TRANSPORTATION INEQUALITIES FOR NON-GLOBALLY DISSIPATIVE SDES WITH JUMPS VIA MALLIAVIN CALCULUS AND COUPLING

MATEUSZ B. MAJKA

Accepted for publication in Annales de l'Institut Henri Poincaré Probabilités et Statistiques

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

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\widetilde{N}(dt, du)\,, \tag{1.1}$$

where $(W_t)_{t\geq 0}$ is an m-dimensional Brownian motion and $\widetilde{N}(dt, du) = N(dt, du) - dt\,\nu(du)$ is a compensated Poisson random measure on $\mathbb{R}_+ \times U$, where $(U, \mathcal{U}, \nu)$ is a σ-finite measure space. Moreover, the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ and $g : \mathbb{R}^d \times U \to \mathbb{R}^d$ 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 $\kappa : \mathbb{R}_+ \to \mathbb{R}$ such that for all x, $y \in \mathbb{R}^d$ we have

$$\langle b(x) - b(y), x - y\rangle + \frac{1}{2}\int_U |g(x, u) - g(y, u)|^2 \nu(du) + \|\sigma(x) - \sigma(y)\|_{HS}^2 \leq -\kappa(|x - y|)|x - y|^2\,, \tag{1.2}$$

where $\|\sigma\|_{HS} = \sqrt{\operatorname{tr} \sigma\sigma^T}$ 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)

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

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)

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

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

$$\alpha(W_{1,\rho}(\eta, \mu)) \leq H(\eta|\mu)\,. \tag{1.3}$$

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 $f : E \to \mathbb{R}$ bounded and Lipschitz with $\|f\|_{\mathrm{Lip}} \leq 1$ we have

$$\int e^{\lambda(f-\mu(f))} d\mu \leq e^{\alpha^*(\lambda)} \text{ for any } \lambda > 0\,, \tag{1.4}$$

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 $f : E \to \mathbb{R}$ bounded and Lipschitz with $\|f\|_{\mathrm{Lip}} \leq 1$ we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) - \mu(f) > r\right) \leq e^{-n\alpha(r)} \text{ for any } r > 0, n \geq 1\,. \tag{1.5}$$

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

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t\,, \tag{1.6}$$

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

$$dX_t = b(X_t)dt + \int_U g(X_{t-}, u)\widetilde{N}(dt, du)\,, \tag{1.7}$$

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 $\widetilde{N}$.

Remark 1.1. Assumption E is quite restrictive. In particular, let us consider the case where $U \subset \mathbb{R}^d$ and $g(x, u) = \widetilde{g}(x)u$ for some $\mathbb{R}^{d\times d}$-valued function $\widetilde{g}$ and hence the equation (1.7) is driven by a d-dimensional Lévy process $(L_t)_{t\geq 0}$ (i.e., we have $dX_t = b(X_t)dt + \widetilde{g}(X_{t-})dL_t$). 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

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

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

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

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

$$D_{t,u}f(N) := f(N + \delta_{(t,u)}) - f(N)\,. \tag{1.8}$$

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

$$\mathbb{E}\int_0^T |\nabla_t F|^2 dt + \mathbb{E}\int_0^T \int_U |D_{t,u}F|^2 \nu(du)dt < \infty\,, \tag{1.9}$$

we have

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

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 (2.7) 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

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-},u)\widetilde{N}(dt,du) \tag{2.1}$$

is to be able to control perturbations of $(X_t)_{t\geq 0}$ via a coupling, with respect to changes in initial conditions (see (2.3) below) as well as changes of the drift (see (2.4)). 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 $(\widetilde{X}_t)_{t\geq 0}$ a solution to

$$d\widetilde{X}_t = b(\widetilde{X}_t)dt + \sigma(\widetilde{X}_t)h_t dt + \sigma(\widetilde{X}_t)dW_t + \int_U g(\widetilde{X}_{t-},u)\widetilde{N}(dt,du)\,. \tag{2.2}$$

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 $\|\sigma(x)\| \le \sigma_\infty$, where $\|\cdot\|$ is the operator norm, and there exists a measurable function $g_\infty : U \to \mathbb{R}$ such that $|g(x,u)| \le g_\infty(u)$ 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 $(X_t, Y_t)_{t\ge 0}$ of solutions to (2.1) and a function $c_1 : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $0 \le s \le t$ we have*

$$\mathbb{E}[|X_t - Y_t|/\mathcal{F}_s] \le c_1(t-s)|X_s - Y_s|\,. \tag{2.3}$$

Furthermore, assume that there exists a coupling $(X_t, Y_t')_{t\ge 0}$ of solutions to (2.1) and functions c_2, $c_3 : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $0 \le s \le t$ we have

$$\mathbb{E}[|\widetilde{X}_t - Y_t'|/\mathcal{F}_s] \le c_2(t-s)\mathbb{E}\int_s^t c_3(r)|\sigma(\widetilde{X}_r)h_r|dr\,. \tag{2.4}$$

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*

$$\alpha_T(W_1(\eta, \delta_x p_T)) \le H(\eta|\delta_x p_T) \tag{2.5}$$

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

$$\alpha_T(r) := \sup_{\lambda\ge 0}\left\{ r\lambda - \int_0^T \beta(c_1(T-t)\lambda)dt - \frac{\sigma_\infty^2 c_2^2(T)\lambda^2}{2}\int_0^T c_3^2(t)dt \right\}.$$

(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 cádlág $\mathbb{R}^d$-valued functions on $[0,T]$ satisfies*

$$\alpha_T^P(W_{1,d_{L^1}}(Q, \mathbb{P}_{x,[0,T]})) \le H(Q|\mathbb{P}_{x,[0,T]}) \tag{2.6}$$

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

$$\alpha_T^P(r) := \sup_{\lambda\ge 0}\left\{ r\lambda - \int_0^T \beta\left(\lambda\int_t^T c_1(s-t)ds\right)dt - \frac{\sigma_\infty^2\lambda^2}{2}\int_0^T c_3^2(t)\left(\int_t^T c_2(r)dr\right)^2 dt \right\}.$$

Note that in (2.4) the process $(Y_t')_{t\ge 0}$ is coupled with $(X_t)_{t\ge 0}$, but the estimated distance is between $(Y_t')_{t\ge 0}$ and $(\widetilde{X}_t)_{t\ge 0}$ given by (2.2). In other words, we need to consider a process $(Y_t')_{t\ge 0}$ whose law is determined by the dynamics (2.1), but it behaves in such a way that we can control its distance to a process following the modified dynamics given by (2.2). An example of such a construction can be found in the proof of Theorem 2.8.

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 $f : \mathbb{R}^d \to \mathbb{R}$, we have*

$$\begin{aligned}&\mathbb{E}\phi\Big(f(X_T(x)) - p_T f(x)\Big)\\ &\quad\le \mathbb{E}\phi\left(\|f\|_{\mathrm{Lip}}\left(\int_0^T\int_U c_1(T-t)g_\infty(u)\widetilde{N}(dt,du) + c_2(T)\int_0^T c_3(t)j(t)dW_t\right)\right),\end{aligned} \tag{2.7}$$

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 $F : \mathbb{D}([0,T];\mathbb{R}^d) \to \mathbb{R}$ we have

$$\begin{aligned}
&\mathbb{E}\phi\Big(F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x))\Big) \\
&\leq \mathbb{E}\phi\Bigg(\|F\|_{\mathrm{Lip}}\Bigg(\int_0^T\int_U\left(\int_t^T c_1(r-t)dr\right)g_\infty(u)\widetilde{N}(dt,du) \\
&+ \int_0^T c_3(t)\left(\int_t^T c_2(r)dr\right)j(t)dW_t\Bigg)\Bigg).
\end{aligned} \tag{2.8}$$

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 (2.3) and (2.4), 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 (2.3) and (2.4) 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

$$\inf_{x\in\mathbb{R}^d: 0<|x|\leq\delta}\int_{\{|v|\leq m\}\cap\{|v+x|\leq m\}} q(v)\wedge q(v+x)dv > 0\,. \tag{2.9}$$

Assumption L4. (Positive mass in a neighbourhood of zero) There exists a constant $\varepsilon > 0$ such that $\varepsilon \leq \delta$ (with δ defined via (2.9) 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 $\kappa : \mathbb{R}_+ \to \mathbb{R}$ 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

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq Le^{-\theta t}f(|x-y|)\,. \tag{2.10}$$

However, the function f used in [27] is discontinuous. It is actually of the form

$$f = a\mathbf{1}_{(0,\infty)} + f_1 \tag{2.11}$$

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 $a_1 x \leq f_1(x) \leq a_2 x$ for some a_1, $a_2 > 0$). Hence we obtain

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq \widetilde{L}e^{-\theta t}(|x-y|+1)\,, \tag{2.12}$$

for some $\widetilde{L} > 0$, 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 (2.10) but with a continuous function f (i.e., with $a = 0$ in (2.11)). To this end, we define

$$C_\varepsilon := 2\int_0^{\varepsilon/4} |y|^2 \nu_1^L(dy)\,, \tag{2.13}$$

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 $\varepsilon \to 0$.

