3), respectively (see Section 5 for details).

Now we present another result, which will consequently lead us to some examples of equations for which the inequalities (2.2) and (2.3) actually hold. First, however, we need to formulate some additional assumptions. We will need a pure jump Lévy process $(L_t)_{t \geq 0}$ with a Lévy measure ν^L satisfying the following set of conditions.

Assumption L1. (Rotational invariance of the Lévy measure) ν^L is rotationally invariant, i.e.,

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

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

Assumption L2. (Absolute continuity of the Lévy measure) ν^L is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d with a density q that is continuous almost everywhere on \mathbb{R}^d .

Under Assumptions L1-L2 it has been proved in [27] (see Theorem 1.1 therein) that there exists a coupling $(X_t, Y_t)_{t \geq 0}$ of solutions to

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

defined as a unique strong solution to the $2d$ -dimensional SDE given in the sequel by (3.2) and (3.3). Moreover, consider two additional conditions on the jump density q .

Assumption L3. (Positive mass of the overlap of the jump density and its translation) There exist constants $m, \delta > 0$ such that $\delta < 2m$ and

$$(2.8) \quad \inf_{x \in \mathbb{R}^d: 0 < |x| \leq \delta} \int_{\{|v| \leq m\} \cap \{|v+x| \leq m\}} q(v) \wedge q(v+x) dv > 0.$$

Assumption L4. (Positive mass in a neighbourhood of zero) There exists a constant $\varepsilon > 0$ such that $\varepsilon \leq \delta$ (with δ defined via (2.8) above) and

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

Suppose now that all the Assumptions L1-L4 are satisfied. Let us define a continuous function $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}$ so that for any $x, y \in \mathbb{R}^d$ the condition $\langle b(x) - b(y), x - y \rangle \leq -\kappa(|x - y|)|x - y|^2$ is satisfied and suppose that Assumption D1 holds. Then we get that, by the inequality (1.8) in Theorem 1.1 in [27], there exist explicitly given $L, \theta > 0$ and a function f such that

$$(2.9) \quad \mathbb{E}|X_t(x) - Y_t(y)| \leq Le^{-\theta t} f(|x - y|).$$

However, the function f used in [27] is discontinuous. It is actually of the form

$$(2.10) \quad f = a\mathbf{1}_{(0,\infty)} + f_1$$

with $a > 0$ and f_1 being a continuous, concave function, extended in an affine way from some point $R_1 > 0$ (and thus we have $a_1x \leq f_1(x) \leq a_2x$ for some $a_1, a_2 > 0$). Hence we obtain

$$(2.11) \quad \mathbb{E}|X_t(x) - Y_t(y)| \leq \tilde{L}e^{-\theta t}(|x - y| + 1),$$

for some $\tilde{L} > 0$, which is, however, undesirable since in order to be able to apply Theorem 2.1 we would like to have $|x - y|$ and not $|x - y| + 1$ on the right hand side (cf. Remark 2.6). Thus we need to improve on the result from [27] and get an inequality like (2.9) but with a continuous function f (i.e., with $a = 0$ in (2.10)). To this end, we define

$$(2.12) \quad C_\varepsilon := 2 \int_0^{\varepsilon/4} |y|^2 \nu_1^L(dy),$$

where ν_1^L is the first marginal of the rotationally invariant measure ν^L . The choice of $\varepsilon/4$ as the upper integration limit is motivated by the calculations in the proof of Theorem 1.1 in [27], see also the proof of Theorem 3.1 below. Now consider a new condition.

Assumption L5. (Sufficient concentration of ν^L around zero) For any $\lambda > 0$ there exists a $K(\lambda) > 0$ such that for all $\varepsilon < \lambda$ we have $\varepsilon \leq K(\lambda)C_\varepsilon$. In other words, $\varepsilon/C_\varepsilon$ is bounded near zero or, using the big O notation, $\varepsilon = O(C_\varepsilon)$ as $\varepsilon \rightarrow 0$.

Intuitively, it is an assumption about sufficient concentration of the Lévy measure ν^L around zero (sufficient small jump activity). It is satisfied e.g. for α -stable processes with $\alpha \in [1, 2)$ since in this case $C_\varepsilon = A\varepsilon^{2-\alpha}$ for some constant $A = A(\alpha)$ and we have $\varepsilon/C_\varepsilon = A\varepsilon^{\alpha-1}$.

It turns out that once we replace Assumptions L3 and L4 in Theorem 1.1 in [27] with Assumption L5, we are able to obtain (2.9) with a continuous function f , which is exactly what we need for Theorem 2.1. This is done in Section 3 in Theorem 3.1. However, we are able to generalize this result even further.

Theorem 2.3. *Consider an SDE of the form*

$$(2.13) \quad dX_t = b(X_t)dt + \sigma_1 dB_t^1 + \sigma(X_t)dB_t^2 + dL_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du),$$

where $(B_t^1)_{t \geq 0}$ and $(B_t^2)_{t \geq 0}$ are d -dimensional Brownian motions, $(L_t)_{t \geq 0}$ is a pure jump Lévy process with Lévy measure ν^L satisfying Assumptions L1-L2 and L5, whereas \tilde{N} is a compensated Poisson random measure on $\mathbb{R}_+ \times U$ with intensity measure $dt \nu(du)$. Assume that all the sources of noise are independent, $\sigma_1 \in \mathbb{R}^{d \times d}$ is a constant matrix and the coefficients $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ and $g : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$ satisfy Assumption D1. If at least one of the following two conditions is satisfied

- (1) $\det \sigma_1 > 0$,
- (2) $L_t \neq 0$ and Assumption D2,

then there exists a coupling $(X_t, Y_t)_{t \geq 0}$ of solutions to (2.13) and constants $\tilde{C}, \tilde{c} > 0$ such that for any $x, y \in \mathbb{R}^d$ and any $t > 0$ we have

$$(2.14) \quad \mathbb{E}|X_t(x) - Y_t(y)| \leq \tilde{C}e^{-\tilde{c}t}|x - y|.$$

Remark 2.4. The reason for the particular form of the equation (2.13) is that in order to construct a coupling leading to the inequality (2.14) we need a suitable additive component of the noise. We can either use $(B_t^1)_{t \geq 0}$ if the condition (1) holds, or $(L_t)_{t \geq 0}$ if the

condition (2) holds. The constants \tilde{C} and \tilde{c} depend on which noise we use. In particular, the constant \tilde{c} is either equal to c defined by (4.5) if we use $(B_t^1)_{t \geq 0}$ or to c_1 defined by (3.16) if we use $(L_t)_{t \geq 0}$. On the other hand, if we have only a multiplicative Gaussian noise but the coefficient σ is such that $\sigma\sigma^T$ is uniformly positive definite, we can use Lemma 4.1 below to decompose this noise and extract an additive component satisfying (1). Without such an assumption on σ , Remark 2 in [12] indicates that it might still be possible to perform a suitable construction, using the so-called Kendall-Cranston coupling, although this might significantly increase the level of sophistication of the proof. In the case of the jump noise, as far as we know there are currently no methods for obtaining couplings leading to inequalities like (2.14) in the case of purely multiplicative noise, and the recent papers treating this kind of problems (see e.g. [38], [27] and [23]) use methods that rely on the noise having at least some additive component.

Remark 2.5. The coupling process $(X_t, Y_t)_{t \geq 0}$ is constructed as a unique strong solution to some $2d$ -dimensional SDE. This allows us to infer that $(X_t, Y_t)_{t \geq 0}$ is in fact a Markov process (see e.g. Theorem 6.4.5 in [2] or Proposition 4.2 in [1], where it is shown how the Markov property follows from the uniqueness in law of solutions to SDEs with jumps). As a consequence, we see that the inequality (2.14) actually implies that for any $0 \leq s \leq t$ we have

$$\mathbb{E}[|X_t - Y_t|/\mathcal{F}_s] \leq \tilde{C}e^{-\tilde{c}(t-s)}|X_s - Y_s|.$$

Remark 2.6. Theorem 2.3 is obtained based on Theorem 3.1 which is presented later in this paper. It is however possible to obtain analogous (but perhaps less useful) result based on the already mentioned Theorem 1.1 in [27], where we have Assumptions L3 and L4 instead of Assumption L5. Then we get an inequality of the form (2.11). It is still possible to obtain some transportation inequalities if in Theorem 2.1 we replace the condition (2.2) with a condition like (2.11), but because of its form it forces us to additionally assume that the underlying intensity measure is finite (see Remark 6.1).

The above result is proved using the coupling methods developed in [27] and [12], and is of independent interest, as it extends some of the results obtained there. In particular, it immediately allows us to obtain exponential (weak) contractivity of the transition semigroup $(p_t)_{t \geq 0}$ associated with the solution to (2.13), with respect to the L^1 -Wasserstein distance W_1 , as shown by the following corollary.

Corollary 2.7. *Under the assumptions of Theorem 2.3,*

$$W_1(\eta p_t, \mu p_t) \leq \tilde{C}e^{-\tilde{c}t}W_1(\eta, \mu)$$

for any probability measures η and μ on \mathbb{R}^d and for any $t > 0$. Moreover, $(p_t)_{t \geq 0}$ has an invariant measure μ_0 and we have

$$W_1(\eta p_t, \mu_0) \leq \tilde{C}e^{-\tilde{c}t}W_1(\eta, \mu_0)$$

for any probability measure η on \mathbb{R}^d and any $t > 0$.

This result follows immediately from (2.14) like in the proof of Corollary 3 in [12] or the beginning of Section 3 in [20]. Using couplings allows us also to prove a related result involving a perturbation of the solution to (2.13) by a change in the drift. This gives us a tool to determine some concrete cases in which the assumption (2.3) from Theorem 2.1 holds.

Theorem 2.8. *Let $(X_t)_{t \geq 0}$ be like in Theorem 2.3 and suppose additionally that Assumption D2 holds, $\det \sigma_1 > 0$ and that the coefficients σ and g are Lipschitz. Consider*

a process $(\tilde{X}_t)_{t \geq 0}$ which is a solution to (2.13) with the drift perturbed by u_t , i.e.,

$$d\tilde{X}_t = b(\tilde{X}_t)dt + u_t dt + \sigma_1 dB_t^1 + \sigma(\tilde{X}_t)dB_t^2 + dL_t + \int_U g(\tilde{X}_{t-}, u)\tilde{N}(dt, du),$$

where u_t is either $\sigma_1 h_t$ or $\sigma(\tilde{X}_t)h_t$ for some adapted d -dimensional process h_t . Then there exists a process $(Y_t)_{t \geq 0}$ such that $(X_t, Y_t)_{t \geq 0}$ is a coupling of solutions to (2.13) and for any $0 \leq s \leq t$ we have

$$(2.15) \quad \mathbb{E}[|\tilde{X}_t - Y_t|/\mathcal{F}_s] \leq C \int_s^t e^{c(r-(t-s))}|u_r|dr,$$

where the constants $C, c > 0$ are given by (4.9) and (4.5), respectively.

Observe that the constants above depend on the function κ and hence to calculate their explicit values we need to apply the right version of κ in the formulas (4.9) and (4.5), i.e., the version that is used in the proof of Theorem 2.8. Now, combining Theorems 2.3 and 2.8 to check validity of assumptions of Theorem 2.1, we get the following result.

Corollary 2.9. *Consider the setup of Theorem 2.3. Suppose all its assumptions and Assumption D2 are satisfied and additionally that $\det \sigma_1 > 0$ and the coefficients σ and g are Lipschitz. Moreover, assume that $(X_t)_{t \geq 0}$ is Malliavin differentiable ($X_t \in \mathbb{D}^{1,2}$ for all $t \geq 0$) and, similarly to Theorem 2.1, that there exists a constant σ_∞ such that for any $x \in \mathbb{R}^d$ we have $\|\sigma(x)\| \leq \sigma_\infty$ and there exists a measurable function $g_\infty : U \rightarrow \mathbb{R}$ such that $|g(x, u)| \leq g_\infty(u)$ for any $x \in \mathbb{R}^d$ and $u \in U$. Assume further that there exists some $\lambda > 0$ such that Assumption E is satisfied and that there exists $\tilde{\lambda} > 0$ such that*

$$\beta^L(\tilde{\lambda}) := \int_U (e^{\tilde{\lambda}u} - \tilde{\lambda}u - 1)\nu^L(du) < \infty.$$

Then the transportation inequality (2.4) from the statement of Theorem 2.1 holds with

$$\alpha_T(r) := \sup_{\lambda \geq 0} \left\{ r\lambda - \int_0^T \beta(\tilde{C}e^{-\tilde{c}(T-t)}\lambda)dt - \int_0^T \beta^L(\tilde{C}e^{-\tilde{c}(T-t)}\lambda)dt - \frac{(\sigma_\infty^2 + \|\sigma_1\|^2)\lambda^2}{2} C^2 \frac{1 - e^{-2cT}}{2c} \right\}.$$

Moreover, for the invariant measure μ_0 we have

$$(2.16) \quad \alpha_\infty(W_1(\eta, \mu_0)) \leq H(\eta|\mu_0)$$

for any probability measure η on \mathbb{R}^d , with α_∞ defined as the pointwise limit of α_T as $T \rightarrow \infty$. Finally, the inequality (2.5) holds with

$$\alpha_T^P(r) := \sup_{\lambda \geq 0} \left\{ r\lambda - \int_0^T \beta \left(\lambda \tilde{C} \frac{1 - e^{-\tilde{c}(T-t)}}{\tilde{c}} \right) dt - \int_0^T \beta^L \left(\lambda \tilde{C} \frac{1 - e^{-\tilde{c}(T-t)}}{\tilde{c}} \right) dt - \frac{(\sigma_\infty^2 + \|\sigma_1\|^2)\lambda^2}{2} C^2 \int_0^T \left(\frac{1 - e^{-c(T-t)}}{c} \right)^2 dt \right\}.$$

The constants \tilde{c}, \tilde{C}, c and C appearing in the definitions of α_T and α_T^P are the same as in (2.14) and (2.15).

This corollary extends the results from Theorem 2.2 in [39] to the case where we drop the global dissipativity assumption required therein, as long as we have an additive component of the noise, which we can use in order to construct a coupling required in

our method. It is easy to notice that the corollaries in Section 2 in [39] (various results regarding concentration of measure for solutions of (2.13) in the pure jump case) hold as well under our assumptions. We also extend Theorem 2.2 from [25], where similar results are proved in the jump diffusion case under assumptions analogous to the ones in [39]. However, in [25] there are additionally stronger assumptions on regularity of the coefficients, which are needed to get bounds on Malliavin derivatives of solutions to (2.13). Here we use a different method of getting such bounds (cf. Remark 2.16) which does not require coefficients to be differentiable and works whenever we have $X_t \in \mathbb{D}^{1,2}$ for all $t \geq 0$.

Example 2.10. To have a jump noise satisfying all the assumptions of Corollary 2.9, we can take a Lévy process whose Lévy measure behaves near the origin like that of an α -stable process with $\alpha \in (1, 2)$ (so that Assumptions L1-L2 and L5 are satisfied), but has exponential moments as well (so that Assumption E is also satisfied). A natural example of such a process is the so called relativistic α -stable process, which is a Lévy process $(L_t)_{t \geq 0}$ with the characteristic function given by

$$\mathbb{E} \exp(i\langle z, L_t \rangle) = \exp\left(-t \left[(m^{1/\beta} + |z|^2)^\beta - m \right]\right)$$

for $z \in \mathbb{R}^d$, with $\beta = \alpha/2$ and some parameter $m > 0$. For more information on this process, see e.g. [9] where Corollary II.2 and Proposition II.5 show that it indeed satisfies Assumption E, or [35] where in Lemma 2 the formula for the density of its Lévy measure is calculated, from which we can easily see that Assumption L5 holds. SDEs driven by relativistic stable processes (and in fact also by a significantly more general type of noise) have been recently studied in [18].

Remark 2.11. Both in [39] and [25], apart from the transportation inequalities for measures $\delta_x p_T$ on \mathbb{R}^d and for measures $\mathbb{P}_{x,[0,T]}$ on the path space with the L^1 metric, there are also inequalities on the path space with the L^∞ metric defined by $d_\infty(\gamma_1, \gamma_2) = \sup_{t \in [0,T]} |\gamma_1(t) - \gamma_2(t)|$ (see Theorem 2.11 in [39] and Theorem 2.8 in [25]). However, the method of proof for these (see the second part of the proof of Lemma 3.3 in [25]) involves proving an inequality of the type

$$\mathbb{E} \sup_{0 \leq s \leq t} |X_s(x) - X_s(y)|^2 \leq e^{\widehat{C}t} |x - y|^2,$$

with some constant \widehat{C} , which requires the integral form of the Gronwall inequality, which can only work if the constant \widehat{C} is positive (cf. Remark 2.3 in [36]). Since this is the case even under the global dissipativity assumption, we have not been able to use couplings to improve on these results in any way and hence we skip them in our presentation, referring the interested reader to [39] and [25].

Remark 2.12. Another possible application of our approach would be to extend the results from [36], where transportation inequalities were studied in the context of regime switching processes, modelled by stochastic differential equations with both Gaussian and Poissonian noise (see (2.1) and (2.2) therein). There a kind of one-sided Lipschitz condition is imposed on the coefficients (see the condition (A3) in [36]) and, as pointed out in Remark 2.2 therein, transportation inequalities on the path space can be obtained without dissipativity. However, in such a case the constants with which those inequalities hold for $\mathbb{P}_{x,[0,T]}$, explode when $T \rightarrow \infty$ (see Theorem 2.1 in [36]). Since the method of proof used in [36] is a direct extension of the one developed by Liming Wu in [39], it should be possible to apply our reasoning to obtain non-exploding constants at least in

(2.13) in [36] under a dissipativity at infinity condition. This is, however, beyond the scope of our present paper.

Remark 2.13. In the present paper, we only explain in detail how to check assumptions of Theorem 2.1 using the approach of Theorems 2.3 and 2.8. However, it may be possible to obtain inequalities like (2.2) and (2.3) by other methods. For example, in a recent paper [23], D. Luo and J. Wang obtained an inequality like (2.2) for equations of the type $dX_t = b(X_t)dt + dL_t$ under different than ours assumptions on the Lévy measure and using a different coupling (see Theorem 1.1 therein; (1.6) in [23] follows from an inequality like (2.2) which is needed in the proof of Theorem 3.1 therein). This is sufficient to get the transportation inequalities like in our Theorem 2.1 for an SDE with pure jump noise under their set of assumptions (plus, additionally, Assumption E). On the other hand, Eberle, Guillin and Zimmer in [13] showed an inequality like (2.2) for equations of the type $dX_t = b(X_t)dt + dB_t$ without assuming dissipativity even at infinity, at the cost of multiplying the right hand side of (2.2) by a factor which, however, can possibly be controlled under some suitable integrability assumptions for $(X_t)_{t \geq 0}$, cf. Theorem 2 and formula (28) in [13]. This could lead to obtaining at least some concentration inequalities like (2.6) for solutions of equations of the type $dX_t = b(X_t)dt + dB_t + \int_U g(X_{t-}, u) \tilde{N}(dt, du)$, under some weaker than ours assumptions on the coefficients b and g . These examples show robustness of our formulation of Theorem 2.1, as it allows us to easily obtain transportation or concentration inequalities in many cases where inequalities like (2.2) arise naturally.

The crucial step in the proof of Theorem 2.1 is to find upper bounds for the Malliavin derivatives of X_t . Thus, in the process of proving our main results, we also obtain some bounds that might be interesting on their own in the context of the Malliavin calculus.

Theorem 2.14. *Let $(X_t)_{t \geq 0}$ be a Malliavin differentiable solution to (2.1) such that (2.3) holds. Assume that there exists a constant σ_∞ such that for any $x \in \mathbb{R}^d$ we have $\|\sigma(x)\| \leq \sigma_\infty$. Then for any Lipschitz functional $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$, for any adapted, \mathbb{R}_+ -valued process g and for any $0 \leq s \leq r \leq t$ we have*

$$(2.17) \quad \mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla_u f(X_t) | \mathcal{F}_u]|^2 du \leq c_2^2(t) \sigma_\infty \mathbb{E} \int_s^r g_u c_3^2(u) du.$$

Moreover, we have

$$(2.18) \quad \|\mathbb{E}[\nabla_u f(X_t) | \mathcal{F}_u]\|_{L^\infty(\Omega \times [0, t])} \leq c_2(t) \sigma_\infty \sup_{u \leq t} c_3(u),$$

where the L^∞ norm is the essential supremum on $\Omega \times [0, t]$.

On the other hand, using the condition (2.2), we can obtain related bounds for the Malliavin derivative D of Lipschitz functionals of X_t with respect to the Poisson random measure N (see Section 5.2 for details).

In the same way in which Corollary 2.9 follows from Theorem 2.1 via Theorems 2.3 and 2.8, the following corollary follows from Theorem 2.14 via Theorem 2.8.

Corollary 2.15. *Let $(X_t)_{t \geq 0}$ be a Malliavin differentiable solution to (2.13), satisfying the assumptions of Theorem 2.3 with $\det \sigma_1 > 0$ and $\lim_{r \rightarrow 0} r \kappa(r) = 0$ (i.e., Assumption D2). Moreover, assume that the coefficients σ and g are Lipschitz and that there exists a constant σ_∞ such that for any $x \in \mathbb{R}^d$ we have $\|\sigma(x)\| \leq \sigma_\infty$. Denote by ∇^i the Malliavin derivative with respect to $(B_t^i)_{t \geq 0}$ for $i \in \{1, 2\}$. Then for any functional f and any process g like above and for any $0 \leq s \leq r \leq t$ we have*

$$(2.19) \quad \mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla_u^1 f(X_t) | \mathcal{F}_u]|^2 du \leq C^2 \|\sigma_1\|^2 \mathbb{E} \int_s^r g_u e^{2c(u-t)} du$$

and

$$(2.20) \quad \mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla_u^2 f(X_t) | \mathcal{F}_u]|^2 du \leq C^2 \sigma_\infty^2 \mathbb{E} \int_s^r g_u e^{2c(u-t)} du,$$

where C and c are the same as in (2.15). We also have L^∞ bounds analogous to (2.18) for $\nabla_u^1 f(X_t)$ and $\nabla_u^2 f(X_t)$, with the upper bound being, respectively, $\|\sigma_1\|$ and σ_∞ .

In analogy to our comment below the statement of Theorem 2.14, we observe here that a related corollary for the Malliavin derivatives with respect to $(L_t)_{t \geq 0}$ and N is also true (see the end of Section 5.2, in particular (5.25) and (5.26)).

Remark 2.16. If the global dissipativity assumption is satisfied and the coefficients in the equation are continuously differentiable, it is possible to obtain much stronger bounds than (2.19) and (2.20). Namely, for the multiplicative noise we get

$$\mathbb{E}[\|\nabla_s X_t\|_{HS}^2 | \mathcal{F}_s] \leq \|\sigma(X_s)\|_{HS}^2 e^{2K(s-t)}$$

for any $0 \leq s \leq t$, where $K > 0$ is the constant with which the global dissipativity condition holds (see Lemma 3.4 in [25]). We were not able to obtain such bounds in our case. However, our assumptions are much weaker than the ones in [25] and the bounds (2.19) and (2.20) are sufficient to prove the transportation inequalities in Corollary 2.9. On the other hand, our bounds for the Malliavin derivative D with respect to the Poisson random measure N have the same form as the ones in [39] and [25] (cf. Section 5.2 in the present paper and Section 4.2 in [39]).

The remainder of the paper is organized as follows. In Section 3 we present an extension of the results from [27] regarding couplings of solutions to SDEs driven by pure jump Lévy noise. In Section 4 we explain how to further extend these results to the case of more general jump diffusions and hence we prove Theorem 2.3. In Section 5.1 we introduce our technique of obtaining estimates like (2.3) in Lemma 5.1, which then leads directly to the proofs of Theorem 2.8 and Theorem 2.14, followed by the proof of Corollary 2.15. In Section 5.2 we explain how to show related results in the case of Malliavin derivatives with respect to Poisson random measures. In Section 6 we finally prove the transportation and concentration inequalities, i.e., Theorem 2.1, Theorem 2.2 and Corollary 2.9.

3. COUPLING OF SDEs WITH PURE JUMP NOISE

Here we consider an SDE of the form

$$(3.1) \quad dX_t = b(X_t)dt + dL_t,$$

where $(L_t)_{t \geq 0}$ is a pure jump Lévy process and the drift function b is continuous and satisfies a one-sided Lipschitz condition. In this section, let N be the Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ associated with $(L_t)_{t \geq 0}$ via

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

and let $dt \nu(dv)$ be its intensity measure. Following Section 2.2 in [27], we can replace N with a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d \times [0, 1]$ with intensity $dt \nu(dv) du$, where du is the Lebesgue measure on $[0, 1]$, thus introducing an additional control variable $u \in [0, 1]$. By a slight abuse of notation, we keep denoting this new Poisson random measure by N . We can thus write (3.1) as

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

Without loss of generality, we can choose a constant $m > 1$ and rewrite the equation above as

$$(3.2) \quad dX_t = b(X_t)dt + \int_{\{|v|>m\} \times [0,1]} vN(dt, dv, du) + \int_{\{|v|\leq m\} \times [0,1]} v\tilde{N}(dt, dv, du).$$

Formally we should then change the drift function by an appropriate constant, but since such an operation does not change any relevant properties of the drift, we choose to keep denoting the drift by b . Now we can define a coupling $(X_t, Y_t)_{t \geq 0}$ by putting

$$(3.3) \quad \begin{aligned} dY_t &= b(Y_t)dt + \int_{\{|v|>m\} \times [0,1]} vN(dt, dv, du) \\ &+ \int_{\{|v|\leq m\} \times [0,1]} (X_{t-} - Y_{t-} + v)\mathbf{1}_{\{u < \rho(v, Z_{t-})\}}\tilde{N}(dt, dv, du) \\ &+ \int_{\{|v|\leq m\} \times [0,1]} R(X_{t-}, Y_{t-})v\mathbf{1}_{\{u \geq \rho(v, Z_{t-})\}}\tilde{N}(dt, dv, du), \end{aligned}$$

for $t < T := \inf\{t > 0 : X_t = Y_t\}$ and $Y_t = X_t$ for $t \geq T$, where $Z_t := X_t - Y_t$,

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

and

$$(3.4) \quad R(X_{t-}, Y_{t-}) := I - 2\frac{(X_{t-} - Y_{t-})(X_{t-} - Y_{t-})^T}{|X_{t-} - Y_{t-}|^2} = I - 2e_{t-}e_{t-}^T,$$

with $e_t := (X_t - Y_t)/|X_t - Y_t|$. Intuitively, it is a combination of a modification of the reflection coupling with a positive probability of bringing the marginal processes together instead of performing the reflection (for jumps of size smaller than m) and the synchronous coupling (for jumps larger than m). We can call it the mirror coupling. For the coupling construction itself, m can be chosen arbitrarily. For obtaining convergence rates in Wasserstein distances, we choose m based on assumptions satisfied by the Lévy measure ν of $(L_t)_{t \geq 0}$. For the discussion explaining this construction in detail see Section 2 in [27].

Under Assumptions L1 and L2 it has been proved in [27] (see Theorem 1.1 therein) that the $2d$ -dimensional SDE given by (3.2) and (3.3) has a unique strong solution which is a coupling of solutions to (3.1). Then this coupling was used to prove that, under additional Assumptions L3 and L4 and a dissipativity at infinity condition on the drift, the inequality (2.9) holds with a discontinuous function f , i.e., we have

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq Le^{-\theta t}f(|x - y|)$$

for some constants $L > 1$ and $\theta > 0$.

Now we turn to the proof of a modification of the main result in [27], which will give us an inequality like (2.9), but with a continuous function f . Recall that in the case of an equation of the form (3.1), the function κ is such that for all $x, y \in \mathbb{R}^d$ we have

$$(3.5) \quad \langle b(x) - b(y), x - y \rangle \leq -\kappa(|x - y|)|x - y|^2.$$

Theorem 3.1. *Let $(X_t)_{t \geq 0}$ be a Markov process in \mathbb{R}^d given as a solution to the stochastic differential equation (3.1), where $(L_t)_{t \geq 0}$ is a pure jump Lévy process satisfying Assumptions L1-L2 and Assumption L5 and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a continuous, one-sided Lipschitz vector field satisfying Assumptions D1 and D2. Then there exists a coupling of solutions*

to (3.1) defined as a strong solution to the 2d-dimensional SDE given by (3.2) and (3.3) and a continuous concave function f_1 such that

$$\mathbb{E}f_1(|X_t(x) - Y_t(y)|) \leq e^{-c_1 t} f_1(|x - y|)$$

holds with some constant $c_1 > 0$ for any $t > 0$ and any $x, y \in \mathbb{R}^d$. By the construction of f_1 , we also have

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq L e^{-c_1 t} |x - y|$$

with some constant $L > 0$.

Proof. The existence of the coupling as a strong solution to the system (3.2)-(3.3) has been proved in Section 2 in [27]. Now we will explain how to modify the proof of the inequality (1.8) in Theorem 1.1 in [27] in order to prove the new result presented here. Denote

$$Z_t := X_t - Y_t.$$

Using the expression (3.3) for dY_t , we can write

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

where $A(X_{t-}, Y_{t-}, v, u) := -(Z_{t-} + v - R(X_{t-}, Y_{t-})v) \mathbf{1}_{\{u < \rho(v, Z_{t-})\}}$. Applying the Itô formula (see e.g. Theorem 4.4.10 in [2]) with a function f_1 we get

(3.6)

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

where the last term is obtained from the usual sum over jumps appearing in the Itô formula by applying the Taylor formula and using the fact that in our coupling the vectors Z_{s-} and ΔZ_s are always parallel (see Section 3 in [27] for details). Now we introduce a sequence of stopping times $(\tau_n)_{n=1}^\infty$ defined by

$$(3.7) \quad \tau_n := \inf\{t \geq 0 : |Z_t| \notin (1/n, n)\}.$$

Note that we have $\tau_n \rightarrow T$ as $n \rightarrow \infty$, which follows from non-explosiveness of $(Z_t)_{t \geq 0}$. By some tedious but otherwise easy computations (see the proof of Theorem 1.1 in [27] for details, specifically Lemma 3.1 and Lemma 3.2 therein) we can show that

$$(3.8) \quad \mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} f_1'(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, (I - R(X_{s-}, Y_{s-}))v \rangle \tilde{N}(ds, dv, du) = 0.$$

and

$$(3.9) \quad \mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|v| \leq m\} \times [0,1]} f_1'(|Z_{s-}|) \frac{1}{|Z_{s-}|} \langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u) \rangle \tilde{N}(ds, dv, du) = 0.$$

In [27] it is also shown (see Lemma 3.3 therein) that for any $t > 0$, we have

$$(3.10) \quad \mathbb{E} \sum_{s \in (0, t]} \left(|\Delta Z_s|^2 \int_0^1 (1-u) f_1''(|Z_{s-} + u\Delta Z_s|) du \right) \leq C_\varepsilon \mathbb{E} \int_0^t \bar{f}_\varepsilon(|Z_{s-}|) \mathbf{1}_{\{|Z_{s-}| > \delta\}} ds,$$

where $0 < \delta < 2m$, $\varepsilon \leq \delta$, the constant C_ε is defined as in (2.12) with the first marginal ν_1 of the measure ν and the function \bar{f}_ε is defined by

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

It is important to note that in order for (3.10) to hold, m has to be chosen in such a way that

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

where ν_1^m is the first marginal of the truncated measure $\nu^m(dv) := \mathbf{1}_{\{|v| \leq m\}} \nu(dv)$. This is, however, not a problem, since m can always be chosen large enough, cf. the discussion in Section 2.2 in [27]. The crucial element of the proof in [27], after getting the bounds (3.8), (3.9) and (3.10), is the construction of a function f_1 and a constant $c_1 > 0$ such that

$$(3.11) \quad -f_1'(r) \kappa(r) r + C_\varepsilon \bar{f}_\varepsilon(r) \leq -c_1 f_1(r)$$

holds for all $r > \delta$, where κ is the function satisfying (3.5). Combining this with (3.6) and using Assumption L3 and the discontinuity of the distance function to deal with the case of $r \leq \delta$ (see Lemma 3.7 in [27]), it is shown how to get a bound of the form

$$\mathbb{E} f_1(|Z_{t \wedge \tau_n}|) - \mathbb{E} f_1(|Z_0|) \leq \mathbb{E} \int_0^{t \wedge \tau_n} -c_1 f_1(|Z_s|) ds,$$

which then leads to (2.9). Now we will show a different way of dealing with the case of $r \leq \delta$, using Assumption L5 instead of Assumption L3, which allows us to keep the continuity of f_1 .

