Generators of measure-valued jump diffusions and convergence rate of diffusive mean-field models

Xavier Erny

Study of measure-valued Markov processes. Explicit bounds for the convergence in distribution of mean-field models.

Xavier Erny

SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France

Abstract: The aim of the paper is to prove that the rates of convergence in distribution for N-particle mean-field models are the expected one: N^{-1} in Law of Large Numbers regime, and $N^{-1/2}$ in Central Limit Theorem regime. These proofs require to study empirical measures of McKean-Vlasov particle systems, and conditional laws of McKean-Vlasov processes, as measure-valued Markov processes. In particular, the expressions of the infinitesimal generators of such processes are established for measure-valued processes with general jumps. The generators being differential operators, all the proofs rely on the analytical properties of measure-variable functions, and the differentiation of such functions.

MSC2020 subject classifications: 60B10, 60F05, 60G57, 60J25, 60J60; secondary 46G05, 60B05, 60K35.
Keywords and phrases: Measure-valued Markov process, Mean-field limit, McKean-Vlasov process, Particle system, Propagation of chaos.

Contents

Introduction

The notion of particle systems is widely studied in many frameworks due to the diversity of their applications: biology (e.g. neural networks Fournier and Löcherbach (2016), genomics Reynaud-Bouret and Schbath (2010), ecological networks Billiard et al. (2022)), finance (e.g. order books Lu and Abergel (2018), high frequency data Bauwens and Hautsch (2009)), sociology (e.g. social networks Mitchell and Cates (2009)), physics (e.g. kinetic theory Vedenyapin, Sinitsyn and Dulov (2011)). A particular case of interest concerns mean-field particle systems, meaning systems in which the particles interact with all the others in a similar way. The reason behind this interest is that, the interactions and the dynamics of such systems can be somehow characterized by their empirical measures. The advantage of studying the empirical measure rather than the particle system itself is that, if the dynamics of the particles are characterized by $\mathbb{R}$-valued processes, then the values taken by the empirical measure belong to the set of probability measures on $\mathbb{R}$, which does not depend on the number of particles of the system. In this paper, we study large scale limit behavior, as the number of particles goes to infinity, so it is preferable to work on spaces that are independent of the size of the systems. In particular, the study of mean-field particle systems is closely related to the notion of measure-valued Markov processes.

The study of measure-valued Markov processes goes back at least to Fleming and Viot (1979) to model population genetics. In the seminal course Dawson (1993), Donald Dawson has introduced a notion of differentiation for measure-variable functions, in order to characterize infinitesimal generators and martingale problems in the context of measure-valued Markov processes. A variant of the notion of the derivative introduced by Donald Dawson, has been defined by René Carmona and François Delarue (see e.g. section 5.4.1 of Carmona and Delarue (2018)) and is referred to as the *linear derivative*, and another rather different notion (at first sight) by Pierre-Louis Lions in a course given at Collège de France (see, for example, the lecture notes of Pierre Cardaliaguet Cardaliaguet (2013)). Since then, this topic still draws interest: for instance, recently, Guo, Pham and Wei (2023) and Cox et al. (2021) have established Ito's formulas for measure-valued semimartingales in different frameworks, and the authors of Crisan and McMurray (2018) have proved some regularity properties for stochastic flows of McKean-Vlasov processes. In this paper, we mostly use a notion of *linear derivative* and state precisely the definition at Section 2.1.

There are two main contributions of this paper, that are closely related. The first one concerns the convergence speed for large-scale limits of mean-field particle systems, and aims to improve it for the convergence in distribution. Indeed, it is now of common knowledge that the L^p-convergence speed of N-particle systems is of order $N^{-1/2}$ in Law of Large Numbers regime (see e.g. Theorem 1.4 of Sznitman (1991), or more recently in a more general framework, Proposition 3.1 of Andreis, Dai Pra and Fischer (2018)), whereas, to the best of our knowledge, the fact that the convergence speed in distribution is of order N^{-1} (see Theorem 1.3) is new, even in the simple model studied in Corollary 1.5. In addition, few results exist for the convergence in Central Limit Theorem regime (e.g. Erny, Löcherbach and Loukianova (2021), and, in the more general framework of martingale measures, Erny, Löcherbach and Loukianova (2022a)), and, as far as we know, the only rate of convergence for this kind of model has been obtained in Erny, Löcherbach and Loukianova (2023) for a strong convergence using coupling techniques based on the approximation theorem of Komlós,

Major and Tusnády (1976). This convergence speed is very slow: $(\ln N)^{1/5}/N^{1/10}$. Once again, we prove that the convergence speed in distribution is the expected one, in $N^{-1/2}$ (see Theorem 1.7), and this result is even knew in simple frameworks, like the one of Corollary 1.9.

The second main contribution is the study of some analytical properties of measure-variable functions, and of some Markovian properties of Measure-valued processes. Note that, since the infinitesimal generator is a differential operator, it is required to study a notion of differentiation for measure-variable functions in order to study measure-valued Markov processes. It is the reason why we need to study analytical properties of measure-variable functions to prove our main results (i.e. Theorems 1.3 and 1.7). Most of these properties are not completely new, but are studied further in this paper so that they can be used to obtain the aforementioned rates of convergence. For instance, we prove Taylor-Lagrange's inequality of any order for measure-variable functions, and we use it up to order three (the only related known result in this context is the mean value theorem). This is not as common as it seems since the notion of measure-variable derivative is usually limited to order two because it is sufficient to state martingale problems and express infinitesimal generators (let us mention Cuchiero, Larsson and Svaluto-Ferro (2019) that also studied analytical properties of measure-variable polynomials with derivatives of any order). Concerning the Markovian properties, we prove a Trotter-Kato's formula, and give an explicit expression for the generators of some measure-valued processes. The expressions of the generators extend known results: it is similar to Guo, Pham and Wei (2023) and Cox et al. (2021) that have proved Ito's formula for this kind of processes. However, there are some differences: compared to Cox et al. (2021), the framework is not the same and the dynamics sufficiently differs for the expressions of the generators to have a different form. Our model is closer to the framework of Guo, Pham and Wei (2023) (restricting our study to McKean-Vlasov processes, whereas in Guo, Pham and Wei (2023) general semimartingales are considered), but we somehow generalize their results for two reasons: the first one is that they only consider measure-valued processes that are deterministic processes defined as laws of semimartingales, whereas we obtain a generator for stochastic processes defined as conditional laws of McKean-Vlasov processes given one of the Brownian motion of the SDEs (as a consequence, the expressions of our generators involve an additional term corresponding to the second order linear derivative of the test-function, which is the only property closer to the model of Cox et al. (2021)). The second reason is that we also give the expressions of generators of measure-valued processes defined as empirical measures of particle systems with jumps (this framework and its technical difficulties are mentioned in Remark 3.15 of Guo, Pham and Wei (2023)). In particular, to the best of our knowledge, obtaining Ito's formula or generator for a measure-valued processes with such general jumps is new. The expressions of these generators are given at Theorems 3.6 for the conditional laws of McKean-Vlasov processes, and 3.10 for the empirical measures of McKean-Vlasov particle systems.

Organization of the paper. The next section introduces the notation that are used throughout the paper. Section 1 is dedicated to the statement of the results about the convergence speed in distribution for mean-field particle systems. The section itself is split into two subsections: Section 1.1 for the Law of Large Numbers regime and Section 1.2 for the Central Limit Theorem regime. In each subsection, a general statement is given as a theorem, and a particular case more understandable requiring less technical assumptions is also given as corollary (the results of the corollaries are still new in the literature). The goal of Section 2 is to study some analytical properties of measure-variable functions: in particular, to prove Taylor-Lagrange's inequality, and to study a notion of measure-variable polynomials. Section 3 aims to study some Markovian properties of measure-valued processes: establishing Trotter-Kato's formula, giving an expression for the genera-

tors of such processes (in the cases of conditional law of diffusions, and empirical measures of particle systems), and studying the regularity of their semigroups (which is required to use Trotter-Kato's formula jointly with Taylor-Lagrange's inequality). The proofs of the results stated at Section 1 are given at Section 4. Finally, the Appendix gathers some technical results and technical proofs in order to ease the reading of the paper.

Notation and convention

In the notation below, (E, d) is always a Polish space (or just a set if the notation does not need a particular structure on E).

- For $x = (x_1, ..., x_n) \in \mathbb{R}^n$, we denote $||x||_1$ the classical L^1-norm:
$$||x||_1 = \sum_{k=1}^{n} |x_k|.$$
- $\mathcal{P}(E)$ is the set of probability measures on E, endowed with the topology of Prohorov metric (i.e. the topology of the weak convergence).
- If X is an E-valued random variable, $\mathcal{L}(X) \in \mathcal{P}(E)$ denote its law.
- For $p \geq 1$, $\mathcal{P}_p(E)$ is the set of probability measures on E with a finite p-th order moment, and $\mathcal{P}_\infty(E)$ is the intersection of every $\mathcal{P}_p(E)$ over all $p \geq 1$. Each $\mathcal{P}_p$ is endowed with the p-th order Wasserstein metric.
- For $p \geq 1$ and $m, \mu \in \mathcal{P}_p(E)$, the p-th order Wasserstein metric between m and μ is defined as the best L^p-coupling between m and μ:
$$W_p(m, \mu) = \inf_{\mathcal{L}(X)=m, \mathcal{L}(Y)=\mu} \mathbb{E}\left[d(X, Y)^p\right]^{1/p}.$$
- For $m, \mu \in \mathcal{P}_1(E)$, $D_{KR}(m, \mu)$ denotes the Kantorovich-Rubinstein metric between m and μ:
$$D_{KR}(m, \mu) = \sup_{f \in \mathrm{Lip}_1} \int_E f(x)\, d\,(m - \mu)\,(x),$$
where Lip_1 denotes the set of Lipschitz continuous functions $f : E \to \mathbb{R}$ such that, for all $x, y \in E$, $|f(x) - f(y)| \leq d(x, y)$. By the Kantorovich-Rubinstein duality (e.g. Remark 6.5 of Villani (2009)), $D_{KR} = W_1$. In the paper, we mostly use the characterization of D_{KR} and its notation rather than W_1.
- Note that each $\mathcal{P}_p(\mathbb{R})$ is Polish by Theorem 6.18 of Villani (2009). For $M_1, M_2 \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}))$, we denote specifically by $\mathcal{D}_{KR}(M_1, M_2)$ the Kantorovich-Rubinstein metric between M_1 and M_2.
- A function $f : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^n \to \mathbb{R}$ is said to be Lipschitz continuous if it is Lipschitz continuous w.r.t. both variables together: there exists $C > 0$ such that, for all $m_1, m_2 \in \mathcal{P}_1(\mathbb{R})$ and $x_1, x_2 \in \mathbb{R}^n$,
$$|f(m_1, x_1) - f(m_2, x_2)| \leq C\left(||x_1 - x_2||_1 + D_{KR}(m_1, m_2)\right).$$
- A function $f : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^n \to \mathbb{R}$ is said to be sublinear if: there exists $C > 0$ such that, for all $m \in \mathcal{P}_1(\mathbb{R})$ and $x \in \mathbb{R}^n$,
$$|f(m, x)| \leq C\left(1 + ||x||_1 + \int_{\mathbb{R}} |y| dm(y)\right).$$