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 (2.10) 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*

$$dX_t = b(X_t)dt + \sigma_1 dB_t^1 + \sigma(X_t)dB_t^2 + dL_t + \int_U g(X_{t-},u)\widetilde{N}(dt,du)\,, \tag{2.14}$$

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 $\widetilde{N}$ 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 $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $g : \mathbb{R}^d \times U \to \mathbb{R}^d$ 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 (2.14) and constants $\widetilde{C}$, $\widetilde{c} > 0$ such that for any x, $y \in \mathbb{R}^d$ and any $t > 0$ we have

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq \widetilde{C}e^{-\widetilde{c}t}|x-y|\,. \tag{2.15}$$

Remark 2.4. The reason for the particular form of the equation (2.14) is that in order to construct a coupling leading to the inequality (2.15) 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 $\widetilde{C}$ and $\widetilde{c}$ depend on which noise we use. In particular, the constant $\widetilde{c}$ 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 (2.15) 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 (2.15) actually implies that for any $0 \leq s \leq t$ we have

$$\mathbb{E}[|X_t - Y_t|/\mathcal{F}_s] \leq \widetilde{C}e^{-\widetilde{c}(t-s)}|X_s - Y_s|\,.$$

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 (2.12). It is still possible to obtain some transportation inequalities if in Theorem 2.1 we replace the condition (2.3) with a condition like (2.12), 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 (2.14), 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,*

$$W_1(\eta p_t, \mu p_t) \leq \widetilde{C}e^{-\widetilde{c}t}W_1(\eta, \mu)$$

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

$$W_1(\eta p_t, \mu_0) \leq \widetilde{C}e^{-\widetilde{c}t}W_1(\eta, \mu_0)$$

for any probability measure η on $\mathbb{R}^d$ and any $t > 0$.

This result follows immediately from (2.15) 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 (2.14) by a change in the drift. This gives us a tool to determine some concrete cases in which the assumption (2.4) 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 $(\widetilde{X}_t)_{t\geq 0}$ which is a solution to (2.14) with the drift perturbed by u_t, i.e.,*

$$d\widetilde{X}_t = b(\widetilde{X}_t)dt + u_t dt + \sigma_1 dB^1_t + \sigma(\widetilde{X}_t)dB^2_t + dL_t + \int_U g(\widetilde{X}_{t-}, u)\widetilde{N}(dt, du)\,,$$

where u_t is either $\sigma_1 h_t$ or $\sigma(\widetilde{X}_t)h_t$ 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 (2.14) and for any $0 \leq s \leq t$ we have

$$\mathbb{E}[|\widetilde{X}_t - Y_t|/\mathcal{F}_s] \leq C\int_s^t e^{c(r-(t-s))}|u_r|dr\,, \tag{2.16}$$

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 $g_\infty : U \to \mathbb{R}$ 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 $\widetilde{\lambda} > 0$ such that*

$$\beta^L(\widetilde{\lambda}) := \int_U (e^{\widetilde{\lambda}u} - \widetilde{\lambda}u - 1)\nu^L(du) < \infty\,.$$

Then the transportation inequality (2.5) from the statement of Theorem 2.1 holds with

$$\begin{aligned}\alpha_T(r) := \sup_{\lambda\geq 0}\Bigg\{ r\lambda &- \int_0^T \beta(\widetilde{C}e^{-\widetilde{c}(T-t)}\lambda)dt - \int_0^T \beta^L(\widetilde{C}e^{-\widetilde{c}(T-t)}\lambda)dt \\ &- \frac{(\sigma_\infty^2 + \|\sigma_1\|^2)\lambda^2}{2}C^2\frac{1-e^{-2cT}}{2c}\Bigg\}\,.\end{aligned}$$

Moreover, for the invariant measure μ_0 we have

$$\alpha_\infty(W_1(\eta, \mu_0)) \leq H(\eta|\mu_0) \tag{2.17}$$

for any probability measure η on $\mathbb{R}^d$, with α_∞ defined as the pointwise limit of α_T as $T \to \infty$. Finally, the inequality (2.6) holds with

$$\begin{aligned}\alpha^P_T(r) := \sup_{\lambda\geq 0}\Bigg\{ r\lambda &- \int_0^T \beta\left(\lambda\widetilde{C}\frac{1-e^{-\widetilde{c}(T-t)}}{\widetilde{c}}\right)dt - \int_0^T \beta^L\left(\lambda\widetilde{C}\frac{1-e^{-\widetilde{c}(T-t)}}{\widetilde{c}}\right)dt \\ &- \frac{(\sigma_\infty^2 + \|\sigma_1\|^2)\lambda^2}{2}C^2\int_0^T\left(\frac{1-e^{-c(T-t)}}{c}\right)^2 dt\Bigg\}\,.\end{aligned}$$

The constants $\widetilde{c}$, $\widetilde{C}$, c and C appearing in the definitions of α_T and α^P_T are the same as in (2.15) and (2.16).

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 (2.14) 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 (2.14). 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

$$\mathbb{E}\exp\left(i\langle z, L_t\rangle\right) = \exp\left(-t\left[(m^{1/\beta} + |z|^2)^{\beta} - m\right]\right)$$

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 $T \to \infty$ (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 (2.3) and (2.4) by other methods. For example, in a recent paper [23], D. Luo and J. Wang obtained an inequality like (2.3) 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 (2.3) 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 (2.3) 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 (2.3) 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 (2.7) for solutions of equations of the type $dX_t = b(X_t)dt + dB_t + \int_U g(X_{t-}, u)\widetilde{N}(dt, du)$, 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 (2.3) 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 (2.4) 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 $f : \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_{\mathrm{Lip}} \leq 1$, for any adapted, $\mathbb{R}_+$-valued process g and for any $0 \leq s \leq r \leq t$ we have*

$$\mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla_u f(X_t)|\mathcal{F}_u]|^2 du \leq c_2^2(t)\sigma_\infty \mathbb{E} \int_s^r g_u c_3^2(u) du \,. \tag{2.18}$$

Moreover, we have

$$\|\mathbb{E}[\nabla_u f(X_t)|\mathcal{F}_u]\|_{L^\infty(\Omega\times[0,t])} \leq c_2(t)\sigma_\infty \sup_{u\leq t} c_3(u) \,, \tag{2.19}$$

where the L^∞ norm is the essential supremum on $\Omega \times [0, t]$.

On the other hand, using the condition (2.3), 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 (2.14), satisfying the assumptions of Theorem 2.3 with $\det \sigma_1 > 0$ and $\lim_{r\to 0} r\kappa(r) = 0$ (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 $(B^i_t)_{t\geq 0}$ 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*

$$\mathbb{E} \int_s^r g_u |\mathbb{E}[\nabla^1_u f(X_t)|\mathcal{F}_u]|^2 du \leq C^2 \|\sigma_1\|^2 \mathbb{E} \int_s^r g_u e^{2c(u-t)} du \tag{2.20}$$

and

$$\mathbb{E}\int_s^r g_u|\mathbb{E}[\nabla_u^2 f(X_t)|\mathcal{F}_u]|^2 du \le C^2\sigma_\infty^2 \mathbb{E}\int_s^r g_u e^{2c(u-t)}du\,, \tag{2.21}$$

where C and c are the same as in (2.16). We also have L^∞ bounds analogous to (2.19) 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 $(L_t)_{t\ge0}$ 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 (2.20) and (2.21). Namely, for the multiplicative noise we get

$$\mathbb{E}[\|\nabla_s X_t\|_{HS}^2|\mathcal{F}_s] \le \|\sigma(X_s)\|_{HS}^2 e^{2K(s-t)}$$

for any $0 \le s \le t$, 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 (2.20) and (2.21) 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 (2.4) 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

$$dX_t = b(X_t)dt + dL_t\,, \tag{3.1}$$

where $(L_t)_{t\ge0}$ 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 $(L_t)_{t\ge0}$ via

$$L_t = \int_0^t\int_{\{|v|>1\}} vN(ds,dv) + \int_0^t\int_{\{|v|\le1\}} v\widetilde{N}(ds,dv)$$

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

$$dX_t = b(X_t)dt + \int_{\{|v|>1\}\times[0,1]} vN(dt,dv,du) + \int_{\{|v|\le1\}\times[0,1]} v\widetilde{N}(dt,dv,du)\,.$$

Without loss of generality, we can choose a constant $m > 1$ and rewrite the equation above as

$$(3.2) \qquad dX_t = b(X_t)dt + \int_{\{|v|>m\}\times[0,1]} vN(dt,dv,du) + \int_{\{|v|\le m\}\times[0,1]} v\widetilde{N}(dt,dv,du)\,.$$

