

Transportation inequalities for stochastic differential equations of pure jumps

Liming Wu

Laboratoire de Mathématiques Appliquées, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France and Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, 100190 Beijing, China. E-mail: Li-Ming.Wu@math.univ-bpclermont.fr

Received 15 September 2008; revised 3 April 2009; accepted 20 April 2009

Abstract. For stochastic differential equations of pure jumps, though the Poincaré inequality does not hold in general, we show that W_1H transportation inequalities hold for its invariant probability measure and for its process-level law on right continuous paths space in the L^1 -metric or in uniform metrics, under the dissipative condition. Several applications to concentration inequalities are given.

Résumé. Pour une équation différentielle stochastique de pur saut, bien que l'inégalité de Poincaré ne soit pas valide en général, nous pouvons quand même établir, sous la condition de dissipativité, des inégalités de transport W_1H pour sa mesure invariante et pour sa loi (au niveau de processus) sur l'espace des trajectoires càdlàg, muni de la métrique L^1 ou d'une métrique uniforme. Quelques applications aux inégalités de concentration sont présentées.

MSC: 60E15; 60H10; 60H07

Keywords: Transportation inequalities; Stochastic differential equations; Malliavin calculus

1. Introduction

1.1. Object

Let $N(dt, du)(\omega) = \sum_k \delta_{(t_k, u_k)}$ be a random Poisson point process on $\mathbb{R}^+ \times U$ with intensity measure $dt m(du)$, and $\tilde{N} = N - dt m(du)$ the compensated Poisson point process, where (U, \mathcal{U}, m) is some σ -finite measure space and δ_x is the Dirac measure at x . The object of this paper is the following stochastic differential equation (SDE in short)

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \int_U \sigma(X_{s-}, u) \tilde{N}(ds, du), \quad (1.1)$$

where X_0 is some (random) initial point independent of $N(dt, du)$ and (C) the vector field $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\mathbb{R}^d \ni x \rightarrow \sigma(x, \cdot) \in L^2(U, m; \mathbb{R}^d)$ (the space of all m -square integrable \mathbb{R}^d -valued measurable functions on U) are locally Lipschitzian.

Assume the following dissipative condition:

$$\langle x - y, b(x) - b(y) \rangle + \frac{1}{2} \int_U |\sigma(x, u) - \sigma(y, u)|^2 dm(u) \leq -K|x - y|^2, \quad x, y \in \mathbb{R}^d, \quad (1.2)$$

where $K \in \mathbb{R}$ is some constant, $\langle x, y \rangle$ is the Euclidean inner product and $|x| := \sqrt{\langle x, x \rangle}$. If $X_0 = x$, the SDE (1.1) admits a unique solution $X_t(x)$, which is right continuous and has left-limit X_{t-} in t , and $\mathbb{E} \sup_{s \leq t} |X_s|^2 < +\infty$ for

each $t > 0$ [2]. We denote by $P_t(x, dy)$ the distribution of $X_t(x)$ and (P_t) is the transition kernel semigroup of the Markov process (X_t) . For any $f \in C_b^2(\mathbb{R}^d)$, by Itô’s formula, the generator of (X_t) is given by

$$\mathcal{L}f(x) = \langle b(x), \nabla f(x) \rangle + \int_U (f(x + \sigma(x, u)) - f(x) - \langle \nabla f(x), \sigma(x, u) \rangle) m(du)$$

(∇ being the gradient) and the corresponding *carré-du-champs* operator is given by

$$\Gamma(f, f)(x) := \frac{1}{2}(\mathcal{L}(f^2) - 2f\mathcal{L}f) = \frac{1}{2} \int_U [f(x + \sigma(x, u)) - f(x)]^2 m(du).$$

A particular case: $U = \mathbb{R}^{d*} := \mathbb{R}^d \setminus \{0\}$, $\sigma(x, u) = B(x)u$ where $B(x) \in \mathcal{M}(d \times d)$ (the space of $d \times d$ matrices). In such case the SDE (1.1) becomes

$$dX_t = b(X_t) dt + B(X_{t-}) dL_t, \tag{1.3}$$

where $L_t := \int_0^t \int_{\mathbb{R}^{d*}} u \tilde{N}(dt, du)$ is a Lévy process of pure jumps with Lévy’s measure $m(du)$. See Jacod [9] and references therein on this SDE driven by Lévy processes.

1.2. Several known results in the diffusion case

When $B(X_t) dL_t$ is replaced by $\sqrt{2} dW_t$ in (1.3), where (W_t) is a standard Brownian motion valued in \mathbb{R}^d and b is C^1 , the dissipative condition (1.2) is equivalent to

$$\nabla^s b := \left(\frac{1}{2}(\partial_{x_j} b_i + \partial_{x_i} b_j) \right)_{i,j=1,\dots,d} \leq -KI$$

(I being the identity matrix) in the order of definite positiveness of symmetric matrices, which is exactly the Bakry–Emery’s Γ_2 -condition [1]. Assume $K > 0$ from now on. Bakry–Emery’s criterion [1] says that the unique invariant probability measure μ for the semigroup P_t with generator $\Delta f + \langle b(x), \nabla f(x) \rangle$ satisfies the log-Sobolev inequality:

$$\mu(f \log f) - \mu(f) \log \mu(f) \leq \frac{1}{2K} \mu \left(\frac{|\nabla f|^2}{f} \right), \quad 0 < f \in C_b^2(\mathbb{R}^d) \tag{1.4}$$

($\mu(f) := \int f d\mu$), which is equivalent to the exponential decay in entropy:

$$H(v P_t | \mu) \leq e^{-2Kt} H(v | \mu) \quad \forall t > 0, v \in M_1(\mathbb{R}^d), \tag{1.5}$$

where $M_1(\cdot)$ denotes the space of probability measures on \cdot , $H(v | \mu)$ is the relative entropy of v with respect to (w.r.t. in short) μ , defined by

$$H(v | \mu) = \begin{cases} \int \log \frac{dv}{d\mu} dv, & \text{if } v \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.6}$$

By Otto–Villani [12], the log-Sobolev inequality of μ implies the following Talagrand’s transportation inequality

$$W_2(v, \mu)^2 \leq \frac{2}{K} H(v | \mu) \quad \forall v \in M_1(\mathbb{R}^d), \tag{1.7}$$

which in turn implies the Poincaré inequality $K \text{Var}_\mu(f) \leq \mu(|\nabla f|^2)$ or equivalently

$$\text{Var}_\mu(P_t f) \leq e^{-2Kt} \text{Var}_\mu(f). \tag{1.8}$$

Here $W_p(\nu, \mu) := W_{p,|\cdot|}(\nu, \mu)$ and the L^p -Wasserstein distance $W_{p,d}(\nu_1, \nu_2)$ between two probability measures ν_1, ν_2 on a metric space (E, d) , for $1 \leq p \leq +\infty$, is defined by

$$W_{p,d}(\nu_1, \nu_2) := \inf_{X,Y} \|X - Y\|_p, \tag{1.9}$$

where the infimum is taken for all probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and all couples of E -valued random variables X, Y defined thereon, of law ν_1 and ν_2 , respectively.

For the SDE (1.3) with L_t replaced by W_t (but with non-constant diffusion coefficient $B(x)$), Djellout, Guillin and Wu [5], Theorem 5.6, proved that under condition

$$\langle x - y, b(x) - b(y) \rangle + \frac{1}{2} \|B(x) - B(y)\|_{\text{HS}}^2 \leq -K|x - y|^2, \quad x, y \in \mathbb{R}^d \tag{1.10}$$

($\|\cdot\|_{\text{HS}}$ is the Hilbert–Schmidt norm), Talagrand’s T_2 -inequality (1.7) continues to hold with the constant $2/K$ replaced by $\sup_{x \in \mathbb{R}^d, |z|=1} |\sigma(x)z|^2 / K$, and the law \mathbb{P}_x of the solution $X_{[0,T]}(x)$ satisfies on $C([0, T], \mathbb{R}^d)$,

$$\frac{1}{T} W_{1,d_{L^1}}(Q, \mathbb{P}_x)^2 \leq W_{2,d_{L^2}}(Q, \mathbb{P}_x)^2 \leq \frac{2 \sup_{x \in \mathbb{R}^d, |z|=1} |\sigma(x)z|^2}{K^2} H(Q|\mathbb{P}_x), \tag{1.11}$$

where $d_{L^p}(\gamma_1, \gamma_2) := (\int_0^T |\gamma_1(t) - \gamma_2(t)|^p dt)^{1/p}$ for two paths γ_1, γ_2 indexed by $[0, T]$. Those inequalities are sharp as shown in [5]. See the textbooks of Ledoux [11] and Villani [14] about transportation inequalities and their wide applications.