- $C(E)$ is the set of continuous functions $f : E \to \mathbb{R}$.
- For $k \in \mathbb{N} \cup \{\infty\}$ and $n \in \mathbb{N}^*$, $C^k(\mathbb{R}^n)$ denotes the space of functions $f : \mathbb{R}^n \to \mathbb{R}$ of class C^k.
- For $f \in C^k(\mathbb{R}^n)$ and α a multi-index belonging to $\mathbb{N}^n$, the size of α is

$$|\alpha| = \sum_{i=1}^{n} \alpha_i,$$

  and, if $|\alpha| \leq k$, $\partial_\alpha f$ denotes the derivative of f w.r.t. α.
- When no doubt is possible (and to make some expressions more readable), we use usual notation for partial derivatives, for example, if $g \in C^3(\mathbb{R}^2)$, for all $x, y \in \mathbb{R}$,

$$\partial^3_{xxy} g(y, x) = \partial_{(1,2)} g(y, x) \text{ and } \left(\partial^2_{y_1 y_2} g(y_1, y_2)\right)_{|y_1 = y_2 = x} = \partial_{(1,1)} g(x, x),$$

  and the same notation replacing the derived variables by their indices:

$$\partial^3_{xxy} g(y, x) = \partial^3_{2,2,1} g(y, x) = \partial_{(1,2)} g(y, x).$$

- For a function $f : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ and $m \in \mathcal{P}_1(\mathbb{R}), x \in \mathbb{R}$, we use the notation

$$\partial_{(0,1)} f(m, x) = \partial_x f(m, x) \text{ and } \partial_{(0,2)} f(m, x) = \partial^2_{xx} f(m, x),$$

  but, in this context where f depends on variables of different nature (i.e. belonging to $\mathcal{P}_1(\mathbb{R})$ and $\mathbb{R}$) the index of the measure-variable is always zero. For technical reasons, the meaning would not be clear otherwise in general (see Corollary 2.25 for a statement where it makes sense). However, to state one of our assumptions, we need to introduce a notion of mixed derivatives at Definition 2.11 related to this remark.
- $C^k_b(\mathbb{R}^n)$ is the subspace of $C^k(\mathbb{R}^n)$ containing the functions f such that: for all multi-index α of size non-greater than k,

$$\sup_{x \in \mathbb{R}^n} |\partial_\alpha f(x)| < \infty.$$

- $C^k_c(\mathbb{R}^n)$ is the subspace of $C^k_b(\mathbb{R}^n)$ containing the compactly supported functions.
- For $f \in C^k_b(\mathbb{R}^n)$, we denote

$$||f||_k = \sum_{|\alpha| \leq k} \sup_{x \in \mathbb{R}^n} |\partial_\alpha f(x)|.$$

- A tuple $x \in E^n$ is always indexed from 1 to n: $x = (x_1, ..., x_n)$,
- For $x \in E^n$ and $1 \leq k \leq n$, $x\backslash_k$ denotes the tuple x without the index k:

$$x\backslash_k = (x_1, ..., x_{k-1}, x_{k+1}, ..., x_n) \in E^{n-1},$$

  and, for $y \in E$, $(x\backslash_k y)$ denotes the tuple x where x_k is replaced by y:

$$x\backslash_k y = (x_1, ..., x_{k-1}, y, x_{k+1}, ..., x_n) \in E^n.$$

  In addition, for $k \neq l$, and $y_1, y_2 \in E$, we use the notation $x\backslash_{(k,l)}$ for x without the indices k and l, and $(x\backslash_{(k,l)}(y_1, y_2))$ for the tuple x where x_k is replaced by y_1 and x_l by y_2.

For example, this notation is used in the following context: if $f \in C_b^0(\mathbb{R}^3)$, $m \in \mathcal{P}(\mathbb{R})$ and $y \in \mathbb{R}$,

$$\sum_{k=1}^{3} \int_{\mathbb{R}^2} f(x\backslash_k y) dm^{\otimes 2}(x\backslash_k) = \int_{\mathbb{R}^2} f(y, x_2, x_3) dm^{\otimes 2}(x_2, x_3) + \int_{\mathbb{R}^2} f(x_1, y, x_3) dm^{\otimes 2}(x_1, x_3) + \int_{\mathbb{R}^2} f(x_1, x_2, y) dm^{\otimes 2}(x_1, x_2).$$

- For $n \in \mathbb{N}^*$ and $F : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathbb{R}^n$, we denote by F_x the following function
$$F_x : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto F_x(m) = F(m, x).$$
- For $m \in \mathcal{P}(\mathbb{R})$ and $\lambda \in \mathbb{R}$, we denote by $\mathcal{S}(m, \lambda)$ the probability measure m shifted by λ:
$$\mathcal{S}(m, \lambda) : A \in \mathcal{B}(\mathbb{R}) \longmapsto m\left(\{x - \lambda : x \in A\}\right).$$
- We use C to denote any arbitrary positive constant. The value of C can change from line to line inside an equation. In addition, if C depends on non-model parameters θ, we denote it by C_θ.

1. Main results on the convergence speed of mean-field models

Our main results consist in obtaining new explicit convergence speeds in distribution for well-known mean-field particle systems. Section 1.1 treats the case of the linear regime (i.e. Law of Large Numbers regime) where the interaction strength for the N-particle system is of order N^{-1}. Section 1.2 handles the diffusive regime (i.e. Central Limit Theorem regime) with centered interaction strength of order $N^{-1/2}$. As it can be expected, the convergence speed is N^{-1} for the linear regime, and $N^{-1/2}$ for the diffusive one.

In each of both following subsections, a theorem is given for a general model, and a corollary is stated for a simpler model. The assumptions of the theorems require some technical definitions given at Section 2 to be stated formally, which is not the case for the hypotheses of the corollaries. Despite concerning well-known models, which can be considered elementary, the results of the corollaries are new.

1.1. Mean-field model in linear regime

In this section, we introduce for any $N \in \mathbb{N}^*$, the processes $X^{N,k}$ $(1 \leq k \leq N)$ solutions to

$$\begin{aligned} dX_t^{N,k} =& b\left(\mu_t^N, X_t^{N,k}\right) dt + \sigma\left(\mu_t^N, X_t^{N,k}\right) dB_t^k + \varsigma\left(\mu_t^N, X_t^{N,k}\right) dW_t \\ &+ \frac{1}{N} \sum_{l=1}^{N} \int_{\mathbb{R}_+} h\left(\mu_{t-}^N, X_{t-}^{N,l}\right) \mathbb{1}_{\left\{z \leq f\left(\mu_{t-}^N, X_{t-}^{N,l}\right)\right\}} d\pi^l(t, z), \end{aligned} \tag{1}$$

where B^k $(k \geq 1)$ and W are standard Brownian motions of dimension one, π^l $(l \geq 1)$ are Poisson measures on $\mathbb{R}_+^2$ with Lebesgue intensity, and $X_0^{N,k}$ $(1 \leq k \leq N)$ are i.i.d. such that W, B^k $(k \geq 1)$,

π^l $(l \geq 1)$ and $X_0^{N,k}$ $(1 \leq k \leq N)$ are mutually independent, and

$$\mu_t^N = \frac{1}{N} \sum_{k=1}^{N} \delta_{X_t^{N,k}}.$$

The main result of this section concerns the convergence of the particle systems above to the limit process $\bar{X}$ solution to the SDE below as N goes to infinity.

$$d\bar{X}_t = b\left(\bar{\mu}_t, \bar{X}_t\right) dt + \sigma\left(\bar{\mu}_t, \bar{X}_t\right) dB_t + \varsigma\left(\bar{\mu}_t, \bar{X}_t\right) dW_t + \int_{\mathbb{R}} h\left(\bar{\mu}_t, x\right) f\left(\bar{\mu}_t, x\right) d\bar{\mu}_t(x) dt, \tag{2}$$

where B, W are independent standard Brownian motions of dimension one, $\bar{\mu}_0$ is $\mathcal{P}_1(\mathbb{R})$-valued random variable (so the dynamics in (2) is defined conditionally on $\bar{\mu}_0$) independent of the Brownian motions such that, conditionally on $\bar{\mu}_0$, $\bar{X}_0$ is $\bar{\mu}_0$-distributed, and

$$\bar{\mu}_t = \mathcal{L}\left(\bar{X}_t | \mathcal{W}_t\right),$$

with $\mathcal{W}$ the filtration of W.

The reason of the technical phrasing concerning the initial condition of the SDE (2) is explained formally in Remark 1.4. Roughly speaking, if the probability measure $\bar{\mu}_0$ was deterministic and not a Dirac mass, then, the rate of convergence in distribution in Theorem 1.3 would be at best $N^{-1/2}$, what would make the result useless compared to the existing literature (see Remark 1.2 below).

Note that, since our main result handles convergence in distribution, the process $\bar{X}$ does not need to be defined on the same space as the particle systems X^N, and the Brownian motions may be different or not w.l.o.g.

For the previous equations to be well-posed, we assume the following.

Assumption 1. *For any $m_1, m_2 \in \mathcal{P}_1(\mathbb{R})$, and $x_1, x_2 \in \mathbb{R}$,*

$$|b(m_1, x_1) - b(m_2, x_2)| + |\sigma(m_1, x_1) - \sigma(m_2, x_2)| + |\varsigma(m_1, x_1) - \varsigma(m_2, x_2)| \\ \leq C\left(|x_1 - x_2| + D_{KR}(m_1, m_2)\right),$$

and

$$\int_0^{+\infty} \left|h(m_1, x_1)\mathbb{1}_{\{z \leq f(m_1, x_1)\}} - h(m_2, x_2)\mathbb{1}_{\{z \leq f(m_2, x_2)\}}\right| dz \\ \leq C\left(|x_1 - x_2| + D_{KR}(m_1, m_2)\right).$$

In addition, we also suppose that, for any $p \in \mathbb{N}^*$, there exists some $C_p > 0$ such that, for all $m \in \mathcal{P}_1(\mathbb{R}), x \in \mathbb{R}$,

$$\left|(x + h(m, x))^p - x^p\right| f(m, x) \leq C_p \left(1 + |x|^p + \int_{\mathbb{R}} |y|^p dm(y)\right).$$

Remark 1.1. *The last condition of Assumption 1 is guaranteed if, for example, one of the two following properties hold:*

- *the function h is sublinear and the function f is bounded,*
- *or the function h is bounded and the function f is sublinear.*

We also make the following hypothesis concerning the initial conditions.