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 $(X_t, Y_t)_{t\ge 0}$ by putting

$$(3.3) \qquad \begin{aligned} dY_t = b(Y_t)dt &+ \int_{\{|v|>m\}\times[0,1]} vN(dt,dv,du) \\ &+ \int_{\{|v|\le m\}\times[0,1]} (X_{t-} - Y_{t-} + v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}}\widetilde{N}(dt,dv,du) \\ &+ \int_{\{|v|\le m\}\times[0,1]} R(X_{t-},Y_{t-})v\mathbf{1}_{\{u\ge\rho(v,Z_{t-})\}}\widetilde{N}(dt,dv,du)\,, \end{aligned}$$

for $t < T := \inf\{t > 0 : X_t = Y_t\}$ and $Y_t = X_t$ for $t \ge T$, where $Z_t := X_t - Y_t$,

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

if $q(v) \neq 0$ and $\rho(v, Z_{t-}) := 1$ if $q(v) = 0$, where $\nu(dv) = q(v)dv$ and

$$(3.4) \qquad R(X_{t-},Y_{t-}) := I - 2\frac{(X_{t-} - Y_{t-})(X_{t-} - Y_{t-})^T}{|X_{t-} - Y_{t-}|^2} = I - 2e_{t-}e_{t-}^T\,,$$

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 Assumption L3 satisfied by the Lévy measure ν of $(L_t)_{t\ge 0}$. 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 (2.10) holds with a discontinuous function f, i.e., we have

$$\mathbb{E}|X_t(x) - Y_t(y)| \le Le^{-\theta t}f(|x-y|)$$

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 (2.10), 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

$$(3.5) \qquad \langle b(x) - b(y), x - y\rangle \le -\kappa(|x-y|)|x-y|^2\,.$$

Theorem 3.1. *Let $(X_t)_{t\ge 0}$ be a Markov process in $\mathbb{R}^d$ given as a solution to the stochastic differential equation (3.1), where $(L_t)_{t\ge 0}$ is a pure jump Lévy process satisfying Assumptions L1-L2 and Assumption L5 and $b : \mathbb{R}^d \to \mathbb{R}^d$ 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

$$\mathbb{E}|X_t(x) - Y_t(y)| \leq Le^{-c_1 t}|x-y|$$

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

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

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)

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

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

$$\tau_n := \inf\{t \geq 0 : |Z_t| \notin (1/n, n)\}\,. \tag{3.7}$$

Note that we have $\tau_n \to T$ as $n \to \infty$, 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

$$\mathbb{E}\int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, (I - R(X_{s-}, Y_{s-}))v\rangle \widetilde{N}(ds, dv, du) = 0\,. \tag{3.8}$$

and

$$\mathbb{E}\int_0^{t\wedge\tau_n} \int_{\{|v|\leq m\}\times[0,1]} f_1'(|Z_{s-}|)\frac{1}{|Z_{s-}|}\langle Z_{s-}, A(X_{s-}, Y_{s-}, v, u)\rangle \widetilde{N}(ds, dv, du) = 0\,. \tag{3.9}$$

In [27] it is also shown (see Lemma 3.3 therein) that for any $t > 0$, we have
(3.10)

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

where $0 < \delta < 2m$, $\varepsilon \le \delta$, the constant C_ε is defined as in (2.13) with the first marginal ν_1 of the measure ν and the function $\bar{f}_\varepsilon$ 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

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

where ν_1^m is the first marginal of the truncated measure $\nu^m(dv) := \mathbf{1}_{\{|v|\le m\}}\nu(dv)$. 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

$$-f_1'(r)\kappa(r)r + C_\varepsilon \bar{f}_\varepsilon(r) \le -c_1 f_1(r) \tag{3.11}$$

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 $r \le \delta$ (see Lemma 3.7 in [27]), it is shown how to get a bound of the form

$$\mathbb{E} f_1(|Z_{t\wedge\tau_n}|) - \mathbb{E} f_1(|Z_0|) \le \mathbb{E} \int_0^{t\wedge\tau_n} -c_1 f_1(|Z_s|) ds \,,$$

which then leads to (2.10). Now we will show a different way of dealing with the case of $r \le \delta$, 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

$$f_1''(|Z_{s-} + u\Delta Z_s|) \le \sup_{x\in(|Z_{s-}|,|Z_{s-}|+\varepsilon)} f_1''(x) \mathbf{1}_{\{|Z_s|\in(|Z_{s-}|,|Z_{s-}|+\varepsilon)\}} \,.$$

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

$$\mathbb{E} \sum_{s\in(0,t]} \left(|\Delta Z_s|^2 \int_0^1 (1-u) f_1''(|Z_{s-} + u\Delta Z_s|) du \right) \le C_\varepsilon \mathbb{E} \int_0^t \hat{f}_\varepsilon(|Z_{s-}|) ds \,, \tag{3.12}$$

where

$$\hat{f}_\varepsilon(y) := \sup_{x\in(y-\varepsilon,y)} f_1''(x) \mathbf{1}_{\{y>\delta\}} + \sup_{x\in(y,y+\varepsilon)} f_1''(x) \mathbf{1}_{\{y\le\delta\}} \,.$$

Now we will show that under Assumption L5, after a small modification in the formulas from [27], the inequality

$$-f_1'(r)\kappa(r)r + C_\varepsilon \hat{f}_\varepsilon(r) \le -c_1 f_1(r) \tag{3.13}$$

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

$$f_1(r) = \int_0^r \phi(s)g(s)ds\,, \tag{3.14}$$

where

$$\phi(r) := \exp\left(-\int_0^r \frac{\bar{h}(t)}{C_\varepsilon}dt\right), \qquad g(r) := \begin{cases} 1 - \frac{c_1}{C_\varepsilon}\int_0^r \frac{\Phi(t+\varepsilon)}{\phi(t)}dt\,, & r \leq R_1\,, \\ \frac{1}{2}\,, & r \geq R_1\,. \end{cases}$$

Here $R_1 > 0$ is given by formulas

$$\begin{aligned} R_0 &= \inf\left\{R \geq 0 : \forall r \geq R : \kappa(r) \geq \frac{2M}{R}\right\}\,, \\ R_1 &= \inf\left\{R \geq R_0 + \varepsilon : \forall r \geq R : \kappa(r) \geq \frac{2C_\varepsilon}{(R-R_0)R} + \frac{2M}{R}\right\}\,, \end{aligned} \tag{3.15}$$

but can be chosen arbitrarily large if necessary and c_1 is a positive constant given by

$$c_1 := \frac{C_\varepsilon}{2}\left(\int_0^{R_1} \frac{\Phi(t+\varepsilon)}{\phi(t)}dt\right)^{-1} \tag{3.16}$$

(cf. (3.29) in [27]). Moreover, we have

$$\bar{h}(r) := \sup_{t \in (r, r+\varepsilon)} h^-(t)\,, \qquad \Phi(r) := \int_0^r \phi(s)ds$$

and $h^- = -\min\{h, 0\}$ is the negative part of the function

$$h(r) := r\kappa(r) - 2M\,, \tag{3.17}$$

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

$$-f_1'(r)\kappa(r)r + 2f_1'(r)M + C_\varepsilon\bar{f}_\varepsilon(r) \leq -c_1 f_1(r)\,. \tag{3.18}$$

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 $r \le s$ we have $f_1(r) \le f_1(s) \le \Phi(s) \le \Phi(s+\varepsilon)$ and thus

$$f_1''(s) \le -\frac{c_1}{C_\varepsilon}f_1(r) - \frac{\bar{h}(r)}{C_\varepsilon}f_1'(s)\,.$$

Therefore it remains to be shown that

$$-\frac{\bar{h}(r)}{C_\varepsilon}f_1'(s) \le f_1'(r)\frac{r\kappa(r)}{C_\varepsilon}\,.$$

Actually, we will just show that

$$\frac{1}{C_\varepsilon}f_1'(s)h(s) \le f_1'(r)\frac{r\kappa(r)}{C_\varepsilon}\,. \tag{3.19}$$

Then, since $-h^- \le h$ and $s \in (r, r+\varepsilon)$, we will get

$$\begin{aligned} -\frac{1}{C_\varepsilon}f_1'(s)\sup_{t\in(r,r+\varepsilon)} h^-(t) &= \frac{1}{C_\varepsilon}f_1'(s)\inf_{t\in(r,r+\varepsilon)}(-h^-(t)) \\ &\le \frac{1}{C_\varepsilon}f_1'(s)\inf_{t\in(r,r+\varepsilon)} h(t) \\ &\le f_1'(r)\frac{r\kappa(r)}{C_\varepsilon}\,. \end{aligned}$$

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

$$\begin{aligned} f_1'(s)(s\kappa(s) - 2M) &= f_1'(r)r\kappa(r) - f_1'(r)r\kappa(r) \\ &\quad + f_1'(s)r\kappa(r) - f_1'(s)r\kappa(r) \\ &\quad + f_1'(s)s\kappa(s) - 2Mf_1'(s) \\ &\le f_1'(r)r\kappa(r) + r\kappa(r)(f_1'(s) - f_1'(r)) \\ &\quad + f_1'(s)(s\kappa(s) - r\kappa(r)) - 2Mf_1'(s)\,. \end{aligned}$$