It is quite easy to see (using once again the fact that Z_{s-} and ΔZ_s are parallel, cf. the proof of Lemma 3.3 in [27]) that for any $u \in (0, 1)$ we have

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

We also have

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

and the condition $|Z_s| > |Z_{s-}|$ is equivalent to $\langle \Delta Z_s, 2Z_{s-} + \Delta Z_s \rangle > 0$. Therefore, mimicking the argument in the proof of Lemma 3.3 in [27] we get that

$$(3.12) \quad \mathbb{E} \sum_{s \in (0, t]} \left(|\Delta Z_s|^2 \int_0^1 (1-u) f_1''(|Z_{s-} + u\Delta Z_s|) du \right) \leq C_\varepsilon \mathbb{E} \int_0^t \hat{f}_\varepsilon(|Z_{s-}|) ds,$$

where

$$\hat{f}_\varepsilon(y) := \sup_{x \in (y-\varepsilon, y)} f_1''(x) \mathbf{1}_{\{y > \delta\}} + \sup_{x \in (y, y+\varepsilon)} f_1''(x) \mathbf{1}_{\{y \leq \delta\}}.$$

Now we will show that under Assumption L5, after a small modification in the formulas from [27], the inequality

$$(3.13) \quad -f_1'(r) \kappa(r) r + C_\varepsilon \hat{f}_\varepsilon(r) \leq -c_1 f_1(r)$$

holds for all $r > 0$ (note that here we have $\hat{f}_\varepsilon(r)$ in place of $\bar{f}_\varepsilon(r)$ in (3.11)). The function f_1 , constructed in Lemma 3.6 in [27] in order to satisfy (3.11), is such that $f_1' \geq 0$, $f_1'' \leq 0$ and is defined in the following way

$$(3.14) \quad f_1(r) = \int_0^r \phi(s)g(s)ds,$$

where

$$\phi(r) := \exp\left(-\int_0^r \frac{\bar{h}(t)}{C_\varepsilon} dt\right), \quad g(r) := \begin{cases} 1 - \frac{c_1}{C_\varepsilon} \int_0^r \frac{\Phi(t+\varepsilon)}{\phi(t)} dt, & r \leq R_1, \\ \frac{1}{2}, & r \geq R_1. \end{cases}$$

Here $R_1 > 0$ is given by formulas

$$(3.15) \quad \begin{aligned} R_0 &= \inf \left\{ R \geq 0 : \forall r \geq R : \kappa(r) \geq \frac{2M}{R} \right\}, \\ R_1 &= \inf \left\{ R \geq R_0 + \varepsilon : \forall r \geq R : \kappa(r) \geq \frac{2C_\varepsilon}{(R - R_0)R} + \frac{2M}{R} \right\}, \end{aligned}$$

but can be chosen arbitrarily large if necessary and c_1 is a positive constant given by

$$(3.16) \quad c_1 := \frac{C_\varepsilon}{2} \left(\int_0^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)} dt \right)^{-1}$$

(cf. (3.29) in [27]). Moreover, we have

$$\bar{h}(r) := \sup_{t \in (r, r+\varepsilon)} h^-(t), \quad \Phi(r) := \int_0^r \phi(s)ds$$

and $h^- = -\min\{h, 0\}$ is the negative part of the function

$$(3.17) \quad h(r) := r\kappa(r) - 2M,$$

with some $M > 0$ to be chosen later. Actually, in Lemma 3.6 in [27] the function h is given by $h(r) := r\kappa(r)$, whereas R_0 and R_1 are chosen with $M = 0$ and this already gives (3.11). However, it is easy to check that by taking $M > 0$ we get

$$(3.18) \quad -f_1'(r)\kappa(r)r + 2f_1'(r)M + C_\varepsilon\bar{f}_\varepsilon(r) \leq -c_1f_1(r).$$

Indeed, all the calculations in the proof of Lemma 3.6 in [27] are expressed in terms of a function h , which can be modified if necessary. It is enough to ensure that we choose R_0 such that $h^-(r) = 0$ for $r \geq R_0$ and then R_1 such that $(-r\kappa(r) + 2M)/2 \leq -C_\varepsilon r / (R_1 - R_0)R_1$ for $r \geq R_1$, which obviously holds for the choice of h , R_0 and R_1 presented above. Moreover, we obviously have

$$-f_1'(r)\kappa(r)r + C_\varepsilon\bar{f}_\varepsilon(r) \leq -f_1'(r)\kappa(r)r + 2f_1'(r)M + C_\varepsilon\bar{f}_\varepsilon(r),$$

and thus if we have (3.18) with some $M > 0$ for $r > \delta$, then (3.11) is still valid for $r > \delta$. The reason we introduce the constant M is that it is needed to show that

$$\sup_{x \in (r, r+\varepsilon)} f_1''(x) \leq -\frac{c_1}{C_\varepsilon} f_1(r) + f_1'(r) \frac{r\kappa(r)}{C_\varepsilon}$$

holds for all $r \leq \delta$, which, combined with (3.11) for $r > \delta$, will give us (3.13). Hence we need to show that for any $s \in (r, r + \varepsilon)$ we have

$$f_1''(s) \leq -\frac{c_1}{C_\varepsilon} f_1(r) + f_1'(r) \frac{r\kappa(r)}{C_\varepsilon}.$$

First let us calculate (recall that we can choose R_1 large enough so that $s < \delta + \varepsilon < R_1$)

$$\begin{aligned} f_1''(s) &= \phi(s) \left(-\frac{c_1}{C_\varepsilon} \frac{\Phi(s+\varepsilon)}{\phi(s)} \right) + \left(-\frac{\bar{h}(r)}{C_\varepsilon} \phi(s) g(s) \right) \\ &= -\frac{c_1}{C_\varepsilon} \Phi(s+\varepsilon) - \frac{\bar{h}(r)}{C_\varepsilon} f_1'(s). \end{aligned}$$

Observe now that for $r \leq s$ we have $f_1(r) \leq f_1(s) \leq \Phi(s) \leq \Phi(s+\varepsilon)$ and thus

$$f_1''(s) \leq -\frac{c_1}{C_\varepsilon} f_1(r) - \frac{\bar{h}(r)}{C_\varepsilon} f_1'(s).$$

Therefore it remains to be shown that

$$-\frac{\bar{h}(r)}{C_\varepsilon} f_1'(s) \leq f_1'(r) \frac{r\kappa(r)}{C_\varepsilon}.$$

Actually, we will just show that

$$(3.19) \quad \frac{1}{C_\varepsilon} f_1'(s) h(s) \leq f_1'(r) \frac{r\kappa(r)}{C_\varepsilon}.$$

Then, since $-h^- \leq h$ and $s \in (r, r+\varepsilon)$, we will get

$$\begin{aligned} -\frac{1}{C_\varepsilon} f_1'(s) \sup_{t \in (r, r+\varepsilon)} h^-(t) &= \frac{1}{C_\varepsilon} f_1'(s) \inf_{t \in (r, r+\varepsilon)} (-h^-(t)) \\ &\leq \frac{1}{C_\varepsilon} f_1'(s) \inf_{t \in (r, r+\varepsilon)} h(t) \\ &\leq f_1'(r) \frac{r\kappa(r)}{C_\varepsilon}. \end{aligned}$$

In order to show (3.19) we observe that straight from the definition of h we have

$$\frac{1}{C_\varepsilon} f_1'(s) h(s) = \frac{1}{C_\varepsilon} f_1'(s) (s\kappa(s) - 2M)$$

and then we calculate

$$\begin{aligned} f_1'(s) (s\kappa(s) - 2M) &= f_1'(r) r\kappa(r) - f_1'(r) r\kappa(r) \\ &\quad + f_1'(s) r\kappa(r) - f_1'(s) r\kappa(r) \\ &\quad + f_1'(s) s\kappa(s) - 2M f_1'(s) \\ &\leq f_1'(r) r\kappa(r) + r\kappa(r) (f_1'(s) - f_1'(r)) \\ &\quad + f_1'(s) (s\kappa(s) - r\kappa(r)) - 2M f_1'(s). \end{aligned}$$

Now it is enough to show that it is possible to choose ε , δ and M in such a way that the sum of the last three terms is bounded by some non-positive quantity. Since we have Assumption D2, for any $\lambda > 0$ there exists some $K(\lambda) > 0$ such that for all $|r| < \lambda$ we have $|r\kappa(r)| \leq K(\lambda)$. Since $s < r + \varepsilon \leq \delta + \varepsilon$, we obtain

$$s\kappa(s) - r\kappa(r) \leq 2K(\delta + \varepsilon).$$

We also know that f_1' is non-increasing and thus $f_1'(s) \leq f_1'(r)$, but the sign of $r\kappa(r)$ is unknown so we cannot just bound $r\kappa(r)(f_1'(s) - f_1'(r))$ by zero. We will deal with this term in a more complicated way. We have

$$\begin{aligned} f_1'(s) - f_1'(r) &= \phi(s)g(s) - \phi(s)g(r) + \phi(s)g(r) - \phi(r)g(r) \\ &= \phi(s)(g(s) - g(r)) + (\phi(s) - \phi(r))g(r). \end{aligned}$$

We also have

$$|\phi(s)(g(s) - g(r))| \leq 2\phi(s) \leq 4f_1'(s),$$

since $1/2 \leq g \leq 1$. Furthermore

$$\begin{aligned} |(\phi(s) - \phi(r))g(r)| &= |\phi(s)(1 - \phi(s)^{-1}\phi(r))g(r)| \\ &= \left| \phi(s) \left(1 - \exp\left(\int_r^s \frac{\bar{h}(t)}{C_\varepsilon} dt\right) \right) g(r) \right| \\ &\leq 2f_1'(s) \int_r^s \frac{\bar{h}(t)}{C_\varepsilon} dt \exp\left(\int_r^s \frac{\bar{h}(t)}{C_\varepsilon} dt\right) \\ &\leq \frac{2f_1'(s)\varepsilon}{C_\varepsilon} (2M + K(\delta + 2\varepsilon)) \exp\left(\frac{\varepsilon}{C_\varepsilon} (2M + K(\delta + 2\varepsilon))\right), \end{aligned}$$

where in the first inequality we have used the fact that $|1 - e^x| \leq |xe^x|$ for all $x \geq 0$ and that $g \leq 1$ and $\phi(s) \leq 2f_1'(s)$. In the second inequality we used

$$\int_r^s \frac{\bar{h}(t)}{C_\varepsilon} dt \leq \frac{\varepsilon}{C_\varepsilon} (2M + K(\delta + 2\varepsilon)),$$

which holds since $|s - r| < \varepsilon$. Thus if we find δ , ε and M such that

$$K(\delta) \left(4 + 2\frac{\varepsilon}{C_\varepsilon} (2M + K(\delta + 2\varepsilon)) \exp\left(\frac{\varepsilon}{C_\varepsilon} (2M + K(\delta + 2\varepsilon))\right) \right) + 2K(\delta + \varepsilon) \leq 2M,$$

then (3.19) holds and we prove our statement. This is indeed possible since we assume that $\varepsilon/C_\varepsilon$ is bounded in a neighbourhood of zero. \square

4. COUPLING OF JUMP DIFFUSIONS

Here we study jump diffusions of more general form (2.13) and we prove Theorem 2.3. In order to do this, we first recall results obtained by Eberle in [12] for diffusions of the form

$$dX_t = b(X_t)dt + \sigma_1 dB_t,$$

where σ_1 is a constant non-degenerate $d \times d$ matrix and $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion. Eberle used the coupling by reflection $(X_t, Y_t)_{t \geq 0}$, defined by

$$(4.1) \quad dY_t = \begin{cases} b(Y_t)dt + \sigma_1 R_{\sigma_1}(X_t, Y_t)dB_t & \text{for } t < T, \\ dX_t & \text{for } t \geq T, \end{cases}$$

where $T := \inf\{t \geq 0 : X_t = Y_t\}$ is the coupling time and

$$(4.2) \quad R_{\sigma_1}(X_t, Y_t) := I - 2e_t e_t^T$$

with

$$(4.3) \quad e_t := \sigma_1^{-1}(X_t - Y_t) / |\sigma_1^{-1}(X_t - Y_t)|.$$

Using this coupling, Eberle constructed a concave continuous function f given by

$$(4.4) \quad f(r) := \int_0^r \varphi(s)g(s)ds,$$

where

$$\varphi(r) := \exp\left(-\frac{1}{2} \int_0^r s\kappa^-(s)ds\right), \quad g(r) := \begin{cases} 1 - \frac{\alpha c}{2} \int_0^r \frac{\Phi(t)}{\varphi(t)} dt, & r \leq R_1, \\ \frac{1}{2}, & r \geq R_1, \end{cases}$$

with $\Phi(r) := \int_0^r \varphi(s)ds$ and some constant $R_1 > 0$ defined by (9) in [12]. Here $c > 0$ is a constant given by

$$(4.5) \quad c = \frac{1}{\alpha} \left(\int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} ds \right)^{-1}$$

where $\alpha := \sup\{|\sigma_1^{-1}z|^2 : z \in \mathbb{R}^d \text{ with } \|z\| = 1\}$ (cf. the formula (12) in [12]) and κ is defined by

$$(4.6) \quad \kappa(r) = \inf \left\{ -\frac{|\sigma_1^{-1}(x-y)|^2}{|x-y|^4} \langle b(x) - b(y), x-y \rangle : x, y \in \mathbb{R}^d \text{ s.t. } |x-y| = r \right\}.$$

In other words, κ is the largest quantity satisfying

$$(4.7) \quad \langle b(x) - b(y), x-y \rangle \leq -\kappa(|x-y|)|x-y|^4/|\sigma_1^{-1}(x-y)|^2$$

for all $x, y \in \mathbb{R}^d$, although for our purposes we can consider any continuous function κ such that (4.7) holds. Then it is possible to prove that

$$(4.8) \quad 2f''(r) - r\kappa(r)f'(r) \leq -c\alpha f(r) \text{ for all } r > 0.$$

Note that our definition of κ differs from the one in [12] by a factor 2 to make the notation more consistent with our results for the pure jump noise case presented in the previous Section (cf. formulas in Section 2.1 in [12]). By the methods explained in the proof of Theorem 1 in [12] (see also Corollary 2 therein) we get

$$\mathbb{E}f(|X_t(x) - Y_t(y)|) \leq e^{-ct}f(|x-y|)$$

and, by the choice of f (which is comparable with the identity function, since it is extended in an affine way from $R_1 > 0$), we also get

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq Ce^{-ct}|x-y|$$

with a constant $C > 0$ defined by (cf. (14) and (8) in [12])

$$(4.9) \quad C := 2\varphi(R_0)^{-1}, \text{ where } R_0 := \inf\{R \geq 0 : \forall r \geq R \kappa(r) \geq 0\}.$$

Now we will explain how to combine the results from [12] and [27] to get analogous results for equations involving both the Gaussian and the Poissonian noise. The general idea is, similarly to [12] and [27], to use an appropriate coupling $(X_t, Y_t)_{t \geq 0}$, to write an SDE for the difference process $Z_t = X_t - Y_t$, to use the Itô formula to evaluate $df(|Z_t|)$ and then to choose f in such a way that $df(|Z_t|) \leq dM_t - \tilde{c}f(|Z_t|)dt$ for some constant $\tilde{c} > 0$, where $(M_t)_{t \geq 0}$ is a local martingale.

Proof of Theorem 2.3. We consider an equation of the form

$$dX_t = b(X_t)dt + \sigma_1 dB_t^1 + \sigma(X_t)dB_t^2 + dL_t + \int_U g(X_{t-}, u) \tilde{N}(dt, du),$$

where $\sigma_1 > 0$ is a constant and all the other coefficients and the sources of noise are like in the formulation of Theorem 2.3 (in particular, here we denote the underlying Poisson random measure of $(L_t)_{t \geq 0}$ by N^L and its associated Lévy measure by ν^L). Restricting ourselves to a real constant in front of $(B_t^1)_{t \geq 0}$ instead of a matrix helps us to slightly reduce the notational complexity and seems in fact quite natural at least for the equations for which Lemma 4.1 applies. Recall that κ is such that for all $x, y \in \mathbb{R}^d$ we have

$$(4.10) \quad \langle b(x) - b(y), x-y \rangle + \frac{1}{2} \int_U |g(x, u) - g(y, u)|^2 \nu(du) + \|\sigma(x) - \sigma(y)\|_{HS}^2 \leq -\kappa(|x-y|)|x-y|^2$$

and that it satisfies Assumption D1. Now we will apply the mirror coupling from [27] to $(L_t)_{t \geq 0}$, by using the “mirror operator” $M(\cdot, \cdot)$, i.e., recalling the notation used in the equations (3.2) and (3.3), we define

$$\begin{aligned} M(X_{t-}, Y_{t-})L_t &:= \int_{\{|v| > m\} \times [0, 1]} v N^L(dt, dv, du) \\ &+ \int_{\{|v| \leq m\} \times [0, 1]} (X_{t-} - Y_{t-} + v) \mathbf{1}_{\{u < \rho(v, Z_{t-})\}} \widetilde{N}^L(dt, dv, du) \\ &+ \int_{\{|v| \leq m\} \times [0, 1]} R(X_{t-}, Y_{t-})v \mathbf{1}_{\{u \geq \rho(v, Z_{t-})\}} \widetilde{N}^L(dt, dv, du), \end{aligned}$$

with the reflection operator R defined by (3.4). We will also use the reflection coupling (4.1) from [12], with the reflection operator R_{σ_1} defined by (4.2) and apply it to $(B_t^1)_{t \geq 0}$. Note that if the coefficient near the Brownian motion is just a positive constant and not a matrix, the formulas from [12] become a bit simpler, in particular the unit vector e_t defined by (4.3) becomes just $(X_t - Y_t)/|X_t - Y_t|$. Thus the two reflection operators we defined coincide and we can keep denoting them both by R . Moreover, we apply the synchronous coupling to the other two noises and hence we have

$$dY_t = b(Y_t)dt + \sigma_1 R(X_t, Y_t)dB_t^1 + \sigma(Y_t)dB_t^2 + M(X_{t-}, Y_{t-})dL_t + \int_U g(Y_{t-}, u)\widetilde{N}(dt, du).$$

Since all the sources of noise are independent, it is easy to see that $(X_t, Y_t)_{t \geq 0}$ is indeed a coupling (this follows from the fact that R applied to $(B_t^1)_{t \geq 0}$ gives a Brownian motion and M applied to $(L_t)_{t \geq 0}$ gives the same Lévy process, whereas the solution to the equation above is unique in law). We can now write the equation for $Z_t := X_t - Y_t$ as

$$\begin{aligned} dZ_t &= (b(X_t) - b(Y_t))dt + 2\sigma_1 e_t e_t^T dB_t^1 + (\sigma(X_t) - \sigma(Y_t))dB_t^2 \\ &+ (I - M(X_{t-}, Y_{t-}))dL_t + \int_U (g(X_{t-}, u) - g(Y_{t-}, u))\widetilde{N}(dt, du), \end{aligned}$$

where we evaluated $\sigma_1(I - R(X_{t-}, Y_{t-}))$ as $2\sigma_1 e_t e_t^T$ and we will later use the fact that $d\widetilde{W}_t := e_t^T dB_t^1$ is a one-dimensional Brownian motion in order to simplify our calculations. We apply the Itô formula to get

$$(4.11) \quad df(|Z_t|) = \sum_{j=1}^9 I_j,$$

where

$$\begin{aligned} I_1 &:= f'(|Z_t|) \frac{1}{|Z_t|} \langle b(X_t) - b(Y_t), Z_t \rangle dt, & I_3 &:= f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, (\sigma(X_t) - \sigma(Y_t))dB_t^2 \rangle, \\ I_2 &:= 2f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, \sigma_1 e_t e_t^T dB_t^1 \rangle, & I_4 &:= f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \langle Z_{t-}, (I - M(X_{t-}, Y_{t-})) \rangle dL_t \end{aligned}$$

and

$$I_5 := f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \int_U \langle g(X_{t-}, u) - g(Y_{t-}, u), Z_{t-} \rangle \widetilde{N}(dt, du)$$

constitute the drift and the local martingale terms, while

$$I_6 := \frac{1}{2} \sigma_1^2 \sum_{i,j=1}^d \left[f''(|Z_{t-}|) \frac{Z_{t-}^i Z_{t-}^j}{|Z_{t-}|^2} + f'(|Z_{t-}|) (\delta_{ij} \frac{1}{|Z_{t-}|} - \frac{Z_{t-}^i Z_{t-}^j}{|Z_{t-}|^3}) \right] 4 \frac{Z_{t-}^i Z_{t-}^j}{|Z_{t-}|^2} dt$$

and

$$I_7 := \sum_{i,j=1}^d \left[f''(|Z_{t-}|) \frac{Z_{t-}^i Z_{t-}^j}{|Z_{t-}|^2} + f'(|Z_{t-}|) (\delta_{ij} \frac{1}{|Z_{t-}|} - \frac{Z_{t-}^i Z_{t-}^j}{|Z_{t-}|^3}) \right] \cdot \left[\sum_{k=1}^m (\sigma_{ik}(X_{t-}) - \sigma_{ik}(Y_{t-})) (\sigma_{jk}(X_{t-}) - \sigma_{jk}(Y_{t-})) \right] dt$$

come from the quadratic variation of the Brownian noises, whereas

$$I_8 := \int_U \left[f(|Z_{t-} + (I - M(X_{t-}, Y_{t-})v)|) - f(|Z_{t-}|) - \langle (I - M(X_{t-}, Y_{t-})v), \nabla f(|Z_{t-}|) \rangle \right] N^L(dt, du)$$

and

$$(4.12) \quad I_9 := \int_U \left[f(|Z_{t-} + g(X_{t-}, u) - g(Y_{t-}, u)|) - f(|Z_{t-}|) - \langle g(X_{t-}, u) - g(Y_{t-}, u), \nabla f(|Z_{t-}|) \rangle \right] N(dt, du)$$

are the jump components.

Now we proceed similarly to [12] and [27]. Since we want to obtain an estimate of the form $df(|Z_t|) \leq dM_t - \tilde{c}f(|Z_t|)dt$ and we assume that the function f is concave, we should use its second derivative to obtain a negative term on the right hand side of (4.11). In order to do this, we can use the additive Brownian noise $(B_t^1)_{t \geq 0}$ to get a negative term from I_6 (it is easy to see that it reduces to $2\sigma_1^2 f''(|Z_t|)$) and then use the function f from [12] given by (4.4), aiming to obtain an inequality like (4.8) (then we can just use the synchronous coupling for $(L_t)_{t \geq 0}$ and the terms I_4 and I_8 disappear). Alternatively, we can use the additive jump noise $(L_t)_{t \geq 0}$ to get a negative term from I_8 . As we already mentioned in Section 3 under the formula (3.6), the integral I_8 reduces to the left hand side of (3.12), see Section 3 in [27] for details. Then we can use the function f_1 from [27], aiming to obtain an inequality like (3.13) (then we use the synchronous coupling for $(B_t^1)_{t \geq 0}$ and the terms I_2 and I_6 disappear). In either case, I_3 and I_5 can be controlled via κ , since the coefficients σ and g are included in its definition. If we are only interested in finding any constant $\tilde{c} > 0$ such that (2.14) holds, then it is sufficient to use one of the two additive noises and to apply the synchronous coupling to the other (if both noises are present it is recommendable to use $(B_t^1)_{t \geq 0}$ since the formulas in [12] are simpler than the ones in [27]). If we are interested in finding the best (largest) possible constant \tilde{c} , then we can use both noises, but then we would also need to redefine the function f and this would be technically quite sophisticated (whereas by using only one noise we can essentially just use the formulas that are already available in either [12] or [27]).

We should still explain how to control I_7 and I_9 . We can control I_9 following the ideas from [23] and controlling I_7 is also quite straightforward.

First observe that $\nabla f(|Z_{t-}|) = f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} Z_{t-}$ and, following Section 5.2 in [23], note that since f is concave and differentiable, we have

$$f(a) - f(b) \leq f'(b)(a - b)$$

for any $a, b > 0$. Thus

$$\begin{aligned} & f(|Z_{t-} + g(X_{t-}, u) - g(Y_{t-}, u)|) - f(|Z_{t-}|) - f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \langle g(X_{t-}, u) - g(Y_{t-}, u), Z_{t-} \rangle \\ & \leq f'(|Z_{t-}|) \left(|Z_{t-} + g(X_{t-}, u) - g(Y_{t-}, u)| - |Z_{t-}| - \frac{1}{|Z_{t-}|} \langle g(X_{t-}, u) - g(Y_{t-}, u), Z_{t-} \rangle \right). \end{aligned}$$

Next we will need the inequality

$$|x + y| - |x| - \frac{1}{|x|} \langle y, x \rangle \leq \frac{1}{2|x|} |y|^2,$$

which holds for any $x, y \in \mathbb{R}^d$ since

$$|x||x + y| \leq \frac{1}{2}(|x|^2 + |x + y|^2) = \frac{1}{2}(|x|^2 + |x|^2 + 2\langle x, y \rangle + |y|^2).$$

Hence we obtain

$$I_9 \leq f'(|Z_{t-}|) \frac{1}{2|Z_{t-}|} \int_U |g(X_{t-}, u) - g(Y_{t-}, u)|^2 N(dt, du).$$

On the other hand, if we denote by σ^k the k -th column of the matrix σ , then

$$\begin{aligned} I_7 &= \sum_{k=1}^m f''(|Z_{t-}|) \frac{|\langle Z_t, \sigma^k(X_t) - \sigma^k(Y_t) \rangle|^2}{|Z_{t-}|^2} + \sum_{k=1}^m \sum_{i=1}^d f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} (\sigma_{ik}(X_{t-}) - \sigma_{ik}(Y_{t-}))^2 \\ &\quad - \sum_{k=1}^m f'(|Z_{t-}|) \frac{|\langle Z_t, \sigma^k(X_t) - \sigma^k(Y_t) \rangle|^2}{|Z_{t-}|^3} \leq f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \|\sigma(X_{t-}) - \sigma(Y_{t-})\|_{HS}^2. \end{aligned}$$

Hence we get a bound on $df(|Z_t|)$, which allows us to bound $\mathbb{E}f(|Z_t|) - \mathbb{E}f(|Z_s|)$ for any $0 \leq s < t$. Using a localization argument with a sequence of stopping times $(\tau_n)_{n=1}^\infty$ like in (3.7), we can get rid of the expectations of the local martingale terms. Then we can use the inequality (4.10) multiplied by $f'(|x - y|) \frac{1}{|x - y|}$ to see that, if we are using the additive Lévy noise $(L_t)_{t \geq 0}$ to get our bounds, then after handling I_8 like in (3.12) and using the estimates (3.8) and (3.9), we need to choose a function f_1 such that

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

and this is exactly (3.13), so we can handle further calculations like in the proof of Theorem 3.1. Alternatively, if we are using the additive Gaussian noise $(B_t^1)_{t \geq 0}$, we can modify the definition of κ to include the σ_1 factor (cf. (4.6)) and then we need to choose a function f such that

$$2f''(r) - r\kappa(r)f'(r) \leq -\frac{c}{\sigma_1^2} f(r),$$

hence $1/\sigma_1^2$ plays the role of α in the calculations in [12] (cf. (4.8) earlier in this Section and for the details see the proof of Theorem 1 in [12], specifically the formula (63), while remembering about the change of the factor 2 in our definition of κ compared to the one in [12]).

Either way we obtain some constant $\tilde{c} > 0$ and a function \tilde{f} such that

$$(4.13) \quad \mathbb{E}\tilde{f}(|Z_{t \wedge \tau_n}|) - \mathbb{E}\tilde{f}(|Z_{s \wedge \tau_n}|) \leq -\tilde{c} \int_s^t \mathbb{E}\tilde{f}(|Z_{r \wedge \tau_n}|) dr$$

holds for any $0 \leq s < t$. Here \tilde{c} and \tilde{f} are equal either to c and f defined by (4.5) and (4.4) or c_1 and f_1 defined by (3.16) and (3.14), respectively, depending on whether we

used $(B_t^1)_{t \geq 0}$ or $(L_t)_{t \geq 0}$ in the step above. Thus we can use the differential version of the Gronwall inequality to get

$$\mathbb{E}\tilde{f}(|Z_{t \wedge \tau_n}|) \leq \mathbb{E}\tilde{f}(|Z_0|)e^{-\tilde{c}t} \text{ for any } t > 0,$$

and after using the Fatou lemma, the fact that $\tau_n \rightarrow T$ and that $Z_t = 0$ for $t \geq T$, we get

$$\mathbb{E}\tilde{f}(|Z_t|) \leq \mathbb{E}\tilde{f}(|Z_0|)e^{-\tilde{c}t} \text{ for any } t > 0.$$

Since we can compare our function \tilde{f} with the identity function from both sides, this finishes the proof. Note that in the last step one has to be careful and use the differential version of the Gronwall formula, since the integral version does not work when the term on the right hand side is negative (cf. Remark 2.3 in [36]).

Note also that if we are only dealing with the Gaussian noise, then we can reason like in [12], i.e., having proved that

$$df(|Z_t|) \leq dM_t - cf(|Z_t|)dt$$

(by choosing an appropriate function f) for some local martingale $(M_t)_{t \geq 0}$, we can see that this implies $d(e^{ct}f(|Z_t|)) \leq dM_t$, so by using a localization argument we can directly get $\mathbb{E}[e^{ct}f(|Z_t|)] \leq \mathbb{E}f(|Z_0|)$ without using the Gronwall inequality. However, in the jump case this is not possible, since we have to first take the expectation in order to deal with I_8 and I_9 by transforming the stochastic integrals with respect to N^L and N into deterministic integrals with respect to ν^L and ν , respectively. Only then can we use the definition of κ (4.10) to find an appropriate function \tilde{f} such that (4.13) holds. \square

We will now show how, starting from an equation of the form (2.1) with one multiplicative Gaussian noise, we can obtain an SDE of the form (2.13) with two independent Gaussian noises, one of which is still multiplicative, but the other additive (and the additive one has just a real constant as a coefficient, cf. the comments in the proof of Theorem 2.3 earlier in this section).

Lemma 4.1. *If $(X_t)_{t \geq 0}$ is the unique strong solution to the SDE*

$$(4.14) \quad dX_t = b(X_t)dt + \sigma(X_t)dB_t,$$

where $(B_t)_{t \geq 0}$ is a Brownian motion and $\sigma\sigma^T$ is uniformly positive definite, then $(X_t)_{t \geq 0}$ can also be obtained as a solution to

$$(4.15) \quad dX_t = b(X_t)dt + CdB_t^1 + \tilde{\sigma}(X_t)dB_t^2$$

with two independent Brownian motions $(B_t^1)_{t \geq 0}$ and $(B_t^2)_{t \geq 0}$, some constant $C > 0$ and a diffusion coefficient $\tilde{\sigma}$ such that if $\|\sigma(x)\|_{HS} \leq M$ for all $x \in \mathbb{R}^d$ with some constant $M > 0$, then

$$(4.16) \quad \|\tilde{\sigma}(x) - \tilde{\sigma}(y)\|_{HS} \leq \frac{M}{\sqrt{\lambda^2 - C^2}} \|\sigma(x) - \sigma(y)\|_{HS},$$

where the constants $\lambda > C > 0$ are as indicated in the proof.