1.3. Poincaré inequality: A counter-example in jumps case

Almost everything changes in the pure jumps case, drastically. For instance, Gross’ log-Sobolev inequality for the Wiener measure no longer holds for the Poisson process; the equivalence between the log-Sobolev inequality for the invariant measure μ and the exponential convergence in entropy (1.5) is lost [16]. Below we give a simple counter-example for which the Poincaré inequality does not hold.

Example 1.1 (Ornstein–Uhlenbeck process driven by Poisson process). Let $d = 1$,

$$dX_t = -X_t dt + d\tilde{N}_t, \quad X_0 = x,$$

where N_t is a Poisson process with parameter 1 and $\tilde{N}_t := N_t - t$. Its solution is given by $X_t(x) = e^{-t}x + \int_0^t e^{-(t-s)} d\tilde{N}_s$. Its unique invariant measure μ is the law of $\int_0^\infty e^{-t} d\tilde{N}_t$, that is, an infinitely divisible law of pure jumps with the Lévy measure $m_\infty(dz) = 1_{(0,1)}(z) \frac{1}{z} dz$.

The Poincaré inequality that is equivalent to the exponential decay (1.8) in $L^2(\mu)$ reads now as:

$$\text{Var}_\mu(f) \leq c_P \langle f, -\mathcal{L}f \rangle_\mu = c_P \int_{\mathbb{R}} \Gamma(f) d\mu = \frac{c_P}{2} \int_{\mathbb{R}} (f(x+1) - f(x))^2 d\mu(x), \tag{1.12}$$

which does not hold for periodic smooth non-constant functions f with period 1.

The failure of the Poincaré inequality for this example is rooted at the fact that the Lévy measure δ_1 of \tilde{N}_t is too degenerate. For the linear equation $dX_t = AX_t + dL_t$ ($A \in \mathcal{M}(d \times d)$) satisfying (1.2), under some strong non-degenerate condition on the Lévy measure m of L_t , some positive results are known about the Poincaré inequality (Röckner–Wang [15]) and the stronger Φ -entropy inequality (Gentil–Imbert [6]). When L_t contains the Gaussian noise, Röckner–Wang [15] proved stronger functional inequalities even in the infinite dimension setting.

In the non-linear drift (b) case, actually very little is known about functional inequalities to the knowledge of the author.

1.4. Purpose and organization

Though the Poincaré inequality does not hold in the pure jumps case in general, but some kind of transportation inequalities can be saved. That is the purpose of this paper.

This paper is organized as follows. In the next section after recalling Gozlan–Léonard’s characterization for general W_1H -transportation inequalities, we first state W_1H inequalities for the kernel P_t , the invariant μ as well as for the law of the process $X_{[0,T]}$ in the L^1 -metric (Theorem 2.2). Several applications are given for showing the sharpness and usefulness of those W_1H -transportation inequalities. A generalization to the uniform metrics is also stated (Theorem 2.11) and explained.

In Section 3, by means of the Malliavin calculus on the Poisson space and Klein–Ma–Privault’s forward-backward martingale method, we show a crucial concentration inequality for functionals on the Poisson space. Furthermore some estimates on the SDE (1.1) under (1.2) are established. Having those preparations we prove quite easily Theorems 2.2 and 2.11, respectively in Sections 4 and 5.

We keep the notations in this introduction.

2. Main results and applications

2.1. Gozlan–Léonard’s characterization for W_1H transportation inequality

Let $\mu \in M_1(E)$ be fixed, where E is a metric space with metric d .

Lemma 2.1 (Gozlan–Léonard [7]). *Let $\alpha : \mathbb{R}^+ \rightarrow [0, +\infty]$ be a non-decreasing left continuous convex function with $\alpha(0) = 0$. The following properties are equivalent:*

(i) *the W_1H inequality below holds*

$$\alpha(W_{1,d}(v, \mu)) \leq H(v|\mu) \quad \forall v \in M_1(E); \quad (\alpha\text{-}W_1H)$$

(ii) *for every $f : (E, d) \rightarrow \mathbb{R}$ bounded and Lipschitzian with $\|f\|_{\text{Lip}} \leq 1$,*

$$\int e^{\lambda(f - \mu(f))} d\mu \leq e^{\alpha^*(\lambda)}, \quad \lambda > 0, \quad (2.1)$$

where $\alpha^*(\lambda) := \sup_{r \geq 0} (r\lambda - \alpha(r))$ is the semi-Legendre transformation;

(iii) *let $(\xi_k)_{k \geq 1}$ be a sequence of i.i.d.r.v. valued in E of common law μ , for every $f : E \rightarrow \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$,*

$$\mathbb{P}\left(\frac{1}{n} \sum_{k=1}^n f(\xi_k) - \mu(f) > r\right) \leq e^{-n\alpha(r)} \quad \forall r > 0, n \geq 1. \quad (2.2)$$

In such a case α is called a W_1H -deviation function of μ .

The equivalence of (i) and (ii) is a generalization of Bobkov–Götze’s criterion [3] for quadratic α . And their new characterization (iii) gives a very strong probabilistic meaning to the W_1H -inequality ($\alpha\text{-}W_1H$).

2.2. Main result

The main results of this paper are the following counterparts of (1.7) and (1.11) in the pure jumps case.

Theorem 2.2. *Assume (C) and the dissipative condition (1.2) with $K > 0$. Suppose that there is some \mathcal{U} -measurable function $\sigma_\infty(u)$ on U such that $|\sigma(x, u)| \leq \sigma_\infty(u)$, m -a.e. for every $x \in \mathbb{R}^d$ and*

$$\exists \lambda > 0: \beta(\lambda) := \int_U (e^{\lambda\sigma_\infty(u)} - \lambda\sigma_\infty(u) - 1)m(du) < +\infty. \quad (2.3)$$

Let $\mathbb{P}_{x,[0,T]}$ be the law of $X_{[0,T]}(x) := (X_t(x))_{t \in [0,T]}$, the solution of the SDE (1.1) with $X_0 = x$. The following properties hold true:

- (1) (X_t) admits a unique invariant probability measure μ , and for any $p \in [1, 2]$,

$$W_p(v|P_t, \mu) \leq e^{-Kt} W_p(v, \mu) \quad \forall t > 0, v \in M_1(\mathbb{R}^d). \quad (2.4)$$

- (2) For each $T > 0$, $P_T(x, dy)$ satisfies the following W_1H transportation inequality

$$\alpha_T(W_1(v, P_T(x, dy))) \leq H(v|P_T(x, dy)) \quad \forall v \in M_1(\mathbb{R}^d), \quad (2.5)$$

where

$$\alpha_T(r) := \sup_{\lambda \geq 0} \left\{ r\lambda - \int_0^T \beta(e^{-Kt}\lambda) dt \right\} \geq \frac{1}{K} \beta^*(Kr), \quad r \geq 0, T \in [0, +\infty)$$

($\beta^*(r) := \sup_{\lambda \geq 0} (r\lambda - \beta(\lambda))$). In particular for the invariant measure μ ,

$$\frac{1}{K} \beta^*(K W_1(v, \mu)) \leq \alpha_\infty(W_1(v, \mu)) \leq H(v|\mu) \quad \forall v \in M_1(\mathbb{R}^d). \quad (2.6)$$

- (3) For each $T > 0$, $\mathbb{P}_{x,[0,T]}$ satisfies on the space $\mathbb{D}([0, T], \mathbb{R}^d)$ of right-continuous left-limit \mathbb{R}^d -valued functions on $[0, T]$,

$$\alpha_T^P(W_{1,d_{L^1}}(Q, \mathbb{P}_{x,[0,T]})) \leq H(Q|\mathbb{P}_{x,[0,T]}) \quad \forall Q \in M_1(\mathbb{D}([0, T], \mathbb{R}^d)), \quad (2.7)$$

where $d_{L^1}(\gamma_1, \gamma_2) := \int_0^T |\gamma_1(t) - \gamma_2(t)| dt$ is the L^1 -metric, and

$$\alpha_T^P(r) := \sup_{\lambda \geq 0} \left(\lambda r - \int_0^T \beta((1 - e^{-Kt})\lambda/K) dt \right) \geq T \beta^*\left(\frac{Kr}{T}\right). \quad (2.8)$$

- (4) If the jumps of X_t are bounded in size by some constant $M > 0$, i.e., $|\sigma(x, u)| \leq \sigma_\infty(u) \leq M$, m-a.e. for all x , then $\beta^*(r) \geq \frac{r}{2M} \log(1 + \frac{Mr}{\vartheta^2})$ and in particular

$$\begin{aligned} & \frac{K W_{1,d_{L^1}}(Q, \mathbb{P}_{x,[0,T]})}{2M} \log\left(1 + \frac{K M W_{1,d_{L^1}}(Q, \mathbb{P}_{x,[0,T]})}{T \vartheta^2}\right) \\ & \leq H(Q|\mathbb{P}_{x,[0,T]}) \quad \forall Q \in M_1(\mathbb{D}([0, T], \mathbb{R}^d)), \end{aligned} \quad (2.9)$$

where $\vartheta^2 = \int \sigma_\infty^2(u) dm(u)$.