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 $|r\kappa(r)| \le K(\lambda)$. Since $s < r + \varepsilon \le \delta + \varepsilon$, we obtain

$$s\kappa(s) - r\kappa(r) \le 2K(\delta + \varepsilon)\,.$$

We also know that f_1' is non-increasing and thus $f_1'(s) \le f_1'(r)$, 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. □

4. Coupling of jump diffusions

Here we study jump diffusions of more general form (2.14) 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

$$dY_t = \begin{cases} b(Y_t)dt + \sigma_1 R_{\sigma_1}(X_t,Y_t)dB_t & \text{for } t < T\,, \\ dX_t & \text{for } t \geq T\,, \end{cases} \tag{4.1}$$

where $T := \inf\{t \geq 0 : X_t = Y_t\}$ is the coupling time and

$$R_{\sigma_1}(X_t,Y_t) := I - 2e_t e_t^T \tag{4.2}$$

with

$$e_t := \sigma_1^{-1}(X_t - Y_t)/|\sigma_1^{-1}(X_t-Y_t)|\,. \tag{4.3}$$

Using this coupling, Eberle constructed a concave continuous function f given by

$$f(r) := \int_0^r \varphi(s)g(s)ds\,, \tag{4.4}$$

where

$$\varphi(r) := \exp\left(-\frac{1}{2}\int_0^r s\kappa^-(s)ds\right)\,, \qquad g(r) := \begin{cases} 1 - \frac{\alpha c}{2}\int_0^r \frac{\Phi(t)}{\varphi(t)}dt\,, & r \leq R_1\,, \\ \frac{1}{2}\,, & r \geq R_1\,, \end{cases}$$

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

$$c = \frac{1}{\alpha}\left(\int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} ds\right)^{-1} \tag{4.5}$$

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

$$\kappa(r) = \inf\left\{-\frac{|\sigma_1^{-1}(x-y)|^2}{|x-y|^4}\langle b(x)-b(y), x-y\rangle : x, y \in \mathbb{R}^d \text{ s.t. } |x-y| = r\right\}. \tag{4.6}$$

In other words, κ is the largest quantity satisfying

$$\langle b(x)-b(y), x-y\rangle \leq -\kappa(|x-y|)|x-y|^4/|\sigma_1^{-1}(x-y)|^2 \tag{4.7}$$

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

$$2f''(r) - r\kappa(r)f'(r) \leq -c\alpha f(r) \text{ for all } r > 0\,. \tag{4.8}$$

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])

$$C := 2\varphi(R_0)^{-1}, \text{ where } R_0 := \inf\{R \geq 0 : \forall r \geq R\ \kappa(r) \geq 0\}\,. \tag{4.9}$$

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 $df(|Z_t|) \leq dM_t - \widetilde{c}f(|Z_t|)dt$ for some constant $\widetilde{c} > 0$, where $(M_t)_{t\geq 0}$ is a local martingale.

Proof of Theorem 2.3. We consider an equation of the form

$$dX_t = b(X_t)dt + \sigma_1 dB_t^1 + \sigma(X_t)dB_t^2 + dL_t + \int_U g(X_{t-}, u)\widetilde{N}(dt, du)\,,$$

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

$$\langle b(x)-b(y), x-y\rangle + \frac{1}{2}\int_U |g(x,u)-g(y,u)|^2\nu(du) + \|\sigma(x)-\sigma(y)\|_{HS}^2 \leq -\kappa(|x-y|)|x-y|^2 \tag{4.10}$$

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

$$\begin{aligned} M(X_{t-},Y_{t-})L_t := &\int_{\{|v|>m\}\times[0,1]} vN^L(dt,dv,du) \\ &+ \int_{\{|v|\leq m\}\times[0,1]} (X_{t-}-Y_{t-}+v)\mathbf{1}_{\{u<\rho(v,Z_{t-})\}}\widetilde{N^L}(dt,dv,du) \\ &+ \int_{\{|v|\leq m\}\times[0,1]} R(X_{t-},Y_{t-})v\mathbf{1}_{\{u\geq\rho(v,Z_{t-})\}}\widetilde{N^L}(dt,dv,du)\,, \end{aligned}$$

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 $(B^1_t)_{t\geq 0}$. 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

$$dY_t = b(Y_t)dt + \sigma_1 R(X_t,Y_t)dB^1_t + \sigma(Y_t)dB^2_t + M(X_{t-},Y_{t-})dL_t + \int_U g(Y_{t-},u)\widetilde{N}(dt,du)\,.$$

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 $(B^1_t)_{t\geq 0}$ 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

$$\begin{aligned} dZ_t = &(b(X_t)-b(Y_t))dt + 2\sigma_1 e_t e_t^T dB^1_t + (\sigma(X_t)-\sigma(Y_t))dB^2_t \\ &+ (I - M(X_{t-},Y_{t-}))dL_t + \int_U (g(X_{t-},u)-g(Y_{t-},u))\widetilde{N}(dt,du)\,, \end{aligned}$$

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 $d\widetilde{W_t} := e_t^T dB^1_t$ is a one-dimensional Brownian motion in order to simplify our calculations. We apply the Itô formula to get

$$df(|Z_t|) = \sum_{j=1}^{9} I_j\,, \tag{4.11}$$

where

$$I_1 := f'(|Z_t|)\frac{1}{|Z_t|}\langle b(X_t)-b(Y_t), Z_t\rangle dt\,, \quad I_3 := f'(|Z_t|)\frac{1}{|Z_t|}\langle Z_t, (\sigma(X_t)-\sigma(Y_t))dB^2_t\rangle\,,$$

$$I_2 := 2f'(|Z_t|)\frac{1}{|Z_t|}\langle Z_t, \sigma_1 e_t e_t^T dB^1_t\rangle\,, \qquad I_4 := f'(|Z_{t-}|)\frac{1}{|Z_{t-}|}\langle Z_{t-}, (I - M(X_{t-},Y_{t-}))\rangle dL_t$$

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

$$I_6 := \frac{1}{2}\sigma_1^2 \sum_{i,j=1}^{d}\left[f''(|Z_{t-}|)\frac{Z^i_{t-}Z^j_{t-}}{|Z_{t-}|^2} + f'(|Z_{t-}|)(\delta_{ij}\frac{1}{|Z_{t-}|} - \frac{Z^i_{t-}Z^j_{t-}}{|Z_{t-}|^3})\right]4\frac{Z^i_{t-}Z^j_{t-}}{|Z_{t-}|^2}dt$$

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

$$I_8 := \int_U \Big[f(|Z_{t-} + (I - M(X_{t-}, Y_{t-})v|) - f(|Z_{t-}|)$$
$$- \langle (I - M(X_{t-}, Y_{t-})v, \nabla f(|Z_{t-}|)\rangle \Big] N^L(dt, du)$$

and

$$\begin{aligned} I_9 := \int_U \Big[& f(|Z_{t-} + g(X_{t-}, u) - g(Y_{t-}, u)|) - f(|Z_{t-}|) \\ & - \langle g(X_{t-}, u) - g(Y_{t-}, u), \nabla f(|Z_{t-}|)\rangle \Big] N(dt, du) \end{aligned} \tag{4.12}$$

are the jump components.

Now we proceed similarly to [12] and [27]. Since we want to obtain an estimate of the form $df(|Z_t|) \leq dM_t - \widetilde{c} f(|Z_t|)dt$ 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 $\widetilde{c} > 0$ such that (2.15) 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 $\widetilde{c}$, 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

$$f(|Z_{t-} + g(X_{t-},u) - g(Y_{t-},u)|) - f(|Z_{t-}|) - f'(|Z_{t-}|)\frac{1}{|Z_{t-}|}\langle g(X_{t-},u) - g(Y_{t-},u), Z_{t-}\rangle$$

$$\leq f'(|Z_{t-}|)\left(|Z_{t-} + g(X_{t-},u) - g(Y_{t-},u)| - |Z_{t-}| - \frac{1}{|Z_{t-}|}\langle g(X_{t-},u) - g(Y_{t-},u), Z_{t-}\rangle\right).$$

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

$$I_7 = \sum_{k=1}^m f''(|Z_{t-}|)\frac{|\langle Z_t, \sigma^k(X_t) - \sigma^k(Y_t)\rangle|^2}{|Z_{t-}|^2} + \sum_{k=1}^m\sum_{i=1}^d f'(|Z_{t-}|)\frac{1}{|Z_{t-}|}(\sigma_{ik}(X_{t-}) - \sigma_{ik}(Y_{t-}))^2$$

$$-\sum_{k=1}^m f'(|Z_{t-}|)\frac{|\langle Z_t, \sigma^k(X_t) - \sigma^k(Y_t)\rangle|^2}{|Z_{t-}|^3} \leq f'(|Z_{t-}|)\frac{1}{|Z_{t-}|}\|\sigma(X_{t-}) - \sigma(Y_{t-})\|_{HS}^2\,.$$

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 $\widetilde{c} > 0$ and a function $\widetilde{f}$ such that

$$\mathbb{E}\widetilde{f}(|Z_{t\wedge\tau_n}|) - \mathbb{E}\widetilde{f}(|Z_{s\wedge\tau_n}|) \leq -\widetilde{c}\int_s^t \mathbb{E}\widetilde{f}(|Z_{r\wedge\tau_n}|)dr \tag{4.13}$$

holds for any $0 \leq s < t$. Here $\widetilde{c}$ and $\widetilde{f}$ 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 $(B^1_t)_{t\geq 0}$ or $(L_t)_{t\geq 0}$ in the step above. Thus we can use the differential version of the Gronwall inequality to get

$$\mathbb{E}\widetilde{f}(|Z_{t\wedge\tau_n}|) \leq \mathbb{E}\widetilde{f}(|Z_0|)e^{-\widetilde{c}t} \text{ for any } t > 0\,,$$

and after using the Fatou lemma, the fact that $\tau_n \to T$ and that $Z_t = 0$ for $t \geq T$, we get

$$\mathbb{E}\widetilde{f}(|Z_t|) \leq \mathbb{E}\widetilde{f}(|Z_0|)e^{-\widetilde{c}t} \text{ for any } t > 0\,.$$

Since we can compare our function $\widetilde{f}$ 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 via (4.10) to find an appropriate function $\widetilde{f}$ such that (4.13) holds.