Proof. Observe that if the diffusion coefficient σ is such that $\sigma\sigma^T$ is uniformly positive definite, i.e., there exists $\lambda > 0$ such that for any $x, h \in \mathbb{R}^d$ we have

$$\langle \sigma(x)\sigma(x)^T h, h \rangle \geq \lambda^2 |h|^2,$$

then $\sigma(x)\sigma(x)^T - \lambda^2 I$ is nonnegative definite for any $x \in \mathbb{R}^d$ and thus we can consider

$$a(x) := \sqrt{\sigma(x)\sigma(x)^T - \lambda^2 I},$$

which is the unique (symmetric) nonnegative definite matrix such that $a(x)a(x)^T = \sigma(x)\sigma(x)^T - \lambda^2 I$. Note that if we now define $\tilde{\sigma}$ as

$$\tilde{\sigma}(x) := \sqrt{\sigma(x)\sigma(x)^T - C^2 I}$$

for some constant $0 < C^2 < \lambda^2$, we can get

$$\langle \tilde{\sigma}(x)^2 h, h \rangle = \langle \sigma(x)\sigma(x)^T h, h \rangle - C^2 \langle h, h \rangle \geq (\lambda^2 - C^2) |h|^2,$$

and thus we can assume that $\tilde{\sigma}(x)$ is also uniformly positive definite. Therefore Lemma 3.3 in [32] applies (our $\tilde{\sigma}$ corresponds to σ in [32] and our $\sigma\sigma^T$ corresponds to q therein). Thus we get

$$\|\tilde{\sigma}(x) - \tilde{\sigma}(y)\|_{HS} \leq \frac{1}{2\sqrt{\lambda^2 - C^2}} \|\sigma(x)\sigma(x)^T - \sigma(y)\sigma(y)^T\|_{HS}.$$

(all eigenvalues of $\sigma(x)\sigma(x)^T - C^2 I$ are not less than $\lambda^2 - C^2$, which is the condition that needs to be checked in the proof of Lemma 3.3 in [32]). This shows that whenever $\sigma\sigma^T$ is Lipschitz with a constant L , the function $\tilde{\sigma}$ is Lipschitz with $L/2\sqrt{\lambda^2 - C^2}$. In particular, if σ is Lipschitz with a constant L and bounded with a constant M , then $\sigma\sigma^T$ is Lipschitz with the constant $2LM$ and thus $\tilde{\sigma}$ is Lipschitz with $LM/\sqrt{\lambda^2 - C^2}$. Hence we prove (4.16).

Now assume that $(X_t)_{t \geq 0}$ is a solution to (4.15) and consider the process

$$A_t := CB_t^1 + \int_0^t \sqrt{\sigma(X_s)\sigma(X_s)^T - C^2 I} dB_s^2 = X_t - X_0 - \int_0^t b(X_s) ds.$$

We can easily calculate

$$(4.17) \quad [A^i, A^j]_t = \int_0^t (\sigma\sigma^T)_{ij}(X_s) ds.$$

Hence, if we write

$$dX_t = dA_t + b(X_t)dt = \sigma(X_t)d\tilde{B}_t + b(X_t)dt,$$

where $d\tilde{B}_t = \sigma^{-1}(X_t)dA_t$, then using (4.17) and the Lévy characterization theorem, we infer that $(\tilde{B}_t)_{t \geq 0}$ is a Brownian motion. Thus $(X_t)_{t \geq 0}$ is a solution to (4.14). \square

The proof of (4.16) is based on the reasoning in [32], Section 3 (the matrix $q(x)$ used there is our $\sigma(x)\sigma(x)^T$; the difference in notation follows from the fact that the starting point for studying diffusions in [32] is the generator and not the SDE). Due to (4.16) we see that if the coefficients in (4.14) satisfy Assumption D1, then the coefficients in the modified equation (4.15) also do (after a suitable change in the definition of κ). More generally, Lemma 4.1 allows us to replace an equation of the form (2.1) with

$$dX_t = b(X_t)dt + CdB_t^1 + \sqrt{\sigma(X_t)\sigma(X_t)^T - C^2 I} dB_t^2 + \int_U g(X_{t-}, u) \tilde{N}(dt, du),$$

as long as $\sigma\sigma^T$ is uniformly positive definite.

5. BOUNDS ON MALLIAVIN DERIVATIVES

5.1. Brownian case. In this section we first prove Theorem 2.8 and then we show how to obtain bounds on Malliavin derivatives using the inequality (2.3). As a consequence we prove Theorem 2.14 and Corollary 2.15. We begin with proving the following crucial result.

Lemma 5.1. *Let $(X_t)_{t \geq 0}$ be a d -dimensional jump diffusion process given by*

$$(5.1) \quad dX_t = b(X_t)dt + \sigma dB_t + \int_U g(X_{t-}, u) \tilde{N}(dt, du),$$

where σ is a $d \times d$ matrix with $\det \sigma > 0$ and $(B_t)_{t \geq 0}$ is a d -dimensional Brownian motion, whereas b and g satisfy Assumption D1 and Assumption D2 and g is Lipschitz. Let h_t be an adapted d -dimensional process and consider a jump diffusion $(\tilde{X}_t)_{t \geq 0}$ with the drift perturbed by h_t , i.e.,

$$(5.2) \quad d\tilde{X}_t = b(\tilde{X}_t)dt + h_t dt + \sigma dB_t + \int_U g(\tilde{X}_{t-}, u) \tilde{N}(dt, du).$$

Then there exists a d -dimensional process $(Y_t)_{t \geq 0}$ such that $(X_t, Y_t)_{t \geq 0}$ is a coupling and we have

$$(5.3) \quad \mathbb{E}|\tilde{X}_t - Y_t| \leq C \mathbb{E} \int_0^t e^{c(s-t)} |h_s| ds$$

for some constant $C > 0$.

Proof. The arguments we use here are based on ideas from Sections 6 and 7 in [12], where interacting diffusions (without the jump noise) were studied. Here the most important part of the argument also concerns the Gaussian noise, however, we include the jump noise too in order to show how to handle the additional terms, which is important for the proof of Theorem 2.8. On the other hand, in order to slightly simplify the notation, we assume from now on that $\sigma = I$. Denote

$$Z_t := \tilde{X}_t - Y_t,$$

where $(Y_t)_{t \geq 0}$ will be defined below by (5.6), and consider Lipschitz continuous functions $\lambda, \pi : \mathbb{R}^d \rightarrow [0, 1]$ such that for some fixed $\delta > 0$ we have

$$(5.4) \quad \begin{aligned} \lambda^2(z) + \pi^2(z) &= 1 \text{ for any } z \in \mathbb{R}^d, \\ \lambda(z) &= 0 \text{ if } |z| \leq \delta/2 \\ \lambda(z) &= 1 \text{ if } |z| \geq \delta. \end{aligned}$$

Now fix a unit vector $u \in \mathbb{R}^d$ and define $R(\tilde{X}_t, Y_t) := I - 2e_t e_t^T$, where

$$e_t := \begin{cases} \frac{Z_t}{|Z_t|}, & \text{if } \tilde{X}_t \neq Y_t, \\ u, & \text{if } \tilde{X}_t = Y_t. \end{cases}$$

We will see from the proof that the exact value of u is irrelevant. Let us notice that the equation (5.2) for the process $(\tilde{X}_t)_{t \geq 0}$ can be rewritten as

$$(5.5) \quad d\tilde{X}_t = b(\tilde{X}_t)dt + h_t dt + \lambda(Z_t)dB_t^1 + \pi(Z_t)dB_t^2 + \int_U g(\tilde{X}_{t-}, u) \tilde{N}(dt, du),$$

where $(B_t^1)_{t \geq 0}$ and $(B_t^2)_{t \geq 0}$ are independent Brownian motions, and define

$$(5.6) \quad dY_t = b(Y_t)dt + \lambda(Z_t)R(\tilde{X}_t, Y_t)dB_t^1 + \pi(Z_t)dB_t^2 + \int_U g(Y_{t-}, u) \tilde{N}(dt, du).$$

Using the Lévy characterization theorem and the fact that $\lambda^2 + \pi^2 = 1$, we can show that the processes defined by

$$\begin{aligned} d\tilde{B}_t &:= \lambda(Z_t)dB_t^1 + \pi(Z_t)dB_t^2, \\ d\bar{B}_t &:= \lambda(Z_t)R(\tilde{X}_t, Y_t)dB_t^1 + \pi(Z_t)dB_t^2 \end{aligned}$$

are both d -dimensional Brownian motions and hence the process $(Y_t)_{t \geq 0}$ defined by (5.6) has the same finite dimensional distributions as $(X_t)_{t \geq 0}$ defined by (5.1) with $\sigma = I$, while both (5.2) with $\sigma = I$ and (5.5) also define the same (in law) process, which follows from the uniqueness in law of solutions to equations of the form (5.1). Thus $(X_t, Y_t)_{t \geq 0}$ is a coupling. Note that obviously in this case $(\tilde{X}_t, Y_t)_{t \geq 0}$ is not a coupling, but we do not need this to prove (5.3). Consider the equation for $Z_t = \tilde{X}_t - Y_t$, which is given by

$$dZ_t = (b(\tilde{X}_t) - b(Y_t))dt + h_t dt + 2\lambda(Z_t)e_t e_t^T dB_t^1 + \int_U (g(\tilde{X}_{t-}, u) - g(Y_{t-}, u))\tilde{N}(dt, du),$$

and observe that the process

$$d\tilde{W}_t := e_t^T dB_t^1$$

is a one-dimensional Brownian motion. Now we would like to apply the Itô formula to calculate $df(|Z_t|)$ for the function f given by (4.4), just like we did in the proof of Theorem 2.3. However, the function $x \mapsto f(|x|)$ is not differentiable at zero. In the proof of Theorem 2.3 this was not a problem, since we started the marginal processes of our coupling at two different initial points and were only interested in the behaviour of $f(|Z_t|)$ until Z_t reaches zero for the first time. Here on the other hand we will actually want to start both the marginal processes at the same point. Moreover, because of the modified construction of the coupling, which now behaves like a synchronous coupling for small values of $|Z_t|$, it can keep visiting zero infinitely often. A way to rigorously deal with this is to apply the version of the Meyer-Itô formula that can be found e.g. as Theorem 71 in Chapter IV in [33]. We begin with computing the formula for $d|Z_t|$, by calculating $d|Z_t|^2$ first and then applying the Itô formula once again to a smooth approximation of the square root function, given e.g. by

$$S(r) := \begin{cases} -(1/8)\varepsilon^{-3/2}r^2 + (3/4)\varepsilon^{-1/2}r + (3/8)\varepsilon^{1/2}, & r < \varepsilon, \\ \sqrt{r}, & r \geq \varepsilon. \end{cases}$$

A related argument was given by Zimmer in [40] in the context of infinite-dimensional diffusions, see Lemmas 2-5 therein. In our case, after two applications of the Itô formula, we get

$$\begin{aligned} dS(|Z_t|^2) &= 4S'(|Z_t|^2)\lambda(Z_t)|Z_t|d\tilde{W}_t + 2S'(|Z_t|^2)\langle Z_t, h_t + b(\tilde{X}_t) - b(Y_t) \rangle dt \\ &\quad + 4S'(|Z_t|^2)\lambda^2(Z_t)dt + 8S''(|Z_t|^2)\lambda^2(Z_t)|Z_t|^2 dt \\ &\quad + \int_U \left(S(|Z_{t-} + g(\tilde{X}_{t-}, u) - g(Y_{t-}, u)|^2) - S(|Z_{t-}|^2) \right) N(dt, du) \\ &\quad - 2 \int_U S'(|Z_{t-}|^2)\langle Z_{t-}, g(\tilde{X}_{t-}, u) - g(Y_{t-}, u) \rangle \nu(du) dt. \end{aligned}$$

Since for any $r \in [0, \infty)$ we have $S(r) \rightarrow \sqrt{r}$ when $\varepsilon \rightarrow 0$, we can also show almost sure convergence of the integrals appearing in the formula above. For example, using the fact that S is concave, for any $a, b \geq 0$ we have $S(a) - S(b) \leq S'(b)(b - a)$ and hence

$$\begin{aligned} &\mathbb{E} \int_0^T \int_U \left(S(|Z_{t-} + g(\tilde{X}_{t-}, u) - g(Y_{t-}, u)|^2) - S(|Z_{t-}|^2) \right) N(dt, du) \\ &\leq \mathbb{E} \int_0^T \int_U S'(|Z_{t-}|^2) \left(|g(\tilde{X}_{t-}, u) - g(Y_{t-}, u)|^2 + 2\langle Z_{t-}, g(\tilde{X}_{t-}, u) - g(Y_{t-}, u) \rangle \right) \nu(du) dt. \end{aligned}$$

Using the fact that g is Lipschitz and that $\sup_{r \leq \varepsilon} S'(r) \lesssim \varepsilon^{-1/2}$, we see that the integral

$$\int_0^T \int_U \left(S \left(|Z_{t-} + g(\tilde{X}_{t-}, u) - g(Y_{t-}, u)|^2 \right) - S \left(|Z_{t-}|^2 \right) \right) N(dt, du)$$

converges to

$$\int_0^T \int_U \mathbf{1}_{\{Z_{t-} \neq 0\}} \left(|Z_{t-} + g(\tilde{X}_{t-}, u) - g(Y_{t-}, u)| - |Z_{t-}| \right) N(dt, du)$$

in L^1 and hence, via a subsequence, almost surely when $\varepsilon \rightarrow 0$. Dealing with the other integrals is even easier, cf. Lemmas 2 and 3 in [40] for analogous arguments. Thus we are able to get

$$\begin{aligned} d|Z_t| &= 2\lambda(Z_t)d\tilde{W}_t + \mathbf{1}_{\{Z_t \neq 0\}} \frac{1}{|Z_t|} \langle Z_t, h_t + b(\tilde{X}_t) - b(Y_t) \rangle dt \\ &\quad + \int_U \mathbf{1}_{\{Z_{t-} \neq 0\}} \left(|Z_{t-} + g(\tilde{X}_{t-}, u) - g(Y_{t-}, u)| - |Z_{t-}| \right) N(dt, du) \\ &\quad - \int_U \mathbf{1}_{\{Z_{t-} \neq 0\}} \frac{1}{|Z_{t-}|} \langle Z_{t-}, g(\tilde{X}_{t-}, u) - g(Y_{t-}, u) \rangle \nu(du) dt. \end{aligned}$$

Now observe that the function f defined by (4.4) is twice continuously differentiable at all points except for R_1 , whereas f' exists and is continuous even at R_1 . Therefore we can apply the Meyer-Itô formula in its version given as Theorem 71 in Chapter IV in [33] to the process $(|Z_t|)_{t \geq 0}$ and the function f . For any $0 \leq s \leq r$ we get

(5.7)

$$\begin{aligned} f(|Z_r|) - f(|Z_s|) &= 2 \int_s^r f'(|Z_t|) \lambda(Z_t) d\tilde{W}_t \\ &\quad + \int_s^r \mathbf{1}_{\{Z_t \neq 0\}} f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, h_t + b(\tilde{X}_t) - b(Y_t) \rangle dt \\ &\quad + \int_s^r \int_U \mathbf{1}_{\{Z_{t-} \neq 0\}} f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \langle Z_{t-}, g(\tilde{X}_{t-}, u) - g(Y_{t-}, u) \rangle \tilde{N}(dt, du) \\ &\quad + \int_s^r \int_U \mathbf{1}_{\{Z_{t-} \neq 0\}} \left[f(|Z_{t-} + g(X_{t-}, u) - g(Y_{t-}, u)|) - f(|Z_{t-}|) \right. \\ &\quad \left. - f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \langle Z_{t-}, g(X_{t-}, u) - g(Y_{t-}, u) \rangle \right] N(dt, du) \\ &\quad + 2 \int_s^r f''(|Z_t|) \lambda^2(Z_t) dt. \end{aligned}$$

We can see that the integrand in the integral with respect to $(\tilde{W}_t)_{t \geq 0}$ in (5.7) is bounded (since f' and λ are bounded) and the integrand in the integral with respect to \tilde{N} is square integrable with respect to $\nu(du)dt$. Thus the expectations of both these integrals are zero. Moreover, the expectation of the integral with respect to N in (5.7) can be dealt with in the same way as the expectation of the term I_9 in the proof of Theorem 2.3, see (4.12). Thus, after taking the expectation everywhere in (5.7) and using the definition of κ , we

get

$$(5.8) \quad \begin{aligned} \mathbb{E}f(|Z_r|) - \mathbb{E}f(|Z_s|) &\leq \mathbb{E} \int_s^r \mathbf{1}_{\{Z_t \neq 0\}} |h_t| dt - \mathbb{E} \int_s^r \mathbf{1}_{\{Z_t \neq 0\}} f'(|Z_t|) |Z_t| \kappa(|Z_t|) dt \\ &\quad + \mathbb{E} \int_s^r 2f''(|Z_t|) \lambda^2(Z_t) dt. \end{aligned}$$

We will now want to use the fact that the function f defined by (4.4) satisfies

$$(5.9) \quad 2f''(r) - r\kappa(r)f'(r) \leq -cf(r).$$

In particular, denoting $r_t := |Z_t|$, we get

$$2f''(r_t)\lambda^2(Z_t) - r_t\kappa(r_t)f'(r_t)\lambda^2(Z_t) + r_t\kappa(r_t)f'(r_t) - r_t\kappa(r_t)f'(r_t) \leq -cf(r_t)\lambda^2(Z_t)$$

and thus

$$(5.10) \quad -r_t\kappa(r_t)f'(r_t) + 2f''(r_t)\lambda^2(Z_t) \leq -cf(r_t)\lambda^2(Z_t) + r_t\kappa(r_t)f'(r_t)(\lambda^2(Z_t) - 1).$$

Now observe that

$$-cf(r_t)\lambda^2(Z_t) = cf(r_t)(1 - \lambda^2(Z_t)) - cf(r_t) \leq c\delta - cf(r_t),$$

which holds since if $|Z_t| \geq \delta$, then $1 - \lambda^2(Z_t) = 0$ and if $|Z_t| \leq \delta$, then $1 - \lambda^2(Z_t) \leq 1$ and $cf(r_t) \leq c\delta$, which follow from the properties (5.4) of the function λ and the fact that $f(x) \leq x$ for any $x \in [0, \infty)$. Since obviously $-\kappa \leq \kappa^-$, we can further bound the right hand side of (5.10) by

$$(5.11) \quad c\delta - cf(r_t) + \kappa^-(r_t)(1 - \lambda^2(Z_t))r_t f'(r_t) \leq c\delta - cf(r_t) + \sup_{r \leq \delta} r\kappa^-(r),$$

where the last inequality follows from the fact that $1 - \lambda^2(Z_t) = 0$ when $|Z_t| \geq \delta$ and that $f' \leq 1$. If we denote

$$m(\delta) := c\delta + \sup_{r \leq \delta} r\kappa^-(r),$$

then, from (5.10) and (5.11) we obtain

$$(5.12) \quad -r_t\kappa(r_t)f'(r_t) + 2f''(r_t)\lambda^2(Z_t) \leq -cf(r_t) + m(\delta).$$

Hence, combining (5.8) with (5.12) multiplied by $\mathbf{1}_{\{r_t \neq 0\}}$, we get

$$\begin{aligned} \mathbb{E}f(|Z_r|) - \mathbb{E}f(|Z_s|) &\leq -c \int_s^r \mathbb{E} \mathbf{1}_{\{Z_t \neq 0\}} f(|Z_t|) dt + \mathbb{E} \int_s^r \mathbf{1}_{\{Z_t \neq 0\}} (|h_t| + m(\delta)) dt \\ &\leq -c \int_s^r \mathbb{E}f(|Z_t|) dt + \mathbb{E} \int_s^r |h_t| dt + \int_s^r m(\delta) dt. \end{aligned}$$

Now observe that due to Assumption D2, we have $m(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. We can also choose $\tilde{X}_0 = Y_0$ so that $Z_0 = 0$. Eventually, applying the Gronwall inequality, we obtain

$$\mathbb{E}[f(|Z_t|)] \leq \mathbb{E} \int_0^t e^{c(s-t)} |h_s| ds,$$

which finishes the proof, since there exists a constant $C > 0$ such that for any $x \geq 0$ we have $x \leq Cf(x)$. \square

Proof of Theorem 2.8. Once we have Lemma 5.1, extending its result to the equation (2.13) is quite straightforward. In comparison to the proof of Lemma 5.1, the key step is to redefine κ in order to include the additional coefficient σ of the multiplicative Brownian noise (so that κ satisfies (4.10)) and then perform the same procedure as we did earlier (mixed reflection-synchronous coupling) only on the additive Brownian noise in order to construct processes like (5.5) and (5.6), where to the other noises we apply just the

synchronous coupling. This way we can still get the inequality (5.9) with the same function f as in the proof of Lemma 5.1. The details are left to the reader, as they are just a repetition of what we have already presented. Once we obtain an inequality of the form (5.3) for the equation (2.13), we can use the Markov property of the process $(\tilde{X}_t, Y_t)_{t \geq 0}$ to get (2.15), cf. Remark 2.5. \square

Proof of Theorem 2.14. In order to keep notational simplicity, assume that we are dealing here with the equation $dX_t = b(X_t)dt + \sigma(X_t)dW_t$, i.e., the coefficient of the jump noise is zero. It does not influence our argument in any way, since we will need to perturb only the Gaussian noise. Recall that for a functional $f : \mathbb{R}^d \rightarrow \mathbb{R}$, the Malliavin derivative $\nabla_s f(X_t)$ is an m -dimensional vector $(\nabla_{s,1}f(X_t), \dots, \nabla_{s,m}f(X_t))$, where $\nabla_{s,k}f(X_t)$ can be thought of as a derivative with respect to $(W_t^k)_{t \geq 0}$, where $W_t = (W_t^1, \dots, W_t^m)$ is the driving m -dimensional Brownian motion.

We know that if F is a random variable of the form $F = f(\int_0^T g_s^1 dW_s, \dots, \int_0^T g_s^N dW_s)$ for some smooth function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ and $g^1, \dots, g^N \in L^2([0, T]; \mathbb{R}^m)$ (i.e., $F \in \mathcal{S}$), then for any element $h \in H = L^2([0, T]; \mathbb{R}^m)$ we have

$$(5.13) \quad \langle \nabla F, h \rangle_{L^2([0, T]; \mathbb{R}^m)} = \int_0^t \langle \nabla_s F, h_s \rangle ds = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left(F(W. + \varepsilon \int_0^\cdot h_s ds) - F(W.) \right),$$

where convergence is in $L^2(\Omega)$. However, it is unclear whether (5.13) holds also for arbitrary $F \in \mathbb{D}^{1,2}$ and in particular for X_t (see the discussion in Appendix A in [10], specifically Definitions A.10 and A.13). Nevertheless, for $F \in \mathbb{D}^{1,2}$ we can still prove that

$$(5.14) \quad \mathbb{E} \langle \nabla F, h \rangle_{L^2([0, T]; \mathbb{R}^m)} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left(F(W. + \varepsilon \int_0^\cdot h_s ds) - F(W.) \right),$$

even if we replace $h \in H$ with an adapted stochastic process $(\omega, t) \mapsto h_t(\omega)$ such that $\mathbb{E} \int_0^T |h_s|^2 ds < \infty$ and the Girsanov theorem applies (e.g. the Novikov condition for h is satisfied).

Indeed, we know that for any $F \in \mathbb{D}^{1,2}$ and for any adapted square integrable h we have

$$(5.15) \quad \mathbb{E} \langle \nabla F, h \rangle_{L^2([0, T]; \mathbb{R}^m)} = \mathbb{E} \left[F \int_0^T h_s dW_s \right].$$

We recall now the proof of this fact, as we need to slightly modify it in order to get (5.14). As a reference, see e.g. Lemma A.15. in [10], where (5.15) is proved only for $F \in \mathcal{S}$ and for deterministic h , but the argument can be easily generalized, or Theorems 1.1 and 1.2 in Chapter VIII of [5]. For now assume that h is adapted and bounded (and thus it satisfies the assumptions of the Girsanov theorem). Then, starting from the right hand side of (5.15), we have

$$(5.16) \quad \begin{aligned} \mathbb{E} \left[F \int_0^T h_s dW_s \right] &= \mathbb{E} \left[F \frac{d}{d\varepsilon} \exp \left(\varepsilon \int_0^T h_s dW_s - \frac{1}{2} \varepsilon^2 \int_0^T |h_s|^2 ds \right) \Big|_{\varepsilon=0} \right] \\ &= \mathbb{E} \left[F \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left[\exp \left(\varepsilon \int_0^T h_s dW_s - \frac{1}{2} \varepsilon^2 \int_0^T |h_s|^2 ds \right) - 1 \right] \right] \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left[F \exp \left(\varepsilon \int_0^T h_s dW_s - \frac{1}{2} \varepsilon^2 \int_0^T |h_s|^2 ds \right) - F \right] \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left[F(W. + \varepsilon \int_0^\cdot h_s ds) - F(W.) \right], \end{aligned}$$

where in the last step we use the Girsanov theorem. In order to explain the third step, notice that the process

$$Z_t^\varepsilon := \exp \left(\varepsilon \int_0^t h_s dW_s - \frac{1}{2} \varepsilon^2 \int_0^t |h_s|^2 ds \right)$$

is the stochastic exponential of $\varepsilon \int_0^t h_s dW_s$ and thus it satisfies $dZ_t^\varepsilon = \varepsilon Z_s^\varepsilon h_s dW_s$, from which we get

$$\frac{1}{\varepsilon} [Z_T^\varepsilon - 1] = \int_0^T Z_s^\varepsilon h_s dW_s.$$

Now it is easy to see that since for any $\omega \in \Omega$ we have $Z_t^\varepsilon(\omega) \rightarrow 1$ with $\varepsilon \rightarrow 0$ and Z_t^ε is uniformly bounded in $L^2(\Omega \times [0, T])$, there is a subsequence such that

$$\frac{1}{\varepsilon} [Z_T^\varepsilon - 1] = \int_0^T Z_s^\varepsilon h_s dW_s \rightarrow \int_0^T h_s dW_s \text{ as } \varepsilon \rightarrow 0, \text{ in } L^2(\Omega).$$

Thus the third step in (5.16) holds for any $F \in L^2(\Omega)$ and in particular for any $F \in \mathbb{D}^{1,2}$. If F is smooth, then the last expression in (5.16) is equal to $\mathbb{E} \langle \nabla F, h \rangle_{L^2([0, T]; \mathbb{R}^m)}$, which proves (5.15) for any smooth F and adapted, bounded h . Then (5.15) can be extended by approximation to any $F \in \mathbb{D}^{1,2}$ and any adapted, square integrable h .

Now in order to prove (5.14), observe that the calculations in (5.16) still hold when applied directly to an $F \in \mathbb{D}^{1,2}$ and an adapted, bounded h (note that the argument does not work for general adapted, square integrable h as we need to use the Girsanov theorem in the last step). Thus for any $F \in \mathbb{D}^{1,2}$ and any adapted, bounded h we get

$$\mathbb{E} \langle \nabla F, h \rangle_{L^2([0, T]; \mathbb{R}^m)} = \mathbb{E} \left[F \int_0^T h_s dW_s \right] = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left[F(W. + \varepsilon \int_0^\cdot h_s ds) - F(W.) \right].$$

Since $X_t \in \mathbb{D}^{1,2}$ and f is Lipschitz, we have $f(X_t) \in \mathbb{D}^{1,2}$ (cf. [28], Proposition 1.2.4), and hence

$$\mathbb{E} \langle \nabla f(X_t), h \rangle_{L^2([0, T]; \mathbb{R}^m)} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left(f(X_t)(W. + \varepsilon \int_0^\cdot h_s ds) - f(X_t)(W.) \right)$$

holds for any adapted, bounded process h . From now on, we fix $t > 0$ and take $T = t$.

Recall that the process $(X_t)_{t \geq 0}$ is now given by $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ and thus

$$(5.17) \quad X_t(W. + \varepsilon \int_0^\cdot h_s ds) = \int_0^t b(X_s) ds + \varepsilon \int_0^t \sigma(X_s) h_s ds + \int_0^t \sigma(X_s) dW_s.$$

Hence, using the assumption (2.3) from Theorem 2.1 (taking $\varepsilon \sigma(X_s) h_s$ as the adapted change of drift and denoting the solution to (5.17) by $(\tilde{X}_t)_{t \geq 0}$) we obtain

$$\begin{aligned} \mathbb{E} \left(f(X_t)(W. + \varepsilon \int_0^\cdot h_s ds) - f(X_t)(W.) \right) &= \mathbb{E} \left(f(X_t)(W. + \varepsilon \int_0^\cdot h_s ds) - f(Y'_t)(W.) \right) \\ &\leq \varepsilon c_2(t) \mathbb{E} \int_0^t c_3(s) |\sigma(\tilde{X}_s) h_s| ds, \end{aligned}$$

where $\mathbb{E} f(X_t) = \mathbb{E} f(Y'_t)$, since $(X_t, Y'_t)_{t \geq 0}$ is a coupling. This in turn implies, together with our above calculations, that we have

$$(5.18) \quad \mathbb{E} \langle \nabla f(X_t), h \rangle_{L^2([0, t]; \mathbb{R}^m)} \leq c_2(t) \mathbb{E} \int_0^t c_3(s) |\sigma(\tilde{X}_s) h_s| ds \leq c_2(t) \sigma_\infty \mathbb{E} \int_0^t c_3(s) |h_s| ds.$$

Now by approximation we can show that the above inequality holds for any adapted process h such that $\mathbb{E} \int_0^t |h_s|^2 ds < \infty$. Then, using the Cauchy-Schwarz inequality for $L^2(\Omega \times [0, t])$, we get

$$\mathbb{E} \langle \nabla f(X_t), h \rangle_{L^2([0, t]; \mathbb{R}^m)} \leq c_2(t) \sigma_\infty \left(\mathbb{E} \int_0^t c_3^2(s) ds \right)^{1/2} \left(\mathbb{E} \int_0^t |h_s|^2 ds \right)^{1/2}.$$

Moreover, observe that since h is adapted, we have
(5.19)

$$\mathbb{E} \langle \nabla f(X_t), h \rangle_{L^2([0, t]; \mathbb{R}^m)} = \mathbb{E} \int_0^t \langle \mathbb{E}[\nabla_s f(X_t) | \mathcal{F}_s], h_s \rangle ds =: \mathbb{E} \langle \mathbb{E}[\nabla \cdot f(X_t) | \mathcal{F}], h \cdot \rangle_{L^2([0, t]; \mathbb{R}^m)}.$$

If we replace h above with $h \sqrt{g}$ for some adapted, integrable, \mathbb{R}_+ -valued process g , we get (by coming back to (5.18) and splitting h and \sqrt{g} via the Cauchy-Schwarz inequality)

$$\begin{aligned} \mathbb{E} \langle \sqrt{g} \cdot \mathbb{E}[\nabla \cdot f(X_t) | \mathcal{F}], h \cdot \rangle_{L^2([0, t]; \mathbb{R}^m)} &= \mathbb{E} \langle \mathbb{E}[\nabla \cdot f(X_t) | \mathcal{F}], h \cdot \sqrt{g} \cdot \rangle_{L^2([0, t]; \mathbb{R}^m)} \\ &\leq c_2(t) \sigma_\infty \left(\mathbb{E} \int_0^t g_s c_3^2(s) ds \right)^{1/2} \left(\mathbb{E} \int_0^t |h_s|^2 ds \right)^{1/2}. \end{aligned}$$

Since this holds for an arbitrary adapted, square integrable process h , we have

$$(5.20) \quad \mathbb{E} \int_0^t g_u |\mathbb{E}[\nabla_u f(X_t) | \mathcal{F}_u]|^2 du \leq c_2^2(t) \sigma_\infty^2 \mathbb{E} \int_0^t g_u c_3^2(u) du.$$

Observe that in the inequality above we can integrate on any interval $[s, r] \subset [0, t]$. We can also approximate an arbitrary adapted, \mathbb{R}_+ -valued process g with processes $g \wedge n$ for $n \geq 1$, for which we have (5.20). Then, by the Fatou lemma on the left hand side and the dominated convergence theorem on the right hand side, we get

$$\begin{aligned} \mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla_u f(X_t) | \mathcal{F}_u]|^2 du &\leq \lim_{n \rightarrow \infty} \mathbb{E} \int_s^r (g_u \wedge n) |\mathbb{E}[\nabla_u f(X_t) | \mathcal{F}_u]|^2 du \\ &\leq c_2^2(t) \sigma_\infty^2 \mathbb{E} \int_s^r g_u c_3^2(u) du. \end{aligned}$$

Hence we finally obtain (2.17). In order to get (2.18), we just need to go back to (5.18) and notice that it implies

$$\mathbb{E} \langle \nabla f(X_t), h \rangle_{L^2([0, t]; \mathbb{R}^m)} \leq c_2(t) \sigma_\infty \sup_{u \leq t} c_3(u) \mathbb{E} \int_0^t |h_s| ds.$$

Since we can show that this holds for an arbitrary adapted h from $L^1(\Omega \times [0, t])$, using (5.19) and the fact that the dual of L^1 is L^∞ , we finish the proof. \square

Proof of Corollary 2.15. Note that from Theorem 2.8 we obtain an inequality of the form (2.15), where on the right hand side we have either the coefficient σ_1 or σ , depending on whether we want to consider ∇^1 or ∇^2 . Recall from the proof of Theorem 2.8 that in order to get (2.15) we need to use the additive Brownian noise $(B_t^1)_{t \geq 0}$, regardless of which change of the drift we consider in the equation defining $(\tilde{X}_t)_{t \geq 0}$ that appears therein. Therefore we need to assume $\det \sigma_1 > 0$ even if we are only interested in bounding the Malliavin derivative with respect to the multiplicative Brownian noise $(B_t^2)_{t \geq 0}$. Once we have (2.15), it is sufficient to apply Theorem 2.14 with $c_2(t) = Ce^{-ct}$ and $c_3(s) = e^{cs}$. \square

5.2. **Poissonian case.** Consider the solution $(X_t(x))_{t \geq 0}$ to

$$(5.21) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du)$$

with initial condition $x \in \mathbb{R}^d$ as a functional of the underlying Poisson random measure $N = \sum_{j=1}^{\infty} \delta_{(\tau_j, \xi_j)}$. Then define

$$X^{(t,u)}(x) = X^{(t,u)}(x, N) := X(x, N + \delta_{(t,u)}),$$

which means that we add a jump of size $g(X_{t-}, u)$ at time t to every path of X . Then

$$X_s^{(t,u)}(x) = X_s(x) \text{ for } s < t$$

and

$$\begin{aligned} X_s^{(t,u)}(x) &= X_t(x) + g(X_{t-}, u) + \int_t^s b(X_r^{(t,u)}(x))dr \\ &\quad + \int_t^s \sigma(X_r^{(t,u)}(x))dW_r + \int_t^s \int_U g(X_{r-}^{(t,u)}(x), u)\tilde{N}(dr, du) \text{ for } s \geq t. \end{aligned}$$

This means that after time t , the process $(X_s^{(t,u)}(x))_{s \geq t}$ is a solution of the same SDE but with different initial condition, i.e., $X_t^{(t,u)}(x) = X_t(x) + g(X_{t-}, u)$.