Assumption 2. *For every $N \in \mathbb{N}^*$, the variables $X_0^{N,k}$ $(1 \leq k \leq N)$ are i.i.d.*

All the initial conditions $\bar{X}_0$ and $X_0^{N,k}$ $(N \in \mathbb{N}^*$ and $1 \leq k \leq N)$ have a finite fourth order moment, uniformly bounded w.r.t. N:

$$\mathbb{E}\left[\int_{\mathbb{R}} |x|^4 d\bar{\mu}_0(x)\right] < \infty \quad \text{and} \quad \sup_{N\in\mathbb{N}^*} \mathbb{E}\left[\left|X_0^{N,1}\right|^4\right] < \infty.$$

Under Assumptions 1 and 2, it is known that the SDEs (1) and (2) are well-posed (see e.g. Theorem 2.1 of Graham (1992)) in the following sense: there exist unique strong solutions $(X^{N,k})_k$ and $\bar{X}$ of the respective SDEs that satisfy, for all $T > 0$, $1 \leq k \leq N$,

$$\int_{\mathcal{P}_1(\mathbb{R})} \mathbb{E}_m\left[\sup_{t\leq T} \left|\bar{X}_t\right|^4\right] d\left(\mathcal{L}\left(\bar{\mu}_0\right)\right)(m) < \infty \quad \text{and} \quad \sup_{N\in\mathbb{N}^*} \mathbb{E}\left[\sup_{t\leq T} \left|X_t^{N,k}\right|^4\right] < \infty, \tag{3}$$

with $\mathbb{E}_m$ the expectation under which the law of $\bar{X}_0$ is m.

Note that, a priori, uniqueness only holds for processes that satisfy the above controls, which are not a priori estimates. The controls (3) can be guaranteed on the solutions that are built via Banach-Picard schemes for example (see the *Step 1* of the proof of Lemma 3.8 at Appendix B for the SDE (2)).

Remark 1.2. *It is now classical (see e.g. Proposition 3.1 of Andreis, Dai Pra and Fischer (2018)) that, if we define $\bar{X}_k$ as the solution of (2) w.r.t. the same Brownian motion W as in (1) and B^k instead of B, then, the particle systems $(X^{N,k})_k$ converges in L^1 to the limit system $(\bar{X}^k)_k$ in the following sense: for all $T > 0$, there exists $C_T > 0$ such that, for all $1 \leq k \leq N$,*

$$\mathbb{E}\left[\sup_{t\leq T} \left|X_t^{N,k} - \bar{X}_t^k\right|\right] \leq C_T N^{-1/2}.$$

In particular, this result can be used to prove the convergence in distribution of μ^N to $\bar{\mu}$, but it cannot allow to prove that the convergence speed is N^{-1}.

Our first main result requires an additional technical assumption. This assumption relies on the notion of differentiability of measure-variable functions as it is defined in Section 2.1. Corollary 1.5 is a particular case of Theorem 1.3 that does not need this technical assumption.

Assumption 3. *The functions $b, \sigma, \varsigma, \beta$ and $f \cdot h$ admit fifth order mixed derivatives (in the sense of Definition 2.11) and all their mixed derivatives of orders from one to five are bounded. In addition, if $g : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ is any of the previous functions, then, there exists $C > 0$ such that for all $x \in \mathbb{R}$, the function g_x belongs to $C_b^2(\mathcal{P}_1(\mathbb{R}))$ (the set is introduced in Definition 2.8), and*

$$\sup_{\substack{m\in\mathcal{P}_1(\mathbb{R})\\ y_1,y_2\in\mathbb{R}}} \left|\partial^2_{y_1y_2}\delta^2(g_x)(m, y_1, y_2)\right| \leq C(1 + |x|).$$

Theorem 1.3. *Grant Assumptions 1, 2 and 3. Then, for all $T > 0$, there exists $C_T > 0$ such that, for any measure-variable polynomial G of order four (in the sense of Definition 2.30), for all $N \in \mathbb{N}^*$,*

$$\sup_{t\leq T} \left|\mathbb{E}\left[G\left(\mu_t^N\right)\right] - \mathbb{E}\left[G\left(\bar{\mu}_t\right)\right]\right| \leq C_T ||G||_2 \left(\mathcal{D}_{KR}\left(\mathcal{L}\left(\frac{1}{N}\sum_{k=1}^N \delta_{X_0^{N,k}}\right), \mathcal{L}(\bar{\mu}_0)\right) + \frac{1}{N}\right),$$

where $||G||_2$ is a bound of the mixed derivatives of G up to order two (defined at Definition 2.14).

Remark 1.4. *Let us explain precisely why considering a standard phrasing for the initial condition of (2) would make Theorem 1.3 useless. Usually, one would have considered an initial condition $\bar{X}_0$ for the SDE (2) with deterministic law $\bar{\mu}_0$. However, if $\bar{\mu}_0$ was a deterministic element of $\mathcal{P}_1(\mathbb{R})$, then, we would have*

$$\mathcal{L}(\bar{\mu}_0) = \delta_{\bar{\mu}_0},$$

which would imply

$$\mathcal{D}_{KR}\left(\mathcal{L}\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{X_0^{N,k}}\right), \mathcal{L}(\bar{\mu}_0)\right) = \mathbb{E}\left[D_{KR}\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{X_0^{N,k}}, \bar{\mu}_0\right)\right].$$

Then, by Theorem 3.1 of Bobkov and Ledoux (2019), the rate of convergence of the quantity above would necessarily be, at best, $N^{-1/2}$ (except in the trivial case, where $\mathcal{L}(\bar{X}_0) = \mathcal{L}(X_0^{N,k}) = \delta_{x_0}$ for some $x_0 \in \mathbb{R}$, where the quantity would be zero), which would make the convergence speed of Theorem 1.3 not better than what can be proved with existing techniques.

As mentioned previously, let us study a particular case of this model and state a corollary of Theorem 1.3 that can be understood without the technical definitions of Section 2. So let us introduce the following particle systems: for $N \in \mathbb{N}^*$ and $1 \leq k \leq N$,

$$dY_t^{N,k} = \tilde{b}\left(Y_t^{N,k}\right)dt + \tilde{\sigma}\left(Y_t^{N,k}\right)dB_t^k + \frac{1}{N}\sum_{l=1}^{N}\int_{\mathbb{R}_+}\mathbb{1}_{\left\{z\leq\tilde{f}\left(X_{t-}^{N,l}\right)\right\}}d\pi^l(t,z), \tag{4}$$

where B^k $(k \geq 1)$ are standard Brownian motions, π^l $(l \geq 1)$ are Poisson measures with Lebesgue intensity, such that the Brownian motions and the Poisson measures are independent, and $Y_0^{N,k} = x_0$ $(1 \leq k \leq N)$ for some deterministic $x_0 \in \mathbb{R}$.

And let us define the following limit process

$$d\bar{Y}_t = \tilde{b}\left(\bar{Y}_t\right)dt + \tilde{\sigma}\left(\bar{Y}_t\right)dB_t + \mathbb{E}\left[\tilde{f}(\bar{Y}_t)\right]dt, \tag{5}$$

with B a standard Brownian motion and $\bar{Y}_0 = x_0$.

Corollary 1.5. *Assume that the functions $\tilde{b}, \tilde{\sigma}$ and $\tilde{f}$ are C^5 and that their derivatives of orders from one to five are bounded. Then, for all $T > 0$, there exists $C_T > 0$ such that, for any $\varphi \in C_b^4(\mathbb{R})$, for all $N \in \mathbb{N}^*$,*

$$\sup_{t\leq T}\left|\mathbb{E}\left[\varphi\left(Y_t^{N,1}\right)\right] - \mathbb{E}\left[\varphi\left(\bar{Y}_t\right)\right]\right| \leq C_T\left(||\varphi||_\infty + ||\varphi'||_\infty + ||\varphi''||_\infty\right)N^{-1}.$$

1.2. Mean-field model in diffusive regime

This section is similar as the previous one, for a different type of model. In the N-particle system (1), the interaction strength is N^{-1}, whereas, in the model of this section, it is of order $N^{-1/2}$. For this kind of model to converge as the number of particles tends to infinity, the interactions between

particles need to be centered. So, the dynamics depends on a probability measure called ν, that satisfies

$$\int_{\mathbb{R}} u\, d\nu(u) = 0 \quad \text{and} \quad \int_{\mathbb{R}} |u|^3 d\nu(u) < \infty.$$

Formally, for any $N \in \mathbb{N}^*$, we introduce $X^{N,k}$ $(1 \leq k \leq N)$ as solutions to

$$\begin{aligned} dX_t^{N,k} =& b\left(\mu_t^N, X_t^{N,k}\right) dt + \sigma\left(\mu_t^N, X_t^{N,k}\right) dB_t^k \\ &+ \frac{1}{\sqrt{N}} \sum_{l=1}^{N} \int_{\mathbb{R}_+ \times \mathbb{R}} u \cdot h\left(\mu_{t-}^N, X_{t-}^{N,l}\right) \mathbb{1}_{\left\{z \leq f\left(\mu_{t-}^N, X_{t-}^{N,l}\right)\right\}} d\pi^l(t,z,u), \end{aligned} \tag{6}$$

where B^k $(k \geq 1)$ are standard Brownian motions of dimension one, π^l $(l \geq 1)$ are Poisson measures on $\mathbb{R}_+^2 \times \mathbb{R}$ with intensity $dt \cdot dz \cdot d\nu(u)$, and $X_0^{N,k}$ $(1 \leq k \leq N)$ are i.i.d. such that B^k $(k \geq 1)$, π^l $(l \geq 1)$ and $X_0^{N,k}$ $(1 \leq k \leq N)$ are mutually independent, and

$$\mu_t^N = \frac{1}{N} \sum_{k=1}^{N} \delta_{X_t^{N,k}}.$$

The main result of this section states the rate of convergence in distribution of the N-particle system (6) to the limit process $\bar{X}$ solution to the following SDE:

$$\begin{aligned} d\bar{X}_t =& b\left(\bar{\mu}_t, \bar{X}_t\right) dt + \sigma\left(\bar{\mu}_t, \bar{X}_t\right) dB_t \\ &+ \left(\int_{\mathbb{R}} u^2 d\nu(u)\right)^{1/2} \left(\int_{\mathbb{R}} h(\bar{\mu}_t, x)^2 f(\bar{\mu}_t, x) d\bar{\mu}_t(x)\right)^{1/2} dW_t, \end{aligned} \tag{7}$$

where B, W are standard Brownian motions of dimension one, $\bar{\mu}_0$ is $\mathcal{P}_1(\mathbb{R})$-valued random variable such that $B, W, \bar{\mu}_0$ are independent, conditionally on $\bar{\mu}_0$, $\bar{X}_0$ is $\bar{\mu}_0$-distributed, and

$$\bar{\mu}_t = \mathcal{L}\left(\bar{X}_t | \mathcal{W}_t\right),$$

with $\mathcal{W}$ the filtration of W.