□

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 (2.14) 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*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t\,, \tag{4.14}$$

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

$$dX_t = b(X_t)dt + CdB^1_t + \widetilde{\sigma}(X_t)dB^2_t \tag{4.15}$$

with two independent Brownian motions $(B^1_t)_{t\geq 0}$ and $(B^2_t)_{t\geq 0}$, some constant $C > 0$ and a diffusion coefficient $\widetilde{\sigma}$ such that if $\|\sigma(x)\|_{HS} \leq M$ for all $x \in \mathbb{R}^d$ with some constant $M > 0$, then

$$\|\widetilde{\sigma}(x) - \widetilde{\sigma}(y)\|_{HS} \leq \frac{M}{\sqrt{\lambda^2 - C^2}}\|\sigma(x) - \sigma(y)\|_{HS}\,, \tag{4.16}$$

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 $\widetilde{\sigma}$ as

$$\widetilde{\sigma}(x) := \sqrt{\sigma(x)\sigma(x)^T - C^2 I}$$

for some constant $0 < C^2 < \lambda^2$, we can get

$$\langle \widetilde{\sigma}(x)^2 h, h\rangle = \langle \sigma(x)\sigma(x)^T h, h\rangle - C^2\langle h, h\rangle \geq (\lambda^2 - C^2)|h|^2 ,$$

and thus we can assume that $\widetilde{\sigma}(x)$ is also uniformly positive definite. Therefore Lemma 3.3 in [32] applies (our $\widetilde{\sigma}$ corresponds to σ in [32] and our $\sigma\sigma^T$ corresponds to q therein). Thus we get

$$\|\widetilde{\sigma}(x) - \widetilde{\sigma}(y)\|_{HS} \leq \frac{1}{2\sqrt{\lambda^2 - C^2}}\|\sigma(x)\sigma(x)^T - \sigma(y)\sigma(y)^T\|_{HS} .$$

(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 $\widetilde{\sigma}$ 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 $\widetilde{\sigma}$ 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

$$[A^i, A^j]_t = \int_0^t \left(\sigma\sigma^T\right)_{ij}(X_s)ds . \tag{4.17}$$

Hence, if we write

$$dX_t = dA_t + b(X_t)dt = \sigma(X_t)d\widetilde{B}_t + b(X_t)dt ,$$

where $d\widetilde{B}_t = \sigma^{-1}(X_t)dA_t$, then using (4.17) and the Lévy characterization theorem, we infer that $(\widetilde{B}_t)_{t\geq 0}$ is a Brownian motion. Thus $(X_t)_{t\geq 0}$ is a solution to (4.14). □

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

$$dX_t = b(X_t)dt + CdB_t^1 + \sqrt{\sigma(X_t)\sigma(X_t)^T - C^2 I}\, dB_t^2 + \int_U g(X_{t-}, u)\widetilde{N}(dt, du) ,$$

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 (2.4). 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*

$$dX_t = b(X_t)dt + \sigma dB_t + \int_U g(X_{t-}, u)\widetilde{N}(dt, du)\,, \tag{5.1}$$

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 $(\widetilde{X}_t)_{t\geq 0}$ with the drift perturbed by h_t, i.e.,

$$d\widetilde{X}_t = b(\widetilde{X}_t)dt + h_t dt + \sigma dB_t + \int_U g(\widetilde{X}_{t-}, u)\widetilde{N}(dt, du)\,. \tag{5.2}$$

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

$$\mathbb{E}|\widetilde{X}_t - Y_t| \leq C\mathbb{E}\int_0^t e^{c(s-t)}|h_s|ds \tag{5.3}$$

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

$$Z_t := \widetilde{X}_t - Y_t\,,$$

where $(Y_t)_{t\geq 0}$ will be defined below by (5.6), and consider Lipschitz continuous functions λ, $\pi : \mathbb{R}^d \to [0,1]$ such that for some fixed $\delta > 0$ we have

$$\begin{aligned} \lambda^2(z) + \pi^2(z) &= 1 \text{ for any } z \in \mathbb{R}^d\,, \\ \lambda(z) &= 0 \text{ if } |z| \leq \delta/2 \\ \lambda(z) &= 1 \text{ if } |z| \geq \delta\,. \end{aligned} \tag{5.4}$$

Now fix a unit vector $u \in \mathbb{R}^d$ and define $R(\widetilde{X}_t, Y_t) := I - 2e_t e_t^T$, where

$$e_t := \begin{cases} \frac{Z_t}{|Z_t|}\,, & \text{if } \widetilde{X}_t \neq Y_t\,, \\ u\,, & \text{if } \widetilde{X}_t = Y_t\,. \end{cases}$$

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 $(\widetilde{X}_t)_{t\geq 0}$ can be rewritten as

$$d\widetilde{X}_t = b(\widetilde{X}_t)dt + h_t dt + \lambda(Z_t)dB_t^1 + \pi(Z_t)dB_t^2 + \int_U g(\widetilde{X}_{t-}, u)\widetilde{N}(dt, du)\,, \tag{5.5}$$

where $(B_t^1)_{t\geq 0}$ and $(B_t^2)_{t\geq 0}$ are independent Brownian motions, and define

$$dY_t = b(Y_t)dt + \lambda(Z_t)R(\widetilde{X}_t, Y_t)dB_t^1 + \pi(Z_t)dB_t^2 + \int_U g(Y_{t-}, u)\widetilde{N}(dt, du)\,. \tag{5.6}$$

Using the Lévy characterization theorem and the fact that $\lambda^2 + \pi^2 = 1$, we can show that the processes defined by

$$\begin{aligned} d\widetilde{B}_t &:= \lambda(Z_t)dB_t^1 + \pi(Z_t)dB_t^2\,, \\ d\bar{B}_t &:= \lambda(Z_t)R(\widetilde{X}_t, Y_t)dB_t^1 + \pi(Z_t)dB_t^2 \end{aligned}$$

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 $(\widetilde{X}_t, Y_t)_{t\geq 0}$ is not a coupling, but we do not need this to prove (5.3). Consider the equation for $Z_t = \widetilde{X}_t - Y_t$, which is given by

$$dZ_t = (b(\widetilde{X}_t) - b(Y_t))dt + h_t dt + 2\lambda(Z_t)e_t e_t^T dB_t^1 + \int_U (g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u))\widetilde{N}(dt, du)\,,$$

and observe that the process

$$d\widetilde{W}_t := e_t^T dB_t^1$$

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

$$\begin{aligned} dS(|Z_t|^2) &= 4S'(|Z_t|^2)\lambda(Z_t)|Z_t|d\widetilde{W}_t + 2S'(|Z_t|^2)\langle Z_t, h_t + b(\widetilde{X}_t) - b(Y_t)\rangle dt \\ &\quad + 4S'(|Z_t|^2)\lambda^2(Z_t)dt + 8S''(|Z_t|^2)\lambda^2(Z_t)|Z_t|^2 dt \\ &\quad + \int_U \left(S\left(|Z_{t-} + g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)|^2\right) - S\left(|Z_{t-}|^2\right)\right) N(dt, du) \\ &\quad - 2\int_U S'(|Z_{t-}|^2)\langle Z_{t-}, g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)\rangle \nu(du)dt\,. \end{aligned}$$

Since for any $r \in [0, \infty)$ we have $S(r) \to \sqrt{r}$ when $\varepsilon \to 0$, 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

$$\begin{aligned} &\mathbb{E}\int_0^T \int_U \left(S\left(|Z_{t-} + g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)|^2\right) - S\left(|Z_{t-}|^2\right)\right) N(dt, du) \\ &\leq \mathbb{E}\int_0^T \int_U S'(|Z_{t-}|^2)\left(|g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)|^2 + 2\langle Z_{t-}, g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)\rangle\right) \nu(du)dt\,. \end{aligned}$$