If the global dissipativity assumption is satisfied (like in [39] and [25]), it is easy to show that the solution $(X_t)_{t \geq 0}$ to (5.21) satisfies for any x and $y \in \mathbb{R}^d$ the inequality

$$\mathbb{E}|X_t(x) - X_t(y)| \leq e^{-Kt}|x - y|$$

with some constant $K > 0$. Then we easily see that for any Lipschitz function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$, if $t < T$ we have

$$\begin{aligned} \mathbb{E}[D_{t,u}f(X_T(x))|\mathcal{F}_t] &\leq \mathbb{E}\left[\left|f(X_T^{(t,u)}(x)) - f(X_T(x))\right| \middle| \mathcal{F}_t\right] \\ &\leq \mathbb{E}\left[\left|X_T^{(t,u)}(x) - X_T(x)\right| \middle| \mathcal{F}_t\right] \\ &\leq e^{-K(T-t)}|g(X_{t-}, u)|. \end{aligned}$$

In order to improve this result we will work under the assumption (2.2) from Theorem 2.1 stating that there exists a coupling $(X_t, Y_t)_{t \geq 0}$ of solutions to (5.21) such that

$$(5.22) \quad \mathbb{E}[|X_T - Y_T|/\mathcal{F}_t] \leq c_1(T-t)|X_t - Y_t|$$

holds for any $T \geq t \geq 0$ with some function $c_1 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. We fix $t > 0$ and we express the process $(X_s^{(t,u)}(x))_{s \geq 0}$ as

$$X_s^{(t,u)}(x) := \begin{cases} X_s(x) & \text{for } s < t, \\ \bar{X}_s & \text{for } s \geq t, \end{cases}$$

where $(\bar{X}_s)_{s \geq t}$ is a solution to (5.21) started at t with initial point $X_t(x) + g(X_{t-}, u)$. Obviously both $(X_s)_{s \geq 0}$ and $(\bar{X}_s)_{s \geq t}$ have the same transition probabilities (since they are solutions to the same SDE satisfying sufficient conditions for uniqueness of its solutions in law). Thus we can apply our coupling to the process $(\bar{X}_s)_{s \geq t}$ to get a process $(\bar{Y}_s)_{s \geq t}$ with initial point $X_t(x)$ and the same transition probabilities as $(Y_s)_{s \geq 0}$ (and thus also $(X_s)_{s \geq 0}$). Now if we define the coupling time $\tau := \inf\{r > t : \bar{X}_r = \bar{Y}_r\}$ then we can put

$$\widehat{Y}_s(x) := \begin{cases} X_s(x) & \text{for } s < t, \\ \bar{Y}_s & \text{for } t \leq s < \tau, \\ \bar{X}_s & \text{for } s \geq \tau, \end{cases}$$

and we obtain a process with the same transition probabilities as $(X_s^{(t,u)}(x))_{s \geq 0}$ and thus also $(X_s(x))_{s \geq 0}$. This follows from a standard argument about gluing couplings at stopping times, see e.g. Subsection 2.2 in [38] for a possible approach. This way we get a coupling $(X_s(x), \widehat{Y}_s(x))_{s \geq 0}$ such that

$$\mathbb{E} \left[\left| X_T^{(t,u)}(x) - \widehat{Y}_T(x) \right| \middle| \mathcal{F}_t \right] \leq c_1(T-t) |g(X_{t-}, u)|$$

holds for any $T \geq t$ (from our construction we see that $X_t^{(t,u)}(x) - \widehat{Y}_t(x) = g(X_{t-}, u)$ and we use (5.22)). Now we can easily compute

$$\begin{aligned} \mathbb{E}[D_{t,u}f(X_T(x)) | \mathcal{F}_t] &= \mathbb{E}[f(X_T^{(t,u)}(x)) - f(X_T(x)) | \mathcal{F}_t] \\ &= \mathbb{E}[f(X_T^{(t,u)}(x)) - f(\widehat{Y}_T(x)) | \mathcal{F}_t] \\ (5.23) \quad &\leq \mathbb{E} \left[\left| X_T^{(t,u)}(x) - \widehat{Y}_T(x) \right| \middle| \mathcal{F}_t \right] \\ &\leq c_1(T-t) |g(X_{t-}, u)|, \end{aligned}$$

where we used the coupling property in the second step. In particular, if there exists a measurable function $g_\infty : U \rightarrow \mathbb{R}$ such that $|g(x, u)| \leq g_\infty(u)$ for any $x \in \mathbb{R}^d$ and $u \in U$, then we obviously get

$$(5.24) \quad \mathbb{E}[D_{t,u}f(X_T(x)) | \mathcal{F}_t] \leq c_1(T-t)g_\infty(u).$$

To end this section, let us consider briefly the case of the equation (2.13), where we have two jump noises, given by a Lévy process $(L_t)_{t \geq 0}$ and a Poisson random measure N . Then we can easily obtain analogous bounds on the Malliavin derivatives with respect to $(L_t)_{t \geq 0}$ and N , which we denote by D^L and D , respectively. Namely, in the framework of Theorem 2.3 we obtain a coupling $(X_t, Y_t)_{t \geq 0}$ such that

$$\mathbb{E} [|X_T - Y_T| | \mathcal{F}_t] \leq \widetilde{C} e^{-\widetilde{c}(T-t)} |X_t - Y_t|$$

holds for any $T \geq t \geq 0$ with some constants $\widetilde{C}, \widetilde{c} > 0$. Then, repeating the reasoning above, we easily get

$$(5.25) \quad \mathbb{E}[D_{t,u}^L f(X_T(x)) | \mathcal{F}_t] \leq \widetilde{C} e^{-\widetilde{c}(T-t)} u$$

and

$$(5.26) \quad \mathbb{E}[D_{t,u} f(X_T(x)) | \mathcal{F}_t] \leq \widetilde{C} e^{-\widetilde{c}(T-t)} g_\infty(u).$$

6. PROOFS OF TRANSPORTATION AND CONCENTRATION INEQUALITIES

Proof of Theorem 2.1 and Theorem 2.2. We first briefly recall the method of the proof of Theorem 2.2 in [39] and its extension from [25] (however, we denote certain quantities differently from [25] to make the notation more consistent with the original one from [39]). We will make use of the elements of Malliavin calculus described in Section 1. Specifically, we work on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ equipped with a Brownian motion $(W_t)_{t \geq 0}$ and a Poisson random measure N , on which we define the Malliavin derivative ∇ with respect to $(W_t)_{t \geq 0}$ (a differential operator) and the Malliavin derivative D with respect to N (a difference operator). We use the Clark-Ocone formula, i.e., if F is a functional such that the integrability condition (1.9) is satisfied, then

$$(6.1) \quad F = \mathbb{E}F + \int_0^T \mathbb{E}[\nabla_t F | \mathcal{F}_t] dW_t + \int_0^T \int_U \mathbb{E}[D_{t,u} F | \mathcal{F}_t] \widetilde{N}(dt, du).$$

From the proof of Lemma 3.2 in [25] we know that if we show that there exists a deterministic function $h : [0, T] \times U \rightarrow \mathbb{R}$ such that $\int_0^T \int_U h(t, u)^2 \nu(du) dt < \infty$ and

$$(6.2) \quad \mathbb{E}[D_{t,u}F|\mathcal{F}_t] \leq h(t, u)$$

and there exists a deterministic function $j : [0, T] \rightarrow \mathbb{R}^m$ such that $\int_0^T |j(t)|^2 dt < \infty$ and

$$(6.3) \quad |\mathbb{E}[\nabla_t F|\mathcal{F}_t]| \leq |j(t)|,$$

then for any C^2 convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ such that ϕ' is also convex, we have

$$(6.4) \quad \mathbb{E}\phi(F - \mathbb{E}F) \leq \mathbb{E}\phi\left(\int_0^T \int_U h(t, u)\tilde{N}(dt, du) + \int_0^T j(t)dW_t\right).$$

In particular, for any $\lambda > 0$ we have

$$(6.5) \quad \mathbb{E}e^{\lambda(F - \mathbb{E}F)} \leq \exp\left(\int_0^T \int_U (e^{\lambda h(t, u)} - \lambda h(t, u) - 1)\nu(du)dt + \int_0^T \frac{\lambda^2}{2}|j(t)|^2 dt\right).$$

The way to prove this is based on the forward-backward martingale method developed by Klein, Ma and Privault in [19]. On the product space $(\Omega^2, \mathcal{F}^2, \mathbb{P}^2)$ for any $(\omega, \omega') \in \Omega^2$ we can define

$$(6.6) \quad M_t(\omega, \omega') := \int_0^t \int_U \mathbb{E}[D_{s,u}F|\mathcal{F}_s](\omega)\tilde{N}(\omega, ds, du) + \int_0^t \mathbb{E}[\nabla_s F|\mathcal{F}_s](\omega)dW_s(\omega),$$

which is a forward martingale with respect to the increasing filtration $\mathcal{F}_t \otimes \mathcal{F}$ on Ω^2 and

$$(6.7) \quad M_t^*(\omega, \omega') := \int_t^T \int_U h(s, u)\tilde{N}(\omega', ds, du) + \int_t^T j(s)dW_s(\omega'),$$

which is a backward martingale with respect to the decreasing filtration $\mathcal{F} \otimes \mathcal{F}_t^*$, where \mathcal{F}_t^* is the σ -field generated by $N([r, \infty), A)$ and W_r for $r \geq t$ where A are Borel subsets of U . Application of the forward-backward Itô formula (see Section 8 in [19]) to $\phi(M_t + M_t^*)$ and comparison of the characteristics of M_t and M_t^* shows that for any $s \leq t$ we have

$$\mathbb{E}\phi(M_t + M_t^*) \leq \mathbb{E}\phi(M_s + M_s^*).$$

This follows from Theorem 3.3 in [19]. However, it is important to note that if we replace (6.3) with a weaker assumption, stating that for any adapted, \mathbb{R}_+ -valued process g and for any $[s, r] \subset [0, T]$ we have

$$(6.8) \quad \mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla_u F|\mathcal{F}_u]|^2 du \leq \mathbb{E} \int_s^r g_u |j(u)|^2 du,$$

then the argument from [19] still holds (check the page 493 in [19] and observe that what we need for the proof of Theorem 3.3 therein is that the integral of the process $\phi''(M_u + M_u^*)$ appearing there is non-positive and that is indeed the case if M and M^* are given by (6.6) and (6.7), respectively, and the condition (6.8) holds). Now we will use the fact that by the Clark-Ocone formula (6.1) we know that $M_t + M_t^* \rightarrow F - \mathbb{E}F$ in L^2 as $t \rightarrow T$. Observe that since ϕ is convex, we have

$$\phi(M_t + M_t^*) - \phi(0) \geq \phi'(0)(M_t + M_t^*)$$

and thus we can apply the Fatou lemma for $\phi(M_t + M_t^*) - \phi(0) - \phi'(0)(M_t + M_t^*)$ to get

$$\mathbb{E}\phi(F - \mathbb{E}F) - \phi'(0)\mathbb{E}(M_T) \leq \liminf_{t \rightarrow T} \mathbb{E}\phi(M_t + M_t^*).$$

Here $\phi(0)$ cancels since it appears on both sides and by (6.1) we know that $\mathbb{E}(M_T) = \mathbb{E}(F - \mathbb{E}F) = 0$. Thus we get

$$\mathbb{E}\phi(F - \mathbb{E}F) \leq \liminf_{t \rightarrow T} \mathbb{E}\phi(M_t + M_t^*) \leq \lim_{t \rightarrow T} \mathbb{E}\phi(M_0^*) = \mathbb{E}\phi(M_0^*),$$

which proves (6.4).

Now we can return to the equation (2.1). Using the assumption (2.2) we can get a bound on the Malliavin derivative D of a Lipschitz functional of $X_T(x)$, i.e., for any $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$ we have

$$(6.9) \quad \mathbb{E}[D_{t,u}f(X_T(x))|\mathcal{F}_t] \leq c_1(T-t)|g(X_{t-}, u)| \leq c_1(T-t)g_\infty(u)$$

(see the discussion in Section 5.2, in particular (5.23) and (5.24)). Note that the square integrability condition on the upper bound required in (6.2) is satisfied due to our assumptions on g_∞ . On the other hand, due to the assumption (2.3), via Theorem 2.14, for any adapted \mathbb{R}_+ -valued process g and any $[s, r] \subset [0, T]$ we get

$$(6.10) \quad \mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla_u f(X_T)|\mathcal{F}_u]|^2 du \leq c_2^2(T)\sigma_\infty^2 \mathbb{E} \int_s^r g_u c_3^2(u) du.$$

It is easy to see that with our bounds, directly from (6.4) we obtain (2.6). Note that as the integrand in the Brownian integral appearing in (2.6) we can take any m -dimensional function whose norm coincides with our upper bound in (6.10). For the inequalities on the path space $\mathbb{D}([0, T]; \mathbb{R}^d)$ we can still use our coupling $(X_s(x), \hat{Y}_s(x))_{s \geq 0}$ which we discussed in Section 5.2. Denote by $\hat{Y}_{[0, T]}$ a path of the process $(\hat{Y}_s(x))_{t \in [0, T]}$. Then for any Lipschitz functional $F : \mathbb{D}([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}$ (where we consider $\mathbb{D}([0, T]; \mathbb{R}^d)$ equipped with the L^1 metric $d_{L^1}(\gamma_1, \gamma_2) := \int_0^T |\gamma_1(t) - \gamma_2(t)| dt$) such that $\|F\|_{\text{Lip}} \leq 1$ we have

$$\begin{aligned} \mathbb{E}[D_{t,u}F(X_{[0, T]}(x))|\mathcal{F}_t] &= \mathbb{E}[F(X_{[0, T]}^{(t,u)}(x)) - F(X_{[0, T]}(x))|\mathcal{F}_t] \\ &= \mathbb{E} \left[F(X_{[0, T]}^{(t,u)}(x)) - F(\hat{Y}_{[0, T]}(x)) \middle| \mathcal{F}_t \right] \leq \mathbb{E} \left[\int_0^T |X_r^{(t,u)}(x) - \hat{Y}_r(x)| dr \middle| \mathcal{F}_t \right] \\ &= \int_t^T \mathbb{E} \left[|X_r^{(t,u)}(x) - \hat{Y}_r(x)| \middle| \mathcal{F}_t \right] dr \leq \int_t^T c_1(r-t)|g(X_{t-}, u)| dr \\ &\leq g_\infty(u) \int_t^T c_1(r-t) dr. \end{aligned}$$

In order to get a bound on $\mathbb{E}[\nabla F(X_{[0, T]}(x))|\mathcal{F}_t]$, we proceed similarly as in the proof of Theorem 2.14, using again the coupling $(X_t, Y'_t)_{t \geq 0}$ satisfying the assumption (2.3). Namely, we can show that for any bounded, adapted process h we have

$$\begin{aligned} &\mathbb{E} \langle \nabla F(X_{[0, T]}(x)), h \rangle_{L^2([0, T]; \mathbb{R}^m)} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left(F(X_{[0, T]}(x))(W. + \varepsilon \int_0^\cdot h_u du) - F(Y'_{[0, T]}(x))(W.) \right) \\ &\leq \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_0^T \mathbb{E} \left| X_r(x)(W. + \varepsilon \int_0^\cdot h_u du) - Y'_r(x)(W.) \right| dr \\ &\leq \int_0^T \left(c_2(r) \int_0^r c_3(u) \sigma_\infty |h_u| du \right) dr = \int_0^T \left(\int_u^T c_2(r) c_3(u) \sigma_\infty |h_u| dr \right) du \\ &\leq \left(\int_0^T \left(\int_u^T c_2(r) dr \right)^2 c_3^2(u) \sigma_\infty^2 du \right)^{1/2} \left(\int_0^T |h_u|^2 du \right)^{1/2}. \end{aligned}$$

Then we can extend this argument to obtain for any adapted \mathbb{R}_+ -valued process g and any $[s, t] \subset [0, T]$ the inequality

$$\mathbb{E} \int_s^t g_u |\mathbb{E}[\nabla_u F(X_{[0, T]}(x))|\mathcal{F}_u]|^2 du \leq \sigma_\infty^2 \mathbb{E} \int_s^t g_u c_3^2(u) \left(\int_u^T c_2(r) dr \right)^2 du.$$

This, due to (6.4), gives (2.7). This finishes the proof of Theorem 2.2. Notice that the inequalities therein are true even if the expectation on the right hand side is infinite. However, if we want to obtain transportation inequalities from Theorem 2.1, we need the Assumption E. Then we can apply our reasoning and the inequality (6.4) with the function $\phi(x) = \exp(\lambda x)$ and after simple calculations we obtain (6.5), which in the case of our bounds on Malliavin derivatives reads as

$$(6.11) \quad \mathbb{E}e^{\lambda(f(X_T(x)) - p_T f(x))} \leq \exp\left(\int_0^T \beta(\lambda c_1(T-t))dt + \frac{\lambda^2}{2}\sigma_\infty^2 c_2^2(T) \int_0^T c_3^2(t)dt\right)$$

and on the path space as

$$(6.12) \quad \mathbb{E}e^{\lambda(F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x)))} \leq \exp\left(\int_0^T \beta\left(\lambda \int_t^T c_1(r-t)dr\right)dt + \frac{\lambda^2}{2}\sigma_\infty^2 \int_0^T c_3^2(t) \left(\int_t^T c_2(r)dr\right)^2 dt\right).$$

Then, by the Gozlan-Léonard characterization (1.4) and the Fenchel-Moreau theorem, we easily get (2.4) from (6.11) and (2.5) from (6.12). \square

Remark 6.1. Note that if instead of (2.2) we have an inequality like

$$(6.13) \quad \mathbb{E}[|X_t - Y_t|/\mathcal{F}_s] \leq c_1(t-s)(|X_s - Y_s| + 1),$$

then, by the same reasoning as in Section 5.2, instead of (6.9) we get

$$\mathbb{E}[D_{t,u}f(X_T(x))|\mathcal{F}_t] \leq c_1(T-t)(g_\infty(u) + 1).$$

Then, if we want to obtain transportation or concentration inequalities, $g_\infty(u) + 1$ has to be square integrable with respect to the measure ν . However, if ν is a Lévy measure, this implies that ν has to be finite. This could still allow us to obtain some interesting results in certain cases that are not covered by Corollary 2.9, where Assumption L5 is required, which we do not need to obtain (6.13) (cf. Remark 2.6). For the sake of brevity, we skip the details.

Proof of Corollary 2.9. In the presence of two Gaussian and two jump noises, we use the Clark-Ocone formula of the form

$$\begin{aligned} F &= \mathbb{E}F + \int_0^T \mathbb{E}[\nabla_t^1 F|\mathcal{F}_t]dB_t^1 + \int_0^T \mathbb{E}[\nabla_t^2 F|\mathcal{F}_t]dB_t^2 \\ &+ \int_0^T \int_U \mathbb{E}[D_{t,u}^L F|\mathcal{F}_t]\tilde{N}^L(dt, du) + \int_0^T \int_U \mathbb{E}[D_{t,u}F|\mathcal{F}_t]\tilde{N}(dt, du), \end{aligned}$$

which holds for square integrable functionals F , where ∇^1 , ∇^2 , D^L and D are the Malliavin derivatives with respect to $(B_t^1)_{t \geq 0}$, $(B_t^2)_{t \geq 0}$, N^L and N , respectively (see e.g. Theorem 12.20 in [10]). Then we proceed as in the proof of Theorem 2.1, using the fact that under our assumptions, Theorem 2.3 and Theorem 2.8 provide us with couplings such that the conditions (2.14) and (2.15) are satisfied and this allows us to obtain the required bounds on the Malliavin derivatives (of the type (6.9) and (6.10)). More precisely, under our assumptions we obtain (2.19) and (2.20) from Corollary 2.15, whereas (5.25) and (5.26) follow from our reasoning at the end of Section 5.2. Combining all these bounds and using (6.4), just like in the proof of Theorem 2.1, allows us to obtain the desired transportation inequalities. Furthermore, taking $T \rightarrow \infty$ in the α_T - W_1H inequality, we obtain (2.16) by the argument from the proof of Lemma 2.2 in [11]. \square

ACKNOWLEDGEMENT

I would like to thank Arnaud Guillin and Liming Wu for suggesting the topic and for their hospitality and fruitful discussions during my visit to Clermont-Ferrand. This visit was financed by DAAD and took place while I was a PhD student funded by the Bonn International Graduate School of Mathematics. I am also grateful to my PhD advisor, Andreas Eberle, for many helpful suggestions and constant support. This research was partially supported by the ERC grant no. 694405.

REFERENCES

- [1] S. Albeverio, Z. Brzeźniak, J.-L. Wu, Existence of global solutions and invariant measures for stochastic differential equations driven by Poisson type noise with non-Lipschitz coefficients, *J. Math. Anal. Appl.* 371 (2010), no. 1, 309-322.
- [2] D. Applebaum, *Lévy Processes and Stochastic Calculus*, second ed., Cambridge University Press, 2009.
- [3] D. Bakry, M. Émery, Diffusions hypercontractives, *Séminaire de probabilités, XIX, 1983/84*, 177-206, *Lecture Notes in Math.*, 1123, Springer, Berlin, 1985.
- [4] R. F. Bass, M. Cranston, The Malliavin calculus for pure jump processes and applications to local time, *Ann. Probab.* 14 (1986), no. 2, 490-532.
- [5] R. F. Bass, *Diffusions and Elliptic Operators*, Springer-Verlag, New York, 1998.
- [6] K. Bichteler, J.-B. Gravereaux, J. Jacod, *Malliavin calculus for processes with jumps*, *Stochastics Monographs*, 2. Gordon and Breach Science Publishers, New York, 1987.
- [7] J.-M. Bismut, Calcul des variations stochastique et processus de sauts, *Z. Wahrsch. Verw. Gebiete* 63 (1983), no. 2, 147-235.
- [8] S. G. Bobkov, F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, *J. Funct. Anal.* 163 (1999), no. 1, 1-28.
- [9] R. Carmona, W. C. Masters, B. Simon, Relativistic Schrödinger operators: asymptotic behavior of the eigenfunctions, *J. Funct. Anal.* 91 (1990), no. 1, 117-142.
- [10] G. Di Nunno, B. Øksendal, F. Proske, *Malliavin Calculus for Lévy Processes with Applications to Finance*, Universitext. Springer-Verlag, Berlin, 2009.
- [11] H. Djellout, A. Guillin, L. Wu, Transportation cost-information inequalities and applications to random dynamical systems and diffusions, *Ann. Probab.* 32 (2004), no. 3B, 2702-2732.
- [12] A. Eberle, Reflection couplings and contraction rates for diffusions, *Probab. Theory Related Fields* 166 (2016), no. 3-4, 851-886.
- [13] A. Eberle, A. Guillin, R. Zimmer, Quantitative Harris type theorems for diffusions and McKean-Vlasov processes, preprint, arXiv:1606.06012.
- [14] N. Gozlan, C. Léonard, A large deviation approach to some transportation cost inequalities, *Probab. Theory Related Fields* 139 (2007), no. 1-2, 235-283.
- [15] N. Gozlan, C. Léonard, Transport inequalities. A survey, *Markov Process. Related Fields* 16 (2010), no. 4, 635-736.
- [16] A. Guillin, C. Léonard, L. Wu, N. Yao, Transportation-information inequalities for Markov processes, *Probab. Theory Related Fields* 144 (2009), no. 3-4, 669-695.
- [17] I. Gyöngy, N. V. Krylov, On stochastic equations with respect to semimartingales. I., *Stochastics* 4 (1980/81), no. 1, 1-21.
- [18] L. Huang, Density estimates for SDEs driven by tempered stable processes, preprint, arXiv:1504.04183.
- [19] T. Klein, Y. Ma, N. Privault, Convex concentration inequalities and forward-backward stochastic calculus, *Electron. J. Probab.* 11 (2006), no. 20, 486-512.
- [20] T. Komorowski, A. Walczuk, Central limit theorem for Markov processes with spectral gap in the Wasserstein metric, *Stochastic Process. Appl.* 122 (2012), no. 5, 2155-2184.
- [21] G. Last, M. Penrose, Martingale representation for Poisson processes with applications to minimal variance hedging, *Stochastic Process. Appl.* 121 (2011), no. 7, 1588-1606.
- [22] T. Lindvall, L. C. G. Rogers, Coupling of multidimensional diffusions by reflection, *Ann. Probab.* 14 (1986), no. 3, 860-872.
- [23] D. Luo, J. Wang, Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises, preprint, arXiv:1604.07206.

- [24] A. Løkka, Martingale representation of functionals of Lévy processes, *Stochastic Anal. Appl.* 22 (2004), no. 4, 867-892.
- [25] Y. Ma, Transportation inequalities for stochastic differential equations with jumps, *Stochastic Process. Appl.* 120 (2010), no. 1, 2-21.
- [26] Y. Ma, N. Privault, Convex concentration for some additive functionals of jump stochastic differential equations, *Acta Math. Sin. (Engl. Ser.)* 29 (2013), no. 8, 1449-1458.
- [27] M. B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, *Stochastic Process. Appl.* (2017), in press, <https://doi.org/10.1016/j.spa.2017.03.020>.
- [28] D. Nualart, *The Malliavin Calculus and Related Topics*, second ed., Springer-Verlag, Berlin, 2006.
- [29] F. Otto, C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, *J. Funct. Anal.* 173 (2000), no. 2, 361-400.
- [30] J. Picard, Formules de dualité sur l'espace de Poisson, *Ann. Inst. H. Poincaré Probab. Statist.* 32 (1996), no. 4, 509-548.
- [31] J. Picard, On the existence of smooth densities for jump processes, *Probab. Theory Related Fields* 105 (1996), no. 4, 481-511.
- [32] E. Priola, F. Y. Wang, Gradient estimates for diffusion semigroups with singular coefficients, *J. Funct. Anal.* 236 (2006), no. 1, 244-264.
- [33] P. Protter, *Stochastic Integration and Differential Equations*, Second edition. Version 2.1., Springer-Verlag, Berlin, 2005.
- [34] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, 1999.
- [35] M. Ryznar, Estimates of Green function for relativistic α -stable process, *Potential Anal.* 17 (2002), no. 1, 1-23.
- [36] J. Shao, C. Yuan, Transportation-cost inequalities for diffusions with jumps and its application to regime-switching processes, *J. Math. Anal. Appl.* 425 (2015), no. 2, 632-654.
- [37] C. Villani, *Optimal Transport. Old and New.*, Springer-Verlag, Berlin, 2009.
- [38] J. Wang, L^p -Wasserstein distance for stochastic differential equations driven by Lévy processes, *Bernoulli* 22 (2016), no. 3, 1598-1616.
- [39] L. Wu, Transportation inequalities for stochastic differential equations of pure jumps, *Ann. Inst. Henri Poincaré Probab. Stat.* 46 (2010), no. 2, 465-479.
- [40] R. Zimmer, Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions, *Stoch. Partial Differ. Equ. Anal. Comput.* (2017), in press, <https://doi.org/10.1007/s40072-017-0091-8>.

INSTITUTE FOR APPLIED MATHEMATICS, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY

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

6 A note on existence of global solutions and invariant measures for jump SDEs with locally one-sided Lipschitz drift

A NOTE ON EXISTENCE OF GLOBAL SOLUTIONS AND INVARIANT MEASURES FOR JUMP SDES WITH LOCALLY ONE-SIDED LIPSCHITZ DRIFT

MATEUSZ B. MAJKA

ABSTRACT. We extend some methods developed by Albeverio, Brzeźniak and Wu and we show how to apply them in order to prove existence of global strong solutions of stochastic differential equations with jumps, under a local one-sided Lipschitz condition on the drift (also known as a monotonicity condition) and a local Lipschitz condition on the diffusion and jump coefficients, while an additional global one-sided linear growth assumption is satisfied. Then we use these methods to prove existence of invariant measures for a broad class of such equations.