Remark 1.6. *The convergence in distribution of μ^N to $\bar{\mu}$ has already been proved in Erny, Löcherbach and Loukianova (2022a) in a more general framework (cf Theorem 2.6 for the general case, and Example 1 for a model closer to the one of this section). But the technics used in the proofs did not allow to obtain an explicit convergence speed.*

For the main result in the diffusive regime, we need the following assumption that slightly differs from Assumption 3.

Assumption 4. *the functions b, σ and $m \mapsto \left(\int_{\mathbb{R}} h(m,x)^2 f(m,x) dm(x)\right)^{1/2}$ admit fifth order mixed derivatives (in the sense of Definition 2.11) and all their mixed derivatives of orders from one to five are bounded. In addition, if $g : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ is any of the previous functions, then, there exists $C > 0$ such that for all $x \in \mathbb{R}$, the function g_x belongs to $C_b^2(\mathcal{P}_1(\mathbb{R}))$ (the set is introduced in Definition 2.8), and*

$$\sup_{\substack{m \in \mathcal{P}_1(\mathbb{R}) \\ y_1, y_2 \in \mathbb{R}}} \left|\partial_{y_1 y_2}^2 \delta^2(g_x)(m, y_1, y_2)\right| \leq C(1 + |x|).$$

As previously, a particular case (see Corollary 1.9) of the theorem below is given to avoid the use of the technical assumption above.

Theorem 1.7. *Grant Assumptions 1, 2 and 4. Assume that the probability measure ν is centered with a finite third moment. Then, for all $T > 0$, there exists $C_T > 0$ such that, for any measure-variable polynomial G of order four (in the sense of Definition 2.30), for all $N \in \mathbb{N}^*$,*

$$\sup_{t\leq T}\left|\mathbb{E}\left[G\left(\mu_t^N\right)\right]-\mathbb{E}\left[G\left(\bar{\mu}_t\right)\right]\right| \leq C_T||G||_3\left(\mathcal{D}_{KR}\left(\mathcal{L}\left(\frac{1}{N}\sum_{k=1}^N \delta_{X_0^{N,k}}\right),\mathcal{L}(\bar{\mu}_0)\right)+\frac{1}{\sqrt{N}}\right),$$

where $||G||_3$ is a bound of the mixed derivatives of G up to order three (defined at Definition 2.14).

Remark 1.8. *Contrary to the situation in the linear regime, for the diffusive regime, it is possible to state the dynamics of (7) as usual concerning the initial condition (i.e. with $\bar{\mu}_0 = \mathcal{L}(\bar{X}_0)$ deterministic). Indeed, in Theorem 1.7, the rate of convergence is $N^{-1/2}$ because it corresponds to a Central Limit Theorem regime. So, the constraint about the convergence speed of the initial condition (i.e. the fact that the term $\mathcal{D}_{KR}$ is of order $N^{-1/2}$), does not weaken our result. This is the reason why, in the simple model below, it is not necessary to consider deterministic initial conditions $\bar{Y}_0$ and $Y_0^{N,k}$ $(1 \leq k \leq N)$, as it was done in Corollary 1.5.*

Now, let us give, as a corollary, a consequence of Theorem 1.7 that is easier to state. So let us introduce: for $N \in \mathbb{N}^*$ and $1 \leq k \leq N$,

$$dY_t^{N,k} = \tilde{b}\left(Y_t^{N,k}\right)dt + \tilde{\sigma}\left(Y_t^{N,k}\right)dB_t^k + \frac{1}{\sqrt{N}}\sum_{l=1}^N \int_{\mathbb{R}_+\times\mathbb{R}} u\cdot \mathbb{1}_{\left\{z\leq \tilde{f}\left(X_{t-}^{N,l}\right)\right\}} d\pi^l(t,z,u), \tag{8}$$

where B^k $(k \geq 1)$ are standard Brownian motions, π^l $(l \geq 1)$ are Poisson measures with intensity $dt\cdot dz\cdot d\nu(u)$ and, the variables $Y_0^{N,k}$ $(1 \leq k \leq N)$ are i.i.d. with finite p-th order moment for all $p \in \mathbb{N}^*$. The Brownian motions, Poisson measures and initial conditions are assumed to be independent.

And let $\bar{Y}$ be solution of

$$d\bar{Y}_t = \tilde{b}\left(\bar{Y}_t\right)dt + \tilde{\sigma}\left(\bar{Y}_t\right)dB_t + \left(\int_{\mathbb{R}} u^2 d\nu(u)\right)^{1/2}\sqrt{\mathbb{E}\left[\tilde{f}(\bar{Y}_t)\middle|\mathcal{W}_t\right]}dW_t, \tag{9}$$

with B, W independent standard Brownian motions, $\mathcal{W}$ the filtration of W, and $\bar{Y}_0$ having the same law as the variables $Y_0^{N,k}$. The Brownian motions B, W are assumed to be independent of $\bar{Y}_0$.

Corollary 1.9. *Assume that the functions $\tilde{b}, \tilde{\sigma}$ and $\tilde{f}$ are C^5 and that their derivatives of orders from one to five are bounded. In addition, we assume that $\tilde{f}$ is bounded and lower-bounded by some positive constants. We also assume that ν is centered with a finite third order moment. Then, for any $\varphi \in C_b^4(\mathbb{R})$, for all $T > 0$, there exists $C_{T,\varphi} > 0$ such that, for all $N \in \mathbb{N}^*$,*

$$\sup_{t\leq T}\left|\mathbb{E}\left[\varphi\left(Y_t^{N,1}\right)\right]-\mathbb{E}\left[\varphi\left(\bar{Y}_t\right)\right]\right| \leq C_T\left(||\varphi||_\infty+||\varphi'||_\infty+||\varphi''||_\infty+||\varphi'''||_\infty\right)N^{-1/2}.$$

2. Analysis of functions defined on spaces of probability measures

2.1. Definitions and some elementary properties

The following notion of differentiability is a generalization of the one introduced in Dawson (1993), and often referred to as the *linear derivative*.

Definition 2.1. *For $m_0 \in \mathcal{P}_1(\mathbb{R})$, a function $F : \mathcal{P}_1(\mathbb{R}) \longrightarrow \mathbb{R}$ is said to be differentiable at m_0 if there exists a measurable sublinear function $H_{m_0} : \mathbb{R} \to \mathbb{R}$ such that, for all $m \in \mathcal{P}_1(\mathbb{R})$,*

$$F(m) - F(m_0) = \int_{\mathbb{R}} H_{m_0}(x)\, d(m - m_0)(x) + \varepsilon_{m_0}(m), \tag{10}$$

where $\varepsilon_{m_0}(m)/D_{KR}(m, m_0)$ vanishes as m converges to m_0 for the metric D_{KR}. The function H_{m_0} is called a version of the derivative of F at m_0. We define the canonical derivative of F at m_0 as the only version of the derivative that satisfies

$$\int_{\mathbb{R}} H_{m_0}(x)\, dm_0(x) = 0,$$

and denote it by, for all $x \in \mathbb{R}$,

$$\delta F(m_0, x) = H_{m_0}(x).$$

In the rest of the paper, we may omit "canonical" and "linear" and just mention "the derivative". For most of the properties that are studied in this paper, the canonical version of the derivative is not particularly needed, but it allows the definitions to be unquestionably well-posed, particularly for the notion of many times differentiable. Note that, the set $C_b^n(\mathcal{P}_1(\mathbb{R}))$ (see Definition 2.8) depends on the version of the derivative, and so Assumptions 3 and 4 also depend on our choice of canonical derivative. However, for example, the expressions of the infinitesimal generators given at Theorems 3.6 and 3.10 do not depend on this choice.

This particular choice of canonical derivative is motivated by the fact that it allows to recover "natural properties" (see Lemma 2.4), and it is close to the notion of derivative that is used in Dawson (1993). Let us remark that it is also the convention that was used by François Delarue in his course at the Summer School "Mean Field Models" organized by Centre Henri Lebesgue at Rennes in 2023.

Remark 2.2. *It is easy to see that the versions of the derivative of F are never unique when they exist: if some function $H_{m_0} : \mathbb{R} \to \mathbb{R}$ satisfies (10), then any function of the form $x \mapsto H_{m_0}(x) + G(m_0)$ also satisfies (10). It can be noted (cf Lemma 2.3 below) that this notion of derivative is still unique for the following equivalence relation : two functions $H_1, H_2 : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ are equivalent if for all m, x, $H_1(m, x) - H_2(m, x)$ is independent of x. Hence the canonical derivative of a differentiable function always exists and is unique.*

The following result allows to guarantee the uniqueness of the canonical derivative of differentiable functions.

Lemma 2.3. *Let $m_0 \in \mathcal{P}_1(\mathbb{R})$ and $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ be measurable and sublinear. Assume that, for all $m \in \mathcal{P}_1(\mathbb{R})$,*

$$\int_{\mathbb{R}} h_1(x)\, d(m - m_0)(x) = \int_{\mathbb{R}} h_2(x)\, d(m - m_0)(x) + \varepsilon_{m_0}(m),$$

where $\varepsilon_{m_0}(m)/D_{KR}(m,m_0)$ vanishes as m converges to m_0. Then, for all $x,y\in\mathbb{R}$,

$$h_1(x)-h_2(x)=h_1(y)-h_2(y).$$

In particular, if a function $F:\mathcal{P}_1(\mathbb{R})\to\mathbb{R}$ is differentiable at some $m\in\mathcal{P}_1(\mathbb{R})$, then its canonical derivative at m exists and is unique.

Lemma 2.3 being not particularly useful for the understanding of the paper, we postpone its proof to Appendix C.

Let us also state an elementary lemma guaranteeing expected properties for a differential operator. This lemma is implicitly used in the computation throughout the paper. For this lemma, the choice of the canonical derivative is important.

Lemma 2.4.