Remark 2.3. In [5] the transportation inequality (1.11) for diffusions is proved by means of the Girsanov transformation. We have not succeeded in adapting this last approach to the jumps case. Our proof here, based on the Malliavin calculus on the Poisson space, is completely different (but simple, too).

Remark 2.4. The W_1H inequalities (2.5)–(2.7) are sharp. Indeed for Example 1.1, we have $K = 1$, $\beta(\lambda) = e^\lambda - \lambda - 1$, and

$$\mathbb{E} e^{\lambda(X_T(x) - \mathbb{E}X_T(x))} = \mathbb{E} \exp\left(\lambda \int_0^T e^{-(T-t)} d(N_t - t)\right) = \exp\left(\int_0^T \beta(\lambda e^{-t}) dt\right).$$

Similarly its invariant measure μ being the law of $\int_0^\infty e^{-t} d(N_t - t)$ verifies $\int e^{\lambda x} d\mu(x) = \exp(\int_0^\infty \beta(e^{-t}\lambda) dt)$. Then (2.5) and (2.6) are optimal by Lemma 2.1. Furthermore,

$$\mathbb{E} e^{\lambda(\int_0^T X_t(x) dt - \mathbb{E}\int_0^T X_t(x) dt)} = \mathbb{E} \exp\left(\lambda \int_0^T (1 - e^{-t}) d\tilde{N}_t\right) = \exp\left(\int_0^T \beta((1 - e^{-t})\lambda) dt\right),$$

which shows that (2.7) is optimal, too.

Remark 2.5. The exponential integrability condition (2.3) on the jumps size, in the case of SDE (1.3) driven by Lévy process with Lévy measure m on \mathbb{R}^{d*} , reads as

$$\exists \lambda > 0, \quad \int_{\mathbb{R}^{d*}} (e^{\lambda \|B\|_\infty |u|} - \lambda \|B\|_\infty |u| - 1) m(du) < +\infty, \quad \|B\|_\infty := \sup_{x \in \mathbb{R}^d} \sup_{|z|=1} |B(x)z| < \infty.$$

This condition is indispensable for the W_1H transportation inequalities in this theorem. Indeed in the special case $dX_t = -X_t dt + dL_t$, if $P_t(x, \cdot)$ (resp. μ) satisfies $(\alpha\text{-}W_1H)$ for some non-zero deviation function α , then there is some $\delta > 0$ such that $\mathbb{E}e^{\delta|X_t|} < +\infty$ (resp. $\int e^{\delta|x|} d\mu(x) < +\infty$), by Lemma 2.1(ii) (and a monotone convergence argument). Either of those two conditions is equivalent to (2.3) by simple explicit calculus.

However without the exponential integrability condition (2.3) some convex concentration inequalities for Lipschitzian functionals still hold, see Remark 5.2, where the natural interpretation of this theorem in comparison form is also presented.

Remark 2.6. In the symmetric case, that is, P_t is symmetric on $L^2(\mu)$, the exponential decay (2.4) of P_t to μ in the Wasserstein metric W_p implies the Poincaré inequality ([18], Lemma 5.4): $K \text{Var}_\mu(f) \leq \int \Gamma(f, f) d\mu, \forall f \in C_b^2(\mathbb{R}^d)$.

2.3. Applications to concentration of empirical measure

We now explain parts (3) and (4) of Theorem 2.2 by the following:

Corollary 2.7. In the framework of Theorem 2.2, let \mathcal{A} be a (non-empty) family of real Lipschitzian functions f on \mathbb{R}^d with $\|f\|_{\text{Lip}} \leq 1$, and $Z_T := \sup_{f \in \mathcal{A}} (\frac{1}{T} \int_0^T f(X_s(x)) ds - \mu(f))$. We have for all $r, T > 0$,

$$\log \mathbb{P}(Z_T > \mathbb{E}Z_T + r) \leq -\alpha_T^P(T r) \leq -T\beta^*(K r). \tag{2.10}$$

In particular in the situation of part (4) of Theorem 2.2 (bounded jumps case),

$$\mathbb{P}(Z_T > \mathbb{E}Z_T + r) \leq \exp\left(-\frac{TKr}{2M} \log\left(1 + \frac{KMr}{\vartheta^2}\right)\right) \quad \forall r, T > 0. \tag{2.11}$$

The same inequalities hold for $Z_T = W_1(L_T, \mu)$, where $L_T := \frac{1}{T} \int_0^T \delta_{X_s(x)} ds$ is the empirical measure.

Proof. We show at first Z_T is measurable. Indeed we may assume that $f(0) = 0$ for all $f \in \mathcal{A}$. Then for any closed ball $\bar{B}(0, R)$ centered at 0 of radius $R > 0$, $\{f_{\bar{B}(0, R)}; f \in \mathcal{A}\}$ is compact in $C_b(\bar{B}(0, R))$ (by Arzela–Ascoli). Then Z_T is measurable on the event $\sup_{s \leq t} |X_s(x)| \leq R$. It remains to let $R \rightarrow +\infty$.

Consider $F(\gamma) := \sup_{f \in \mathcal{A}} |\frac{1}{T} \int_0^T (f(\gamma_t) dt - \mu(f))|, \|F\|_{\text{Lip}} \leq 1/T$ w.r.t. the d_{L^1} -metric. Then $Z_T = F(X_{[0, T]}(x))$ satisfies (2.10) and (2.11) by parts (3) and (4) of Theorem 2.2 and Lemma 2.1.

Finally when \mathcal{A} is the whole family of all f with $\|f\|_{\text{Lip}} \leq 1$, then $Z_T = W_1(L_T, \mu)$ by Kantorovitch–Rubinstein’s identity. □

Remark 2.8. The Poisson behavior in the concentration inequality (2.11) in the bounded jumps case is well known for the sequence of bounded i.i.d.r.v. (Bennett’s inequality) or for the stochastic integral $\int_0^T \int_U f(t, u) d\tilde{N}(dt, du)$ with bounded $f \in L^2(dt dm)$.

For statistical applications of the previous explicit inequality we should estimate the asymptotic bias $\mathbb{E}W_1(L_T, \mu)$. When $T \rightarrow \infty$, since $L_T \rightarrow \mu$, a.s. weakly and $\int |x|^2 dL_T(x) \rightarrow \mu(|x|^2)$, a.s., we have $W_1(L_T, \mu) \rightarrow 0$, a.s. [14], then in $L^1(\mathbb{P})$ by dominated convergence. But transportation inequalities W_1H do not directly give information about the rate of that convergence (even not the much stronger log-Sobolev inequality, see Ledoux [11]). When the central limit theorem for $\sqrt{T}(L_T(f) - \mu(f))$ holds uniformly over $\{f; \|f\|_{\text{Lip}} \leq 1\}$, we have $\mathbb{E}W_1(L_T, \mu) = O(T^{-1/2})$.