1. EXISTENCE OF GLOBAL SOLUTIONS UNDER LOCAL LIPSCHITZ CONDITIONS

Consider a stochastic differential equation in \mathbb{R}^d of the form

$$(1.1) \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\tilde{N}(dt, du).$$

Here $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ and $g : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$, where $(W_t)_{t \geq 0}$ is a d -dimensional Wiener process, (U, \mathcal{U}, ν) is a σ -finite measure space and $N(dt, du)$ is a Poisson random measure on $\mathbb{R}^+ \times U$ with intensity measure $dt \nu(du)$, while $\tilde{N}(dt, du) = N(dt, du) - dt \nu(du)$ is the compensated Poisson random measure. We denote by $(X_t(x))_{t \geq 0}$ a solution to (1.1) with initial condition $x \in \mathbb{R}^d$ and $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm of a matrix. The main result of the present paper is Theorem 2.1, where we prove existence of invariant measures for a certain class of such equations. However, we would first like to discuss the matter of existence of strong solutions to (1.1), in the context of the paper [1] by Albeverio, Brzeźniak and Wu. We claim that the following result holds.

Theorem 1.1. *Assume that the coefficients in (1.1) satisfy the following local one-sided Lipschitz condition, i.e., for every $R > 0$ there exists $C_R > 0$ such that for any $x, y \in \mathbb{R}^d$ with $|x|, |y| \leq R$ we have*

$$(1.2) \quad \langle b(x) - b(y), x - y \rangle + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq C_R |x - y|^2.$$

Moreover, assume a global one-sided linear growth condition, i.e., there exists $C > 0$ such that for any $x \in \mathbb{R}^d$ we have

$$(1.3) \quad \langle b(x), x \rangle + \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq C(1 + |x|^2).$$

Under (1.2) and (1.3) and an additional assumption that $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuous, there exists a unique global strong solution to (1.1).

2010 *Mathematics Subject Classification.* 60H10, 60G51.

Key words and phrases. Stochastic differential equations, Lévy processes, invariant measures.

The one-sided Lipschitz condition (1.2) above is sometimes called a monotonicity condition (see e.g. [6] or [9]) or a dissipativity condition ([10], [11] or [14]), although the term “dissipativity” is often reserved for the case in which (1.2) is satisfied with a negative constant $C_R < 0$. We keep using the latter convention, calling (1.2) one-sided Lipschitz regardless of the sign of the constant and using the term dissipativity only if the constant is negative. Note that the above theorem is a generalization of the following classic result.

Theorem 1.2. *Assume that the coefficients in (1.1) satisfy a global Lipschitz condition, i.e., there exists $C > 0$ such that for any $x, y \in \mathbb{R}^d$ we have*

$$(1.4) \quad |b(x) - b(y)|^2 + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq C|x - y|^2.$$

Moreover, assume a global linear growth condition, i.e., there exists $L > 0$ such that for any $x \in \mathbb{R}^d$ we have

$$(1.5) \quad |b(x)|^2 + \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq L(1 + |x|^2).$$

Under (1.4) and (1.5) there exists a unique strong solution to (1.1).

Theorem 1.2 is very well-known and its proof can be found in many textbooks, see e.g. Theorem IV-9.1 in [8] or Theorem 6.2.3 in [2]. However, Theorem 1.1 is not so widespread in the literature and we had significant problems with finding a suitable reference for such a result. We finally learned that Theorem 1.1 can be inferred from Theorem 2 in [6], where a more general result is proved for equations driven by locally square integrable càdlàg martingales taking values in Hilbert spaces.

Nevertheless, many authors use existence of solutions to equations like (1.1) under a one-sided Lipschitz condition for the drift (see e.g. [10], [11], [13], [14] for examples of some recent papers) claiming that this result is well-known, without giving any reference or while referring to positions that do not contain said result. Books that appear in this context include e.g. [3] and [12] which, admittedly, contain various interesting extensions of the classic Theorem 1.2, but not the extension in which the Lipschitz condition is replaced with a one-sided Lipschitz condition and the linear growth with a one-sided linear growth.

Moreover, in a quite recent paper [1], Albeverio, Brzeźniak and Wu proved the following result (see Theorem 3.1 therein).

Theorem 1.3. *Assume that the coefficients in (1.1) are such that for any $R > 0$ there exists $C_R > 0$ such that for any $x, y \in \mathbb{R}^d$ with $|x|, |y| \leq R$ we have*

$$(1.6) \quad |b(x) - b(y)|^2 + \|\sigma(x) - \sigma(y)\|_{HS}^2 \leq C_R|x - y|^2.$$

Moreover, there exists $L > 0$ such that for any $x, y \in \mathbb{R}^d$ we have

$$(1.7) \quad \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq L|x - y|^2.$$

Finally, we assume a global one-sided linear growth condition exactly like (1.3), i.e., there exists $C > 0$ such that for any $x \in \mathbb{R}^d$ we have

$$\langle b(x), x \rangle + \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq C(1 + |x|^2).$$

Then there exists a unique global strong solution to (1.1).

It is clear that Theorem 1.3 is less general than Theorem 1.1 and thus it is also a special case of Theorem 2 in [6]. Nevertheless, the proof in [1] is clearer and more direct than the one in [6], where the authors consider a much more general case. The main idea in [1] is to modify the locally Lipschitz coefficients in such a way as to obtain globally Lipschitz functions that agree with the given coefficients on a ball of fixed radius. Then using the classic Theorem 1.2 it is possible to obtain a solution in every such ball and then to “glue” such local solutions by using the global one-sided linear growth condition to obtain a global solution. It is important to mention that the authors of [1] also use their methods to prove existence of invariant measures for a broad class of equations of the form (1.1).

In view of all the above comments, we feel that it is necessary to give a direct proof of Theorem 1.1. Following the spirit of the proof in [1], we show how to extend the classic result (Theorem 1.2) in a step-by-step way in order to obtain Theorem 1.1. Then we explain how to use the methods from [1] to obtain existence of invariant measures in our case, see Theorem 2.1 in Section 2. The latter is an original result with potential applications in the theory of SPDEs, see Example 2.7.

For proving both Theorem 1.1 and 2.1 we need the following auxiliary result regarding a possible modification of the coefficients in (1.1).

Lemma 1.4. *Assume that the coefficients in (1.1) satisfy the local one-sided Lipschitz condition (1.2) and that they are locally bounded in the sense that for every $R > 0$ there exists an $M_R > 0$ such that for all $x \in \mathbb{R}^d$ with $|x| \leq R$ we have*

$$(1.8) \quad |b(x)|^2 + \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq M_R.$$

Then for every $R > 0$ there exist truncated functions $b_R : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma_R : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ and $g_R : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$ such that for all $x \in \mathbb{R}^d$ with $|x| \leq R$ we have

$$(1.9) \quad b_R(x) = b(x), \sigma_R(x) = \sigma(x) \text{ and } g_R(x, u) = g(x, u) \text{ for all } u \in \mathbb{R}^d.$$

Moreover, b_R , σ_R and g_R satisfy a global one-sided Lipschitz condition, i.e., there exists a constant $C(R) > 0$ such that for all $x, y \in \mathbb{R}^d$ we have

$$(1.10) \quad \langle b_R(x) - b_R(y), x - y \rangle + \|\sigma_R(x) - \sigma_R(y)\|_{HS}^2 + \int_U |g_R(x, u) - g_R(y, u)|^2 \nu(du) \leq C(R)|x - y|^2$$

and they are globally bounded, which means that there exists $M(R) > 0$ such that for all $x \in \mathbb{R}^d$ we have

$$(1.11) \quad |b_R(x)|^2 + \|\sigma_R(x)\|_{HS}^2 + \int_U |g_R(x, u)|^2 \nu(du) \leq M(R).$$

Then, combining Theorem 1.2 and Lemma 1.4, we are able to prove existence of solutions while the coefficients in (1.1) are bounded and satisfy a global one-sided Lipschitz condition.

Theorem 1.5. *Assume that b is continuous and that the coefficients in (1.1) satisfy a global one-sided Lipschitz condition, i.e., there exists $K > 0$ such that for all $x, y \in \mathbb{R}^d$ we have*

$$(1.12) \quad \langle b(x) - b(y), x - y \rangle + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq K|x - y|^2.$$

Additionally, assume that the coefficients are globally bounded, i.e., there exists $M > 0$ such that for all $x \in \mathbb{R}^d$ we have

$$(1.13) \quad |b(x)|^2 + \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq M.$$

Then there exists a unique strong solution to (1.1).

The proofs of Lemma 1.4 and Theorem 1.5 can be found in Section 3. Having proved the above two results, we proceed with the proof of Theorem 1.1 as in [1] (see Proposition 2.9 and Theorem 3.1 therein, see also [6], page 14, for a similar reasoning). More details can be found at the end of Section 3 below.

2. EXISTENCE OF INVARIANT MEASURES

The existence of an invariant measure for the solution of (1.1) is shown using the Krylov-Bogoliubov method, see e.g. Theorem III-2.1 in [7] and the discussion in the introduction to [5]. It follows from there that for the existence of an invariant measure for a process $(X_t)_{t \geq 0}$ with a Feller semigroup $(p_t)_{t \geq 0}$ it is sufficient to show that for some $x \in \mathbb{R}^d$ the process $(X_t(x))_{t \geq 0}$ is bounded in probability at infinity in the sense that for any $\varepsilon > 0$ there exist $R > 0$ and $t > 0$ such that for all $s \geq t$ we have

$$(2.1) \quad \mathbb{P}(|X_s(x)| > R) < \varepsilon.$$

Therefore if we show that there exist constants $M, K > 0$ such that

$$(2.2) \quad \mathbb{E}|X_t(x)|^2 \leq |x|^2 e^{-Kt} + M/K$$

holds for all $t \geq 0$, then (2.1) follows easily by the Chebyshev inequality and we obtain the existence of an invariant measure. Based on this idea, we can prove the following result.

Theorem 2.1. *Assume that the coefficients in (1.1) satisfy the local one-sided Lipschitz condition (1.2) and that there exist constants $K, M > 0$ such that for all $x \in \mathbb{R}^d$ we have*

$$(2.3) \quad \langle b(x), x \rangle + \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq -K|x|^2 + M.$$

Assume also that there exists a constant $L > 0$ such that for all $x \in \mathbb{R}^d$ we have

$$(2.4) \quad \|\sigma(x)\|_{HS}^2 + \int_U |g(x, u)|^2 \nu(du) \leq L(1 + |x|^2).$$

Finally, let the drift coefficient b in (1.1) be continuous. Then there exists an invariant measure for the solution of (1.1).

We can compare this result with the one proved in [1] (see Theorem 4.5 therein).

Theorem 2.2. *Assume that the coefficients b and σ in (1.1) satisfy the local Lipschitz condition (1.6) and that g satisfies the global Lipschitz condition (1.7). Assume also the condition (2.3) as in the Theorem 2.1 above. Then there exists an invariant measure for the solution of (1.1).*

Remark 2.3. Observe that our additional condition (2.4) in Theorem 2.1 does not follow from (2.3) since $\langle b(x), x \rangle$ can be negative. Therefore it would seem that our result is not a straightforward generalization of Theorem 4.5 in [1]. However, we believe that the condition (2.4) is also necessary to prove Theorem 4.5 in [1], at least we were not able to retrace the proof of Proposition 4.3 therein (which is crucial for the proof of Theorem 4.5) without this additional condition. Therefore we are convinced that (2.4) should be

added to the list of assumptions of Theorem 4.5 in [1] and that our result is indeed its strict generalization. This has been confirmed in our private communication with one of the authors of [1].

For the proof of Theorem 2.1 we first need the following fact, which can be proved exactly like in [1].

Lemma 2.4. *The solution $(X_t)_{t \geq 0}$ to the equation (1.1) is a strong Markov process and thus it generates a Markov semigroup $(p_t)_{t \geq 0}$.*

Proof. See Proposition 4.1 and Proposition 4.2 in [1]. □

Now we need the following lemma, which is a generalization of Proposition 4.3 in [1] (see Remark 2.3 about inclusion of the assumption (2.4)).

Lemma 2.5. *Under the assumptions (1.2), (1.3), (2.4) and if b is continuous, the semigroup $(p_t)_{t \geq 0}$ associated with the solution $(X_t)_{t \geq 0}$ of (1.1) is Feller.*

Having proved the above lemma, we can easily conclude the proof of Theorem 2.1, following the proof of Theorem 4.5 in [1], i.e., we just use the condition (2.3) to show (2.2) and then use the Krylov-Bogoliubov method presented above. More details can be found in Section 3.

Before concluding this section, let us look at some examples.

Example 2.6. Consider an SDE of the form (1.1) with the drift given by

$$b(x) := -x|x|^{-\alpha} \mathbf{1}_{\{x \neq 0\}},$$

where $\alpha \in (0, 1)$. Equations of this type are considered in Example 171 in [12]. It is easy to check that the function b defined above is not locally Lipschitz, since it does not satisfy a Lipschitz condition in any neighbourhood of zero. However, we can show that it satisfies a one-sided Lipschitz condition globally with constant zero. Indeed, following the calculations in Example 171 in [12], for any nonzero $x, y \in \mathbb{R}^d$ we have

$$\begin{aligned} \langle x - y, -x|x|^{-\alpha} + y|y|^{-\alpha} \rangle &= -|x|^{2-\alpha} + \langle y, x|x|^{-\alpha} \rangle + \langle x, y|y|^{-\alpha} \rangle - |y|^{2-\alpha} \\ &\leq -|x|^{2-\alpha} - |y|^{2-\alpha} + |y||x|^{1-\alpha} + |x||y|^{1-\alpha} \\ &= (|x| - |y|)(|y|^{1-\alpha} - |x|^{1-\alpha}) \leq 0, \end{aligned}$$

where the last inequality holds since $1 - \alpha \in (0, 1)$. Thus, if we consider an equation of the form (1.1) with the drift b and any locally Lipschitz coefficients σ and g , the condition (1.2) is satisfied. Moreover, if σ and g satisfy the global linear growth condition (2.4) with some constant $L > 0$, then by replacing the drift b defined above with

$$\tilde{b}(x) := b(x) - Kx,$$

where $K > L$, we obtain coefficients that satisfy (2.3). More generally, we can take

$$\tilde{b}(x) := b(x) - \nabla U(x),$$

where U is a strongly convex function with convexity constant $K > L$. This way we obtain a class of examples of equations for which our Theorem 2.1 applies, but Theorem 4.5 in [1] does not, since the local Lipschitz assumption is not satisfied.

Example 2.7. Our results may have applications in the study of stochastic evolution equations with Lévy noise on infinite dimensional spaces, where the coefficients are often not Lipschitz, see e.g. [4] and the references therein. In particular, in [4] the authors consider SPDEs with drifts satisfying a local monotonicity condition and use their finite

dimensional approximations, which may lead to SDEs satisfying our condition (1.2), cf. the condition (H2) and the formula (4.4) in [4].

3. PROOFS

In order to keep our presentation compact, we will only present the proof of Theorem 1.1 in a slightly less general setting than that presented in the first section. Namely, we will additionally assume that the diffusion coefficient σ and the jump coefficient g in the equation (1.1) satisfy a local Lipschitz condition separately from the drift b , i.e., for every $R > 0$ there exists $S_R > 0$ such that for any $x, y \in \mathbb{R}^d$ with $|x|, |y| \leq R$ we have

$$(3.1) \quad \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq S_R |x - y|^2.$$

Obviously, (3.1) does not follow from (1.2), since the values of $\langle b(x) - b(y), x - y \rangle$ can be negative. However, requiring the condition (3.1) to be satisfied seems to be rather natural in many cases. It is possible to weaken this assumption and prove the exact statement of Theorem 1.1 using methods from Section 3 of Chapter II in [9] (see also Section 3 in [6]), but this creates additional technical difficulties and thus we decided to omit this extension here, aiming at a clear and straightforward presentation.

The consequence of adding the assumption (3.1) is that the coefficients of (1.1) automatically satisfy the local boundedness condition (1.8) required in Lemma 1.4 (remember that b is assumed to be continuous and thus it is locally bounded anyway). It also means that from Lemma 1.4 we obtain coefficients σ_R and g_R that satisfy a separate global Lipschitz condition, i.e., the condition (1.10) without the term involving b_R . Hence we can prove Theorem 1.5 under an additional assumption, i.e., we can use the fact that there exists $S > 0$ such that for all $x, y \in \mathbb{R}^d$ we have

$$(3.2) \quad \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq S |x - y|^2.$$

However, the assumption (3.1) is not needed for the proof of Theorem 2.1, where we also use Lemma 1.4, but we do not need to obtain truncated coefficients σ_R and g_R satisfying a separate global Lipschitz condition and the assumption about local boundedness is guaranteed by the separate linear growth condition (2.4) and the continuity of b . Thus the reasoning presented below gives a complete proof of the exact statement of our Theorem 2.1.

Proof of Lemma 1.4. For a related reasoning, see the proof of Lemma 4 in [6] or Lemma 172 in [12]. Note that the method of truncating the coefficients of (1.1) which was used in the proof of Proposition 2.7 in [1] and which works in the case of Lipschitz coefficients, does not work for a one-sided Lipschitz drift and thus we need a different approach. For any $R > 0$, we can consider a smooth, non-negative function with compact support $\eta_R \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ such that

$$\eta_R(x) = \begin{cases} 1, & \text{if } |x| \leq R, \\ 0, & \text{if } |x| > R + 1 \end{cases}$$

and $\eta_R(x) \leq 1$ for all $x \in \mathbb{R}^d$. Then we can define

$$b_R(x) := \eta_R(x)b(x), \quad \sigma_R(x) := \eta_R(x)\sigma(x) \quad \text{and} \quad g_R(x, u) := \eta_R(x)g(x, u) \quad \text{for all } u \in U.$$

Then it is obvious that the condition (1.9) is satisfied and the condition (1.11) immediately follows from (1.8). Therefore it remains to be shown that the functions b_R , σ_R and g_R

satisfy the global one-sided Lipschitz condition (1.10). We have

$$\begin{aligned}
(3.3) \quad & \langle b_R(x) - b_R(y), x - y \rangle + \|\sigma_R(x) - \sigma_R(y)\|_{HS}^2 + \int_U |g_R(x, u) - g_R(y, u)|^2 \nu(du) \\
& = \langle \eta_R(x)b(x) - \eta_R(y)b(y), x - y \rangle + \|\eta_R(x)\sigma(x) - \eta_R(y)\sigma(y)\|_{HS}^2 \\
& + \int_U |\eta_R(x)g(x, u) - \eta_R(y)g(y, u)|^2 \nu(du) \\
& \leq \eta_R(x)\langle b(x) - b(y), x - y \rangle + \langle (\eta_R(x) - \eta_R(y))b(y), x - y \rangle \\
& + |\eta_R(x)|^2 \|\sigma(x) - \sigma(y)\|_{HS}^2 + \|(\eta_R(x) - \eta_R(y))\sigma(y)\|_{HS}^2 \\
& + \int_U |\eta_R(x)|^2 |g(x, u) - g(y, u)|^2 \nu(du) + \int_U |\eta_R(x) - \eta_R(y)|^2 |g(y, u)|^2 \nu(du).
\end{aligned}$$

Now assume x and y are such that

$$(3.4) \quad \eta_R(y) \geq \eta_R(x) > 0.$$

The case when $\eta_R(x) = 0$ is simpler and the case $\eta_R(y) \leq \eta_R(x)$ can be handled by changing the role of x and y in the calculations above. From (3.4) it follows that $|y| \leq R+1$ and $|x| \leq R+1$ and thus we can use the local one-sided Lipschitz condition (1.2) with $R+1$ to get

$$\langle b(x) - b(y), x - y \rangle + \|\sigma(x) - \sigma(y)\|_{HS}^2 + \int_U |g(x, u) - g(y, u)|^2 \nu(du) \leq C_{R+1}|x - y|^2$$

with some constant C_{R+1} . Combining this with the fact that $\eta_R \leq 1$ (and thus $\eta_R^2 \leq \eta_R$) allows us to bound the sum of the first, the third and the fifth term on the right hand side of (3.3) by $C_{R+1}|x - y|^2$. Observe now that the function η_R is Lipschitz (with a constant, say, $C_{Lip(\eta_R)}$) and thus

$$\eta_R(x) - \eta_R(y) \leq C_{Lip(\eta_R)}|x - y|.$$

Since $|y| \leq R+1$, we can use the local boundedness condition (1.8) with some constant M_{R+1} . We first bound $|b(y)|$ by the square root of the left hand side of (1.8) in order to get

$$\langle (\eta_R(x) - \eta_R(y))b(y), x - y \rangle \leq \sqrt{M_{R+1}} C_{Lip(\eta_R)} |x - y|^2.$$

Then we use (1.8) once again in order to bound the sum of the fourth and the sixth term on the right hand side of (3.3) by $M_{R+1} C_{Lip(\eta_R)}^2 |x - y|^2$. Combining all these facts together, we can bound the right hand side of (3.3) by

$$C_{R+1}|x - y|^2 + \sqrt{M_{R+1}} C_{Lip(\eta_R)} |x - y|^2 + M_{R+1} C_{Lip(\eta_R)}^2 |x - y|^2.$$

Therefore the global one-sided Lipschitz condition for b_R , σ_R and g_R is satisfied with a constant

$$C(R) := C_{R+1} + \sqrt{M_{R+1}} C_{Lip(\eta_R)} + M_{R+1} C_{Lip(\eta_R)}^2,$$

which finishes the proof. \square

Before we proceed to the proof of Theorem 1.5, let us formulate a crucial technical lemma. Its proof is just a slightly altered second part of the proof of Lemma 3.3 in [10], but we include the full calculations here for completeness and, more importantly, because we need to use a related, but modified reasoning in the proof of Theorem 1.5. The lemma itself will be used in the proof of Lemma 2.5 later on.

Lemma 3.1. *Assume that the coefficients of the equation (1.1) with an initial condition $x \in \mathbb{R}^d$ satisfy the global one-sided linear growth condition (1.3) and that σ and g additionally satisfy the separate linear growth condition (2.4). Then there exist constants $\tilde{C} > 0$ and $\tilde{K} > 0$ such that*

$$\mathbb{E} \sup_{s \leq t} |X_s|^2 \leq \tilde{K} e^{2\tilde{C}t} (1 + |x|^2),$$

where $(X_t)_{t \geq 0} = (X_t(x))_{t \geq 0}$ is a solution to (1.1) with initial condition $x \in \mathbb{R}^d$.

Proof. By the Itô formula, we have

$$(3.5) \quad \begin{aligned} |X_t|^2 &= |x|^2 + 2 \int_0^t \langle X_s, b(X_s) \rangle ds + 2 \int_0^t \langle X_s, \sigma(X_s) dW_s \rangle \\ &+ \int_0^t \|\sigma(X_s)\|_{HS}^2 ds + 2 \int_0^t \int_U \langle X_s, g(X_{s-}, u) \rangle \tilde{N}(ds, du) \\ &+ \int_0^t \int_U |g(X_{s-}, u)|^2 N(ds, du). \end{aligned}$$

Now let us consider the process

$$M_t := \int_0^t \langle X_s, \sigma(X_s) dW_s \rangle + \int_0^t \int_U \langle X_s, g(X_{s-}, u) \rangle \tilde{N}(ds, du),$$

which is a local martingale. Thus, by the Burkholder-Davis-Gundy inequality, there exists a constant $C_1 > 0$ such that

$$\begin{aligned} \mathbb{E} \sup_{s \leq t} |M_s| &\leq C_1 \mathbb{E} \left[\int_0^t |\sigma^*(X_s) X_s|^2 ds + \int_0^t \int_U |\langle X_s, g(X_{s-}, u) \rangle|^2 N(ds, du) \right]^{\frac{1}{2}} \\ &\leq C_1 \mathbb{E} \left[\left(\sup_{s \leq t} |X_s|^2 \right) \left(\int_0^t \|\sigma^*(X_s)\|^2 ds + \int_0^t \int_U |g(X_{s-}, u)|^2 N(ds, du) \right) \right]^{\frac{1}{2}} \\ &\leq C_1 \left(\mathbb{E} \sup_{s \leq t} |X_s|^2 \right)^{\frac{1}{2}} \left(\mathbb{E} \left[\int_0^t \|\sigma^*(X_s)\|^2 ds + \int_0^t \int_U |g(X_{s-}, u)|^2 N(ds, du) \right] \right)^{\frac{1}{2}} \\ &\leq \frac{C_1}{2} a \mathbb{E} \sup_{s \leq t} |X_s|^2 + \frac{C_1}{2a} \mathbb{E} \left[\int_0^t \|\sigma^*(X_s)\|^2 ds + \int_0^t \int_U |g(X_{s-}, u)|^2 N(ds, du) \right] \\ &\leq \frac{C_1}{2} a \mathbb{E} \sup_{s \leq t} |X_s|^2 + \frac{C_1}{2a} L \mathbb{E} \int_0^t (|X_s|^2 + 1) ds. \end{aligned}$$

Here $\|\cdot\|$ denotes the operator norm and σ^* is a transposed σ . In the third step we used the Hölder inequality in the form $\mathbb{E} A^{\frac{1}{2}} B^{\frac{1}{2}} \leq (\mathbb{E} A)^{\frac{1}{2}} (\mathbb{E} B)^{\frac{1}{2}}$, in the fourth step we used $(AB)^{\frac{1}{2}} \leq \frac{1}{2} a A + \frac{1}{2a} B$ for any $a > 0$, which can be chosen later, and in the fifth step we used the separate global linear growth condition (2.4) for σ and g along with the fact that $\|\cdot\| \leq \|\cdot\|_{HS}$. Now we can use the formula (3.5) to get

$$(3.6) \quad \begin{aligned} \mathbb{E} \sup_{s \leq t} |X_s|^2 &\leq |x|^2 + 2 \mathbb{E} \sup_{s \leq t} |M_s| + 2 \mathbb{E} \sup_{s \leq t} \int_0^s \langle X_r, b(X_r) \rangle dr \\ &+ \mathbb{E} \sup_{s \leq t} \left[\int_0^s \|\sigma(X_r)\|_{HS}^2 dr + \int_0^s \int_U |g(X_{r-}, u)|^2 N(dr, du) \right]. \end{aligned}$$

Observe that obviously

$$(3.7) \quad \langle X_r, b(X_r) \rangle \leq \langle X_r, b(X_r) \rangle + \|\sigma(X_r)\|_{HS}^2 + \int_U |g(X_{r-}, u)|^2 \nu(du)$$

and thus from the global one-sided linear growth condition (1.3) we get

$$\mathbb{E} \sup_{s \leq t} \int_0^s \langle X_r, b(X_r) \rangle dr \leq C \mathbb{E} \sup_{s \leq t} \int_0^s (|X_r|^2 + 1) dr \leq C \mathbb{E} \int_0^t (|X_r|^2 + 1) dr.$$

On the other hand, using the separate linear growth condition (2.4) we get

$$\begin{aligned} & \mathbb{E} \sup_{s \leq t} \left[\int_0^s \|\sigma(X_r)\|_{HS}^2 dr + \int_0^s \int_U |g(X_{r-}, u)|^2 N(dr, du) \right] \\ &= \mathbb{E} \left[\int_0^t \|\sigma(X_r)\|_{HS}^2 dr + \int_0^t \int_U |g(X_{r-}, u)|^2 N(dr, du) \right] \\ &= \mathbb{E} \left[\int_0^t \|\sigma(X_r)\|_{HS}^2 dr + \int_0^t \int_U |g(X_{r-}, u)|^2 \nu(du) dr \right] \\ &\leq L \mathbb{E} \int_0^t (|X_r|^2 + 1) dr. \end{aligned}$$

Combining all the above estimates, we get from (3.6) that

$$\mathbb{E} \sup_{s \leq t} |X_s|^2 \leq |x|^2 + C_1 a \mathbb{E} \sup_{s \leq t} |X_s|^2 + \left(\frac{C_1}{a} L + 2C + L \right) \mathbb{E} \int_0^t (|X_r|^2 + 1) dr.$$

Now, choosing $a = 1/(2C_1)$ we obtain

$$\mathbb{E} \sup_{s \leq t} |X_s|^2 \leq 2|x|^2 + 2(2C_1^2 L + 2C + L) \mathbb{E} \int_0^t \sup_{w \leq r} (|X_w|^2 + 1) dr.$$

Hence, using the Gronwall inequality for the function $\mathbb{E} \sup_{s \leq t} |X_s|^2 + 1$ we get

$$\mathbb{E} \sup_{s \leq t} |X_s|^2 + 1 \leq 2(|x|^2 + 1) \exp(2(2C_1^2 L + 2C + L)t),$$

which finishes the proof. \square

Proof of Theorem 1.5. Let $j \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ be a smooth function with a compact support contained in $B(0, 1)$, such that $\int_{\mathbb{R}^d} j(z) dz = 1$. Then, for any $k \geq 1$, define

$$b^k(x) := \int_{\mathbb{R}^d} b\left(x - \frac{z}{k}\right) j(z) dz.$$

Now we can consider a sequence of equations

$$(3.8) \quad dX_t^k = b^k(X_t^k) dt + \sigma(X_t^k) dW_t + \int_U g(X_{t-}^k, u) \tilde{N}(dt, du).$$

Note that we have replaced only the drift coefficient b with b^k while σ and g remain unchanged. This is due to the fact that we decided to prove Theorem 1.1 with an additional assumption of separate local Lipschitz condition (3.1) for σ and g . Thanks to this, we can work in the present proof under an additional assumption that σ and g are globally Lipschitz, i.e. they satisfy (3.2), cf. the discussion at the beginning of this section. Now observe that the function b^k defined above is also globally Lipschitz.

Indeed, for any $x, y \in \mathbb{R}^d$ we have

$$\begin{aligned}
|b^k(x) - b^k(y)| &= \left| \int_{\mathbb{R}^d} b\left(x - \frac{z}{k}\right) j(z) dz - \int_{\mathbb{R}^d} b\left(y - \frac{z}{k}\right) j(z) dz \right| \\
&= \left| k^d \int_{\mathbb{R}^d} b(w) j(k(x - w)) dz - k^d \int_{\mathbb{R}^d} b(w) j(k(y - w)) dw \right| \\
&\leq k^d \int_{\mathbb{R}^d} |b(w)| |j(k(x - w)) - j(k(y - w))| dw \\
&\leq k^{d+1} \sqrt{M} |x - y| \int_{\mathbb{R}^d} \sup_{w \in \mathbb{R}^d} |\nabla j(w)| dw,
\end{aligned}$$

where in the last step we use the fact that b is bounded by \sqrt{M} (cf. (1.13)) and j is Lipschitz with the Lipschitz constant given by the supremum of the norm of its gradient (which is obviously integrable since $j \in \mathcal{C}_c^\infty(\mathbb{R}^d)$). Having proved that b^k is globally Lipschitz, we can use Theorem 1.2 to ensure existence of a unique strong solution $(X_t^k)_{t \geq 0}$ to the equation (3.8). We will prove now that the sequence of solutions $\{(X_t^k)_{t \geq 0}\}_{k=1}^\infty$ has a limit (in the sense of almost sure convergence, uniform on bounded time intervals) and that this limit is in fact a solution to (1.1). To this end, we will make use of the calculations from Lemma 3.1.