Using the fact that g is Lipschitz and that $\sup_{r\leq\varepsilon} S'(r) \lesssim \varepsilon^{-1/2}$, we see that the integral

$$\int_0^T \int_U \left(S\left(|Z_{t-} + g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)|^2 \right) - S\left(|Z_{t-}|^2 \right) \right) N(dt, du)$$

converges to

$$\int_0^T \int_U \mathbf{1}_{\{Z_{t-}\neq 0\}} \left(|Z_{t-} + g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)| - |Z_{t-}| \right) N(dt, du)$$

in L^1 and hence, via a subsequence, almost surely when $\varepsilon \to 0$. Dealing with the other integrals is even easier, cf. Lemmas 2 and 3 in [40] for analogous arguments. Thus we are able to get

$$\begin{aligned} d|Z_t| &= 2\lambda(Z_t) d\widetilde{W}_t + \mathbf{1}_{\{Z_t\neq 0\}} \frac{1}{|Z_t|} \langle Z_t, h_t + b(\widetilde{X}_t) - b(Y_t) \rangle dt \\ &\quad + \int_U \mathbf{1}_{\{Z_{t-}\neq 0\}} \left(|Z_{t-} + g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u)| - |Z_{t-}| \right) N(dt, du) \\ &\quad - \int_U \mathbf{1}_{\{Z_{t-}\neq 0\}} \frac{1}{|Z_{t-}|} \langle Z_{t-}, g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u) \rangle \nu(du) dt \,. \end{aligned}$$

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)

$$\begin{aligned} f(|Z_r|) - f(|Z_s|) &= 2\int_s^r f'(|Z_t|)\lambda(Z_t) d\widetilde{W}_t \\ &\quad + \int_s^r \mathbf{1}_{\{Z_t\neq 0\}} f'(|Z_t|) \frac{1}{|Z_t|} \langle Z_t, h_t + b(\widetilde{X}_t) - b(Y_t) \rangle dt \\ &\quad + \int_s^r \int_U \mathbf{1}_{\{Z_{t-}\neq 0\}} f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \langle Z_{t-}, g(\widetilde{X}_{t-}, u) - g(Y_{t-}, u) \rangle \widetilde{N}(dt, du) \\ &\quad + \int_s^r \int_U \mathbf{1}_{\{Z_{t-}\neq 0\}} \Bigg[f(|Z_{t-} + g(X_{t-}, u) - g(Y_{t-}, u)|) - f(|Z_{t-}|) \\ &\quad - f'(|Z_{t-}|) \frac{1}{|Z_{t-}|} \langle Z_{t-}, g(X_{t-}, u) - g(Y_{t-}, u) \rangle \Bigg] N(dt, du) \\ &\quad + 2\int_s^r f''(|Z_t|) \lambda^2(Z_t) dt \,. \end{aligned}$$

We can see that the integrand in the integral with respect to $(\widetilde{W}_t)_{t\geq 0}$ in (5.7) is bounded (since f' and λ are bounded) and the integrand in the integral with respect to $\widetilde{N}$ 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

$$
\mathbb{E}f(|Z_r|) - \mathbb{E}f(|Z_s|) \leq \mathbb{E}\int_s^r \mathbf{1}_{\{Z_t \neq 0\}} |h_t| dt - \mathbb{E}\int_s^r \mathbf{1}_{\{Z_t \neq 0\}} f'(|Z_t|)|Z_t|\kappa(|Z_t|) dt \\ + \mathbb{E}\int_s^r 2f''(|Z_t|)\lambda^2(Z_t) dt \,. \tag{5.8}
$$

We will now want to use the fact that the function f defined by (4.4) satisfies

$$
2f''(r) - r\kappa(r)f'(r) \leq -cf(r) \,. \tag{5.9}
$$

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

$$
-r_t\kappa(r_t)f'(r_t) + 2f''(r_t)\lambda^2(Z_t) \leq -cf(r_t)\lambda^2(Z_t) + r_t\kappa(r_t)f'(r_t)(\lambda^2(Z_t) - 1) \,. \tag{5.10}
$$

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

$$
c\delta - cf(r_t) + \kappa^-(r_t)(1 - \lambda^2(Z_t))r_t f'(r_t) \leq c\delta - cf(r_t) + \sup_{r \leq \delta} r\kappa^-(r) \,, \tag{5.11}
$$

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

$$
-r_t\kappa(r_t)f'(r_t) + 2f''(r_t)\lambda^2(Z_t) \leq -cf(r_t) + m(\delta) \,. \tag{5.12}
$$

Hence, combining (5.8) with (5.12) multiplied by $\mathbf{1}_{\{r_t \neq 0\}}$, we get

$$
\begin{aligned}
\mathbb{E}f(|Z_r|) - \mathbb{E}f(|Z_s|) &\leq -c\int_s^r \mathbb{E}\mathbf{1}_{\{Z_t \neq 0\}} f(|Z_t|) dt + \mathbb{E}\int_s^r \mathbf{1}_{\{Z_t \neq 0\}} \left(|h_t| + m(\delta)\right) dt \\
&\leq -c\int_s^r \mathbb{E}f(|Z_t|) dt + \mathbb{E}\int_s^r |h_t| dt + \int_s^r m(\delta) dt \,.
\end{aligned}
$$

Now observe that due to Assumption D2, we have $m(\delta) \to 0$ as $\delta \to 0$. We can also choose $\widetilde{X}_0 = Y_0$ 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 $x \leq Cf(x)$. □

Proof of Theorem 2.8. Once we have Lemma 5.1, extending its result to the equation (2.14) 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 (2.14), we can use the Markov property of the process $(\widetilde{X}_t, Y_t)_{t\geq 0}$ to get (2.16), 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 $f : \mathbb{R}^d \to \mathbb{R}$, the Malliavin derivative $\nabla_s f(X_t)$ is an m-dimensional vector $(\nabla_{s,1} f(X_t), \ldots, \nabla_{s,m} f(X_t))$, where $\nabla_{s,k} f(X_t)$ can be thought of as a derivative with respect to $(W_t^k)_{t\geq 0}$, where $W_t = (W_t^1, \ldots, W_t^m)$ is the driving m-dimensional Brownian motion.

We know that if F is a random variable of the form $F = f(\int_0^T g_s^1 dW_s, \ldots, \int_0^T g_s^N dW_s)$ for some smooth function $f : \mathbb{R}^N \to \mathbb{R}$ and $g^1, \ldots, g^N \in L^2([0,T]; \mathbb{R}^m)$ (i.e., $F \in \mathcal{S}$), then for any element $h \in H = L^2([0,T]; \mathbb{R}^m)$ we have

$$\langle \nabla F, h \rangle_{L^2([0,T];\mathbb{R}^m)} = \int_0^t \langle \nabla_s F, h_s \rangle ds = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(F(W_\cdot + \varepsilon \int_0^\cdot h_s ds) - F(W_\cdot) \right), \tag{5.13}$$

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

$$\mathbb{E} \langle \nabla F, h \rangle_{L^2([0,T];\mathbb{R}^m)} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \mathbb{E} \left(F(W_\cdot + \varepsilon \int_0^\cdot h_s ds) - F(W_\cdot) \right), \tag{5.14}$$

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

$$\mathbb{E} \langle \nabla F, h \rangle_{L^2([0,T];\mathbb{R}^m)} = \mathbb{E} \left[F \int_0^T h_s dW_s \right]. \tag{5.15}$$

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

$$\begin{aligned} \mathbb{E} \left[F \int_0^T h_s dW_s \right] &= \mathbb{E} \left[F \frac{d}{d\varepsilon} \exp \left(\varepsilon \int_0^T h_s dW_s - \frac{1}{2} \varepsilon^2 \int_0^T |h_s|^2 ds \right) |_{\varepsilon = 0} \right] \\ &= \mathbb{E} \left[F \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\exp \left(\varepsilon \int_0^T h_s dW_s - \frac{1}{2} \varepsilon^2 \int_0^T |h_s|^2 ds \right) - 1 \right] \right] \\ &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \mathbb{E} \left[F \exp \left(\varepsilon \int_0^T h_s dW_s - \frac{1}{2} \varepsilon^2 \int_0^T |h_s|^2 ds \right) - F \right] \\ &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \mathbb{E} \left[F(W_\cdot + \varepsilon \int_0^\cdot h_s ds) - F(W_\cdot) \right], \end{aligned} \tag{5.16}$$

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

$$\frac{1}{\varepsilon}\left[Z_T^\varepsilon - 1\right] = \int_0^T Z_s^\varepsilon h_s dW_s\,.$$

Now it is easy to see that since for any $\omega \in \Omega$ we have $Z_t^\varepsilon(\omega) \to 1$ with $\varepsilon \to 0$ and Z_t^ε is uniformly bounded in $L^2(\Omega \times [0,T])$, there is a subsequence such that

$$\frac{1}{\varepsilon}\left[Z_T^\varepsilon - 1\right] = \int_0^T Z_s^\varepsilon h_s dW_s \to \int_0^T h_s dW_s \text{ as } \varepsilon \to 0\,, \text{ in } L^2(\Omega)\,.$$