There is another widely used method to approach μ (which is unknown in general): instead of one sample $X_{[0,T]}(x)$, we produce n independent copies $N^1(dt, du), \dots, N^n(dt, du)$ of the Poisson point processes N , and consider the solution $X_{[0,T]}^k(x_k)$ ($k = 1, \dots, n$) of the SDE (1.1) driven by N^k (instead of N). Then we approach μ by one of

$$\hat{L}_{n,T} := \frac{1}{n} \sum_{k=1}^n \delta_{X_T^k(x_k)}, \quad \tilde{L}_{n,T} := \frac{1}{n} \sum_{k=1}^n \frac{1}{T} \int_0^T \delta_{X_t^k(x)} dt.$$

By the tensorization of $(\alpha\text{-}W_1H)$ in [7], that is, if α is a W_1H -deviation function for $\mu_k \in M_1(E, d)$ for all $k = 1, \dots, n$, then $n\alpha(r/n)$ is a W_1H -deviation function for $\prod_{k=1}^n \mu_k$ w.r.t. the metric $d_{l_1}(x, y) := \sum_{k=1}^n d(x_k, y_k)$ ($x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in E^n$), and using the facts:

- $W_1(\hat{L}_{n,T}, \mu)$ is a functional of $(X_T^1(x_1), \dots, X_T^n(x_n))$ whose Lipschitzian coefficient w.r.t. the metric d_{l_1} is $\leq 1/n$;
- $W_1(\tilde{L}_{n,T}, \mu)$ is a functional of $(X_{[0,T]}^1(x_1), \dots, X_{[0,T]}^n(x_n))$ whose Lipschitzian coefficient w.r.t. the sum- L^1 metric $\sum_{k=1}^n d_{L^1}(\gamma_k, \tilde{\gamma}_k)$ (for $\gamma, \tilde{\gamma} \in (\mathbb{D}([0, T]; \mathbb{R}^d))^n$) is $\leq 1/(nT)$;

we get by parts (2) and (3) of Theorem 2.2 (and Lemma 2.1):

Corollary 2.9. *Under the assumptions of Theorem 2.2, we have*

$$\mathbb{P}(W_1(\hat{L}_{n,T}, \mu) > \mathbb{E}W_1(\hat{L}_{n,T}, \mu) + r) \leq \exp\left(-\frac{n}{K}\beta^*(Kr)\right), \quad r, T > 0, n \geq 1,$$

and similarly

$$\mathbb{P}(W_1(\tilde{L}_{n,T}, \mu) > \mathbb{E}W_1(\tilde{L}_{n,T}, \mu) + r) \leq \exp(-n\alpha_T^P(Tr)), \quad r, T > 0, n \geq 1.$$

Remark 2.10. *Application of transportation inequalities W_1H to concentration of empirical measure was explored amply by Bolley, Guillin and Villani [4] for i.i.d. sequences and for an interacting system of particles issued of granular media.*

2.4. Generalization to uniform metrics

The L^1 -metric d_{L^1} in Theorem 2.2 may be too weak for some issues: for instance, Theorem 2.2 cannot be applied to concentration of functionals $\sup_{t \in [0, T]} |f(X_t)|$, $(1/n) \sum_{k=0}^{n-1} f(X_{t_k})$ where $t_k \in [kT/n, (k+1)T/n]$ may be random (chosen in practice according to the sample path $X_{[kT/n, (k+1)T/n]}$, this Riemannian sum is more practical than the theoretic empirical mean $L_T(f)$), etc. To cover those functionals, consider the uniform metric $d_\infty(\gamma_1, \gamma_2) = \sup_{t \in [0, T]} |\gamma_1(t) - \gamma_2(t)|$ and the stronger metric (as in [5]):

$$d_{\infty, n}(\gamma_1, \gamma_2) := \sum_{k=0}^{n-1} \sup_{t \in [kT/n, (k+1)T/n]} |\gamma_1(t) - \gamma_2(t)|, \quad \gamma_1, \gamma_2 \in \mathbb{D}([0, T]; \mathbb{R}^d).$$

Theorem 2.11. *Assume (C), (1.2) and (2.3). Assume that $x \rightarrow \sigma(x, \cdot) \in L^2(U, m)$ is globally Lipschitzian, that is, $\|\sigma\|_{\text{Lip}} := \sup_{x \neq y} \frac{\|\sigma(x, \cdot) - \sigma(y, \cdot)\|_{L^2(U, m)}}{|x - y|} < +\infty$. Then there is some constant $C \geq 0$ depending only on $K^- := \max\{-K, 0\}$ and $\|\sigma\|_{\text{Lip}}$ [C is given explicitly in Lemma 3.3 (3.5)] such that:*

- (1) for each $T > 0$, $\mathbb{P}_{x, [0, T]}$ satisfies w.r.t. the uniform metric d_∞ on $\mathbb{D}([0, T]; \mathbb{R}^d)$

$$T\beta^*\left(\frac{W_{1, d_\infty}(Q, \mathbb{P}_{x, [0, T]})}{\sqrt{2} T e^{CT}}\right) \leq H(Q | \mathbb{P}_{x, [0, T]}) \quad \forall Q \in M_1(\mathbb{D}([0, T]; \mathbb{R}^d)); \quad (2.12)$$

(2) if $K > 0$, then $\mathbb{P}_{x,[0,T]}$ satisfies w.r.t. the metric $d_{\infty,n}$ on $\mathbb{D}([0, T]; \mathbb{R}^d)$,

$$T\beta^* \left(\frac{W_{1,d_{\infty,n}}(Q, \mathbb{P}_{x,[0,T]})}{Tc_{T,n}} \right) \leq H(Q|\mathbb{P}_{x,[0,T]}) \quad \forall Q \in M_1(\mathbb{D}([0, T]; \mathbb{R}^d)) \tag{2.13}$$

for each $T > 0, n \geq 1$, where $c_{T,n} := \frac{2\sqrt{2} e^{CT/n}}{1 - e^{-KT/n}}$.

Remark 2.12. In the case that $\sigma(x, u) = \sigma(u)$ is independent of x , (2.12) and (2.13) hold with $C = 0$ and without the factor $\sqrt{2}$ (see Remark 5.1).

Remark 2.13. The inequality (2.12) w.r.t. d_{∞} is bad for large T , but sharp in order for small T . Indeed in the bounded jumps case, for $T = \varepsilon$ small, (2.12) together with part (4) of Theorem 2.2, implies that for $Z_{\varepsilon} := \sup_{0 \leq t \leq \varepsilon} |X_t(x) - x|$ and $r > 0$ fixed,

$$\mathbb{P}(Z_{\varepsilon} > \mathbb{E}Z_{\varepsilon} + r) \leq \exp\left(- (1 + o(1)) \frac{r}{2\sqrt{2}M} \log\left(1 + \frac{Mr}{\sqrt{2} \varepsilon \vartheta^2}\right)\right),$$

where $o(1)$ is an infinitesimal, as $\varepsilon \rightarrow 0+$. This is sharp in order as seen for Example 1.1.

Remark 2.14. To illustrate (2.13), consider for a Lipschitzian observable $f: \mathbb{R}^d \rightarrow \mathbb{R}$, the Riemannian sum $F(\gamma) = (1/n) \sum_{k=0}^{n-1} f(\gamma_{t_k})$, where $t_k = t_k(\gamma) \in [kT/n, (k+1)T/n]$ is chosen so that $|\gamma_1(t_k(\gamma_1)) - \gamma_2(t_k(\gamma_2))| \leq D \sup_{t \in [kT/n, (k+1)T/n]} |\gamma_1(t) - \gamma_2(t)|$. It is easy to see that the $d_{\infty,n}$ -Lipschitzian coefficient of F is $\leq \|f\|_{\text{Lip}} D/n$. Thus we obtain by part (2) of Theorem 2.11 that $F(X_{[0,T]}(x)) = (1/n) \sum_{k=0}^{n-1} f(X_{t_k}(x))$ satisfies the following concentration inequality: for all $r, T > 0, n \geq 1$,

$$\mathbb{P}(F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x)) > r) \leq \exp\left(-T\beta^* \left(\frac{n(1 - e^{-KT/n})r}{2\sqrt{2}Te^{CT/n}\|f\|_{\text{Lip}}D} \right)\right), \tag{2.14}$$

which, as n goes to infinity, gives (2.10) with some slightly worse constant.