Observe that for any $k, l \geq 1$, if we use the Itô formula to calculate $|X_t^k - X_t^l|^2$, we will obtain exactly the formula (3.5) with X_t replaced by the difference $X_t^k - X_t^l$ and the function $b(X_s)$ replaced by $b^k(X_s^k) - b^l(X_s^l)$. Furthermore, we can make the term $|x|^2$ vanish (we can assume that all the solutions $(X_t^k)_{t \geq 0}$ have the same initial condition). Now we can proceed exactly like in the proof of Lemma 3.1, this time using the separate global Lipschitz condition (3.2) for σ and g in the steps where we used the separate linear growth condition (2.4) before, in order to get

$$\begin{aligned}
(3.9) \quad \mathbb{E} \sup_{s \leq t} |X_s^k - X_s^l|^2 &\leq C_1 a \mathbb{E} \sup_{s \leq t} |X_s^k - X_s^l|^2 + \left(\frac{C_1}{a} S + S \right) \mathbb{E} \int_0^t |X_r^k - X_r^l|^2 dr \\
&\quad + 2 \mathbb{E} \sup_{s \leq t} \int_0^s \langle X_r^k - X_r^l, b^k(X_r^k) - b^l(X_r^l) \rangle dr.
\end{aligned}$$

Thus the only term, with which we have to deal in a different way compared to the proof of Lemma 3.1, is the last one. We have

$$\begin{aligned}
&\mathbb{E} \sup_{s \leq t} \int_0^s \langle X_r^k - X_r^l, b^k(X_r^k) - b^l(X_r^l) \rangle dr \\
&= \mathbb{E} \sup_{s \leq t} \int_0^s \left\langle X_r^k - X_r^l, \int_{\mathbb{R}^d} b\left(X_r^k - \frac{z}{k}\right) j(z) dz - \int_{\mathbb{R}^d} b\left(X_r^l - \frac{z}{l}\right) j(z) dz \right\rangle dr \\
&= \mathbb{E} \sup_{s \leq t} \int_0^s \left\{ \int_{\mathbb{R}^d} \left\langle \left(X_r^k - \frac{z}{k}\right) - \left(X_r^l - \frac{z}{l}\right), b\left(X_r^k - \frac{z}{k}\right) - b\left(X_r^l - \frac{z}{l}\right) \right\rangle j(z) dz \right. \\
&\quad \left. + \int_{\mathbb{R}^d} \left\langle \frac{z}{k} - \frac{z}{l}, b\left(X_r^k - \frac{z}{k}\right) - b\left(X_r^l - \frac{z}{l}\right) \right\rangle j(z) dz \right\} dr \\
&=: \mathbb{E} \sup_{s \leq t} \int_0^s (I_r^1 + I_r^2) dr.
\end{aligned}$$

Now observe that since b is assumed to be bounded by \sqrt{M} (see (1.13)), we have

$$\begin{aligned} I_r^2 &\leq 2\sqrt{M} \int_{\mathbb{R}^d} \left| \frac{z}{k} - \frac{z}{l} \right| j(z) dz = 2\sqrt{M} \left| \frac{1}{k} - \frac{1}{l} \right| \int_{\mathbb{R}^d} |z| j(z) dz \\ &=: 2\sqrt{M} \left| \frac{1}{k} - \frac{1}{l} \right| C^1(j) < \infty. \end{aligned}$$

As for I_r^1 , we can use the one-sided Lipschitz condition (1.12) for b similarly like we used one-sided linear growth in (3.7) to get

$$\begin{aligned} I_r^1 &\leq K \int_{\mathbb{R}^d} \left| \left(X_r^k - \frac{z}{k} \right) - \left(X_r^l - \frac{z}{l} \right) \right|^2 j(z) dz \\ &= K \int_{\mathbb{R}^d} \left| \left(X_r^k - X_r^l \right) - \left(\frac{1}{k} - \frac{1}{l} \right) z \right|^2 j(z) dz \\ &\leq 2K \int_{\mathbb{R}^d} |X_r^k - X_r^l|^2 j(z) dz + 2K \int_{\mathbb{R}^d} \left| \frac{1}{k} - \frac{1}{l} \right|^2 |z|^2 j(z) dz \\ &=: 2K |X_r^k - X_r^l|^2 + 2KC^2(j) \left| \frac{1}{k} - \frac{1}{l} \right|^2. \end{aligned}$$

Combining the above estimates, we have

$$\begin{aligned} \mathbb{E} \sup_{s \leq t} \int_0^s (I_r^1 + I_r^2) dr &\leq 2K \mathbb{E} \int_0^t |X_r^k - X_r^l|^2 dr + 2tKC^2(j) \left| \frac{1}{k} - \frac{1}{l} \right|^2 \\ &\quad + 2t\sqrt{M} \left| \frac{1}{k} - \frac{1}{l} \right| C^1(j) \\ &\leq 2K \mathbb{E} \int_0^t \sup_{w \leq r} |X_w^k - X_w^l|^2 dr + \widehat{C}t \left| \frac{1}{k} - \frac{1}{l} \right|, \end{aligned}$$

where the last inequality holds with a constant $\widehat{C} := 2KC^2(j) + 2\sqrt{M}C^1(j)$ for k and l large enough so that $|\frac{1}{k} - \frac{1}{l}| < 1$. Now we can come back to (3.9) and, taking $a = 1/(2C_1)$, similarly like in the proof of Lemma 3.1 we obtain

$$\begin{aligned} \mathbb{E} \sup_{s \leq t} |X_s^k - X_s^l|^2 &\leq 2(2C_1^2S + S) \mathbb{E} \int_0^t \sup_{w \leq r} |X_w^k - X_w^l|^2 dr \\ &\quad + 8K \mathbb{E} \int_0^t \sup_{w \leq r} |X_w^k - X_w^l|^2 dr + 4\widehat{C}t \left| \frac{1}{k} - \frac{1}{l} \right|. \end{aligned}$$

The Gronwall inequality implies

$$\mathbb{E} \sup_{s \leq t} |X_s^k - X_s^l|^2 \leq 4\widehat{C}t \left| \frac{1}{k} - \frac{1}{l} \right| \exp \{ (4C_1^2S + 2S + 8K) t \}.$$

From this we can infer that there exists a process $(X_t)_{t \geq 0}$ such that

$$(3.10) \quad \mathbb{E} \sup_{s \leq t} |X_s - X_s^k|^2 \rightarrow 0 \text{ as } k \rightarrow \infty.$$

It remains to be shown that $(X_t)_{t \geq 0}$ is indeed a solution to (1.1). Observe that, by choosing a subsequence, we have $X_t^k \rightarrow X_t$ almost surely as $k \rightarrow \infty$ and thus, since b is assumed to be continuous, we get

$$b \left(X_t^k - \frac{z}{k} \right) \rightarrow b(X_t) \text{ almost surely as } k \rightarrow \infty.$$

But b is bounded by the constant \sqrt{M} and

$$\int_0^t \int_{\mathbb{R}^d} b\left(X_s^k - \frac{z}{k}\right) j(z) dz ds \leq \sqrt{M} \int_0^t \int_{\mathbb{R}^d} j(z) dz ds < \infty.$$

Therefore we get

$$\int_0^t \int_{\mathbb{R}^d} b\left(X_s^k - \frac{z}{k}\right) j(z) dz ds \rightarrow \int_0^t \int_{\mathbb{R}^d} b(X_s) j(z) dz ds = \int_0^t b(X_s) ds \text{ as } k \rightarrow \infty \text{ a.s.}$$

Moreover, using the Itô isometry and (3.10), we can easily prove that

$$\int_0^t \sigma(X_s^k) dW_s \rightarrow \int_0^t \sigma(X_s) dW_s$$

and

$$\int_0^t \int_U g(X_{s-}^k, u) \tilde{N}(ds, du) \rightarrow \int_0^t \int_U g(X_{s-}, u) \tilde{N}(ds, du)$$

almost surely (by choosing a subsequence), as $k \rightarrow \infty$, which finishes the proof. \square

Now we proceed with the proof of Lemma 2.5, which is needed to ensure existence of an invariant measure for the solution to (1.1).

Proof of Lemma 2.5. First observe that under our assumptions, we can use Lemma 3.1 to get

$$\mathbb{E} \sup_{s \leq t} |X_s|^2 \leq K_1(1 + |x|^2) e^{K_2 t}$$

for some constants $K_1, K_2 > 0$, where $(X_t)_{t \geq 0} = (X_t(x))_{t \geq 0}$ is a solution to (1.1) with initial condition $x \in \mathbb{R}^d$. Hence, by the Chebyshev inequality, for any $\varepsilon > 0$ we can find $R > 0$ large enough so that for any $x \in \mathbb{R}^d$ with $|x| \leq R$ we have

$$(3.11) \quad \mathbb{P} \left[\sup_{s \leq t} |X_s(x)| \geq R \right] < \varepsilon.$$

Now without loss of generality assume that $t \leq 1$ and fix $\varepsilon > 0$ and $R > 0$ like above. We can consider a solution $(X_t^R)_{t \geq 0}$ to the equation (1.1) with the coefficients replaced by the truncated coefficients b_R, σ_R and g_R obtained from Lemma 1.4 (note that the local boundedness assumption (1.8) in Lemma 1.4 is satisfied due to the continuity of b and the separate linear growth condition (2.4) for σ and g , cf. the discussion at the beginning of this section). Then b_R, σ_R and g_R satisfy a global one-sided Lipschitz condition (1.10) with some constant $C(R) > 0$. Moreover, we have $X_s = X_s^R$ for $s \leq \tau_R$ with τ_R defined by

$$(3.12) \quad \tau_R := \inf\{t > 0 : |X_t^R| \geq R\}.$$

Thus for any $x, y \in \mathbb{R}^d$ with $|x| \leq R$ and $|y| \leq R$ and for any $\delta > 0$ we have

$$\mathbb{P}(|X_1(x) - X_1(y)| > \delta) \leq \varepsilon + \mathbb{P}(|X_1^R(x) - X_1^R(y)| > \delta) \leq \varepsilon + \frac{1}{\delta^2} \mathbb{E}|X_1^R(x) - X_1^R(y)|^2,$$

where the first step follows from (3.11) and some straightforward calculations (see page 321 in [1] for details) and the second step is just the Chebyshev inequality. Now from the Itô formula used similarly like in (3.5) (cf. also the proof of Theorem 1.5, although here

we need a different local martingale than in the case where we estimate a supremum) we get

$$\begin{aligned}
|X_1^R(x) - X_1^R(y)|^2 &= |x - y|^2 + 2 \int_0^1 \langle X_s^R(x) - X_s^R(y), b_R(X_s^R(x)) - b_R(X_s^R(y)) \rangle ds \\
&\quad + 2 \int_0^1 \langle X_s^R(x) - X_s^R(y), (\sigma_R(X_s^R(x)) - \sigma_R(X_s^R(y))) dW_s \rangle \\
&\quad + \int_0^1 \|\sigma_R(X_s^R(x)) - \sigma_R(X_s^R(y))\|_{HS}^2 ds \\
&\quad + 2 \int_0^1 \int_U \left\{ \langle X_s^R(x) - X_s^R(y), g_R(X_{s-}^R(x), u) - g_R(X_{s-}^R(y), u) \rangle \right. \\
&\quad \left. + |g_R(X_{s-}^R(x), u) - g_R(X_{s-}^R(y), u)|^2 \right\} \tilde{N}(ds, du) \\
&\quad + \int_0^1 \int_U |g_R(X_{s-}^R(x), u) - g_R(X_{s-}^R(y), u)|^2 \nu(du) ds \\
&\leq 2C(R) \int_0^1 |X_s^R(x) - X_s^R(y)|^2 ds + M_t,
\end{aligned}$$

where we used the global one-sided Lipschitz condition (1.10) for b_R , σ_R and g_R and

$$\begin{aligned}
M_t &:= 2 \int_0^1 \int_U \left\{ \langle X_s^R(x) - X_s^R(y), g_R(X_{s-}^R(x), u) - g_R(X_{s-}^R(y), u) \rangle \right. \\
&\quad \left. + |g_R(X_{s-}^R(x), u) - g_R(X_{s-}^R(y), u)|^2 \right\} \tilde{N}(ds, du) \\
&\quad + 2 \int_0^1 \langle X_s^R(x) - X_s^R(y), (\sigma_R(X_s^R(x)) - \sigma_R(X_s^R(y))) dW_s \rangle
\end{aligned}$$

is a local martingale. Thus by a localization argument and the Gronwall inequality we get

$$(3.13) \quad \mathbb{E}|X_1^R(x) - X_1^R(y)|^2 \leq A|x - y|^2 e^{Bt}$$

for some constants $A, B > 0$ and thus

$$(3.14) \quad \mathbb{P}(|X_1(x) - X_1(y)| > \delta) \leq \varepsilon + \frac{A}{\delta^2}|x - y|^2 e^{Bt}.$$

Once we have (3.14), we proceed exactly like in the proof of Proposition 4.3 in [1]. Namely, we can show that for any sequence $x_n \rightarrow x$ in \mathbb{R}^d we have $X_1(x_n) \rightarrow X_1(x)$ in probability. From this we infer that for any function $f \in \mathcal{C}_b(\mathbb{R}^d)$ we have

$$p_1 f(x_n) \rightarrow p_1 f(x),$$

from which we get the desired Feller property of $(p_t)_{t \geq 0}$. Details of this last step can be found on page 321 in [1]. \square

Proof of Theorem 2.1. Note that (2.3) obviously implies (1.3), hence under the assumptions of Theorem 2.1, the assumptions of Lemma 2.5 are satisfied. Thus the semigroup $(p_t)_{t \geq 0}$ associated with the solution $(X_t)_{t \geq 0}$ of (1.1) is Feller. Hence, if we can show (2.2), then we can just use the Krylov-Bogoliubov method presented at the beginning of Section 2 and conclude the proof. In order to prove (2.2), we apply the Itô formula to $|X_t(x)|^2$

and then proceed like in the proof of Lemma 2.5 presented above, where we apply the Itô formula to obtain (3.13). However, unlike in (3.13), here we need to obtain the term e^{Bt} with a negative constant B in order to guarantee the boundedness in probability at infinity condition (2.1). Thus we need to use the differential version of the Gronwall inequality and not the integral one (cf. Remark 2.3 in [11]). This is however not a problem, since by using (2.3) and choosing a local martingale accordingly, we can obtain

$$\mathbb{E}|X_t(x)|^2 \leq \mathbb{E}|X_s(x)|^2 - 2K \int_s^t \left(\mathbb{E}|X_r(x)|^2 - \frac{M}{K} \right) dr$$

for any $0 \leq s \leq t$. Thus by the differential version of the Gronwall inequality we have

$$\mathbb{E}|X_t(x)|^2 - \frac{M}{K} \leq |x|^2 e^{-2Kt},$$

which gives (2.2) and finishes the proof. \square

Proof of Theorem 1.1. Under our assumptions we can prove that the coefficients of (1.1) are locally bounded in the sense of (1.8), cf. the discussion at the beginning of Section 3. Then combining Lemma 1.4 and Theorem 1.5, we see that for every $R \geq 1$ there exists a unique strong solution $(X_t^R)_{t \geq 0}$ to the equation (1.1) with the coefficients replaced by b_R , σ_R and g_R from Lemma 1.4. If we consider a sequence $\{(X_t^n)_{t \geq 0}\}_{n \in \mathbb{N}}$ of such solutions and define stopping times $\{\tau_n\}_{n \in \mathbb{N}}$ like in (3.12), then we can show that for $n \leq m$ we have $\tau_n \leq \tau_m$ and consequently that

$$\tau := \lim_{n \rightarrow \infty} \tau_n$$

and

$$X_t := \lim_{n \rightarrow \infty} X_t^n \text{ almost surely on } [0, \tau]$$

are well defined. We just need to show that $(X_t)_{t \geq 0}$ is indeed a solution of (1.1). However, by the construction of the coefficients in Lemma 1.4 we can see that

$$X_{t \wedge \tau_n} = X_{t \wedge \tau_n}^n = X_0 + \int_0^{t \wedge \tau_n} b(X_s) ds + \int_0^{t \wedge \tau_n} \sigma(X_s) dW_s + \int_0^{t \wedge \tau_n} \int_U g(X_{s-}, u) \tilde{N}(ds, du)$$

Therefore it remains to be shown that $\tau = \infty$ a.s., which can be done exactly as in Theorem 3.1 in [1]. Namely, using the global one-sided linear growth condition (1.3) we can show that $\mathbb{E}|X_{t \wedge \tau_n}|^2 \leq (\mathbb{E}|X_0|^2 + Kt)e^{Kt}$ for some constant $K > 0$ and then, after showing that $n^2 \mathbb{P}(\tau_n < t) \leq \mathbb{E}|X_{t \wedge \tau_n}|^2$, we see that $\tau_n \rightarrow \infty$ in probability and thus, via a subsequence, almost surely. \square

ACKNOWLEDGEMENT

I would like to thank Zdzisław Brzeźniak for discussions regarding the contents of [1]. This work was financially supported by the Bonn International Graduate School of Mathematics.

REFERENCES

- [1] S. Albeverio, Z. Brzeźniak, J.-L. Wu, Existence of global solutions and invariant measures for stochastic differential equations driven by Poisson type noise with non-Lipschitz coefficients, *J. Math. Anal. Appl.* 371 (2010), no. 1, 309-322.
- [2] D. Applebaum, *Lévy Processes and Stochastic Calculus*, second ed., Cambridge University Press, 2009.
- [3] K. Bichteler, J.-B. Gravereaux, J. Jacod, *Malliavin calculus for processes with jumps*, Stochastics Monographs, 2. Gordon and Breach Science Publishers, New York, 1987.

- [4] Z. Brzeźniak, W. Liu, J. Zhu, Strong solutions for SPDE with locally monotone coefficients driven by Lévy noise, *Nonlinear Anal. Real World Appl.* 17 (2014), 283-310.
- [5] G. Da Prato, D. Gątarek, J. Zabczyk, Invariant measures for semilinear stochastic equations, *Stochastic Anal. Appl.* 10 (1992), no. 4, 387-408.
- [6] I. Gyöngy, N. V. Krylov, On stochastic equations with respect to semimartingales. I., *Stochastics* 4 (1980/81), no. 1, 1-21.
- [7] R. Z. Has'minskii, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Alphen aan den Rijn, 1980.
- [8] N. Ikeda, S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, second ed., North-Holland, Kodansha, 1989.
- [9] N. V. Krylov, B. L. Rozovskii, Stochastic evolution equations, *Current problems in mathematics*, Vol. 14 (Russian), pp. 71-147, 256, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatzii, Moscow, 1979, <http://dx.doi.org/10.1007/BF01084893>.
- [10] Y. Ma, Transportation inequalities for stochastic differential equations with jumps, *Stochastic Process. Appl.* 120 (2010), no. 1, 2-21.
- [11] J. Shao, C. Yuan, Transportation-cost inequalities for diffusions with jumps and its application to regime-switching processes, *J. Math. Anal. Appl.* 425 (2015), no. 2, 632-654.
- [12] R. Situ, *Theory of Stochastic Differential Equations with Jumps and Applications*, Springer, New York, 2005.
- [13] J. Wang, L^p -Wasserstein distance for stochastic differential equations driven by Lévy processes, *Bernoulli* 22 (2016), no. 3, 1598-1616.
- [14] L. Wu, Transportation inequalities for stochastic differential equations of pure jumps, *Ann. Inst. Henri Poincaré Probab. Stat.* 46 (2010), no. 2, 465-479.

INSTITUTE FOR APPLIED MATHEMATICS, UNIVERSITY OF BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY

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

7 Diffusions with inhomogeneous jumps

In this chapter we present some ideas for a further extension of the methods used in the previous part of the thesis. Namely, we would like to study jump diffusions whose jumps are induced by a random measure with intensity which is inhomogeneous in time and space. Our aim is to apply the coupling technique to study stability of such processes in some special cases.

More precisely, we study stochastic differential equations of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t + \int_U v N_\lambda(dt, dv), \quad (7.0.1)$$

where $(W_t)_{t \geq 0}$ is a Brownian motion and N_λ is an integer-valued random measure on $\mathbb{R}_+ \times U$ (where (U, \mathcal{U}) is a measurable space), with a time-dependent and space-dependent compensator $\lambda(t, X_t, v)dt \nu(dv)$ with some Lévy measure ν on (U, \mathcal{U}) and a function $\lambda : [0, \infty) \times \mathbb{R}^d \times U \rightarrow \mathbb{R}$. Solutions of such SDEs can be considered as special cases of time-inhomogeneous processes with generators of the form

$$\mathcal{L}_t = L_t + K_t,$$

with L_t being a time-inhomogeneous diffusion generator and

$$K_t f(x) = \int (f(x+y) - f(x) - \langle y, \nabla f(x) \rangle \mathbf{1}_{\{|y| \leq 1\}}) \nu(t, x, dy),$$

where $\nu : [0, \infty) \times \mathbb{R}^d \times \mathcal{U} \rightarrow [0, \infty]$ is such that for every $(t, x) \in [0, \infty) \times \mathbb{R}^d$ the measure $\nu(t, x, \cdot)$ is a Lévy measure. Such processes were studied by Stroock in [Str75] using the theory of martingale problems, see Section 2.5. Here we will also focus on the approach via generators rather than via SDEs.

Processes of similar type appeared e.g. in [MBP04] in the context of non-linear filtering for jump processes. The authors of [MBP04] considered systems of equations

$$\begin{aligned} dX_t &= b(X_t)dt + \sigma(X_t)dB_t^X, \\ dY_t &= h(t, X_t)dt + dB_t^Y + \int_{\mathbb{R}} v N_\lambda(dt, dv) \end{aligned}$$

with two Brownian motions $(B_t^X)_{t \geq 0}$ and $(B_t^Y)_{t \geq 0}$ and an integer-valued random measure N_λ on $\mathbb{R}_+ \times \mathbb{R}$ with compensator

$$\lambda(t, X_t, v)dt \nu(dv)$$

for some finite Lévy measure ν and a real function λ . They solved a class of filtering problems for such processes, e.g., they were interested in evaluating expressions of the form

$$\mathbb{E}[f(X_t)|\mathcal{F}_t^Y],$$

where \mathcal{F}_t^Y is the σ -field generated by the process $(Y_t)_{t \geq 0}$ up to time $t > 0$ and f is a measurable function.

Here we are interested in processes of similar type, although our motivation is different and comes from the theory of continuous time Feynman-Kac models that appear in the theory of sequential Monte Carlo methods, see e.g. Chapter 5 of [DM13] for an introduction to this topic. Before we discuss connections with these kind of applications, let us formulate precisely our framework and state the result that we would like to prove. Afterwards, in Section 7.2 we will discuss how the kind of models that we consider can relate to the theory of sequential Monte Carlo methods and in Section 7.3 we will present a coupling construction and discuss how it can be applied to prove our stability result. It should be stressed that this part of the thesis is not intended to be fully rigorous and comprehensive. It is meant as an outline of possible future research directions and contains only partial results.

7.1 Formulation of the problem

Let $E \subset \mathbb{R}^d$ be a convex, bounded domain with a smooth boundary and let $R_0 := \text{diam}(E) < \infty$. Suppose we also have C_b^2 (twice continuously differentiable and bounded) functions $H : E \rightarrow \mathbb{R}$ and $V : E \rightarrow \mathbb{R}$.

We would like to consider an E -valued process $(X_t)_{t \in [0, T]}$ which is a solution to the martingale problem for time-inhomogeneous generators given as

$$\begin{aligned} L_{t, \nu_t} f(x) &= \langle -\nabla V(x) - t \nabla H(x), \nabla f(x) \rangle + \Delta f(x) \\ &+ \int_E (H(x) - H(z))_+ (f(z) - f(x)) \nu_t(dz), \end{aligned} \tag{7.1.1}$$

where $\nu_t = \text{Law}(X_t)$ for all $t \in [0, T]$ and $(H(x) - H(z))_+ = \max\{H(x) - H(z), 0\}$. Here functions $f : E \rightarrow \mathbb{R}$ are required to be in $C^2(\bar{E})$ and satisfy Neumann boundary conditions, i.e.,

$$\langle \mathbf{n}(x), \nabla f(x) \rangle = 0 \text{ for all } x \in \partial E,$$

where $\mathbf{n}(x)$ is the interior normal vector at a boundary point x . In other words, we require that for every such f the process

$$f(X_t) - \int_0^t L_{s, \nu_s} f(X_s) ds$$

is a martingale.

The motivation for studying processes of such form in the context of the theory of sequential Monte Carlo methods will be discussed in Section 7.2. However, they are also interesting on their own as examples of a different kind of jump diffusions than the ones

studied in the previous chapters. Even though the jump intensity in (7.1.1) is finite, it is also inhomogeneous in time and space, which poses new challenges in the analysis of such processes. The important thing to know at this point is that $(X_t)_{t \in [0, T]}$ is defined so as to preserve the family of measures $(\mu_t)_{t \in [0, T]}$ on E (in the sense that if $\text{Law}(X_0) = \mu_0$, then for each $t \in [0, T]$ we have $\text{Law}(X_t) = \mu_t$) given by

$$\mu_t(du) = \frac{1}{Z_t} \exp(-tH(u)) \mu_0(du), \quad (7.1.2)$$

where μ_0 is a fixed reference probability measure on E and the normalizing constants Z_t are given as

$$Z_t = \int_E \exp(-tH(u)) \mu_0(du) \quad (7.1.3)$$

for $t \in [0, T]$. As the reference measure μ_0 , we will consider a measure given by

$$\mu_0(du) = \frac{1}{Z} \exp(-V(u)) du \quad (7.1.4)$$

with

$$Z = \int_E \exp(-V(u)) du. \quad (7.1.5)$$

However, $(X_t)_{t \in [0, T]}$ is not a unique process such that $\text{Law}(X_t) = \mu_t$ holds for all $t \in [0, T]$ and the choice of the particular form of the jump intensity in (7.1.1) is motivated by its relation to the theory of sequential Monte Carlo methods, see the discussion in Section 7.2.

Observe that $(X_t)_{t \in [0, T]}$ is a diffusion with jumps, i.e., its generator can be written as

$$L_{t, \nu_t} = L_t + \widehat{L}_{t, \nu_t},$$

where L_t is a diffusion generator corresponding to the solution of the stochastic differential equation

$$dX'_t = (-\nabla V(X'_t) - t\nabla H(X'_t))dt + \sqrt{2}dB_t,$$

i.e.,

$$L_t f = \langle -\nabla V(x) - t\nabla H(x), \nabla f(x) \rangle + \Delta f(x) \quad (7.1.6)$$

and

$$\widehat{L}_{t, \nu_t} f(x) = \int_E (H(x) - H(z))_+ (f(z) - f(x)) \nu_t(dz).$$

Hence existence of such a process can be easily justified by a standard interlacing procedure, since a diffusion process corresponding to the generator (7.1.6) obviously exists due to regularity of the coefficients and the jumps can be added to its paths since their intensity is finite. More precisely, if we consider any probability measure ν_0 on E , we then want to construct an E -valued process $(Y_t)_{t \in [0, T]}$ with initial distribution ν_0 such that its generators are given by L_{t, ν_t} , defined as in (7.1.1), where $\nu_t = \text{Law}(Y_t)$ for $t \in [0, T]$. To see that such a construction is always possible, consider a diffusion process $(X'_t)_{t \geq 0}$ with

7 Diffusions with inhomogeneous jumps

the generator L_t defined by (7.1.6), started with initial distribution ν_0 , and a sequence $(e_n)_{n=0}^\infty$ of i.i.d. exponentially distributed random variables with unit parameter. Define

$$T_1 := \inf \left\{ t \geq 0 : \int_0^t \int_E (H(X'_s) - H(u))_+ \nu'_s(du) ds \geq e_0 \right\},$$

where $\nu'_s := \text{Law}(X'_s)$ for any $s \in [0, T]$. We set $\nu_t := \nu'_t$ for $t \in [0, T_1]$. At time T_1 , we make the process $(X'_t)_{t \geq 0}$ jump to a point chosen according to the distribution

$$\frac{(H(X'_{T_1-}) - H(z))_+ \nu_{T_1-}(dz)}{\int_E (H(X'_{T_1-}) - H(z))_+ \nu_{T_1-}(dz)}$$

(cf. the discussion in Section 5.3.2 in [DM13]). Now we can consider another copy $(X''_t)_{t \geq T_1}$ of the diffusion process described by the generator L_t , this time started at initial point X'_{T_1} . We define

$$T_2 := \inf \left\{ t \geq T_1 : \int_{T_1}^t \int_E (H(X''_s) - H(u))_+ \nu''_s(du) ds \geq e_1 \right\},$$

where $\nu''_s := \text{Law}(X''_s)$ for any $s \in [0, T]$. We set $\nu_t := \nu''_t$ for $t \in (T_1, T_2]$ and make the process $(X''_t)_{t \geq 0}$ jump to a point chosen according to

$$\frac{(H(X''_{T_2-}) - H(z))_+ \nu_{T_2-}(dz)}{\int_E (H(X''_{T_2-}) - H(z))_+ \nu_{T_2-}(dz)}.$$

We iterate our procedure for $n \geq 2$.

Having a fixed nonlinear flow of probability measures $(\mu_t)_{t \in [0, T]}$ given by (7.1.2), we would like to investigate its stability by comparing the processes $(X_t)_{t \in [0, T]}$ defined via (7.1.1) with initial condition μ_0 (so that for each $t \in [0, T]$ we have $\text{Law}(X_t) = \mu_t$) and a process $(Y_t)_{t \in [0, T]}$ defined also by (7.1.1) but with initial condition ν_0 , where $\nu_0 \neq \mu_0$. In other words, we would like to compare the initially given family of measures $(\mu_t)_{t \in [0, T]}$ with the family $(\nu_t)_{t \in [0, T]}$ constructed by applying the process defined by (7.1.1) with initial distribution ν_0 . The idea is that since $(X_t)_{t \in [0, T]}$ preserves $(\mu_t)_{t \in [0, T]}$, the processes $(X_t)_{t \in [0, T]}$ and $(Y_t)_{t \in [0, T]}$ should stay close to each other if we choose ν_0 close to μ_0 and if $(\mu_t)_{t \in [0, T]}$ is indeed stable under perturbations of its initial condition. Here we measure this stability in the standard L^1 -Wasserstein distance and hence we are interested in quantifying the expression $W_1(\mu_t, \nu_t)$ for $t \in [0, T]$.

Below we use the notation M_f and Lip_f to denote, respectively, the upper bound and the Lipschitz constant for any function $f : E \rightarrow \mathbb{R}^d$, i.e., $|f(x)| \leq M_f$ for any $x \in E$ and $|f(x) - f(y)| \leq \text{Lip}_f |x - y|$ for any $x, y \in E$. Recall that $R_0 = \text{diam}(E)$. We would like to prove the following result.

Theorem 7.1.1. *Let the potentials V and $H : E \rightarrow \mathbb{R}$ in (7.1.4) and (7.1.2) be C^2 , bounded, and such that*

$$\frac{2}{\text{Lip}_{\nabla V} + T \text{Lip}_{\nabla H}} \left(e^{\frac{1}{8}(\text{Lip}_{\nabla V} + T \text{Lip}_{\nabla H})R_0^2} - 1 \right) \leq \frac{e^{-\frac{1}{8}(\text{Lip}_{\nabla V} + T \text{Lip}_{\nabla H})R_0^2}}{20R_0 \text{Lip}_H + 4M_H}. \quad (7.1.7)$$

Then there exists a concave, strictly increasing continuous function $f : E \rightarrow \mathbb{R}_+$ and a constant $C > 0$ such that

$$W_f(\mu_t, \nu_t) \leq e^{-Ct} W_f(\mu_0, \nu_0) \quad (7.1.8)$$

for all $t \in [0, T]$. Moreover, the function f is constructed in such a way that

$$W_1(\mu_t, \nu_t) \leq 2e^{\frac{1}{8}(\text{Lip}_{\nabla V} + T \text{Lip}_{\nabla H})R_0^2} e^{-Ct} W_1(\mu_0, \nu_0) \quad (7.1.9)$$

for all $t \in [0, T]$.

Let us briefly discuss the form of the condition (7.1.7).