(*a*) *Let $F:\mathcal{P}_1(\mathbb{R})\to\mathbb{R}$ be a constant function. Then F is differentiable at any $m\in\mathcal{P}_1(\mathbb{R})$, and its canonical derivative is the zero function.*

(*b*) *Let $F,G:\mathcal{P}_1(\mathbb{R})\to\mathbb{R}$ be differentiable at $m_0\in\mathcal{P}_1(\mathbb{R})$, and $\alpha\in\mathbb{R}$. Then, the function $\alpha F+G$ is differentiable at m_0, and, for all $x\in\mathbb{R}$,*

$$\delta(\alpha F+G)(m_0,x)=\alpha\delta F(m_0,x)+\delta G(m_0,x).$$

(*c*) *Let $F,G:\mathcal{P}_1(\mathbb{R})\to\mathbb{R}$ be differentiable at $m_0\in\mathcal{P}_1(\mathbb{R})$ and Lipschitz continuous. Then, the product function FG is differentiable at m_0, with, for all $x\in\mathbb{R}$,*

$$\delta(FG)(m_0,x)=F(m_0)\delta G(m_0,x)+G(m_0)\delta F(m_0,x).$$

Proof. To prove Item (*a*), it is sufficient to notice that (10) is trivially satisfied if H_{m_0} is the zero function, by considering ε_{m_0} to be the zero function. Item (*b*) is a straightforward consequence of the linearity of the integral operator. To prove Item (*c*), let us write, for all $m\in\mathcal{P}_1(\mathbb{R})$,

$$\begin{aligned}F(m)G(m)-F(m_0)G(m_0)=&F(m)G(m)-F(m_0)G(m)\\&+F(m_0)G(m)-F(m_0)G(m_0)\\=&F(m_0)\left(G(m)-G(m_0)\right)+G(m_0)\left(F(m)-F(m_0)\right)\\&+\left(F(m)-F(m_0)\right)\left(G(m)-G(m_0)\right).\end{aligned}\tag{11}$$

Besides, since F and G are assumed to be Lipschitz continuous, we have

$$\left|\left(F(m)-F(m_0)\right)\left(G(m)-G(m_0)\right)\right|\leq CD_{KR}(m,m_0)^2.$$

So, developing (11) using the differentiability of F and G proves the result. □

Let us notice that the derivability in the sense of Definition 2.1 is close to the one used in Cox et al. (2021) (as derivative), and in Guo, Pham and Wei (2023) and Carmona and Delarue (2018) (as linear derivative). It is however different than the derivative introduced by Pierre-Louis Lions, even if it is closely related (see Propositions 5.44, 5.48 and 5.51 of Carmona and Delarue (2018)).

Depending on the properties that are studied, each notion of differentiability can be more or less appropriate to state and prove the results. For the sake of clarity, in all the paper, only one notion of differentiability is used explicitly. We choose to use Definition 2.1 instead of Lions' derivative because it appears to be more appropriate to express the infinitesimal generators of measure-valued

Markov processes, to retrieve properties of real-valued Markov processes (see Remarks 3.7 and 3.11). However, to define the notion of mixed derivatives (that is used in Assumptions 3 and 4, and in the proof of Lemma 3.8), the Lions-derivative is particularly useful as a notation. Instead of stating the original and rather technical definition, we just give the expression of the Lions-derivative using the (linear) derivative. As in Guo, Pham and Wei (2023), we use the notation δ for the (linear) derivative, and ∂ for the Lions-derivative.

Definition 2.5 (Lions-derivative)**.** *Let $F : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ and $m \in \mathcal{P}_1(\mathbb{R})$. The function F is said to be L-differentiable at m if:*

- *F is differentiable at m (in the sense of Definition 2.1),*
- *and $x \in \mathbb{R} \mapsto \delta F(m, x)$ is differentiable on $\mathbb{R}$.*

In this case, the Lions-derivative of F at m is the function $\partial F(m, \bullet)$ defined as

$$\partial F(m, \bullet) : x \in \mathbb{R} \longmapsto \partial F(m, x) = \partial_x \delta F(m, x).$$

As mentioned earlier, the notion of derivative we use (cf Definition 2.1) is related to the definition given at (Dawson 1993, p. 19) in the following sense (which is a straightforward consequence of Lemma 2.15 with $m_0 = m$ and $m_1 = \delta_x$).

Remark 2.6. *Let $F : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ be differentiable at $m \in \mathcal{P}_1(\mathbb{R})$, then, for all $x \in \mathbb{R}$,*

$$\frac{F((1-\eta)m + \eta\delta_x) - F(m)}{\eta}$$

converges, as $\eta > 0$ goes to zero, to the canonical derivative $\delta F(m, x)$ of F. In other words

$$\delta F(m, x) = \left(\partial_\eta F((1-\eta)m + \eta\delta_x)\right)_{|\eta=0} .$$

One can note that the previous remark gives a useful way of characterizing the expression of the derivative of a given function, once it is known it is differentiable.

In the following, it is required to differentiate many times some functions. To define this notion, we use the following notation: if $G : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^k \to \mathbb{R}$ (with $k \in \mathbb{N}^*$), then, for all $x_1, ..., x_k \in \mathbb{R}$

$$G_{(x_1,...,x_k)} : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto G(m, x_1, ..., x_k).$$

Definition 2.7. *A function $F : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ is said to be n-times differentiable if:*

- *F is $(n-1)$-times differentiable,*
- *for all $x \in \mathbb{R}^{n-1}$, the function $(\delta^{n-1}F)_x$ is differentiable.*

In that case we define inductively the canonical n-th derivative of F at m by: for all $x = (x_1, ..., x_n) \in \mathbb{R}^n$,

$$\delta^n F(m, x_1, ..., x_n) = \delta\left((\delta^{n-1}F)_{x_{\setminus n}}\right)(m, x_n).$$

The differentiability w.r.t. the measure-variable is not the only regularity that will be needed to prove our results. We also need regularity w.r.t. the real variables appearing when a function defined on $\mathcal{P}_1(\mathbb{R})$ is differentiated.

Definition 2.8. *We define $C^{n,k}(\mathcal{P}_1(\mathbb{R}))$ (with $n \in \mathbb{N}^*$ and $k \in \mathbb{N} \cup \{\infty\}$) the set of functions $F : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ such that F is n-times differentiable on $\mathcal{P}_1(\mathbb{R})$ and, for every $1 \leq j \leq n$, the function*

$$(m, x) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^j \mapsto \delta^j F(m, x)$$

is C^k w.r.t. x for fixed m, and is continuous w.r.t. m for fixed x.

In addition, we denote $C_b^n(\mathcal{P}_1(\mathbb{R}))$ the subspace of $C^{n,n}(\mathcal{P}_1(\mathbb{R}))$ composed of the (non-necessarily bounded) functions F such that, for all $1 \leq k \leq n$ and $1 \leq j \leq k$,

$$\sup_{m \in \mathcal{P}_1(\mathbb{R}), x \in \mathbb{R}^k} \left| \frac{\partial^k}{\partial x_1 ... \partial x_k} \delta^k F(m, x) \right| < \infty \quad \text{and} \quad \sup_{m \in \mathcal{P}_1(\mathbb{R}), x \in \mathbb{R}^k} \left| \partial_{x_j} \delta^k F(m, x) \right| < \infty.$$

Remark 2.9. *The set $C_b^n(\mathcal{P}_1(\mathbb{R}))$ depends on the choice of the version of the derivatives because of the second boundedness conditions (i.e. boundedness of $\partial_{x_j} \delta F^k(m, x)$). These conditions imply that, for any $\mu_1, ..., \mu_k \in \mathcal{P}_1(\mathbb{R})$, the family of functions $(x \mapsto \delta F^k(m, x))_{m \in \mathcal{P}_1(\mathbb{R})}$ is uniformly integrable w.r.t. $\bigotimes_{i=1}^k \mu_i$.*

One can easily note that $C_b^1(\mathcal{P}_1(\mathbb{R}))$ is included in the set of functions $F : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ that are Lipschitz continuous functions w.r.t. D_{KR} (cf Proposition 2.21 below).

In the following, the notion of many times L-differentiability is also used. The definition is essentially the same as Definition 2.7 for the Lions-derivative instead of the (linear) derivative.

Definition 2.10. *A function $F : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ is said to be n-times L-differentiable if:*

- *F is $(n-1)$-times L-differentiable,*
- *for all $x \in \mathbb{R}^{n-1}$, the function $(\partial^{n-1} F)_x$ is L-differentiable.*

In that case we define inductively the n-th Lions-derivative of F at m by: for all $x = (x_1, ..., x_n) \in \mathbb{R}^n$,

$$\partial^n F(m, x_1, ..., x_n) = \partial \left((\partial^{n-1} F)_{x \backslash_n} \right) (m, x_n).$$

Let us now introduce the notion of *mixed derivatives* for functions defined on $\mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^d$ for any $d \in \mathbb{N}$. The *mixed derivatives* are defined as sets of functions.

Definition 2.11 (Mixed derivatives)**.** *Let $d \in \mathbb{N}$ and $F : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^d \to \mathbb{R}$. We say that F admits mixed derivatives if, for all $x = (x_1, ..., x_d) \in \mathbb{R}^d$:*

- *for any $1 \leq k \leq d$ and $m \in \mathcal{P}_1(\mathbb{R})$, $y \in \mathbb{R} \mapsto F(m, x \backslash_k y)$ is differentiable on $\mathbb{R}$,*
- *and the function F_x is $L-$differentiable on $\mathcal{P}_1(\mathbb{R})$.*

In that case, the mixed derivatives of F at m form the following set of $d+1$ functions:

$$\begin{aligned} &\left\{ (m, x) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^d \longmapsto \partial_{x_k} F(m, x) \ : \ 1 \leq k \leq d \right\} \\ &\cup \left\{ (m, x, y) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^d \times \mathbb{R} \longmapsto \partial(F_x)(m, y) \right\}, \end{aligned}$$

where $\partial(F_x)$ above denotes the Lions-derivative of F_x (cf Definition 2.5).

In addition, we define inductively the n-th order mixed derivatives of F as the union of the mixed derivatives of all the functions belonging to the set corresponding to the $(n-1)$-th order mixed derivatives of F (assuming all these mixed derivatives exist).

Example 2.12. *Let $b : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$. Then, the (first order) mixed derivatives of b are the two following functions:*

$$(m, x) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \longmapsto \partial_x b(m, x),$$
$$(m, x, y) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^2 \longmapsto \partial(b_x)(m, y).$$

The second order mixed derivatives of b are the five following functions

$$(m, x) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \longmapsto \partial^2_{xx} b(m, x),$$
$$(m, x, y) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^2 \longmapsto \partial_x \partial(b_x)(m, y),$$
$$(m, x, y) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^2 \longmapsto \partial_y \partial(b_x)(m, y),$$
$$(m, x, y) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^2 \longmapsto \partial \left(\partial_x b_x\right)(m, y),$$
$$(m, x, y, z) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^3 \longmapsto \partial^2(b_x)(m, y, z).$$

The only way the notion of *mixed derivatives* is used is to write Assumptions 3 and 4 in a compact way, stating that all the mixed derivatives up to order five of some functions are bounded. This kind of assumption is required to prove the regularity of the semigroups of the measure-valued Markov processes that we study in this paper (i.e. Proposition 3.9, and more precisely, the proof of Lemma 3.8).