2.5. An application to transportation-information inequality

Let

$$J(v|\mu) := \sup \left\{ \int -\frac{\mathcal{L}V}{V} dv; 1 \leq V \in C_b^2(\mathbb{R}^d) \right\}, \quad \text{if } v \ll \mu; +\infty \text{ otherwise} \tag{2.15}$$

be the (modified) Donsker–Varadhan information, which is the rate function in the large deviations of $\mathbb{P}_{\mu}(L_t \in \cdot)$. By [17], Theorem B.1, for any bounded $f: \mathbb{R}^d \rightarrow \mathbb{R}$ with $\|f\|_{\text{Lip}} \leq 1$ and $r > 0$, we have for μ -a.s. $x \in \mathbb{R}^d$,

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{P} \left(\frac{1}{T} \int_0^T f(X_s(x)) ds - \mu(f) > r \right) \geq -\inf \{ J(v|\mu); v(f) - \mu(f) > r \}$$

(which is true without the Lipschitzian property of f). But the left-hand side is not greater than $-g(r) := -\beta^*(Kr)$ by applying (2.10) to $\mathcal{A} = \{f\}$. Now for any $v \in M_1(\mathbb{R}^d)$ different from μ and $\varepsilon > 0$, taking some bounded f with $\|f\|_{\text{Lip}} \leq 1$ such that (possible by Kantorovitch–Rubinstein’s identity)

$$v(f) - \mu(f) > W_1(v, \mu) - \varepsilon > 0,$$

and letting $r = W_1(v, \mu) - \varepsilon$, we have $J(v|\mu) \geq g(r) = g(W_1(v, \mu) - \varepsilon)$. Letting $\varepsilon \rightarrow 0$, we get:

Corollary 2.15. *In the framework of Theorem 2.2,*

$$\beta^*(KW_1(\nu, \mu)) \leq J(\nu|\mu) \quad \forall \nu \in M_1(\mathbb{R}^d).$$

The argument above is borrowed from Guillin et al. [8], where the transportation-information inequalities are studied in different aspects.

3. Preparations

3.1. Malliavin calculus on the Poisson space

The *Poisson space* $(\Omega, \mathcal{F}, \mathbb{P})$ over $\mathbb{R}^+ \times U$ with the intensity measure $dt \times m(du)$ (where m is a positive σ -finite measure on (U, \mathcal{U})) is given by

- $\Omega := \{\omega = \sum_i \delta_{(t_i, u_i)} \text{ (at most countable); } (t_i, u_i) \in \mathbb{R}^+ \times U\}$;
- $\mathcal{F} = \sigma(\omega \rightarrow \omega(B) | B \in \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{U})$; $\mathcal{F}_t = \sigma(\omega \rightarrow \omega(B) | B \in \mathcal{B}([0, t]) \otimes \mathcal{U})$;
- $\forall B \in \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{U}, \forall k \in \mathbb{N}: \mathbb{P}(\omega: \omega(B) = k) = e^{-(dt \times m)(B)} \frac{[(dt \times m)(B)]^k}{k!}$;
- $\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{U}$ disjoint, $\omega(B_1), \dots, \omega(B_n)$ are \mathbb{P} -independent.

Under \mathbb{P} , $N(\omega, dt, du) := \omega(dt, du)$ is exactly the Poisson point process on $\mathbb{R}^+ \times U$ with intensity measure $dt m(du)$.

For a real \mathbb{P} -a.s. well-defined measurable function F on Ω , $D_{t,u}F(\omega) := F(\omega + \delta_{(t,u)}) - F(\omega)$ is well defined up to $\mathbb{P}(d\omega) dt m(du)$ -equivalence and this *difference operator* plays in the Malliavin calculus on the Poisson space, the role of the Malliavin gradient on the Wiener space. We recall the following martingale representation [13,16]:

Lemma 3.1. *For any $F \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E} \int_{\mathbb{R}^+} \int_U [\mathbb{E}(D_{t,u}F | \mathcal{F}_t)]^2 dt m(du) < +\infty$ and*

$$F = \mathbb{E}F + \int_0^\infty \int_U f(t, u) \tilde{N}(dt, du),$$

where $f(t, \cdot)$ is the \mathcal{F}_t -predictable $\mathbb{P} \times dt \times m(du)$ version of $\mathbb{E}(D_{t,u}F | \mathcal{F}_t)$.

3.2. Klein, Ma and Privault's convex concentration inequalities on the Poisson space

The following lemma is one key for the results of this paper.

Lemma 3.2. *Let $F: \Omega \rightarrow \mathbb{R}$ be a bounded measurable function. If there is a deterministic measurable function $h(t, u) \in L^2(dt m(du))$ on $[0, T] \times U$ such that*

$$|\mathbb{E}(D_{t,u}F | \mathcal{F}_t)| \leq h(t, u), \mathbb{P}(d\omega) \times dt \times m(du)\text{-a.e.} \tag{3.1}$$

then for every C^2 -convex function $\phi: \mathbb{R} \rightarrow \mathbb{R}$ so that ϕ' is convex,

$$\mathbb{E}\phi(F - \mathbb{E}F) \leq \mathbb{E}\phi\left(\int_0^\infty \int_U h(t, u) \tilde{N}(dt, du)\right). \tag{3.2}$$

In particular

$$\mathbb{E}e^{F - \mathbb{E}F} \leq \exp\left(\int \int (e^h - h - 1) dm(u) dt\right). \tag{3.3}$$

When $\mathbb{E}(D_{t,u}F | \mathcal{F}_t)$ in condition (3.1) is replaced by $D_{t,u}F$, this lemma is proved by the author [16] using the L^1 -log-Sobolev inequality therein.

Proof of Lemma 3.2. We shall use the forward–backward martingale method in Klein, Ma and Privault [10]. Without loss of generality we assume that $\mathbb{E}F = 0$. Let $f(t, u, \omega)$ be an \mathcal{F}_t -predictable $\mathbb{P}(d\omega)dt m(du)$ -version of $\mathbb{E}(D_{t,u}F|\mathcal{F}_t)$. On the product space $(\Omega^2, \mathcal{F}^2, \mathbb{P}^2)$, define

$$M_t(\omega, \omega') := \int_0^t \int_U f(t, u, \omega)(\omega(dt, du) - dt m(du)), \quad (\omega, \omega') \in \Omega^2,$$

which is a forward martingale w.r.t. the (increasing) filtration $\bar{\mathcal{F}}_t := \mathcal{F}_t \otimes \mathcal{F}$ on Ω^2 , and

$$M_t^*(\omega, \omega') := \int_t^\infty \int_U h(t, u)(\omega'(dt, du) - dt m(du)),$$

which is a backward martingale w.r.t. the (decreasing) filtration $\bar{\mathcal{F}}_t^* := \mathcal{F} \otimes \mathcal{F}_t^*$ on Ω^2 , where $\mathcal{F}_t^* := \sigma(\omega \rightarrow \omega(B); B \in \mathcal{B}([t, +\infty)) \otimes \mathcal{U})$. Observe that M_t is $\bar{\mathcal{F}}_t^*$ -adapted, M_t^* is $\bar{\mathcal{F}}_t$ -adapted (the starting condition for [10], Theorem 3.3). Now our condition (3.1) implies condition (3.6) in [10], Theorem 3.3, so [10], Theorem 3.3, says that $\mathbb{E}\phi(M_t + M_t^*)$ is non-increasing in t . Since $M_t + M_t^* \rightarrow F - \mathbb{E}F$ in L^2 as $t \rightarrow +\infty$ by Lemma 3.1, we apply Fatou’s lemma (applicable for $\phi(M_t + M_t^*) \geq \phi'(0)(M_t + M_t^*)$) to get

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

which is (3.2). Taking $\phi(x) = e^x$ in (3.2) gives (3.3). □

3.3. Some estimates under the dissipativity

We return to our SDE.

Lemma 3.3. Assume (C) and (1.2). For two different initial points $x, y \in \mathbb{R}^d$, the solutions $X_t(x), X_t(y)$ of the SDE (1.1) satisfy

$$\mathbb{E}|X_t(x) - X_t(y)|^2 \leq e^{-2Kt}|x - y|^2, \quad t > 0. \tag{3.4}$$

If furthermore $\|\sigma\|_{\text{Lip}} < +\infty$ (see Theorem 2.11 for this notation), then there is some universal constant $C_1 > 0$ such that for $C := 2K^- + (2C_1^2 + 1)\|\sigma\|_{\text{Lip}}^2$,

$$\mathbb{E} \sup_{0 \leq s \leq h} |X_{t+s}(x) - X_{t+s}(y)|^2 \leq 2e^{-2Kt+2Ch}|x - y|^2, \quad t, h > 0. \tag{3.5}$$

Indeed C_1 is the best universal constant in the L^1 Burkholder–Davies–Gundy inequality $\mathbb{E} \sup_{s \leq t} |M_t| \leq C_1 \mathbb{E} \sqrt{[M]_t}$ for local martingale M_t with $M_0 = 0$.