Example 7.1.2. We would like to give examples of functions V, H satisfying the condition (7.1.7), i.e.,

$$\frac{2}{K} \left(e^{\frac{1}{8}KR_0^2} - 1 \right) \leq \frac{e^{-\frac{1}{8}KR_0^2}}{20R_0 \text{Lip}_H + 4M_H}, \quad (7.1.10)$$

where

$$K = \text{Lip}_{\nabla V} + T \text{Lip}_{\nabla H}. \quad (7.1.11)$$

Note that if we fix sufficiently smooth functions V and H , then we can always rescale H in such a way that (7.1.10) holds. Namely, if we replace H by aH for some $a > 0$, then all the constants associated with H are multiplied by a and (7.1.10) becomes

$$\frac{2}{\text{Lip}_{\nabla V} + aT \text{Lip}_{\nabla H}} \left(e^{\frac{1}{8}(\text{Lip}_{\nabla V} + aT \text{Lip}_{\nabla H})R_0^2} - 1 \right) \leq \frac{e^{-\frac{1}{8}(\text{Lip}_{\nabla V} + aT \text{Lip}_{\nabla H})R_0^2}}{20aR_0 \text{Lip}_H + 4aM_H},$$

which is equivalent to

$$\frac{2a}{\text{Lip}_{\nabla V} + aT \text{Lip}_{\nabla H}} \left(e^{\frac{1}{8}(\text{Lip}_{\nabla V} + aT \text{Lip}_{\nabla H})R_0^2} - 1 \right) \leq \frac{e^{-\frac{1}{8}(\text{Lip}_{\nabla V} + aT \text{Lip}_{\nabla H})R_0^2}}{20R_0 \text{Lip}_H + 4M_H}.$$

When we take $a \rightarrow 0$, then the left hand side converges to 0 and the right hand side converges to

$$\frac{e^{-\frac{1}{8} \text{Lip}_{\nabla V} R_0^2}}{20R_0 \text{Lip}_H + 4M_H},$$

hence it is always possible to choose $a > 0$ small enough so that (7.1.10) holds.

The function f in Theorem 7.1.1 is defined as

$$f(r) := \int_0^r \varphi(s)g(s)ds,$$

where

$$\varphi(r) := \exp\left(-\frac{1}{8}Kr^2\right), \quad g(r) := 1 - \frac{c}{4} \int_0^r \frac{\Phi(s)}{\varphi(s)} ds$$

7 Diffusions with inhomogeneous jumps

with

$$c = 2 \left(\int_0^{R_0} \frac{\Phi(s)}{\varphi(s)} ds \right)^{-1}, \quad \Phi(r) := \int_0^r \varphi(s) ds = \int_0^r e^{(-\frac{1}{8}Ks^2)} ds,$$

$R_0 = \text{diam}(E)$ and K defined by (7.1.11).

The constant $C > 0$ appearing in (7.1.8) is given by

$$C := c - (10M_f \text{Lip}_H + 2M_H \text{Lip}_f) 2e^{\frac{1}{8}KR_0^2}.$$

We will present a fairly detailed outline of the proof of Theorem 7.1.1 in Section 7.3. Before we proceed, however, let us discuss briefly the connection between the process $(X_t)_{t \in [0, T]}$ defined via (7.1.1) and the theory of sequential Monte Carlo methods.

7.2 Nonlinear flows of probability measures

Let us consider once again the nonlinear flow of probability measures $(\mu_t)_{t \in [0, T]}$ on E given as

$$\mu_t(du) = \frac{1}{Z_t} \exp(-tH(u)) \mu_0(du) \quad (7.2.1)$$

with normalizing constants Z_t defined by (7.1.3) and

$$\mu_0(du) = \frac{1}{Z} \exp(-V(u)) du \quad (7.2.2)$$

with Z defined by (7.1.5).

Consider an unnormalized reference measure η_0 associated with μ_0 , i.e.,

$$\eta_0(du) = \exp(-V(u)) du.$$

We can also consider a family of unnormalized measures

$$\eta_t(du) = \exp(-tH(u)) \eta_0(du)$$

corresponding to $(\mu_t)_{t \in [0, T]}$ and the associated family of densities $(\rho_t)_{t \in [0, T]}$, i.e.,

$$\rho_t(x) := \exp(-tH(x)) \exp(-V(x))$$

for $x \in E$.

Suppose we are looking for a process $(X_t)_{t \in [0, T]}$ which preserves the family of measures $(\mu_t)_{t \in [0, T]}$. We will now sketch an argument justifying that defining $(X_t)_{t \in [0, T]}$ via (7.1.1) is indeed the right choice.

It is known that if $(X_t)_{t \in [0, T]}$ indeed preserves $(\mu_t)_{t \in [0, T]}$, then the densities $(\rho_t)_{t \in [0, T]}$ should satisfy the Fokker-Planck equation (see e.g. Chapter 4 in [Pav14])

$$\frac{d}{dt} \rho_t = \mathcal{L}_t^* \rho_t, \quad (7.2.3)$$

where $(\mathcal{L}_t)_{t \in [0, T]}$ is the family of generators of $(X_t)_{t \in [0, T]}$ and \mathcal{L}_t^* is the formal adjoint of \mathcal{L}_t , i.e., we should have

$$\langle \mathcal{L}_t f, \mu \rangle = \langle f, \mathcal{L}_t^* \mu \rangle.$$

Of course, since we are working on a bounded domain E , we should specify boundary conditions for \mathcal{L}_t . We choose the Neumann boundary conditions, i.e.,

$$\langle \nabla f(x), \mathbf{n}(x) \rangle = 0$$

for $x \in \partial E$, which corresponds to reflecting the process at the boundary of E .

Assume we are looking for an operator of the form

$$\mathcal{L}_t = L_t + K_t,$$

where L_t is a diffusion operator of the form

$$L_t f(x) = \sum_{i=1}^d b_i(t, x) \partial_i f(x) + \Delta f(x)$$

and K_t is a jump operator. This is motivated by the fact that a measure of the form (7.2.2) is an invariant measure for the Langevin equation. In some sense, we aim to use the jumps to make up for the evolution (7.2.1) that perturbs the measure μ_0 and to obtain this way a process that preserves the family $(\mu_t)_{t \in [0, T]}$. We have

$$\mathcal{L}_t^* = L_t^* + K_t^*$$

and

$$\begin{aligned} L_t^* f(x) &= \nabla \cdot (-b(t, x) f(x) + \nabla f(x)) \\ &= - \sum_{i=1}^d \partial_i b_i^i(x) f(x) - \sum_{i=1}^d b_i^i(x) \partial_i f(x) + \Delta f(x). \end{aligned}$$

Now observe that we have $\rho_0(x) = e^{-V(x)}$ and hence

$$\partial_i \rho_0(x) = -\partial_i V(x) \rho_0(x).$$

We also have $\rho_t(x) = e^{-tH(x)} \rho_0(x)$ and hence

$$\frac{d}{dt} \rho_t(x) = -H(x) \rho_t(x).$$

Now we calculate

$$\partial_i \rho_t(x) = -t e^{-tH(x)} \partial_i H(x) \rho_0(x) + e^{-tH(x)} \partial_i \rho_0(x) = -t \rho_t(x) \partial_i H(x) - \partial_i V(x) \rho_t(x)$$

and

$$\partial_{ii}^2 \rho_t(x) = -t \partial_i \rho_t(x) \partial_i H(x) - t \rho_t(x) \partial_{ii}^2 H(x) - \partial_{ii}^2 V(x) \rho_t(x) - \partial_i V(x) \partial_i \rho_t(x),$$

7 Diffusions with inhomogeneous jumps

which implies

$$\Delta\rho_t(x) = -t\langle\nabla\rho_t(x), \nabla H(x)\rangle - t\rho_t(x)\Delta H(x) - \rho_t(x)\Delta V(x) - \langle\nabla V(x), \nabla\rho_t(x)\rangle.$$

We infer that if we put

$$b_t(x) = -t\nabla H(x) - \nabla V(x),$$

then

$$L_t^*\rho_t(x) = 0$$

and hence we should have

$$K_t^*\rho_t(x) = -H(x)\rho_t(x).$$

This allows us to conjecture that if we set

$$L_t^H f(x) := L_t f(x) - H(x),$$

then the unnormalized measures $(\eta_t)_{t\in[0,T]}$ will satisfy the equation

$$\frac{d}{dt}\eta_t(f) = \eta_t(L_t^H(f)) \tag{7.2.4}$$

for a class of sufficiently regular functions f .

Note that the fact that the densities $(\rho_t)_{t\in[0,T]}$ satisfy (7.2.3) does not automatically imply (7.2.4) and our reasoning is just a heuristic argument allowing us to guess the right form of the generator of a process preserving our given family of measures. A fully rigorous argument justifying (7.2.4) will be given in an upcoming article based on the contents of this chapter.

However, if (7.2.4) indeed holds, then the normalized measures $(\mu_t)_{t\in[0,T]}$ satisfy

$$\frac{d}{dt}\mu_t(f) = \mu_t(L_t(f)) - \mu_t(Hf) + \mu_t(H)\mu_t(f), \tag{7.2.5}$$

see e.g. Section 5.2 in [DM13] and the references therein or Section 1.2 in [EM13]. Suppose now that the generator of $(X_t)_{t\in[0,T]}$ is given by

$$L_{t,\mu_t} = L_t + \widehat{L}_{t,\mu_t}. \tag{7.2.6}$$

Then our goal is to find a generator \widehat{L}_{t,μ_t} such that

$$-\mu_t(Hf) + \mu_t(H)\mu_t(f) = \mu_t(\widehat{L}_{t,\mu_t}(f)). \tag{7.2.7}$$

If (7.2.7) holds, then we can easily check that a process $(X_t)_{t\in[0,T]}$ with the generators $(L_{t,\mu_t})_{t\in[0,T]}$ defined by (7.2.6) indeed preserves the family of measures $(\mu_t)_{t\in[0,T]}$ due to (7.2.5). Namely, it is sufficient to check that for the associated time-inhomogeneous semigroup $(p_{s,t})_{0\leq s\leq t\leq T}$ we have

$$\frac{d}{ds} \int_E f d\mu_s p_{s,t} = 0$$

for sufficiently regular functions f , which then implies that for any $t \in [0, T]$ we have $\mu_0 p_{0,t} = \mu_t p_{t,t} = \mu_t$.

The choice of \widehat{L}_{t,μ_t} satisfying (7.2.7) is not unique, see Section 5.3 in [DM13], but one of the possible choices is

$$\widehat{L}_{t,\mu_t} f(x) := \int_E (H(x) - H(z))_+ (f(z) - f(x)) \mu_t(dz),$$

which is what we choose in (7.1.1). We will now try to briefly explain why this particular choice seems reasonable in the context of possible applications of the process $(X_t)_{t \in [0, T]}$ in the theory of sequential Monte Carlo methods.

Having determined the formula for the generators of the process $(X_t)_{t \in [0, T]}$, we can use it to describe the interacting particle system for which $(X_t)_{t \in [0, T]}$ is its mean field limit. Namely, given a system of a finite number of interacting particles, $(X_t)_{t \in [0, T]}$ is supposed to describe the behaviour of a single particle after the number of particles in the system goes to infinity. See [EM06] for more details about the connection between interacting particle systems and the mean field limit processes such as $(X_t)_{t \in [0, T]}$.

Let us now briefly discuss a particle system that would correspond to the process $(X_t)_{t \in [0, T]}$ defined via (7.1.1). Fix $N \in \mathbb{N}$. We can define a system of N interacting particles as a Markov process on E^N via its generator \mathcal{L}_t^N given as

$$\begin{aligned} \mathcal{L}_t^N f(x_1, \dots, x_N) &= \sum_{i=1}^N \langle -\nabla V(x_i) - t \nabla H(x_i), \nabla_i f(x_1, \dots, x_N) \rangle + \Delta_i f(x_1, \dots, x_N) \\ &\quad + \sum_{i=1}^N \int_E (H(x_i) - H(z))_+ (f(z) - f(x_i)) \eta^N(dz) \\ &= \sum_{i=1}^N \langle -\nabla V(x_i) - t \nabla H(x_i), \nabla_i f(x_1, \dots, x_N) \rangle + \Delta_i f(x_1, \dots, x_N) \\ &\quad + \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (H(x_i) - H(x_j))_+ (f(x_j) - f(x_i)), \end{aligned}$$

where $\eta^N(dz) = \sum_{j=1}^N \delta_{x_j}(dz)$. Hence we have a system of N particles in which between the jumps every particle moves according to the diffusion generator

$$L_t f = \langle -\nabla V(x) - t \nabla H(x), \nabla f(x) \rangle + \Delta f(x)$$

and when a particle jumps, it jumps to a position chosen among the positions of other particles in the system. Having a closer look at the jump intensities in this particle system may be instructive to understand the form of the jump intensity in (7.1.1). Consider as an example a bounded potential $H(x)$ shaped like $|x|^2$ in some neighbourhood of zero and recall that $\mu_t(du) = \frac{1}{Z_t} \exp(-tH(u)) \mu_0(du)$. Then it is clear that in this particle system the particles jump only to the regions of space chosen in such a way that the mass of μ_t increases (observe that when H decreases, μ_t increases).

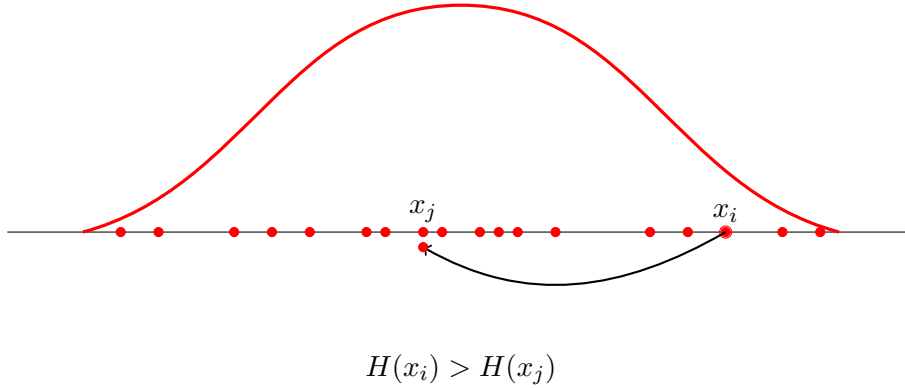


Figure 7.1: Jumps of particles in the interacting particle system associated with $(X_t)_{t \in [0, T]}$.

Interacting particle systems associated with processes of the type (7.1.1) appear in the theory of sequential Markov Chain Monte Carlo methods, where they are used to sample from a given family of measures $(\mu_t)_{t \in [0, T]}$. Thus the choice of the jump intensities presented above seems reasonable, as we would like the particles in our system to explore regions of space where the measure that we want to sample from has a lot of mass.

One possible application where we want to have a sequence of samples $(x_t)_{t \in [0, T]}$ from a given family of measures $(\mu_t)_{t \in [0, T]}$ is when we want to sample from a measure μ_T which is difficult to sample from. We deal with this problem by sampling from a measure μ_0 which is easy to sample from, and then we move from μ_0 to μ_T via a sequence of intermediate measures. Another area where such sequential methods are applied is Bayesian inference, where μ_t is the posterior distribution of some parameter given the data collected until time $t > 0$, see [DDJ06] and the references therein. See also [EM13], [DM13] or [DMM00] for more details on the connection between the interacting particle systems that we presented here and sequential Monte Carlo methods.

It is reasonable to conjecture that certain stability properties of such interacting particle systems should be preserved by their mean field limits. Therefore, we hope that our analysis of the process $(X_t)_{t \in [0, T]}$ will eventually lead us to obtain information on the behaviour of corresponding particle systems, which could subsequently be applied to study sequential Monte Carlo algorithms. This, however, will be the subject of a future work and we do not discuss these topics in any more detail in this thesis outside of the present section.

We now go back to studying the process $(X_t)_{t \in [0, T]}$.

7.3 Stability via coupling

We would like to couple the processes $(X_t)_{t \in [0, T]}$ and $(Y_t)_{t \in [0, T]}$ defined in Section 7.1 in order to obtain upper bounds on the quantity $W_1(\mu_t, \nu_t)$ from Theorem 7.1.1. Note that here we are interested in coupling two different processes and hence the situation is

different from what we discussed everywhere else in this thesis, as both in [Maj15] and [Maj16] we were interested in coupling two copies of the same process. In other words, we do not use Definition 3.1.4 directly, but we consider a coupling of two different random objects in the sense of Definition 3.1.2.

Hence, having two processes $(X_t)_{t \in [0, T]}$ and $(Y_t)_{t \in [0, T]}$ defined in terms of their generators L_{t, μ_t} and L_{t, ν_t} on E , we want to define a process on $E \times E$ with a generator $\bar{L}_{t, \mu_t, \nu_t}$ such that its marginal generators are L_{t, μ_t} and L_{t, ν_t} , respectively.

The coupling that we will use is given by

$$\begin{aligned}
 \bar{L}_{t, \mu_t, \nu_t} f(x, y) &= \langle -\nabla V(x) - t\nabla H(x), \nabla_x f(x, y) \rangle + \langle -\nabla V(y) - t\nabla H(y), \nabla_y f(x, y) \rangle \\
 &+ \Delta_x f(x, y) + \Delta_y f(x, y) \\
 &+ \sum_{i, j=1}^d \left(I - 2 \frac{x-y}{|x-y|} \frac{(x-y)^T}{|x-y|} \right)_{i, j} \frac{\partial^2}{\partial x_i \partial y_j} f(x, y) \\
 &+ \int_E (H(x) - H(z))_+ \wedge (H(y) - H(z))_+ (f(z, z) - f(x, y)) \mu_t(dz) \\
 &+ \int_E [(H(x) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(z, y) - f(x, y)) \mu_t(dz) \\
 &+ \int_E [(H(y) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(x, z) - f(x, y)) \nu_t(dz) \\
 &+ \int_E (H(x) \wedge H(y) - H(z))_+ (f(x, z) - f(x, y)) (\nu_t(dz) - \mu_t(dz))
 \end{aligned} \tag{7.3.1}$$

for functions $f : E \times E \rightarrow \mathbb{R}$ in $C^2(\overline{E \times E})$ with Neumann boundary conditions

$$\langle \mathbf{n}(x), \nabla_x f(x, y) \rangle = 0 \text{ for } x \in \partial E, y \in E \tag{7.3.2}$$

and

$$\langle \mathbf{n}(y), \nabla_y f(x, y) \rangle = 0 \text{ for } x \in E, y \in \partial E. \tag{7.3.3}$$

Existence of a solution to the martingale problem for (7.3.1) will be discussed in detail in a future paper based on the contents of this chapter. Here let us just mention that since the jump intensities in (7.3.1) are finite, we will want to apply an interlacing procedure similar to the one presented in Section 7.1.

While the formula (7.3.1) may look sophisticated, the rationale behind it is very simple. We apply the coupling by reflection to the Brownian motion driving (7.1.1) and we force the two coupled processes to jump to the same point with the maximal possible rate, which is the minimum of the jump rates of the two marginal processes (the first integral in (7.3.1)). The other integrals are just perturbation terms required in order for the generator (7.3.1) to actually define a coupling.

Let us briefly discuss why the process defined by (7.3.1) is indeed a coupling of $(X_t)_{t \in [0, T]}$ and $(Y_t)_{t \in [0, T]}$. It is easy to check that for $f(x, y) = g(x)$ we have

$$\bar{L}_{t, \mu_t, \nu_t} f(x, y) = L_{t, \mu_t} g(x)$$

7 Diffusions with inhomogeneous jumps

and for $f(x, y) = g(y)$ we have

$$\bar{L}_{t, \mu_t, \nu_t} f(x, y) = L_{t, \nu_t} g(y).$$

and hence the marginal generators are as we desired. However, to conclude we need to know that the martingale problem associated with L_{t, ν_t} has a unique solution. Let us briefly sketch an argument justifying this.

Recall from the discussion in Section 2.5 that in [Str75] Stroock investigated the theory of martingale problems associated with generators of the form

$$\mathcal{L}_t = L_t + K_t, \quad (7.3.4)$$

where

$$L_t = \frac{1}{2} \sum_{i,j=1}^d a^{ij}(t, x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b^i(t, x) \frac{\partial}{\partial x_i}$$

is a time-inhomogeneous diffusion generator on \mathbb{R}^d and

$$K_t f(x) = \int \left(f(x+y) - f(x) - \frac{\langle y, \nabla f(x) \rangle}{1+|y|^2} \right) \nu(t, x, dy) \quad (7.3.5)$$

is a time-inhomogeneous Lévy-type generator on \mathbb{R}^d . Here ν is such that for every (t, x) the measure $\nu(t, x, \cdot)$ is a Lévy measure. From Section 3 in [Str75] we can infer that if we have a generator \mathcal{L}_t of the type (7.3.4) for which uniqueness for the martingale problem holds and we consider a generator K'_t of the type (7.3.5) with a finite jump measure, then uniqueness holds also for the martingale problem for $\mathcal{L}_t + K'_t$. Thus in our case it is enough to verify uniqueness for the martingale problem for the generator defined by

$$L_t f(x) = \langle -\nabla V(x) - t \nabla H(x), \nabla f(x) \rangle + \Delta f(x).$$

However, due to Corollary 2.5 in [Kur11], we know that this is equivalent to uniqueness in law of weak solutions to the associated SDE, i.e.,

$$dX_t = (-\nabla V(X_t) - t \nabla H(X_t)) dt + \sqrt{2} dB_t. \quad (7.3.6)$$

Due to regularity of the coefficients of (7.3.6) its solution is indeed unique in law, which finishes our argument.

We are now ready to begin our analysis of the coupling given by (7.3.1). In order to prove Theorem 7.1.1 we will first focus on the diffusion part. Let us write

$$\bar{L}_{t, \mu_t, \nu_t} f(x, y) = \bar{L}_{t, \mu_t, \nu_t}^1 f(x, y) + \bar{L}_{t, \mu_t, \nu_t}^2 f(x, y),$$

where $\bar{L}_{t, \mu_t, \nu_t}^1$ is the diffusion part of $\bar{L}_{t, \mu_t, \nu_t}$ and $\bar{L}_{t, \mu_t, \nu_t}^2$ is its jump part, i.e., we have

$$\begin{aligned} \bar{L}_{t, \mu_t, \nu_t}^1 f(x, y) &= \langle -\nabla V(x) - t \nabla \Phi(x), \nabla_x f(x, y) \rangle + \langle -\nabla V(y) - t \nabla \Phi(y), \nabla_y f(x, y) \rangle \\ &\quad + \Delta_x f(x, y) + \Delta_y f(x, y) \\ &\quad + \sum_{i,j=1}^d \left(I - 2 \frac{x-y}{|x-y|} \frac{(x-y)^T}{|x-y|} \right)_{i,j} \frac{\partial^2}{\partial x_i \partial y_j} f(x, y) \end{aligned} \quad (7.3.7)$$

and

$$\begin{aligned}
 \bar{L}_{t,\mu_t,\nu_t}^2 f(x,y) &= \int_{\mathbb{R}^d} (H(x) - H(z))_+ \wedge (H(y) - H(z))_+ (f(z,z) - f(x,y)) \mu_t(dz) \\
 &+ \int_{\mathbb{R}^d} [(H(x) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(z,y) - f(x,y)) \mu_t(dz) \\
 &+ \int_{\mathbb{R}^d} [(H(y) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(x,z) - f(x,y)) \nu_t(dz) \\
 &+ \int_{\mathbb{R}^d} (H(x) \wedge H(y) - H(z))_+ (f(x,z) - f(x,y)) (\nu_t(dz) - \mu_t(dz)) .
 \end{aligned} \tag{7.3.8}$$

We will use the methods from [Ebe16] (presented also in Section 4 of [Maj16]) to find a concave function f and a constant $c > 0$ such that, formally,

$$\bar{L}_{t,\mu_t,\nu_t}^1 f(|x-y|) \leq -cf(|x-y|) \tag{7.3.9}$$

for all $x, y \in E$. Note that in fact we cannot directly apply $\bar{L}_{t,\mu_t,\nu_t}^1$ to the function $r \mapsto f(|r|)$ as it is not differentiable at zero. However, in the proof of Theorem 7.1.1 we only need to use (7.3.9) for x and $y \in E$ such that $|x-y| \geq \varepsilon > 0$, i.e., we need to have

$$\int_0^{t \wedge \tau_n} \bar{L}_{s,\mu_s,\nu_s}^1 f(|X_s - Y_s|) ds \leq -c \int_0^{t \wedge \tau_n} f(|X_s - Y_s|) ds, \tag{7.3.10}$$

where

$$\tau_n = \inf\{t > 0 : |X_t - Y_t| \notin (1/n, n)\}.$$

Note also that we formulated the martingale problem for \bar{L}_{t,μ_t,ν_t} for functions f satisfying the boundary conditions (7.3.2) and (7.3.3). Here we will apply $\bar{L}_{t,\mu_t,\nu_t}^1$ to a function f which does not necessarily satisfy (7.3.2) and (7.3.3), but it satisfies inequalities

$$\langle \mathbf{n}(x), \nabla_x f(x,y) \rangle \leq 0 \text{ for } x \in \partial E, y \in E$$

and

$$\langle \mathbf{n}(y), \nabla_y f(x,y) \rangle \leq 0 \text{ for } x \in E, y \in \partial E,$$

which will turn out to be sufficient for our purposes.

In order to find a function f for which (7.3.10) holds, let us first consider a process $(\bar{X}_t)_{t \in [0,T]}$ defined by

$$d\bar{X}_t = b(\bar{X}_t)dt + \mathbf{n}(\bar{X}_t)d\ell_t^{\bar{X}} + \sqrt{2}dB_t, \tag{7.3.11}$$

where $\mathbf{n}(x)$ is the interior normal vector for $x \in \partial E$ and $(\ell_t^{\bar{X}})_{t \in [0,T]}$ is the local time of $(\bar{X}_t)_{t \in [0,T]}$ on ∂E . In other words, $(\ell_t^{\bar{X}})_{t \in [0,T]}$ is a non-decreasing continuous process which increases only when $(\bar{X}_t)_{t \in [0,T]}$ is at the boundary, i.e., we have

$$\int_0^\infty \mathbf{1}_{E \setminus \partial E}(\bar{X}_t) d\ell_t^{\bar{X}} = 0 \text{ a.s.}$$

7 Diffusions with inhomogeneous jumps

Such an SDE corresponds to the generator

$$L_t f(x) = \langle b(x), \nabla f(x) \rangle + \Delta f(x)$$

with Neumann boundary conditions

$$\langle \mathbf{n}(x), \nabla f(x) \rangle = 0 \text{ for all } x \in \partial E,$$

see e.g. Section 1.3 in [And09], see also [LS84] or [BCJ06] for more information on diffusion processes with reflection at the boundary of their state space.

We will now prove that for a coupling by reflection $(\bar{X}_t, \bar{Y}_t)_{t \in [0, T]}$ of two copies of the solution to (7.3.11) we have a function f and a constant $c > 0$ such that

$$\mathbb{E}f(|\bar{X}_{t \wedge \tau_n} - \bar{Y}_{t \wedge \tau_n}|) - \mathbb{E}f(|\bar{X}_0 - \bar{Y}_0|) \leq -c \int_0^{t \wedge \tau_n} \mathbb{E}f(|\bar{X}_s - \bar{Y}_s|) ds \quad (7.3.12)$$

with $\tau_n := \inf\{t > 0 : |\bar{X}_t - \bar{Y}_t| \notin (1/n, n)\}$. This will be done by using the methods from [Ebe16], see also the proof of Theorem 2.3 in [Maj16]. Note that showing (7.3.12) with some function f will allow us to infer that (7.3.10) holds with the same function f if we choose the drift in (7.3.11) to be the same as the one in the process defined by (7.1.1), i.e., our coupling $(\bar{X}_t, \bar{Y}_t)_{t \in [0, T]}$ will correspond to the coupling generator $\bar{L}_{t, \mu_t, \nu_t}^1$ defined by (7.3.7).

Following Remark 3 in [Ebe16], the methods from [Ebe16] which we presented in Section 4 in [Maj16] and which were originally applied to diffusions on \mathbb{R}^d of the form $dX_t = b(X_t)dt + dB_t$, are easily adapted to the case of (7.3.11). Namely, the coupling $(\bar{X}_t, \bar{Y}_t)_{t \in [0, T]}$ is specified by choosing $(\bar{Y}_t)_{t \in [0, T]}$ as

$$d\bar{Y}_t = b(\bar{Y}_t)dt + \mathbf{n}(\bar{Y}_t)d\ell_t^{\bar{Y}} + \sqrt{2}R(\bar{X}_t, \bar{Y}_t)dB_t,$$

where $(\ell_t^{\bar{Y}})_{t \in [0, T]}$ is the local time of $(\bar{Y}_t)_{t \in [0, T]}$ on ∂E and $R(\bar{X}_t, \bar{Y}_t)$ is the usual reflection operator that we introduced in Section 3.1.1. Then, by the Itô formula, we have

$$\begin{aligned} df(|Z_t|) &= f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, b(\bar{X}_t) - b(\bar{Y}_t) \rangle dt + f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, \mathbf{n}(\bar{X}_t) \rangle d\ell_t^{\bar{X}} \\ &\quad - f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, \mathbf{n}(\bar{Y}_t) \rangle d\ell_t^{\bar{Y}} + 2\sqrt{2}f'(|Z_t|)d\widetilde{W}_t + 4f''(|Z_t|)dt, \end{aligned}$$

where $Z_t := \bar{X}_t - \bar{Y}_t$ and $(\widetilde{W}_t)_{t \geq 0}$ is a one-dimensional Brownian motion, cf. Section 4 of [Ebe16] or Section 4 of [Maj16]. Now observe that due to convexity of E we have

$$\langle Z_t, \mathbf{n}(\bar{X}_t) \rangle \leq 0$$

and

$$-\langle Z_t, \mathbf{n}(\bar{Y}_t) \rangle \leq 0,$$

which implies that

$$df(|Z_t|) \leq f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, b(\bar{X}_t) - b(\bar{Y}_t) \rangle dt + 2\sqrt{2}f'(|Z_t|)d\widetilde{W}_t + 4f''(|Z_t|)dt.$$

This shows that having the additional term involving the local time in (7.3.11) does not change our estimates and we can proceed exactly as in [Ebe16] or [Maj16], i.e., we need to construct a function f such that

$$4f''(r) - r\kappa(r)f'(r) \leq -cf(r)$$

holds for all $r \geq 0$ with some constant $c > 0$, where κ is a function such that

$$\langle b(x) - b(y), x - y \rangle \leq -\kappa(|x - y|)|x - y|^2. \quad (7.3.13)$$

We can easily deduce a right formula for f in Theorem 7.1.1 by applying some minor modifications to the original formulas from [Ebe16]. As we will see in the sequel, the fact that in our case the drift is time-inhomogeneous does not actually pose a problem since the time horizon is finite. Here we base our presentation on the one from Section 4 in [Maj16]. Note that the case that we are considering here corresponds to taking $\sigma_1 = \sqrt{2}$ and $\alpha = 1/2$ in the formulas in Section 4 of [Maj16]. Recall that the drift function in our process is given by

$$b_t(x) = -\nabla V(x) - t\nabla H(x).$$

Thus we have

$$\begin{aligned} \langle b_t(x) - b_t(y), x - y \rangle &= \langle -\nabla V(x) - t\nabla H(x) + \nabla V(y) + t\nabla H(y), x - y \rangle \\ &\leq |\nabla V(x) - \nabla V(y)||x - y| + T|\nabla H(x) - \nabla H(y)||x - y| \\ &\leq \text{Lip}_{\nabla V} |x - y|^2 + T \text{Lip}_{\nabla H} |x - y|^2. \end{aligned}$$

Hence we see that our setting corresponds to (7.3.13) with the function κ from (7.3.13) being constant and defined as

$$\kappa := -\text{Lip}_{\nabla V} - T \text{Lip}_{\nabla H}.$$

For convenience we denote

$$K := \text{Lip}_{\nabla V} + T \text{Lip}_{\nabla H}.$$

Hence we would like to find a function f such that

$$4f''(r) + rf'(r)K \leq -cf(r)$$

or, equivalently,

$$2f''(r) + \frac{1}{2}rf'(r)K \leq -\frac{c}{2}f(r) \quad (7.3.14)$$

for all $r \geq 0$ with some constant $c > 0$.