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

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

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

$$\mathbb{E}\langle \nabla f(X_t), h\rangle_{L^2([0,T];\mathbb{R}^m)} = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\left(f(X_t)(W_\cdot + \varepsilon\int_0^\cdot h_s ds) - f(X_t)(W_\cdot)\right)$$

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

$$X_t(W_\cdot + \varepsilon\int_0^\cdot h_s ds) = \int_0^t b(X_s)ds + \varepsilon\int_0^t \sigma(X_s)h_s ds + \int_0^t \sigma(X_s)dW_s\,. \tag{5.17}$$

Hence, using the assumption (2.4) 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 $(\widetilde{X}_t)_{t\geq 0}$) we obtain

$$\begin{aligned}\mathbb{E}\left(f(X_t)(W_\cdot + \varepsilon\int_0^\cdot h_s ds) - f(X_t)(W_\cdot)\right) &= \mathbb{E}\left(f(X_t)(W_\cdot + \varepsilon\int_0^\cdot h_s ds) - f(Y_t')(W_\cdot)\right)\\ &\leq \varepsilon c_2(t)\mathbb{E}\int_0^t c_3(s)|\sigma(\widetilde{X}_s)h_s|ds\,,\end{aligned}$$

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

$$\mathbb{E}\langle \nabla f(X_t), h\rangle_{L^2([0,t];\mathbb{R}^m)} \leq c_2(t)\mathbb{E}\int_0^t c_3(s)|\sigma(\widetilde{X}_s)h_s|ds \leq c_2(t)\sigma_\infty\mathbb{E}\int_0^t c_3(s)|h_s|ds\,. \tag{5.18}$$

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)

$$\mathbb{E}\langle \nabla f(X_t), h\rangle_{L^2([0,t];\mathbb{R}^m)} = \mathbb{E}\int_0^t \langle \mathbb{E}[\nabla_s f(X_t)|\mathcal{F}_s], h_s\rangle ds =: \mathbb{E}\langle \mathbb{E}[\nabla_\cdot f(X_t)|\mathcal{F}_\cdot], h_\cdot\rangle_{L^2([0,t];\mathbb{R}^m)} .$$

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)

$$\begin{aligned}\mathbb{E}\langle \sqrt{g_\cdot}\mathbb{E}[\nabla_\cdot f(X_t)|\mathcal{F}_\cdot], h_\cdot\rangle_{L^2([0,t];\mathbb{R}^m)} &= \mathbb{E}\langle \mathbb{E}[\nabla_\cdot f(X_t)|\mathcal{F}_\cdot], h_\cdot\sqrt{g_\cdot}\rangle_{L^2([0,t];\mathbb{R}^m)} \\ &\leq c_2(t)\sigma_\infty \left(\mathbb{E}\int_0^t g_s c_3^2(s)ds\right)^{1/2} \left(\mathbb{E}\int_0^t |h_s|^2 ds\right)^{1/2} .\end{aligned}$$

Since this holds for an arbitrary adapted, square integrable process h, we have

$$\mathbb{E}\int_0^t g_u |\mathbb{E}[\nabla_u f(X_t)|\mathcal{F}_u]|^2 du \leq c_2^2(t)\sigma_\infty^2 \mathbb{E}\int_0^t g_u c_3^2(u) du . \tag{5.20}$$

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

$$\begin{aligned}\mathbb{E}\int_s^r g_u |\mathbb{E}[\nabla_u f(X_t)|\mathcal{F}_u]|^2 du &\leq \lim_{n\to\infty} \mathbb{E}\int_s^r (g_u \wedge n)|\mathbb{E}[\nabla_u f(X_t)|\mathcal{F}_u]|^2 du \\ &\leq c_2^2(t)\sigma_\infty^2 \mathbb{E}\int_s^r g_u c_3^2(u) du .\end{aligned}$$

Hence we finally obtain (2.18). In order to get (2.19), 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.

□

Proof of Corollary 2.15. Note that from Theorem 2.8 we obtain an inequality of the form (2.16), 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 (2.16) 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 $(\widetilde{X}_t)_{t\geq 0}$ 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 (2.16), it is sufficient to apply Theorem 2.14 with $c_2(t) = Ce^{-ct}$ and $c_3(s) = e^{cs}$.

□

5.2. **Poissonian case.** Consider the solution $(X_t(x))_{t\geq 0}$ to

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_U g(X_{t-}, u)\widetilde{N}(dt, du) \tag{5.21}$$

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

$$\begin{aligned} X_s^{(t,u)}(x) &= X_t(x) + g(X_{t-}, u) + \int_t^s b(X_r^{(t,u)}(x))dr \\ &\quad + \int_t^s \sigma(X_r^{(t,u)}(x))dW_r + \int_t^s \int_U g(X_{r-}^{(t,u)}(x), u)\widetilde{N}(dr, du) \text{ for } s \geq t \,. \end{aligned}$$

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 $f : \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_{\mathrm{Lip}} \leq 1$, if $t < T$ we have

$$\begin{aligned} \mathbb{E}[D_{t,u} f(X_T(x))|\mathcal{F}_t] &\leq \mathbb{E}\left[\left|f(X_T^{(t,u)}(x)) - f(X_T(x))\right| |\mathcal{F}_t\right] \\ &\leq \mathbb{E}\left[\left|X_T^{(t,u)}(x) - X_T(x)\right| |\mathcal{F}_t\right] \\ &\leq e^{-K(T-t)}|g(X_{t-}, u)| \,. \end{aligned}$$

In order to improve this result we will work under the assumption (2.3) from Theorem 2.1 stating that there exists a coupling $(X_t, Y_t)_{t\geq 0}$ of solutions to (5.21) such that

$$\mathbb{E}\left[|X_T - Y_T| \,/\mathcal{F}_t\right] \leq c_1(T - t)|X_t - Y_t| \tag{5.22}$$

holds for any $T \geq t \geq 0$ with some function $c_1 : \mathbb{R}_+ \to \mathbb{R}_+$. 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

$$\mathbb{E}\left[\left|X_T^{(t,u)}(x) - \widehat{Y}_T(x)\right| |\mathcal{F}_t\right] \leq c_1(T-t)|g(X_{t-},u)|$$

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

$$\begin{aligned} \mathbb{E}[D_{t,u}f(X_T(x))|\mathcal{F}_t] &= \mathbb{E}[f(X_T^{(t,u)}(x)) - f(X_T(x))|\mathcal{F}_t] \\ &= \mathbb{E}[f(X_T^{(t,u)}(x)) - f(\widehat{Y}_T(x))|\mathcal{F}_t] \\ &\leq \mathbb{E}\left[\left|X_T^{(t,u)}(x) - \widehat{Y}_T(x)\right| |\mathcal{F}_t\right] \\ &\leq c_1(T-t)|g(X_{t-},u)|\,, \end{aligned} \tag{5.23}$$

where we used the coupling property in the second step. In particular, if there exists a measurable function $g_\infty : U \to \mathbb{R}$ such that $|g(x,u)| \leq g_\infty(u)$ for any $x \in \mathbb{R}^d$ and $u \in U$, then we obviously get

$$\mathbb{E}[D_{t,u}f(X_T(x))|\mathcal{F}_t] \leq c_1(T-t)g_\infty(u)\,. \tag{5.24}$$

To end this section, let us consider briefly the case of the equation (2.14), 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

$$\mathbb{E}\left[|X_T - Y_T|\,/\mathcal{F}_t\right] \leq \widetilde{C}e^{-\widetilde{c}(T-t)}|X_t - Y_t|$$

holds for any $T \geq t \geq 0$ with some constants $\widetilde{C}$, $\widetilde{c} > 0$. Then, repeating the reasoning above, we easily get

$$\mathbb{E}[D^L_{t,u}f(X_T(x))|\mathcal{F}_t] \leq \widetilde{C}e^{-\widetilde{c}(T-t)}u \tag{5.25}$$

and

$$\mathbb{E}[D_{t,u}f(X_T(x))|\mathcal{F}_t] \leq \widetilde{C}e^{-\widetilde{c}(T-t)}g_\infty(u)\,. \tag{5.26}$$

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

$$F = \mathbb{E}F + \int_0^T \mathbb{E}[\nabla_t F|\mathcal{F}_t]dW_t + \int_0^T \int_U \mathbb{E}[D_{t,u}F|\mathcal{F}_t]\widetilde{N}(dt,du)\,. \tag{6.1}$$

From the proof of Lemma 3.2 in [25] we know that if we show that there exists a deterministic function $h : [0,T] \times U \to \mathbb{R}$ such that $\int_0^T \int_U h(t,u)^2 \nu(du)dt < \infty$ and

$$\mathbb{E}[D_{t,u}F|\mathcal{F}_t] \leq h(t,u) \tag{6.2}$$

and there exists a deterministic function $j : [0,T] \to \mathbb{R}^m$ such that $\int_0^T |j(t)|^2 dt < \infty$ and

$$|\mathbb{E}[\nabla_t F|\mathcal{F}_t]| \leq |j(t)| \,, \tag{6.3}$$

then for any C^2 convex function $\phi : \mathbb{R} \to \mathbb{R}$ such that ϕ' is also convex, we have

$$\mathbb{E}\phi(F - \mathbb{E}F) \leq \mathbb{E}\phi\left(\int_0^T \int_U h(t,u)\widetilde{N}(dt,du) + \int_0^T j(t)dW_t \right) . \tag{6.4}$$