Proof. Writing $\hat{X}_t := X_t(x) - X_t(y)$, $\hat{b}_t := b(X_t(x)) - b(X_t(y))$, $\hat{\sigma}_t(u) := \sigma(X_t(x), u) - \sigma(X_t(y), u)$, we have by integration by parts and our dissipative condition (1.2),

$$\begin{aligned} d|\hat{X}_t|^2 &= 2\langle \hat{X}_{t-}, \hat{b}_{t-} \rangle dt + 2 \int_U \langle \hat{X}_{t-}, \hat{\sigma}_{t-}(u) \rangle \tilde{N}(dt, du) + \int_U |\hat{\sigma}_{t-}(u)|^2 N(dt, du) \\ &= 2\langle \hat{X}_t(x), \hat{b}_t \rangle dt + \int_U |\hat{\sigma}_t(u)|^2 m(du) dt + \int_U (2\langle \hat{X}_{t-}, \hat{\sigma}_{t-}(u) \rangle + |\hat{\sigma}_{t-}(u)|^2) \tilde{N}(dt, du) \\ &\leq -2K|\hat{X}_t(x)|^2 dt + \int_U (2\langle \hat{X}_{t-}, \hat{\sigma}_{t-}(u) \rangle + |\hat{\sigma}_{t-}(u)|^2) \tilde{N}(dt, du), \end{aligned} \tag{3.6}$$

which implies (3.4) by a localization procedure and Gronwall’s inequality.

For (3.5) let $Z_t := \sup_{s \leq t} |\hat{X}_s|^2$, $M_t := \int_0^t \int_U \langle \hat{X}_{t-}, \hat{\sigma}_{t-}(u) \rangle \tilde{N}(dt, du)$. By Burkholder–Davies–Gundy’s inequality, there is some best universal constant C_1 such that

$$\begin{aligned} \mathbb{E} \sup_{s \leq t} M_t &\leq C_1 \mathbb{E} \sqrt{[M]_t} = C_1 \mathbb{E} \sqrt{\int_0^t \int_U \langle \hat{X}_{s-}, \hat{\sigma}_{s-}(u) \rangle^2 N(ds, du)} \\ &\leq C_1 \mathbb{E} \sqrt{\int_0^t \int_U |\hat{\sigma}_{s-}(u)|^2 N(ds, du)} \leq C_1 \sqrt{\mathbb{E} Z_t \mathbb{E} \int_0^t \int_U |\hat{\sigma}_{s-}(u)|^2 m(du) ds} \\ &\leq \frac{C_1}{2} \left(a \mathbb{E} Z_t + \frac{\|\sigma\|_{\text{Lip}}^2}{a} \mathbb{E} \int_0^t |\hat{X}_s|^2 ds \right), \end{aligned}$$

where $a > 0$ is arbitrary. Now using the first equality in (3.6) and $\langle \hat{X}_t, \hat{b}_t \rangle \leq -2K |\hat{X}_t|^2$,

$$\begin{aligned} \mathbb{E} Z_t &\leq |x - y|^2 + 2K^- \int_0^t |\hat{X}_s|^2 ds + 2 \mathbb{E} \sup_{0 \leq s \leq t} |M_t| + \mathbb{E} \int_0^t \int_U |\hat{\sigma}_{s-}(u)|^2 N(ds, du) \\ &\leq |x - y|^2 + a C_1 \mathbb{E} Z_t + [2K^- + (C_1/a + 1) \|\sigma\|_{\text{Lip}}^2] \int_0^t \mathbb{E} |\hat{X}_s|^2 ds, \end{aligned}$$

letting $a = 1/(2C_1)$ and $C = 2K^- + (2C_1^2 + 1) \|\sigma\|_{\text{Lip}}^2$ we obtain by Gronwall’s inequality

$$\mathbb{E} Z_h \leq 2 \exp(2[2K^- + (2C_1^2 + 1) \|\sigma\|_{\text{Lip}}^2]h) |x - y|^2 = 2e^{2Ch} |x - y|^2, \quad h > 0,$$

which is (3.5) for $t = 0$. Now for (3.5) with $t > 0$, it is enough to notice

$$\mathbb{E} \left(\mathbb{E} \left[\sup_{0 \leq s \leq h} |X_{t+s}(x) - X_{t+s}(y)|^2 / \mathcal{F}_t \right] \right) \leq 2e^{2Ch} \mathbb{E} |X_t(x) - X_t(y)|^2$$

and then to apply (3.4). □

Remark 3.4. In the additive noise case, that is, $\sigma(x, u) = \sigma(u)$ is independent of $x \in \mathbb{R}^d$, we will have $d(X_t(x) - X_t(y)) = [b(X_t(x)) - b(X_t(y))] dt$, which gives

$$d|X_t(x) - X_t(y)|^2 = 2 \langle X_t(x) - X_t(y), b(X_t(x)) - b(X_t(y)) \rangle dt \leq -2K |X_t(x) - X_t(y)|^2 dt.$$

Whence \mathbb{P} -a.s.,

$$|X_t(x) - X_t(y)| \leq e^{-Kt} |x - y| \quad \forall t > 0. \quad (3.7)$$

4. Proof of Theorem 2.2

4.1. Part (1)

Though this part should be well known to specialists, yet we give its proof for the convenience of the reader. Recall that $M_1^p(\mathbb{R}^d) := \{\nu \in M_1(\mathbb{R}^d); \int |x|^p d\nu < +\infty\}$ equipped with the Wasserstein metric W_p is complete [14]. Below let $p \in [1, 2]$.

For any two initial probability measures $\nu_1, \nu_2 \in M_1^p(\mathbb{R}^d)$, let (X_0, Y_0) be a couple of \mathbb{R}^d -valued random variables of law ν_1, ν_2 respectively, independent of the Poisson point process $N(dt, du)$ such that $\mathbb{E} |X_0 - Y_0|^p = W_p(\nu_1, \nu_2)^p$. Let X_t (resp. Y_t) be the solution of the SDE (1.1) with initial condition X_0 (resp. Y_0). (X_t, Y_t) constitutes a coupling of $\nu_1 P_t$ and $\nu_2 P_t$. By Lemma 3.3,

$$\mathbb{E} (|X_t - Y_t|^p | X_0, Y_0) \leq [\mathbb{E} (|X_t - Y_t|^2 | X_0, Y_0)]^{p/2} \leq e^{-pKt} |X_0 - Y_0|^p,$$

whence

$$W_p(v_1 P_t, v_2 P_t) \leq \|X_t - Y_t\|_p \leq e^{-Kt} W_p(v_1, v_2). \tag{4.1}$$

So for each $t > 0$, $v \rightarrow v P_t$ is a contraction on $(M_1^p(\mathbb{R}^d), W_p)$. Thus P_t admits a unique invariant probability measure $\mu_t^{(p)} \in M_1^p(\mathbb{R}^d)$ by the fixed point theorem for contraction mapping. As $\mu_t^{(2)} \in M_1^p(\mathbb{R}^d)$ for all $p \in [1, 2]$, all $\mu_t^{(p)}$, $p \in [1, 2]$ are the same, denoted by μ_t (then $\mu_t = \mu_t^{(2)} \in M_1^2(\mathbb{R}^d)$). Now for $s > 0$,

$$W_1(\mu_t P_s, \mu_t) = W_1(\mu_t P_s P_t, \mu_t P_t) \leq e^{-Kt} W_1(\mu_t P_s, \mu_t),$$

which together with the fact $W_2(\mu_t P_s, \mu_t) < +\infty$ yields $\mu_t P_s = \mu_t$ and so $\mu_t = \mu_s$. Thus μ_t is the same μ for all $t > 0$.

Now for any extreme invariant probability measure $\tilde{\mu}$ of P_t [maybe not belonging to $M_1^1(\mathbb{R}^d)$], where $t > 0$ is fixed, there is some $x_0 \in \mathbb{R}^d$ such that $\frac{1}{n} \sum_{k=1}^n P_{nt}(x_0, dy) \rightarrow \tilde{\mu}$ weakly. But $\frac{1}{n} \sum_{k=1}^n P_{nt}(x_0, dy) \rightarrow \mu$ in W_2 -metric by (4.1), so $\tilde{\mu} = \mu$. Hence μ is the unique invariant probability measure of P_t .

Finally if $v \notin M_1^p(\mathbb{R}^d)$, as $W_p(v, \mu) = +\infty$, (2.4) holds true trivially; if $v \in M_1^p(\mathbb{R}^d)$, (2.4) follows by letting $v_1 = v, v_2 = \mu$ in (4.1).