Let us write now a remark that is a direct consequence of Proposition 2.21, explaining the reason behind the boundedness hypothesis of the mixed order derivatives.

Remark 2.13. *Let F be a function admitting n-th order mixed derivatives (with $n \geq 2$) such that all its n-th order mixed derivatives are bounded. Then all the $(n-1)$-th order mixed derivatives of F are Lipschitz continuous: for any $G : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R}^d$ belonging to the set of the $(n-1)$-th order mixed derivatives, there exists $L > 0$ such that, for all $m, \mu \in \mathcal{P}_1(\mathbb{R})$ and $x, y \in \mathbb{R}^d$,*

$$|G(m, x) - G(\mu, y)| \leq C \left(D_{KR}(m, \mu) + ||x - y||_1\right).$$

Let us end this section with a notation.

Definition 2.14. *Let $n \in \mathbb{N}^*$ and $F : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ bounded and belonging to $C^n_b(\mathcal{P}_1(\mathbb{R}))$. Let us assume that all the mixed derivatives of F of order up to n exist and are bounded. Then we denote $||F||_n$ a bound of all the k-th order mixed derivatives, for all $0 \leq k \leq n$.*

2.2. Chain rules

In this section, we state and prove some elementary results about "chain rules" for the differentiation of measure-variable functions. We want to prove the differentiability of functions of the following form:

$$G \circ h : \mathbb{R} \xrightarrow{h} \mathcal{P}_1(\mathbb{R}) \xrightarrow{G} \mathbb{R}.$$

The general form that one could naturally expect for the derivative of this kind of function is the following: for $t \in \mathbb{R}$,

$$(G \circ h)'(t) = \int_{\mathbb{R}} \delta G(h(t), x)\, d(h'(t))(x), \tag{12}$$

where the sense of the derivative of $h : \mathbb{R} \to \mathcal{P}_1(\mathbb{R})$ has to be defined. In the first particular case, we consider, for two measures $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$ the following function

$$h : t \in [0,1] \longmapsto (1-t)m_0 + tm_1,$$

where the derivative of h has to be understood as the quantity $(h(t) - h(t_0))/(t - t_0)$ which is constant and equal to $m_1 - m_0$.

Lemma 2.15. *Let $G : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ be differentiable, $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$, and*

$$f : t \in [0,1] \longmapsto G((1-t)m_0 + tm_1).$$

Then f is differentiable, and, for all $t \in [0,1]$,

$$f'(t) = \int_{\mathbb{R}} \delta G((1-t)m_0 + tm_1, x)\, d(m_1 - m_0)(x).$$

Proof. We have, for any $t, t_0 \in [0,1]$,

$$f(t) - f(t_0) = (t - t_0) \int_{\mathbb{R}} \delta G((1-t_0)m_0 + t_0 m_1, x) d(m_1 - m_0)(x) + \varepsilon((1-t)m_0 + tm_1), \quad (13)$$

with $\varepsilon(\mu)/D_{KR}(\mu, (1-t_0)m_0 + t_0 m_1)$ vanishing as μ converges to $(1-t_0)m_0 + t_0 m_1$. Noticing that

$$D_{KR}((1-t)m_0 + tm_1, (1-t_0)m_0 + t_0 m_1) \leq 2|t - t_0|,$$

the proof of the lemma follows from (13) dividing it by $(t - t_0)$ and then letting t goes to t_0. □

Remark 2.16. *We have the following consequence of Lemma 2.15 and of the fundamental theorem of calculus: for any function G belonging to $C_b^1(\mathcal{P}_1(\mathbb{R}))$, for all $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$,*

$$G(m_1) - G(m_0) = \int_0^1 \int_{\mathbb{R}} \delta G((1-t)m_0 + tm_1, x) d(m_1 - m_0)(x).$$

The above property is used as the definition of the linear derivative by Carmona and Delarue (2018) (cf Section 4.1), Guo, Pham and Wei (2023) and Cox et al. (2021). In addition, by Proposition 5.44 of Carmona and Delarue (2018), under some generic assumptions, this property can characterize the linear derivative as it is defined at Definition 2.1 when the analysis of measure-variable functions is studied on $\mathcal{P}_2(\mathbb{R})$.

Now we prove a similar result for functions that are many times differentiable with bounded derivatives. In particular, the following result is not strictly stronger than the previous one.

Lemma 2.17. *Let $n \in \mathbb{N}^*$, $G \in C_b^n(\mathcal{P}_1(\mathbb{R}))$, $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$, and*

$$f : t \in [0,1] \longmapsto G((1-t)m_0 + tm_1).$$

Then f is C^n, and, for all $t \in [0,1]$,

$$f^{(n)}(t) = \int_{\mathbb{R}^n} \delta^n G((1-t)m_0 + tm_1, x)\, d(m_1 - m_0)^{\otimes n}(x).$$

Proof. We prove the statement by induction on n. For $n = 1$, Lemma 2.15 implies that f is differentiable. The fact that f is C^1 follows by the dominated convergence theorem since $G \in C_b^1(\mathcal{P}_1(\mathbb{R}))$.

For the general case (i.e. $n \geq 2$), let us assume that f is C^{n-1} with, for all $t \in [0,1]$,

$$f^{(n-1)}(t) = \int_{\mathbb{R}^{n-1}} \delta^{n-1} G((1-t)m_0 + tm_1, x) d(m_1 - m_0)^{\otimes n-1}(x).$$

Note that, since $G \in C_b^n(\mathcal{P}_1(\mathbb{R}))$, the function

$$(t,x) \in [0,1] \times \mathbb{R}^{n-1} \longmapsto \delta^{n-1} G((1-t)m_0 + tm_1, x)$$

is sublinear w.r.t. $||x||_1$ (hence integrable w.r.t. m_1 and m_0) and differentiable w.r.t. t (thanks to Lemma 2.15) with derivative

$$(t,x) \longmapsto \int_{\mathbb{R}} \delta\left(\left(\delta^{n-1} G\right)_x\right)((1-t)m_0 + tm_1, y) d(m_1 - m_0)(y),$$

which is also sublinear w.r.t. $||x||_1$ (with a constant which is uniform w.r.t. $t \in [0,1]$ since $G \in C_b^n(\mathcal{P}_1(\mathbb{R}))$. Then, by differentiation under the integral sign, for all $t \in [0,1]$,

$$f^{(n)}(t) = \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \delta\left(\left(\delta^{n-1} G\right)_x\right)((1-t)m_0 + tm_1, y) d(m_1 - m_0)(y) d(m_1 - m_0)^{\otimes n-1}(x).$$

Once again, the continuity of $f^{(n)}$ is a consequence of the dominated convergence theorem and of the assumption that G belongs to $C_b^n(\mathcal{P}_1(\mathbb{R}))$. □

Another particular case of interest for the chain rule (12) is the one of the function

$$h : \lambda \in \mathbb{R} \longmapsto \mathcal{S}(m, \lambda),$$

for some fixed $m \in \mathcal{P}_1(\mathbb{R})$, with $\mathcal{S}$ a shift operator defined as:

$$\mathcal{S}(m,\lambda) : A \in \mathcal{B}(\mathbb{R}) \longmapsto m(\{x - \lambda : x \in A\}).$$

In that case, the derivative of h could be seen as a reminiscent of the derivative in the sense of Schwartz' distribution, since it requires to differentiate the integrand instead of the measure in (12) (it is actually the opposite of one would expect from the derivative in the distribution sense).

Lemma 2.18. *Let $m \in \mathcal{P}_1(\mathbb{R})$, $G : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ be differentiable such that, for any $\mu \in \mathcal{P}_1(\mathbb{R})$, $x \mapsto \delta G(\mu, x)$ is differentiable. Then the function*

$$f : \lambda \in \mathbb{R} \longmapsto G(\mathcal{S}(m, \lambda))$$

is differentiable and, for all $\lambda \in \mathbb{R}$,

$$f'(\lambda) = \int_{\mathbb{R}} \left(\partial_y \delta G(\mathcal{S}(m,\lambda), x + y)\right)_{|y=\lambda} dm(x).$$

Proof. We have for any $t \in \mathbb{R}$,

$$f(\lambda) - f(\lambda_0) = \int_{\mathbb{R}} \delta G(\mathcal{S}(m,\lambda_0),x) d(\mathcal{S}(m,\lambda) - \mathcal{S}(m,\lambda_0))(x) + \varepsilon(\mathcal{S}(m,\lambda)),$$

with $\varepsilon(\mu)/D_{KR}(\mu, \mathcal{S}(m,\lambda_0))$ vanishing as μ converges to $\mathcal{S}(m,\lambda_0)$ for D_{KR}. Since, writing $m_0 := \mathcal{S}(m,\lambda_0)$,

$$D_{KR}(\mathcal{S}(m,\lambda), \mathcal{S}(m,\lambda_0)) = D_{KR}(\mathcal{S}(m_0, \lambda - \lambda_0), m_0) = |\lambda - \lambda_0|,$$

we know that $\varepsilon(\mathcal{S}(m,\lambda))/|\lambda - \lambda_0|$ vanishes as λ converges to λ_0. And, by the Dominated Convergence Theorem,

$$\begin{aligned}
&\frac{1}{\lambda - \lambda_0} \int_{\mathbb{R}} \delta G(\mathcal{S}(m,\lambda_0),x) d(\mathcal{S}(m,\lambda) - \mathcal{S}(m,\lambda_0))(x) \\
&\qquad = \int_{\mathbb{R}} \frac{\delta G(\mathcal{S}(m,\lambda_0), x+\lambda) - \delta G(\mathcal{S}(m,\lambda_0), x+\lambda_0)}{\lambda - \lambda_0} dm(x) \\
&\qquad \underset{\lambda \to \lambda_0}{\longrightarrow} \int_{\mathbb{R}} \left(\partial_y \delta G(\mathcal{S}(m,\lambda_0), x+y)\right)_{|y=\lambda_0} dm(x),
\end{aligned}$$

which ends the proof. □

The last case for which we want (12) is when the function h is of the form

$$h : x \in \mathbb{R} \longmapsto \mathcal{L}\left(Y(x) | \mathcal{T}\right).$$

with $(Y(x))_x$ a family of random variables and $\mathcal{T}$ some sigma-field.