In particular, for any $\lambda > 0$ we have

$$\mathbb{E}e^{\lambda(F-\mathbb{E}F)} \leq \exp\left(\int_0^T \int_U (e^{\lambda h(t,u)} - \lambda h(t,u) - 1)\nu(du)dt + \int_0^T \frac{\lambda^2}{2}|j(t)|^2 dt \right) . \tag{6.5}$$

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

$$M_t(\omega,\omega') := \int_0^t \int_U \mathbb{E}[D_{s,u}F|\mathcal{F}_s](\omega)\widetilde{N}(\omega,ds,du) + \int_0^t \mathbb{E}[\nabla_s F|\mathcal{F}_s](\omega)dW_s(\omega) \,, \tag{6.6}$$

which is a forward martingale with respect to the increasing filtration $\mathcal{F}_t \otimes \mathcal{F}$ on Ω^2 and

$$M_t^*(\omega,\omega') := \int_t^T \int_U h(s,u)\widetilde{N}(\omega',ds,du) + \int_t^T j(s)dW_s(\omega') \,, \tag{6.7}$$

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

$$\mathbb{E}\int_s^r g_u |\mathbb{E}[\nabla_u F|\mathcal{F}_u]|^2 du \leq \mathbb{E}\int_s^r g_u |j(u)|^2 du \,, \tag{6.8}$$

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 $M_t + M_t^* \to F - \mathbb{E}F$ in L^2 as $t \to T$. 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

$$\mathbb{E}\phi(F - \mathbb{E}F) - \phi'(0)\mathbb{E}(M_T) \leq \lim_{t\to T} \mathbb{E}\phi(M_t + M_t^*) \,.$$

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

$$\mathbb{E}\phi(F - \mathbb{E}F) \leq \lim_{t\to T} \mathbb{E}\phi(M_t + M_t^*) \leq \lim_{t\to T} \mathbb{E}\phi(M_0^*) = \mathbb{E}\phi(M_0^*) \,,$$

which proves (6.4).

Now we can return to the equation (2.1). Using the assumption (2.3) we can get a bound on the Malliavin derivative D of a Lipschitz functional of $X_T(x)$, i.e., for any $f : \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_{\mathrm{Lip}} \le 1$ we have

$$\mathbb{E}[D_{t,u} f(X_T(x))|\mathcal{F}_t] \le c_1(T-t)|g(X_{t-},u)| \le c_1(T-t)g_\infty(u) \tag{6.9}$$

(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 (2.4), via Theorem 2.14, for any adapted $\mathbb{R}_+$-valued process g and any $[s,r] \subset [0,T]$ we get

$$\mathbb{E}\int_s^r g_u |\mathbb{E}[\nabla_u f(X_T)|\mathcal{F}_u]|^2 du \le c_2^2(T)\sigma_\infty^2 \mathbb{E}\int_s^r g_u c_3^2(u) du\,. \tag{6.10}$$

It is easy to see that with our bounds, directly from (6.4) we obtain (2.7). Note that as the integrand in the Brownian integral appearing in (2.7) 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 $(X_s(x),\widehat{Y}_s(x))_{s\ge 0}$ which we discussed in Section 5.2. Denote by $\widehat{Y}_{[0,T]}$ a path of the process $(\widehat{Y}_s(x))_{t\in[0,T]}$. Then for any Lipschitz functional $F : \mathbb{D}([0,T];\mathbb{R}^d) \to \mathbb{R}$ (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 $\|F\|_{\mathrm{Lip}} \le 1$ we have

$$\begin{aligned}
&\mathbb{E}[D_{t,u}F(X_{[0,T]}(x))|\mathcal{F}_t] = \mathbb{E}[F(X^{(t,u)}_{[0,T]}(x)) - F(X_{[0,T]}(x))|\mathcal{F}_t] \\
&\quad = \mathbb{E}\left[F(X^{(t,u)}_{[0,T]}(x)) - F(\widehat{Y}_{[0,T]}(x))|\mathcal{F}_t\right] \le \mathbb{E}\left[\int_0^T \left|X^{(t,u)}_r(x) - \widehat{Y}_r(x)\right| dr|\mathcal{F}_t\right] \\
&\quad = \int_t^T \mathbb{E}\left[\left|X^{(t,u)}_r(x) - \widehat{Y}_r(x)\right| |\mathcal{F}_t\right] dr \le \int_t^T c_1(r-t)|g(X_{t-},u)|dr \\
&\quad \le g_\infty(u)\int_t^T c_1(r-t)dr\,.
\end{aligned}$$

In order to get a bound on $\mathbb{E}[\nabla_\cdot F(X_{[0,T]}(x))|\mathcal{F}_\cdot]$, we proceed similarly as in the proof of Theorem 2.14, using again the coupling $(X_t,Y'_t)_{t\ge 0}$ satisfying the assumption (2.4). Namely, we can show that for any bounded, adapted process h we have

$$\begin{aligned}
&\mathbb{E}\langle \nabla F(X_{[0,T]}(x)), h\rangle_{L^2([0,T];\mathbb{R}^m)} \\
&\qquad = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\left(F(X_{[0,T]}(x))(W_\cdot + \varepsilon\int_0^\cdot h_u du) - F(Y'_{[0,T]}(x))(W_\cdot)\right) \\
&\qquad \le \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^T \mathbb{E}\left|X_r(x)(W_\cdot + \varepsilon\int_0^\cdot h_u du) - Y'_r(x)(W_\cdot)\right| dr \\
&\qquad \le \int_0^T \left(c_2(r)\int_0^r c_3(u)\sigma_\infty|h_u|du\right) dr = \int_0^T \left(\int_u^T c_2(r)c_3(u)\sigma_\infty|h_u|dr\right) du \\
&\qquad \le \left(\int_0^T \left(\int_u^T c_2(r)dr\right)^2 c_3^2(u)\sigma_\infty^2 du\right)^{1/2} \left(\int_0^T |h_u|^2 du\right)^{1/2}.
\end{aligned}$$

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

$$\mathbb{E}\int_s^t g_u|\mathbb{E}[\nabla_u F(X_{[0,T]}(x))|\mathcal{F}_u]|^2 du \le \sigma_\infty^2\mathbb{E}\int_s^t g_u c_3^2(u)\left(\int_u^T c_2(r)dr\right)^2 du\,.$$

This, due to (6.4), gives (2.8). 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

$$\mathbb{E}e^{\lambda(f(X_T(x))-p_Tf(x))} \leq \exp\left(\int_0^T \beta(\lambda c_1(T-t))dt + \frac{\lambda^2}{2}\sigma_\infty^2 c_2^2(T)\int_0^T c_3^2(t)dt\right) \tag{6.11}$$

and on the path space as

$$\begin{aligned}&\mathbb{E}e^{\lambda\left(F(X_{[0,T]}(x))-\mathbb{E}F(X_{[0,T]}(x))\right)}\\ &\leq \exp\left(\int_0^T \beta\left(\lambda\int_t^T c_1(r-t)dr\right)dt + \frac{\lambda^2}{2}\sigma_\infty^2\int_0^T c_3^2(t)\left(\int_t^T c_2(r)dr\right)^2 dt\right).\end{aligned} \tag{6.12}$$

Then, by the Gozlan-Léonard characterization (1.4) and the Fenchel-Moreau theorem, we easily get (2.5) from (6.11) and (2.6) from (6.12).

□

Remark 6.1. Note that if instead of (2.3) we have an inequality like

$$\mathbb{E}[|X_t - Y_t|/\mathcal{F}_s] \leq c_1(t-s)(|X_s - Y_s| + 1)\,, \tag{6.13}$$

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

$$\begin{aligned}F = \mathbb{E}F &+ \int_0^T \mathbb{E}[\nabla_t^1 F|\mathcal{F}_t]dB_t^1 + \int_0^T \mathbb{E}[\nabla_t^2 F|\mathcal{F}_t]dB_t^2\\ &+ \int_0^T\int_U \mathbb{E}[D_{t,u}^L F|\mathcal{F}_t]\widetilde{N^L}(dt,du) + \int_0^T\int_U \mathbb{E}[D_{t,u}F|\mathcal{F}_t]\widetilde{N}(dt,du)\,,\end{aligned}$$

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 (2.15) and (2.16) 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 (2.20) and (2.21) 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 $T \to \infty$ in the α_T-W_1H inequality, we obtain (2.17) by the argument from the proof of Lemma 2.2 in [11].

□

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, Trans. Amer. Math. Soc. (2018), in press, `https://doi.org/10.1090/tran/7576`.

[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, Stochastic Process. Appl. (2018), in press, `https://doi.org/10.1016/j.spa.2018.09.003`.

[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. 127 (2017), no. 12, 4083-4125.
[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. 5 (2017), no. 3, 368-399.

Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: majka@uni-bonn.de