4.2. Part (2)

From now on to the end of this paper we adopt the following convention.

Convention. We assume without loss of generality that our Poisson point process $N(\omega, dt, du) = \omega(dt, du)$ is defined on the Poisson space $(\Omega, \mathcal{F}, \mathbb{P})$.

Under this convention, the solution $X(x, \omega) = (X_t(x, \omega))_{t \in \mathbb{R}^+}$ of the SDE (1.1) is a measurable mapping from the Poisson space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{D}(\mathbb{R}^+, \mathbb{R}^d)$. We have $\mathbb{P}(d\omega) ds m(du)$ -a.e. on $\Omega \times \mathbb{R}^+ \times U$, $X^{(t,u)}(x, \omega) := X(x, \omega + \delta_{(t,u)})$ (i.e., adding one jump $\sigma(X_{t-}, u)$ at time t in the SDE) satisfies

$$\begin{aligned} X_s^{(t,u)}(x, \omega) &= X_s(x, \omega), \quad \text{if } s < t; \\ X_s^{(t,u)}(x, \omega) &= X_t(x, \omega) + \sigma(X_{t-}(x, \omega), u) + \int_t^s b(X_a^{(t,u)}(x, \omega)) da \\ &\quad + \int_t^s \int_U \sigma(X_{a-}^{(t,u)}(x, \omega), u) \tilde{N}(da, du), \quad \text{if } s \geq t. \end{aligned} \tag{4.2}$$

In other words, after time t , $(X_s^{(t,u)}(x, \omega))_{s \geq t}$ is the solution of the same SDE but with $X_t^{(t,u)}(x, \omega) = X_t(x, \omega) + \sigma(X_{t-}(x, \omega))$. Now given a Lipschitzian function f on \mathbb{R}^d with $\|f\|_{\text{Lip}} \leq 1$, $D_{t,u} f(X_T(x)) = 0$ if $t > T$; and if $t < T$,

$$|D_{t,u} f(X_T(x))| = |f(X_T^{(t,u)}(x)) - f(X_T(x))| \leq |X_T^{(t,u)}(x) - X_T(x)|$$

and then by Lemma 3.3,

$$\begin{aligned} \mathbb{E}(|D_{t,u} f(X_T(x))| / \mathcal{F}_t) &\leq \mathbb{E}(|X_T^{(t,u)}(x) - X_T(x)| / \mathcal{F}_t) \leq e^{-K(T-t)} |X_t^{(t,u)}(x) - X_t(x)| \\ &= e^{-K(T-t)} |\sigma(X_{t-}(x), u)| \leq e^{-K(T-t)} \sigma_\infty(u). \end{aligned}$$

Thus by the key Lemma 3.2, we have for every $\lambda > 0$,

$$\begin{aligned} \mathbb{E} e^{\lambda(f(X_T(x)) - P_T f(x))} &\leq \exp\left(\int_0^T \int_U (e^{\lambda e^{-K(T-t)} \sigma_\infty(u)} - \lambda e^{-K(T-t)} \sigma_\infty(u) - 1) dm(u) dt\right) \\ &= \exp\left(\int_0^T \beta(e^{-Kt} \lambda) dt\right), \end{aligned}$$

where the transportation inequality (2.5) follows by Gozlan–Léonard’s Lemma 2.1 and Fenchel’s theorem $\alpha^{**} = \alpha$ under the condition on α in Lemma 2.1.

To show $\alpha_T(r) \geq \beta^*(Kr)/K$, we notice that the convexity of β (and $\beta(0) = 0$) implies $\beta(e^{-Kt}\lambda) \leq \beta(\lambda)e^{-Kt}$. Then

$$\alpha_T(r) = \sup_{\lambda \geq 0} \left\{ r\lambda - \int_0^T \beta(e^{-Kt}\lambda) dt \right\} \geq \sup_{\lambda \geq 0} \{ r\lambda - \beta(\lambda)/K \} = \frac{1}{K} \beta^*(Kr).$$

Letting $T \rightarrow +\infty$ in (2.5) we obtain (2.6) for the invariant measure μ by the argument in [5], Lemma 2.2 (which is only for quadratic deviation functions α).

4.3. Part (3)

Let $F : \mathbb{D}([0, T], \mathbb{R}^d) \rightarrow \mathbb{R}$ be a bounded d_{L^1} -Lipschitzian function with $\|F\|_{\text{Lip}} \leq 1$. Observing that for $dt m(du)$ -a.e. $(t, u) \in [0, T] \times U$, \mathbb{P} -a.s.,

$$|D_{t,u}F(X_{[0,T]}(x))| \leq d_{L^1}(X_{[0,T]}^{t,u}(x), X_{[0,T]}(x)) = \int_t^T |X_s^{t,u}(x) - X_s(x)| ds$$

we get by Lemma 3.3 and (4.2),

$$\mathbb{E}(|D_{t,u}F(X_{[0,T]}(x))|/\mathcal{F}_t) \leq \int_t^T e^{-K(s-t)} |\sigma(X_{t-}(x), u)| dt \leq \frac{\sigma_\infty(u)}{K} (1 - e^{-K(T-t)}).$$

Thus letting $d(t) := (1 - e^{-Kt})/K$ we have by Lemma 3.2 that for all $\lambda > 0$,

$$\mathbb{E}e^{\lambda[F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x))]} \leq \exp\left(\int_0^T \beta(d(t)\lambda) dt\right).$$

By Gozlan–Léonard’s Lemma 2.1 again, the following function

$$\mathbb{R}^+ \ni r \rightarrow \sup_{\lambda \geq 0} \left(\lambda r - \int_0^T \beta(d(t)\lambda) dt \right) = \alpha_T^P(r)$$

is a W_1H -deviation function for $\mathbb{P}_{x,[0,T]}$ (the law of $X_{[0,T]}(x)$) w.r.t. d_{L^1} -metric, that is, exactly what (2.7) says. It remains to bound $\alpha_T(r)$ from below.

Lower bound in (2.8)

As $d(t) \leq 1/K$, we have $\beta(d(t)\lambda) \leq \beta(\lambda/K)$ (since β is increasing) and then $\int_0^T \beta(d(t)\lambda) dt \leq T\beta(\lambda/K)$. Consequently

$$\alpha_T^P(r) \geq \sup_{\lambda \geq 0} (\lambda r - T\beta(\lambda/K)) = T\beta^*(rK/T).$$

4.4. Part (4)

In the actual bounded jumps case, letting $\beta_0(\lambda) := e^\lambda - \lambda - 1$ and using the increasingness of $\beta_0(\lambda x)/x^2$ in $x > 0$ we have $\beta_0(\lambda x) \leq (x^2/M^2)\beta_0(\lambda M)$ for $x \in [0, M]$, and then

$$\beta(\lambda) = \int_U \beta_0(\lambda \sigma_\infty(u)) dm(u) \leq \frac{\vartheta^2}{M^2} \beta_0(\lambda M).$$

Since $\beta_0^*(r) = (1+r) \log(1+r) - r \geq (r/2) \log(1+r)$ ($r \geq 0$), we obtain

$$\beta^*(r) \geq \frac{\vartheta^2}{M^2} \beta_0^*\left(\frac{Mr}{\vartheta^2}\right) \geq \frac{r}{2M} \log\left(1 + \frac{Mr}{\vartheta^2}\right), \quad r \geq 0,$$

then (2.9).

5. Proof of Theorem 2.11

Its proof is very close to that of Theorem 2.2.