Lemma 2.19. *Let $G \in C_b^1(\mathcal{P}_1(\mathbb{R}))$, $\mathcal{T}$ some sigma-field and $Y(x)$ $(x \in \mathbb{R})$ be real-valued random variables such that, almost surely,*

$$x \longmapsto Y(x)$$

belongs to $C^1(\mathbb{R})$, and assume that for all compact set $K \subset \mathbb{R}$,

$$\mathbb{E}\left[\sup_{x \in K} |\partial_x Y(x)|\right] < \infty. \tag{14}$$

Then, almost surely, the function

$$f : x \in \mathbb{R} \longmapsto G\left(\mathcal{L}[Y(x)|\mathcal{T}]\right)$$

belongs to $C^1(\mathbb{R})$ and, for all $x \in \mathbb{R}$,

$$f'(x) = \mathbb{E}\left[\left(\partial_x Y(x)\right) \cdot \left(\partial_y \delta G\left(\mathcal{L}[Y(x)|\mathcal{T}], y\right)\right)_{|y=Y(x)} \middle| \mathcal{T} \right].$$

Proof. For the sake of notation, let us denote, for all $x \in \mathbb{R}$,

$$m_x = \mathcal{L}(Y(x)|\mathcal{T}).$$

For the rest of the proof, let us fix some $x_0 \in \mathbb{R}$ and some compact set K containing an open neighborhood of x_0. Then, for $x \in K$,

$$\begin{aligned}
f(x) - f(x_0) =& G(m_x) - G(m_{x_0}) = \int_{\mathbb{R}} \delta G(m_{x_0}, y) d\left(m_x - m_{x_0}\right)(y) + \varepsilon_{x_0}(m_x) \\
=& \mathbb{E}\left[\delta G(m_{x_0}, Y(x)) - \delta G(m_{x_0}, Y(x_0)) | \mathcal{T}\right] + \varepsilon_{x_0}(m_x),
\end{aligned}$$

with $\varepsilon_{x_0}(\mu)/D_{KR}(\mu, m_{x_0})$ vanishing as μ converges to m_{x_0}. Consequently, for all $x \in K\backslash\{x_0\}$,

$$\frac{f(x)-f(x_0)}{x-x_0} = \mathbb{E}\left[\frac{\delta G(m_{x_0}, Y(x)) - \delta G(m_{x_0}, Y(x_0))}{x-x_0}\middle|\mathcal{T}\right] + \frac{\varepsilon_{x_0}(m_x)}{x-x_0}.$$

Since, for all $x \in K$,

$$D_{KR}(m_x, m_{x_0}) \leq \mathbb{E}\left[|Y(x)-Y(x_0)||\,\mathcal{T}\right] \leq |x-x_0| \cdot \mathbb{E}\left[\sup_{\tilde{x}\in K}|\partial_{\tilde{x}}Y(\tilde{x})|\middle|\mathcal{T}\right],$$

we know, by (14), that $\varepsilon_{x_0}(m_x)/|x-x_0|$ vanishes almost surely as x goes to x_0.

Then, using (14) and the hypothesis that G belongs to $C_b^1(\mathcal{P}_1(\mathbb{R}))$, the dominated convergence theorem concludes the proof. □

Remark 2.20. *With the notation of Lemma 2.19, the expression of f' can be substantially simplified using the Lions-derivative instead of the (linear) derivative: for all $x \in \mathbb{R}$,*

$$f'(x) = \mathbb{E}\left[(\partial_x Y(x)) \cdot \partial G\left(\mathcal{L}[Y(x)|\mathcal{T}], Y(x)\right)|\,\mathcal{T}\right].$$

2.3. Taylor's formulas and some corollaries

The results of this section are mainly consequences of Taylor's theorems applied to the function f of Lemma 2.15 from the previous section. Let us give the statement for the Mean Value Theorem for measure-variable functions. The proof consisting in applying directly the Mean Value Theorem for the real-variable function f of Lemma 2.15, we omit it.

Proposition 2.21 (Mean Value Theorem)**.** *Let $G : \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ be differentiable. Then, for any $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$, there exists some $t \in]0,1[$ such that*

$$G(m_1) - G(m_0) = \int_{\mathbb{R}} \delta G((1-t)m_0 + tm_1, x) d(m_1 - m_0)(x).$$

In particular, every function G belonging to $C_b^1(\mathcal{P}_1(\mathbb{R}))$ is Lipschitz continuous: for all $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$,

$$|G(m_1) - G(m_0)| \leq D_{KR}(m_0, m_1) \cdot \sup_{m\in\mathcal{P}_1(\mathbb{R}), x\in\mathbb{R}} |\partial_x \delta G(m, x)|.$$

Let us now generalize the previous result by stating a version of Taylor's formula with integral remainder for measure-variable functions.

Theorem 2.22 (Taylor's formula with integral remainder)**.** *Let $n \in \mathbb{N}^*$ and $G \in C_b^{n+1}(\mathcal{P}_1(\mathbb{R}))$. Then, for any $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$,*

$$G(m_1) - \sum_{k=0}^{n} \frac{1}{k!} \int_{\mathbb{R}^k} \delta^k G(m_0, x)\, d(m_1 - m_0)^{\otimes k}(x)$$
$$= \frac{1}{n!} \int_0^1 (1-t)^n \int_{\mathbb{R}^{n+1}} \delta^{n+1} G((1-t)m_0 + tm_1, x)\, d(m_1 - m_0)^{\otimes n+1}(x)\, dt.$$

Proof. By Lemma 2.17, the function

$$f : t \in [0,1] \longmapsto G((1-t)m_0 + tm_1)$$

belongs to $C^{n+1}([0,1])$, and for all $0 \leq k \leq n+1$ and $t \in [0,1]$,

$$f^{(k)}(t) = \int_{\mathbb{R}^k} \delta^k G((1-t)m_0 + tm_1, x) d(m_1 - m_0)^{\otimes k}(x).$$

Then, by the (standard) Taylor's formula with remainder term,

$$f(1) - \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} = \frac{1}{n!} \int_0^1 f^{(n+1)}(t)(1-t)^n dt.$$

This last equality is exactly the statement of the theorem. □

In order to prove a version of the Taylor-Lagrange's inequality for measure-variable functions, the following lemma about Kantorovich-Rubinstein metric is required.

Lemma 2.23. *Let $n \in \mathbb{N}^*$ and $g \in C_b^n(\mathbb{R}^n)$. Then, for any $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$,*

$$\left| \int_{\mathbb{R}^n} g(x) d(m_1 - m_0)^{\otimes n}(x) \right| \leq D_{KR}(m_0, m_1)^n \cdot \sup_{x_1,\dots,x_n \in \mathbb{R}} \left| \frac{\partial^n}{\partial x_1 \dots \partial x_n} g(x_1, \dots, x_n) \right|. \tag{15}$$

Proof. We prove the result by induction on n. For $n = 1$, it is a mere consequence of the definition of D_{KR}, and of the fact that, any $g \in C_b^1(\mathbb{R})$ is Lipschitz continuous with Lipschitz constant non-greater than $||g'||_\infty$.

Then, let $n \in \mathbb{N}^*$ and assume that (15) holds true for any $g \in C_b^n(\mathbb{R}^n)$. Let $h \in C_b^{n+1}(\mathbb{R}^{n+1})$, and

$$g : x \in \mathbb{R}^n \longmapsto \int_{\mathbb{R}} h(x, y) d(m_1 - m_0)(y),$$

which belongs to $C_b^{n+1}(\mathbb{R}^n)$. So

$$\begin{aligned} \left| \int_{\mathbb{R}^{n+1}} h(x) d(m_1 - m_0)^{\otimes n+1}(x) \right| &= \left| \int_{\mathbb{R}^n} g(x) d(m_1 - m_0)^{\otimes n}(x) \right| \\ &\leq D_{KR}(m_0, m_1)^n \cdot \sup_{x_1,\dots,x_n \in \mathbb{R}} \left| \frac{\partial^n}{\partial x_1 \dots \partial x_n} g(x_1, \dots, x_n) \right|. \end{aligned}$$

And, thanks to the case $n = 1$, we know that

$$\begin{aligned} \left| \frac{\partial^n}{\partial x_1 \dots \partial x_n} g(x_1, \dots, x_n) \right| &= \left| \int_{\mathbb{R}} \frac{\partial^n}{\partial x_1 \dots \partial x_n} h(x, y) d(m_1 - m_0(y)) \right| \\ &\leq D_{KR}(m_0, m_1) \cdot \sup_{x_1,\dots,x_n,y \in \mathbb{R}} \left| \frac{\partial^{n+1}}{\partial x_1 \dots \partial x_n \partial y} h(x_1, \dots, x_n, y) \right|. \end{aligned}$$

Combining the two previous inequalities proves the statement of the lemma. □

As a straightforward consequence of Theorem 2.22 and Lemma 2.23 we obtain the following result.

Theorem 2.24 (Taylor-Lagrange's inequality)**.** *Let $n \in \mathbb{N}^*$ and $G \in C_b^{n+1}(\mathcal{P}_1(\mathbb{R}))$. Then, for any $m_0, m_1 \in \mathcal{P}_1(\mathbb{R})$,*

$$\left| G(m_1) - \sum_{k=0}^{n} \frac{1}{k!} \int_{\mathbb{R}^k} \delta^k G(m_0, x) d(m_1 - m_0)^{\otimes k}(x) \right| \leq \frac{1}{(n+1)!} D_{KR}(m_0, m_1)^{n+1} \cdot \sup_{x \in \mathbb{R}^{n+1}, m \in \mathcal{P}_1(\mathbb{R})} \left| \frac{\partial^{n+1}}{\partial x_1 ... \partial x_{n+1}} \delta^{n+1} G(m, x) \right|.$$

Another consequence of Theorem 2.22 is that, in some sense, it is possible to interchange the derivative w.r.t. the measure-variable and the partial derivative w.r.t. the real-variable, when it makes sense. For the sake of readability, the result is only stated for second order derivatives, but can be generalized inductively.

Corollary 2.25. *Let $G \in C_b^3(\mathcal{P}_1(\mathbb{R}))$. Then, for all $m \in \mathcal{P}_1(\mathbb{R})$, and $x, y \in \mathbb{R}$,*

$$\partial_x \delta \left((\delta G)_x \right)(m, y) = \delta \left(\partial_x (\delta G)_x \right)(m, y),$$

or, with the Lions-derivative notation,

$$\partial^2_{xy} \delta^2 G(m, x, y) = \partial^2 G(m, x, y).$$

Proof. The proof consists in showing that the function

$$(m, x, y) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \times \mathbb{R} \longmapsto \partial_x \delta \left((\delta G)_x \right)(m, y)$$

is the derivative of the function

$$(m, x) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \longmapsto \partial_x \delta G(m, x)$$

in the sense of Definition 2.1. Let us fix in all the proof, some $m, m_0 \in \mathcal{P}_1(\mathbb{R})$.