We will now guess the formula for such f based on the formulas from [Ebe16]. Afterwards we will show that our f indeed satisfies (7.3.14). Recall that the function $f(r)$ in [Ebe16] is defined for all $r \in [0, \infty)$ and its behaviour is different on intervals $[0, R_0]$, $(R_0, R_1]$ and (R_1, ∞) for some appropriately chosen constants $0 < R_0 < R_1$. Observe that since κ is always negative, in the formula which defines R_0 (see (4.9) in [Maj16] or

7 Diffusions with inhomogeneous jumps

(8) in [Ebe16]) the infimum is taken over an empty set. However, since we are dealing here with a bounded domain E , we can set

$$R_0 = R_1 = \text{diam}(E)$$

and we need to specify the behaviour of the function f only for $r \in [0, \text{diam}(E)]$. This, following the formulas in Section 4 of [Maj16], leads us to define

$$f(r) := \int_0^r \varphi(s)g(s)ds, \quad (7.3.15)$$

where

$$\varphi(r) := \exp\left(-\frac{1}{4}\int_0^r Ksds\right) = \exp\left(-\frac{1}{8}Kr^2\right), \quad g(r) := 1 - \frac{c}{4}\int_0^r \frac{\Phi(s)}{\varphi(s)}ds$$

with

$$c = 2\left(\int_0^{R_0} \frac{\Phi(s)}{\varphi(s)}ds\right)^{-1} \quad (7.3.16)$$

and

$$\Phi(r) := \int_0^r \varphi(s)ds = \int_0^r \exp\left(-\frac{1}{8}Ks^2\right)ds.$$

We can now easily check that f defined this way satisfies (7.3.14) for all $r \in [0, R_0]$. Moreover, it is comparable from above and below with a (rescaled) identity function, which will later allow us to compare the Wasserstein distance W_f with the (rescaled) standard L^1 -Wasserstein distance W_1 .

We obviously have

$$f'(r) = \varphi(r)g(r)$$

and

$$f''(r) = \varphi'(r)g(r) + \varphi(r)g'(r).$$

Hence (7.3.14) can be written as

$$2\varphi'(r)g(r) + 2\varphi(r)g'(r) + \frac{1}{2}Kr\varphi(r)g(r) \leq -\frac{c}{2}f(r). \quad (7.3.17)$$

Note that

$$g'(r) = -\frac{c}{4}\frac{\Phi(r)}{\varphi(r)}$$

and

$$\varphi'(r) = -\frac{1}{4}Kr\varphi(r).$$

Thus the left hand side of (7.3.17) becomes just

$$-\frac{c}{2}\Phi(r).$$

However, since $g(r) \leq 1$ for all $r \in [0, R_0]$, we obviously have $f(r) \leq \Phi(r)$ for all $r \in [0, R_0]$. Hence we see that (7.3.14) indeed holds.

Moreover, for all $r \in [0, R_0]$ we have $\varphi(r) \leq 1$ and thus $f(r) \leq r$.

As for the lower bound on f , since we have

$$\int_0^{R_0} \frac{\Phi(s)}{\varphi(s)} ds \geq \int_0^r \frac{\Phi(s)}{\varphi(s)} ds$$

for any $r \in [0, R_0]$, we see that

$$1 \geq \frac{c}{2} \int_0^r \frac{\Phi(s)}{\varphi(s)} ds$$

and hence

$$g(r) = 1 - \frac{c}{4} \int_0^r \frac{\Phi(s)}{\varphi(s)} ds \geq \frac{1}{2}.$$

Moreover, the function φ is decreasing and hence

$$\varphi(r) \geq \varphi(R_0) = \exp\left(-\frac{1}{8}KR_0^2\right)$$

for all $r \in [0, R_0]$. Therefore

$$f(r) = \int_0^r \varphi(s)g(s)ds \geq \int_0^r \frac{1}{2} \exp\left(-\frac{1}{8}KR_0^2\right) ds = \frac{1}{2}r \exp\left(-\frac{1}{8}KR_0^2\right),$$

which implies

$$r \leq 2 \exp\left(\frac{1}{8}KR_0^2\right) f(r) \tag{7.3.18}$$

for all $r \in [0, R_0]$. Thus we see that f has all the desired properties.

Hence we have proved (7.3.12) and thus (7.3.10) with the function f defined by (7.3.15) and the constant $c > 0$ defined by (7.3.16). Moreover, we have (7.3.18) and

$$f(r) \leq r \tag{7.3.19}$$

for all $r \in [0, R_0]$.

Let us now deal with the jump part (7.3.8) of our coupling (7.3.1). We will apply $\bar{L}_{t, \mu_t, \nu_t}^2$ to the same function f with which (7.3.10) holds.

We have

$$\begin{aligned} & \bar{L}_{t, \mu_t, \nu_t}^2 f(|x - y|) \\ &= \int_E (H(x) - H(z))_+ \wedge (H(y) - H(z))_+ (f(|z - z|) - f(|x - y|)) \mu_t(dz) \\ &+ \int_E [(H(x) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ &+ \int_E [(H(y) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(|x - z|) - f(|x - y|)) \nu_t(dz) \\ &+ \int_E (H(x) \wedge H(y) - H(z))_+ (f(|x - z|) - f(|x - y|)) (\nu_t(dz) - \mu_t(dz)). \end{aligned} \tag{7.3.20}$$

7 Diffusions with inhomogeneous jumps

Obviously, for the first term in (7.3.20) we have

$$\int_E (H(x) - H(z))_+ \wedge (H(y) - H(z))_+ (f(0) - f(|x - y|)) \mu_t(dz) \leq 0,$$

since $f(0) = 0$.

For the last term in (7.3.20), we use the Kantorovich-Rubinstein duality (see Remark 6.5 in [Vil09]), i.e., we have

$$\int g(z) (\nu_t(dz) - \mu_t(dz)) \leq W_1(\nu_t, \mu_t) \text{Lip}_g,$$

for any Lipschitz function g with Lipschitz constant Lip_g . Note that if H is bounded by M_H and is Lipschitz with a constant Lip_H and the distance function f is bounded by M_f and is Lipschitz with a constant Lip_f , then the function

$$z \mapsto (H(x) \wedge H(y) - H(z))_+ (f(|x - z|) - f(|x - y|))$$

is Lipschitz with the constant $2M_f \text{Lip}_H + 2M_H \text{Lip}_f$. Indeed, for any Lipschitz functions g, h we have

$$\begin{aligned} (g \wedge h)(y) &\leq h(y) \leq h(x) + |h(y) - h(x)| \\ &\leq h(x) + \text{Lip}_h |y - x| \\ &\leq h(x) + \max(\text{Lip}_h, \text{Lip}_g) |y - x|. \end{aligned}$$

Analogously,

$$(g \wedge h)(y) \leq g(x) + \max(\text{Lip}_h, \text{Lip}_g) |y - x|$$

and hence

$$(g \wedge h)(y) \leq (g \wedge h)(x) + \max(\text{Lip}_h, \text{Lip}_g) |y - x|.$$

Thus, by symmetry,

$$|(g \wedge h)(y) - (g \wedge h)(x)| \leq \max(\text{Lip}_h, \text{Lip}_g) |y - x|.$$

Hence the function

$$z \mapsto (H(x) - H(z))_+ \wedge (H(y) - H(z))_+ = (H(x) \wedge H(y) - H(z))_+$$

is Lipschitz with Lip_H . It is also bounded by $2M_H$.

On the other hand,

$$|f(|x - z_2|) - f(|x - z_1|)| \leq \text{Lip}_f \left| |x - z_2| - |x - z_1| \right| \leq \text{Lip}_f \left| |x - z_2 - x + z_1| \right|$$

and hence the function

$$z \mapsto f(|x - z|) - f(|x - y|)$$

is Lipschitz with Lip_f . It is also bounded by $2M_f$.

Moreover, it is easy to check that for any two bounded Lipschitz functions g and h we have

$$|(gh)(y) - (gh)(x)| \leq M_g \text{Lip}_h |y - x| + M_h \text{Lip}_g |y - x|.$$

Thus we see that $z \mapsto (H(x) \wedge H(y) - H(z))_+ (f(|x - z|) - f(|x - y|))$ is indeed Lipschitz with the constant $2M_f \text{Lip}_H + 2M_H \text{Lip}_f$.

From the considerations above, we obtain

$$\begin{aligned} \int_E (H(x) \wedge H(y) - H(z))_+ (f(|x - z|) - f(|x - y|)) (\nu_t(dz) - \mu_t(dz)) \\ \leq W_1(\mu_t, \nu_t) (2M_f \text{Lip}_H + 2M_H \text{Lip}_f). \end{aligned}$$

Now we will deal with the second and the third term in (7.3.20). By symmetry, the bound that we will get for the third term will be the same as for the second term, as we do not use here the specific form of the measures μ_t and ν_t . We split the domain of integration by writing

$$\begin{aligned} \int_E [(H(x) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ = \int_{H(y) \leq H(x)} [(H(x) - H(z))_+ - (H(y) - H(z))_+] (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ = \int_{H(y) \leq H(z) \leq H(x)} (H(x) - H(z)) (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ + \int_{H(z) \leq H(y) \leq H(x)} (H(x) - H(y)) (f(|z - y|) - f(|x - y|)) \mu_t(dz) =: I_1 + I_2. \end{aligned}$$

We have

$$I_2 \leq 2M_f |H(x) - H(y)| \leq 2M_f \text{Lip}_H |x - y|$$

and

$$\begin{aligned} I_1 &= \int_{H(y) \leq H(z) \leq H(x), f(|z-y|) < f(|x-y|)} (H(x) - H(z)) (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ &+ \int_{H(y) \leq H(z) \leq H(x), f(|z-y|) \geq f(|x-y|)} (H(x) - H(y) + H(y) - H(z)) \\ &\times (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ &=: I_1' + I_1''. \end{aligned}$$

We easily notice that $I_1' \leq 0$, while

$$\begin{aligned} I_1'' &= \int_{H(y) \leq H(z) \leq H(x), f(|z-y|) \geq f(|x-y|)} (H(x) - H(y)) (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ &+ \int_{H(y) \leq H(z) \leq H(x), f(|z-y|) \geq f(|x-y|)} (H(y) - H(z)) (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ &=: (I_1'')^* + (I_1'')^{**}. \end{aligned}$$

7 Diffusions with inhomogeneous jumps

It is easy to see that $(I_1'')^{**} \leq 0$, whereas

$$(I_1'')^* \leq 2M_f \text{Lip}_H |x - y|.$$

Combining the estimates above, we get an upper bound for the second term in (7.3.20), i.e., we have

$$\begin{aligned} & \int_E [(H(x) - H(z))_+ - (H(x) \wedge H(y) - H(z))_+] (f(|z - y|) - f(|x - y|)) \mu_t(dz) \\ & \leq 4M_f \text{Lip}_H |x - y|. \end{aligned}$$

Finally, we get an upper bound for all the terms in (7.3.20) added together, which is

$$8M_f \text{Lip}_H |x - y| + (2M_f \text{Lip}_H + 2M_H \text{Lip}_f) W_1(\mu_t, \nu_t).$$

We are now ready to combine all the estimates for the diffusion part and for the jump part.

Proof of Theorem 7.1.1. We use a sequence of stopping times $(\tau_n)_{n=1}^\infty$ in order to remove the martingale terms from our calculations, cf. the proof of Theorem 1 in [Ebe16] or the proof of Theorem 1.2 in [JWa16]. Here

$$\tau_n = \inf\{t > 0 : |X_t - Y_t| \notin (1/n, n)\}.$$

We have

$$\mathbb{E}f(|X_{t \wedge \tau_n} - Y_{t \wedge \tau_n}|) = \mathbb{E}f(|X_0 - Y_0|) + \mathbb{E} \int_0^{t \wedge \tau_n} \bar{L}_{s, \mu_s, \nu_s} f(|X_s - Y_s|) ds. \quad (7.3.21)$$

However, from our calculations we know that

$$\begin{aligned} \int_0^{t \wedge \tau_n} \bar{L}_{s, \mu_s, \nu_s} f(|X_s - Y_s|) ds & \leq \int_0^{t \wedge \tau_n} \left(-cf(|X_s - Y_s|) + 8M_f \text{Lip}_H |X_s - Y_s| \right. \\ & \quad \left. + (2M_f \text{Lip}_H + 2M_H \text{Lip}_f) W_1(\mu_s, \nu_s) \right) ds. \end{aligned} \quad (7.3.22)$$

Since the state space E is compact, the integrand on the right hand side is uniformly bounded, which allows us to pass to the limit with $n \rightarrow \infty$ in (7.3.21) by using the Fatou lemma on the left hand side and the dominated convergence theorem on the right hand side. We obtain

$$\begin{aligned} \mathbb{E}f(|X_t - Y_t|) & \leq \mathbb{E}f(|X_0 - Y_0|) + \mathbb{E} \int_0^t \left(-cf(|X_s - Y_s|) + 8M_f \text{Lip}_H |X_s - Y_s| \right. \\ & \quad \left. + (2M_f \text{Lip}_H + 2M_H \text{Lip}_f) W_1(\mu_s, \nu_s) \right) ds. \end{aligned}$$

We want to obtain an upper bound for the right hand side of (7.3.22). We will use the fact that

$$W_1(\mu_t, \nu_t) \leq \mathbb{E}|X_t - Y_t|.$$

We have

$$\begin{aligned} & \mathbb{E} \int_0^t (-cf(|X_s - Y_s|) + 8M_f \text{Lip}_H |X_s - Y_s| + (2M_f \text{Lip}_H + 2M_H \text{Lip}_f) W_1(\mu_s, \nu_s)) ds \\ & \leq \mathbb{E} \int_0^t \left(-cf(|X_s - Y_s|) + 8M_f \text{Lip}_H |X_s - Y_s| \right. \\ & \quad \left. + (2M_f \text{Lip}_H + 2M_H \text{Lip}_f) \mathbb{E}|X_s - Y_s| \right) ds \\ & = \int_0^t \mathbb{E} (-cf(|X_s - Y_s|) + (10M_f \text{Lip}_H + 2M_H \text{Lip}_f) |X_s - Y_s|) ds \end{aligned}$$

Now observe that due to (7.3.18) we have

$$\begin{aligned} & (10M_f \text{Lip}_H + 2M_H \text{Lip}_f) |X_s - Y_s| \\ & \leq (10M_f \text{Lip}_H + 2M_H \text{Lip}_f) f(|X_s - Y_s|) 2e^{\frac{1}{8}KR_0^2}. \end{aligned}$$

Combining all our estimates together we obtain

$$\begin{aligned} & \mathbb{E}f(|X_t - Y_t|) \leq \mathbb{E}f(|X_0 - Y_0|) \\ & \quad - \left(c - (10M_f \text{Lip}_H + 2M_H \text{Lip}_f) 2e^{\frac{1}{8}KR_0^2} \right) \mathbb{E} \int_0^t f(|X_s - Y_s|) ds. \end{aligned} \tag{7.3.23}$$

We can easily see that (7.3.23) holds not only on the time interval $[0, t]$, but on any interval $[s, t] \subset [0, T]$ and hence by the differential version of the Gronwall inequality we obtain

$$\mathbb{E}f(|X_t - Y_t|) \leq \mathbb{E}f(|X_0 - Y_0|) \exp \left(- \left(c - (10M_f \text{Lip}_H + 2M_H \text{Lip}_f) 2e^{\frac{1}{8}KR_0^2} \right) t \right).$$

This implies that for any $t \in [0, T]$ we have

$$W_f(\mu_t, \nu_t) \leq \exp \left(- \left(c - (10M_f \text{Lip}_H + 2M_H \text{Lip}_f) 2e^{\frac{1}{8}KR_0^2} \right) t \right) W_f(\mu_0, \nu_0).$$

We want the constant in the exponent on the right hand side to be positive and hence we need to check that

$$c > 2e^{\frac{1}{8}KR_0^2} (10M_f \text{Lip}_H + 2M_H \text{Lip}_f).$$

First let us simplify the expression above. Observe that for $x > y$ we have

$$|f(x) - f(y)| = \left| \int_y^x \varphi(r)g(r)dr \right| \leq \int_y^x dr = x - y \leq |x - y|$$

7 Diffusions with inhomogeneous jumps

and hence

$$\text{Lip}_f \leq 1.$$

Moreover, $M_f = f(R_0)$, since f is increasing on $[0, R_0]$. Furthermore, $f(R_0) \leq R_0$ (see (7.3.19)). Hence it is sufficient if we have

$$c > 2e^{\frac{1}{8}KR_0^2}(10R_0 \text{Lip}_H + 2M_H).$$

Observe that

$$\begin{aligned} c^{-1} &= \frac{1}{2} \int_0^{R_0} \Phi(s) \varphi(s)^{-1} ds = \frac{1}{2} \int_0^{R_0} e^{\frac{1}{8}Ks^2} \left(\int_0^s e^{-\frac{1}{8}Kt^2} dt \right) ds \\ &\leq \frac{1}{2} \int_0^{R_0} s e^{\frac{1}{8}Ks^2} ds \\ &= \frac{2}{K} \left(e^{\frac{1}{8}KR_0^2} - 1 \right) \end{aligned}$$

and thus we need to have

$$\frac{2}{K} \left(e^{\frac{1}{8}KR_0^2} - 1 \right) \leq \frac{e^{-\frac{1}{8}KR_0^2}}{20R_0 \text{Lip}_H + 4M_H}.$$

This gives us (7.1.7). Moreover, due to (7.3.19) and (7.3.18), directly from (7.1.8) we can infer (7.1.9), which finishes the proof. □

Bibliography

- [ABW10] S. Albeverio, Z. Brzeźniak, J.-L. Wu, Existence of global solutions and invariant measures for stochastic differential equations driven by Poisson type noise with non-Lipschitz coefficients, *J. Math. Anal. Appl.* 371 (2010), no. 1, 309-322.
- [And09] S. Andres, *Diffusion Processes with Reflection*, PhD thesis, Berlin, 2009.
- [App09] D. Applebaum, *Lévy Processes and Stochastic Calculus*, second ed., Cambridge University Press, 2009.
- [AppErr] D. Applebaum, *Lévy Processes and Stochastic Calculus*, online errata, <http://applebaum.staff.shef.ac.uk/Levcorr21.pdf>.
- [BE85] D. Bakry, M. Émery, Diffusions hypercontractives, Séminaire de probabilités, XIX, 1983/84, 177-206, *Lecture Notes in Math.*, 1123, Springer, Berlin, 1985.
- [BLG15] M. Barczy, Z. Li, G. Pap, Yamada-Watanabe results for stochastic differential equations with jumps, *Int. J. Stoch. Anal.* 2015, Art. ID 460472, 23 pp.
- [BLG15b] M. Barczy, Z. Li, G. Pap, Stochastic differential equation with jumps for multi-type continuous state and continuous time branching processes with immigration, *ALEA Lat. Am. J. Probab. Math. Stat.* 12 (2015), no. 1, 129-169.
- [BC86] R. F. Bass, M. Cranston, The Malliavin calculus for pure jump processes and applications to local time, *Ann. Probab.* 14 (1986), no. 2, 490-532.
- [BGJ87] K. Bichteler, J.-B. Gravereaux, J. Jacod, *Malliavin calculus for processes with jumps*, *Stochastics Monographs*, 2. Gordon and Breach Science Publishers, New York, 1987.
- [Bis83] J.-M. Bismut, Calcul des variations stochastique et processus de sauts, *Z. Wahrsch. Verw. Gebiete* 63 (1983), no. 2, 147-235.
- [BG99] S. G. Bobkov, F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, *J. Funct. Anal.* 163 (1999), no. 1, 1-28.
- [BSW11] B. Böttcher, R.L. Schilling, J. Wang, Construction of coupling processes for Lévy processes, *Stochastic Process. Appl.* 121 (2011), no. 6, 1201-1216.
- [Bre68] L. Breiman, *Probability*, Addison-Wesley Publishing Company, Reading, MA, 1968.

Bibliography

- [BLZ14] Z. Brzeźniak, W. Liu, J. Zhu, Strong solutions for SPDE with locally monotone coefficients driven by Lévy noise, *Nonlinear Anal. Real World Appl.* 17 (2014), 283-310.
- [BK00] K. Burdzy, W. S. Kendall, Efficient Markovian couplings: examples and counterexamples, *Ann. Appl. Probab.* 10 (2000), no. 2, 362-409.
- [BCJ06] K. Burdzy, Z.-Q. Chen, P. Jones, Synchronous couplings of reflected Brownian motions in smooth domains, *Illinois J. Math.* 50 (2006), no. 1-4, 189-268.
- [CT04] R. Cont, P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.
- [DPZ14] G. Da Prato, J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, second ed., Cambridge University Press, 2014.
- [DMM00] P. Del Moral, L. Miclo, Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering, *Séminaire de Probabilités, XXXIV*, 1-145, *Lecture Notes in Math.*, 1729, Springer, Berlin, 2000.
- [DDJ06] P. Del Moral, A. Doucet, A. Jasra, Sequential Monte Carlo samplers, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68 (2006), no. 3, 411-436.
- [DM13] P. Del Moral, *Mean Field Simulation for Monte Carlo Integration*, CRC Press, 2013.
- [DØP09] G. Di Nunno, B. Øksendal, F. Proske, *Malliavin Calculus for Lévy Processes with Applications to Finance*, Universitext. Springer-Verlag, Berlin, 2009.
- [DGW04] H. Djellout, A. Guillin, L. Wu, Transportation cost-information inequalities and applications to random dynamical systems and diffusions, *Ann. Probab.* 32 (2004), no. 3B, 2702-2732.
- [DD76] C. Doléans-Dade, On the existence and unicity of solutions of stochastic integral equations, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 36 (1976), no. 2, 93-101.
- [Dör15] L. Döring, A jump-type SDE approach to real-valued self-similar Markov processes, *Trans. Amer. Math. Soc.* 367 (2015), no. 11, 7797-7836.
- [EM06] A. Eberle, C. Marinelli, Convergence of sequential Markov Chain Monte Carlo methods: I. Nonlinear flow of probability measures, preprint, arXiv:math/0612074.
- [Ebe11] A. Eberle, Reflection coupling and Wasserstein contractivity without convexity, *C. R. Math. Acad. Sci. Paris* 349 (2011), no. 19-20, 1101-1104.
- [EM13] A. Eberle, C. Marinelli, Quantitative approximations of evolving probability measures and sequential Markov chain Monte Carlo methods, *Probab. Theory Related Fields* 155 (2013), 665-701.

- [Ebe16] A. Eberle, Reflection couplings and contraction rates for diffusions, *Probab. Theory Related Fields* 166 (2016), no. 3-4, 851-886.
- [EGZ16] A. Eberle, A. Guillin, R. Zimmer, Quantitative Harris type theorems for diffusions and McKean-Vlasov processes, preprint, arXiv:1606.06012.
- [EK86] S. N. Ethier, T. G. Kurtz, *Markov Processes. Characterization and Convergence.*, Wiley, New York, 1986.
- [GL07] N. Gozlan, C. Léonard, A large deviation approach to some transportation cost inequalities, *Probab. Theory Related Fields* 139 (2007), no. 1-2, 235-283.
- [GL10] N. Gozlan, C. Léonard, Transport inequalities. A survey, *Markov Process. Related Fields* 16 (2010), no. 4, 635-736.
- [GM11] B. Grigelionis, R. Mikulevicius, Nonlinear filtering equations for stochastic processes with jumps, *The Oxford Handbook of Nonlinear Filtering*, 95-128, Oxford Univ. Press, Oxford, 2011.
- [GK80] I. Gyöngy, N. V. Krylov, On stochastic equations with respect to semimartingales. I., *Stochastics* 4 (1980/81), no. 1, 1-21.
- [HS13] E.P. Hsu, K.T. Sturm, Maximal coupling of Euclidean Brownian motions, *Commun. Math. Stat.* 1 (2013), no. 1, 93-104.
- [IW89] N. Ikeda, S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, second ed., North-Holland, Kodansha, 1989.
- [Ish13] Y. Ishikawa, *Stochastic Calculus of Variations for Jump Processes*, De Gruyter, Berlin, 2013.
- [Itô51] K. Itô, On stochastic differential equations, *Mem. Amer. Math. Soc.* No. 4 (1951), 51 pp.
- [JMS14] S. D. Jacka, A. Mijatović, D. Širaj, Mirror and synchronous couplings of geometric Brownian motions, *Stochastic Process. Appl.* 124 (2014), no. 2, 1055-1069.
- [Jac79] J. Jacod, *Calcul stochastique et problèmes de martingales*, *Lecture Notes in Mathematics* 714, Springer, Berlin, 1979.
- [KW12] T. Komorowski, A. Walczuk, Central limit theorem for Markov processes with spectral gap in the Wasserstein metric, *Stochastic Process. Appl.* 122 (2012), no. 5, 2155-2184.
- [Kul09] A.M. Kulik, Exponential ergodicity of the solutions to SDE's with a jump noise, *Stochastic Process. Appl.* 119 (2009), no. 2, 602-632.
- [Kuo06] H.-H. Kuo, *Introduction to Stochastic Integration*, Springer, New York, 2006.

Bibliography

- [Kur07] T. G. Kurtz, The Yamada-Watanabe-Engelbert theorem for general stochastic equations and inequalities, *Electron. J. Probab.* 12 (2007), 951-965.
- [Kur11] T. G. Kurtz, Equivalence of stochastic equations and martingale problems, *Stochastic analysis 2010*, 113-130, Springer, Heidelberg, 2011.
- [Küh17] F. Kühn, On martingale problems and Feller processes, preprint, arXiv:1706.04132.
- [LP11] G. Last, M. Penrose, Poisson process Fock space representation, chaos expansion and covariance inequalities, *Probab. Theory Related Fields* 150 (2011), no. 3-4, 663-690.
- [Lin92] T. Lindvall, *Lectures on the Coupling Method*, Wiley, New York, 1992.
- [LR86] T. Lindvall, L.C.G. Rogers, Coupling of multidimensional diffusions by reflection, *Ann. Probab.* 14 (1986), no. 3, 860-872.
- [LW12] H. Lin, J. Wang, Successful couplings for a class of stochastic differential equations driven by Lévy processes, *Sci. China Math.* 55 (2012), no. 8, 1735-1748.
- [LS84] P.-L. Lions, A.-S. Sznitman, Stochastic differential equations with reflecting boundary conditions, *Comm. Pure Appl. Math.* 37 (1984), no. 4, 511-537.
- [Løk04] A. Løkka, Martingale representation of functionals of Lévy processes, *Stochastic Anal. Appl.* 22 (2004), no. 4, 867-892.
- [LW16] D. Luo, J. Wang, Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises, preprint, arXiv:1604.07206.
- [Ma10] Y. Ma, Transportation inequalities for stochastic differential equations with jumps, *Stochastic Process. Appl.* 120 (2010), no. 1, 2-21.
- [Maj15] M. B. Majka, Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes, 2015, preprint, arXiv:1509.08816.
- [Maj16] M. B. Majka, Transportation inequalities for non-globally dissipative SDEs with jumps via Malliavin calculus and coupling, 2016, preprint, arXiv:1610.06916.
- [Maj16b] M. B. Majka, A note on existence of global solutions and invariant measures for jump SDEs with locally one-sided Lipschitz drift, 2016, preprint, arXiv:1612.03824.
- [McC99] R. McCann, Exact solutions to the transportation problem on the line, *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* 455 (1999), no. 1984, 1341-1380.
- [MBP04] T. Meyer-Brandis, F. Proske, Explicit solution of a non-linear filtering problem for Lévy processes with application to finance, *Appl. Math. Optim.* 50 (2004), no. 2, 119-134.

- [Nor88] J. R. Norris, Integration by parts for jump processes. Séminaire de probabilités, XXII, 271-315, Lecture Notes in Math., 1321, Springer, Berlin, 1988.
- [NV90] D. Nualart, J. Vives, Anticipative calculus for the Poisson process based on the Fock space, Séminaire de probabilités, XXIV, 1988/89, 154-165, Lecture Notes in Math., 1426, Springer, Berlin, 1990.
- [OV00] F. Otto, C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, *J. Funct. Anal.* 173 (2000), no. 2, 361-400.
- [PP15] S. Palau, J. C. Pardo, Branching processes in a Lévy random environment, preprint, arXiv:1512.07691.
- [Pav14] G. A. Pavliotis, Stochastic processes and applications. Diffusion processes, the Fokker-Planck and Langevin equations, Springer, New York, 2014.
- [PZ07] S. Peszat, J. Zabczyk, Stochastic Partial Differential Equations with Lévy Noise, Cambridge University Press, 2007.
- [Pic96] J. Picard, Formules de dualité sur l'espace de Poisson, *Ann. Inst. H. Poincaré Probab. Statist.* 32 (1996), no. 4, 509-548.
- [Pic96b] J. Picard, On the existence of smooth densities for jump processes, *Probab. Theory Related Fields* 105 (1996), no. 4, 481-511.
- [PW06] E. Priola, F. Y. Wang, Gradient estimates for diffusion semigroups with singular coefficients, *J. Funct. Anal.* 236 (2006), no. 1, 244-264.
- [PSXZ12] E. Priola, A. Shirikyan, L. Xu, J. Zabczyk, Exponential ergodicity and regularity for equations with Lévy noise, *Stochastic Process. Appl.* 122 (2012), no. 1, 106-133.
- [Pro05] P. Protter, Stochastic Integration and Differential Equations, Second edition. Version 2.1., Springer-Verlag, Berlin, 2005.
- [vRS05] M.-K. von Renesse, K.-T. Sturm, Transport inequalities, gradient estimates, entropy, and Ricci curvature, *Comm. Pure Appl. Math.* 58 (2005), no. 7, 923-940.
- [Sat99] K. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, 1999.
- [SW11] R.L. Schilling, J. Wang, On the coupling property of Lévy processes, *Ann. Inst. Henri Poincaré Probab. Stat.* 47 (2011), no. 4, 1147-1159.
- [SSW12] R.L. Schilling, P. Sztonyk, J. Wang, Coupling property and gradient estimates of Lévy processes via the symbol, *Bernoulli* 18 (2012), no. 4, 1128-1149.
- [SW12] R.L. Schilling, J. Wang, On the coupling property and the Liouville theorem for Ornstein-Uhlenbeck processes, *J. Evol. Equ.* 12 (2012), no. 1, 119-140.

Bibliography

- [Sit05] R. Situ, Theory of Stochastic Differential Equations with Jumps and Applications, Springer, New York, 2005.
- [Son15] Y. Song, Gradient estimates and coupling property for semilinear SDEs driven by jump processes, *Sci. China Math.* 58 (2015), no. 2, 447-458.
- [Str75] D. W. Stroock, Diffusion processes associated with Lévy generators, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 32 (1975), no. 3, 209-244.
- [SV79] D. W. Stroock, S. R. S. Varadhan, Multidimensional Diffusion Processes, Springer-Verlag, Berlin-New York, 1979.
- [Tho00] H. Thorisson, Coupling, Stationarity, and Regeneration, Springer-Verlag, New York, 2000.
- [Vil09] C. Villani, Optimal Transport. Old and New., Springer-Verlag, Berlin, 2009.
- [Wan11] F.Y. Wang, Coupling for Ornstein-Uhlenbeck processes with jumps, *Bernoulli* 17 (2011), no. 4, 1136-1158.
- [JWa16] J. Wang, L^p -Wasserstein distance for stochastic differential equations driven by Lévy processes, *Bernoulli* 22 (2016), no. 3, 1598-1616.
- [Wu10] L. Wu, Transportation inequalities for stochastic differential equations of pure jumps, *Ann. Inst. Henri Poincaré Probab. Stat.* 46 (2010), no. 2, 465-479.
- [Xu14] L. Xu, Exponential mixing of 2D SDEs forced by degenerate Lévy noises, *J. Evol. Equ.* 14 (2014), no. 2, 249-272.