Proof of part (1) of Theorem 2.11. For any $F : \mathbb{D}([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}$ such that its d_∞ -Lipschitzian coefficient $\|F\|_{\text{Lip}(d_\infty)} \leq 1$, we have

$$|D_{t,u}F(X_{[0,T]}(x))| \leq d_\infty(X_{[0,T]}^{t,u}(x), X_{[0,T]}(x)) = \sup_{t \leq s \leq T} |X_s^{t,u}(x) - X_s(x)|.$$

By Lemma 3.3 (3.5) (with t there equal to 0) and (4.2),

$$\mathbb{E}[|D_{t,u}F(X_{[0,T]}(x))|/\mathcal{F}_t] \leq \sqrt{2}e^{C(T-t)}|X_t^{t,u}(x) - X_t(x)| \leq \sqrt{2}e^{CT}\sigma_\infty(u),$$

where it follows by Lemma 3.2,

$$\log \mathbb{E}e^{\lambda(F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x)))} \leq T\beta(\sqrt{2}e^{CT}\lambda), \quad \lambda > 0.$$

This entails (2.12) by Lemma 2.1 with C given in (3.5). \square

Proof of part (2) of Theorem 2.11. Let $h := T/n$. For any $F : \mathbb{D}([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}$ such that its $d_{\infty,n}$ -Lipschitzian coefficient $\|F\|_{\text{Lip}(d_{\infty,n})} \leq 1$, for $t \in [kh, (k+1)h]$ ($0 \leq k \leq n-1$),

$$|D_{t,u}F(X_{[0,T]}(x))| \leq d_{\infty,n}(X_{[0,T]}^{t,u}(x), X_{[0,T]}(x)) = \sum_{j=k}^{n-1} \sup_{jh \leq s \leq (j+1)h} |X_s^{t,u}(x) - X_s(x)|.$$

By (4.2) and Lemma 3.3 (3.5) with t there equal to $(jh - t)^+$, we have for $j > k$,

$$\mathbb{E}\left[\sup_{jh \leq s \leq (j+1)h} |X_s^{t,u}(x) - X_s(x)|/\mathcal{F}_t\right] \leq \sqrt{2}e^{[C-K(j-k-1)]h}|X_t^{t,u}(x) - X_t(x)|$$

and it is $\leq \sqrt{2}e^{Ch}|X_t^{t,u}(x) - X_t(x)|$ for $j = k$. Therefore using $|X_t^{t,u}(x) - X_t(x)| \leq \sigma_\infty(u)$ we get for $t \in [kh, (k+1)h]$ ($k = 0, 1, \dots, n-1$),

$$\mathbb{E}(|D_{t,u}F(X_{[0,T]}(x))|/\mathcal{F}_t) \leq \sqrt{2}e^{Ch}\left(1 + \sum_{j:k+1 \leq j \leq n-1} e^{-K(j-k-1)h}\right)\sigma_\infty(u) \leq \frac{2\sqrt{2}e^{Ch}}{1 - e^{-Kh}}\sigma_\infty(u).$$

The last bound is independent of k , so by Lemma 3.2,

$$\log \mathbb{E}e^{\lambda(F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x)))} \leq T\beta(c_{T,n}\lambda), \quad \lambda > 0; c_{T,n} := \frac{2\sqrt{2}e^{CT/n}}{1 - e^{-KT/n}},$$

which implies (2.13) by Lemma 2.1. \square

Remark 5.1. For the claim in Remark 2.12 in the additive noise case, it is enough to apply (3.7) instead of (3.5), in the proof above.

Remark 5.2 (Concluding remarks). In the case that $\sigma_\infty \in L^2(U, m)$ but is not m -exponentially integrable (i.e., the condition (2.3) on jumps size is violated), though the W_1H -inequalities in this paper do not hold in general (see Remark 2.5), but following the proof of Theorem 2.2 and using Lemma 3.2, we have again the convex concentration inequalities below: under (C) and (1.2) we have for $f : \mathbb{R}^d \rightarrow \mathbb{R}$ Lipschitzian,

$$\mathbb{E}\phi(f(X_T(x)) - P_T f(x)) \leq \mathbb{E}\phi\left(\|f\|_{\text{Lip}} \int_0^T \int_U e^{-Kt} \sigma_\infty(u) \tilde{N}(dt, du)\right) \quad (5.1)$$

and for $F : \mathbb{D}([0, T], \mathbb{R}^d) \rightarrow \mathbb{R}$ d_{L^1} -Lipschitzian,

$$\mathbb{E}\phi(F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x))) \leq \mathbb{E}\phi\left(\|F\|_{\text{Lip}(d_{L^1})} \int_0^T \int_U \frac{1 - e^{-Kt}}{K} \sigma_\infty(u) \tilde{N}(dt, du)\right), \quad (5.2)$$

where $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is convex, C^2 such that ϕ' is convex. When $\phi(x) = e^{\lambda x}$ ($\lambda > 0$), by Lemma 2.1 the two inequalities (5.1) and (5.2) are equivalent to the W_1H -inequalities (2.5) and (2.7), respectively.

The W_1H -inequalities (2.5) and (2.7) in Theorem 2.2 seem to be complicated. But we can interpret them or their generalizations (5.1) and (5.2) in comparison form. Indeed consider the linear Ornstein–Uhlenbeck process

$$dY_t = -KY_t dt + dL_t, \quad Y_0 = 0; L_t := \int_0^t \int_U \sigma_\infty(u) \tilde{N}(ds, du). \quad (5.3)$$

X_t is more dissipative than Y_t by condition (1.2), and the jumps size $|\sigma(X_{t-}, u)|$ of X_t is bounded by the jumps $\sigma_\infty(u)$ of Y_t . Now since $Y_T = \int_0^T \int_U e^{-K(T-t)} \sigma_\infty(u) \tilde{N}(dt, du)$ and $\int_0^T Y_t dt = \int_0^T \int_U \frac{1 - e^{-K(T-t)}}{K} \sigma_\infty(u) \tilde{N}(dt, du)$, thus (5.1) and (5.2) are equivalent to say that: for f, F, ϕ given as above,

$$\mathbb{E}\phi[f(X_T(x)) - P_T f(x)] \leq \mathbb{E}\phi(\|f\|_{\text{Lip}} Y_T)$$

and

$$\mathbb{E}\phi[F(X_{[0,T]}(x)) - \mathbb{E}F(X_{[0,T]}(x))] \leq \mathbb{E}\phi\left(\|F\|_{\text{Lip}(d_{L^1})} \int_0^T Y_t dt\right).$$

In other words the whole work here says nothing else but the fact that (X_t) is more concentrated than the linear Ornstein–Uhlenbeck process (Y_t) given by (5.3).

References

- [1] D. Bakry and M. Emery. Diffusions hypercontractives. In *Séminaire de Probabilités 177–206. Lecture Notes in Math.* **1123**. Springer Berlin, 1985. [MR0889476](#)
- [2] K. Bichteler, J. B. Gravereaux and J. Jacod. *Malliavin Calculus for Processes With Jumps. Stochastics Monographs* **2**. Gordon and Breach, New York, 1987. [MR1008471](#)
- [3] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1999) 1–28. [MR1682772](#)
- [4] F. Bolley, A. Guillin and C. Villani. Quantitative concentration inequalities for empirical measures on non compact spaces. *Probab. Theory Related Fields* **137** (2007) 541–593. [MR2280433](#)
- [5] H. Djellout, A. Guillin and L. Wu. Transportation cost-information inequalities for random dynamical systems and diffusions. *Ann. Probab.* **32** (2004) 2702–2732. [MR2078555](#)
- [6] I. Gentil and C. Imbert. The Lévy–Fokker–Planck equation: Φ -entropies and convergence to equilibrium. Preprint, 2008. [MR2450356](#)
- [7] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields*. To appear, 2009. [MR2322697](#)
- [8] A. Guillin, C. Léonard, L. Wu and N. Yao. Transport-information inequalities for Markov processes (I). *Probab. Theory Related Fields*. Preprint, 2007. To appear.
- [9] J. Jacod. The Euler scheme for Lévy driven stochastic differential equations: Limit theorems. *Ann. Probab.* **32** (2004) 1830–1872. [MR2073179](#)
- [10] T. Klein, Y. Ma and N. Privault. Convex concentration inequalities and forward–backward stochastic calculus. *Electron. J. Probab.* **11** (2006) 486–512. [MR2242653](#)
- [11] M. Ledoux. *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI, 2001. [MR1849347](#)
- [12] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** (2000) 361–400. [MR1760620](#)
- [13] J. Picard. Formules de dualité sur l’espace de Poisson. *Ann. Inst. H. Poincaré Probab. Statist.* **32** (1996) 509–548. [MR1411270](#)
- [14] C. Villani. *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI, 2003. [MR1964483](#)
- [15] M. Röckner and F. Y. Wang. Harnack and functional inequalities for generalized Mehler semigroups. *J. Funct. Anal.* **203** (2003) 237–261. [MR1996872](#)
- [16] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields* **118** (2000) 427–438. [MR1800540](#)
- [17] L. Wu. Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal.* **172** (2000) 301–376. [MR1753178](#)
- [18] L. Wu. Essential spectral radius for Markov semigroups (I): Discrete time case. *Probab. Theory Related Fields* **128** (2004) 255–321. [MR2031227](#)