Firstly, by differentiation under the integral sign (which is permitted since $G \in C_b^3(\mathcal{P}_1(\mathbb{R}))$), we have, for any $x \in \mathbb{R}$,

$$\partial_x \int_{\mathbb{R}} \delta^2 G(m, x, y)\, d(m - m_0)(y) = \int_{\mathbb{R}} \partial_x \delta^2 G(m, x, y)\, d(m - m_0)(y),$$

such that,

$$\begin{aligned}
&\partial_x \delta G(m, x) - \partial_x \delta G(m_0, x) - \int_{\mathbb{R}} \partial_x \delta^2 G(m, x, y)\, d(m - m_0)(y) \\
&= \partial_x \left(\delta G(m, x) - \delta G(m_0, x) - \int_{\mathbb{R}} \delta^2 G(m, x, y)\, d(m - m_0)(y) \right) \\
&= \partial_x \int_0^1 (1 - t) \int_{\mathbb{R}^2} \delta^3 G((1 - t)m_0 + tm, x, y_1, y_2)\, d(m - m_0)^{\otimes 2}(y_1, y_2) dt \\
&= \int_0^1 (1 - t) \int_{\mathbb{R}^2} \partial_x \delta^3 G((1 - t)m_0 + tm, x, y_1, y_2)\, d(m - m_0)^{\otimes 2}(y_1, y_2) dt,
\end{aligned}$$

where we have used the Taylor's formula with integral remainder (cf Theorem 2.22) to obtain the before last equality, and we have once again differentiated under the integral sign to obtain the last one. Then, by Lemma 2.23 and recalling that $G \in C_b^3(\mathcal{P}_1(\mathbb{R}))$, we obtain

$$\left|\partial_x \delta G(m,x) - \partial_x \delta G(m_0,x) - \int_{\mathbb{R}} \partial_x \delta^2 G(m,x,y)\, d(m-m_0)(y)\right| \leq D_{KR}(m,m_0)^2 \cdot \sup_{\mu\in\mathcal{P}_1(\mathbb{R}),y\in\mathbb{R}^3} \left|\frac{\partial^3}{\partial y_1 \partial y_2 \partial y_3}\delta^3 G(\mu,y)\right|,$$

which proves the result. □

Let us end this section with two applications of Taylor-Lagrange's inequality (i.e. Theorem 2.24) that permits to differentiate some particular functions.

Corollary 2.26. *Let $H : \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ and*

$$F : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}} H(m,x) dm(x).$$

Assume that:

(i) for any $x \in \mathbb{R}$, the function H_x belongs to $C_b^2(\mathcal{P}_1(\mathbb{R}))$, and there exists $C > 0$ such that, for all $x \in \mathbb{R}$,

$$\sup_{\substack{m\in\mathcal{P}_1(\mathbb{R})\\ y_1,y_2\in\mathbb{R}}} \left|\partial^2_{y_1y_2}\delta^2(H_x)(m,y_1,y_2)\right| \leq C(1+|x|),$$

(ii) for any $m \in \mathcal{P}_1(\mathbb{R})$ and $x \in \mathbb{R}$, the function $y \mapsto H(m,y)$ is C^1 on $\mathbb{R}$, and $\partial_x H_x$ belongs to $C^{1,1}(\mathcal{P}_1(\mathbb{R}))$, and

$$\sup_{\substack{m\in\mathcal{P}_1(\mathbb{R})\\ x,y\in\mathbb{R}}} \left|\partial_y \delta\left(\partial_x H_x\right)(m,y)\right| < \infty.$$

Then, F is differentiable on $\mathcal{P}_1(\mathbb{R})$, with: for all $m \in \mathcal{P}_1(\mathbb{R})$ and $x \in \mathbb{R}$,

$$\delta F(m,x) = H(m,x) + \int_{\mathbb{R}} \delta H_y(m,x) dm(y) - F(m).$$

Proof. Let $m, m_0 \in \mathcal{P}_1(\mathbb{R})$ be fixed in all the proof.

$$\begin{aligned} F(m) - F(m_0) &= \int_{\mathbb{R}} H(m,x) d(m-m_0)(x) + \int_{\mathbb{R}} \left(H(m,x) - H(m_0,x)\right) dm_0(x) \\ &= \int_{\mathbb{R}} H(m_0,x) d(m-m_0)(x) + \int_{\mathbb{R}} \left(H(m,x) - H(m_0,x)\right) dm_0(x) \\ &\quad + \int_{\mathbb{R}} \left(H(m,x) - H(m_0,x)\right) d(m-m_0)(x). \end{aligned}$$

Consequently,

$$\begin{aligned}\left|F(m)-F(m_0)-\int_{\mathbb{R}}\left(H(m_0,x)-\int_{\mathbb{R}}\delta H_y(m_0,x)dm_0(y)\right)d(m-m_0)(x)\right| \\ \leq \int_{\mathbb{R}}\left|H_x(m)-H_x(m_0)-\int_{\mathbb{R}}\delta H_x(m_0,y)d(m-m_0)(y)\right|dm_0(x) \qquad (16) \\ +\left|\int_{\mathbb{R}}\left(H(m,x)-H(m_0,x)\right)d(m-m_0)(x)\right|. \qquad (17)\end{aligned}$$

Then, by the hypothesis (i) of the corollary and Theorem 2.24, the integrand of (16) is bounded by

$$C(1+|x|)D_{KR}(m,m_0)^2$$

for any $x\in\mathbb{R}$ (with $C>0$ independent of x,m,m_0), whence the quantity at (16) is bounded by

$$C\left(1+\int_{\mathbb{R}}|x|dm_0(x)\right)D_{KR}(m,m_0)^2$$

which is negligible compared to $D_{KR}(m,m_0)$ when m converges to m_0 (for a fixed m_0).

In addition, the term (17) is non-greater than

$$D_{KR}(m,m_0)\cdot\sup_{x\in\mathbb{R}}\left|\partial_x\left(H(m,x)-H(m_0,x)\right)\right|.$$

And, by the Mean Value Theorem (i.e. Proposition 2.21) and the hypothesis (ii) of the corollary, for all $x\in\mathbb{R}$,

$$\begin{aligned}\left|\partial_x\left(H(m,x)-H(m_0,x)\right)\right| =& \left|\partial_x H_x(m)-\partial_x H_x(m_0)\right| \\ \leq& D_{KR}(m,m_0)\cdot\sup_{\substack{m\in\mathcal{P}_1(\mathbb{R})\\ x,y\in\mathbb{R}}}\left|\partial_y\delta\left(\partial_x H_x\right)(m,y)\right|,\end{aligned}$$

which entails that the term at (17) is bounded by

$$CD_{KR}(m,m_0)^2$$

for some constant $C>0$ independent of m,m_0.

So we have proved that F is differentiable on $\mathcal{P}_1(\mathbb{R})$, and that the following function is one version of its derivative:

$$(m,x)\in\mathcal{P}_1(\mathbb{R})\times\mathbb{R}\longmapsto H(m,x)+\int_{\mathbb{R}}\delta H_y(m,x)dm(y).$$

It is then sufficient to subtract $F(m)$ to the quantity above to obtain the canonical derivative of F. □

The next result is similar as the previous one for a larger class of functions.

Corollary 2.27. *Let* $H:\mathcal{P}_1(\mathbb{R})\times\mathbb{R}^2\to\mathbb{R}$ *and*

$$F:m\in\mathcal{P}_1(\mathbb{R})\longmapsto\int_{\mathbb{R}^2}H(m,x)dm^{\otimes 2}(x).$$

Assume that:

(*i*) *for any* $x \in \mathbb{R}^2$, *the function* H_x *belongs to* $C_b^2(\mathcal{P}_1(\mathbb{R}))$, *and there exists* $C > 0$ *such that, for all* $x_1, x_2 \in \mathbb{R}$,

$$\sup_{\substack{m\in\mathcal{P}_1(\mathbb{R})\\ y_1,y_2\in\mathbb{R}}} \left|\partial^2_{y_1y_2}\delta^2(H_x)(m,y_1,y_2)\right| \leq C(1+|x_1|+|x_2|+|x_1x_2|),$$

(*ii*) *for any* $m \in \mathcal{P}_1(\mathbb{R})$ *and* $x_1, x_2 \in \mathbb{R}$, *the functions* $x \mapsto H(m,x)$ *is* C^1 *on* $\mathbb{R}^2$, *and both* $\partial_{x_1}H_x$ *and* $\partial_{x_2}H_x$ *belong to* $C^{1,1}(\mathcal{P}_1(\mathbb{R}))$, *with*

$$\sup_{\substack{m\in\mathcal{P}_1(\mathbb{R})\\ x_1,x_2,y\in\mathbb{R}}} |\partial_y\delta\left(\partial_{x_1}H_x\right)(m,y)| + |\partial_y\delta\left(\partial_{x_2}H_x\right)(m,y)| < \infty,$$

(*iii*) *for all* $m \in \mathcal{P}_1(\mathbb{R})$, *the function* $x \mapsto H(m,x)$ *belongs to* $C_b^2(\mathbb{R}^2)$.

Then, F *is differentiable on* $\mathcal{P}_1(\mathbb{R})$, *with: for all* $m \in \mathcal{P}_1(\mathbb{R})$ *and* $y \in \mathbb{R}$,

$$\delta F(m,y) = \int_{\mathbb{R}^2} \delta(H_x)(m,y)dm^{\otimes 2}(x) + \int_{\mathbb{R}} H(m,x_1,y)dm(x_1) + \int_{\mathbb{R}} H(m,y,x_2)dm(x_2) - 2F(m).$$

The proof of this result uses the following Lemma whose proof is given at Appendix C.

Lemma 2.28. *Let* $d \in \mathbb{N}^*$ *and* $m_1, ..., m_d, \mu_1, ..., \mu_d \in \mathcal{P}_1(\mathbb{R})$. *Then,*

$$D_{KR}\left(\bigotimes_{k=1}^d m_k, \bigotimes_{k=1}^d \mu_k\right) \leq \sum_{k=1}^d D_{KR}(m_k,\mu_k).$$

Proof of Corollary 2.27. Let us fix some $m, m_0 \in \mathcal{P}_1(\mathbb{R})$ in the proof.

$$\begin{aligned}
F(m) - F(m_0) &= \int_{\mathbb{R}^2} H(m,x)d\left(m^{\otimes 2} - m_0^{\otimes 2}\right)(x) \\
&\quad + \int_{\mathbb{R}^2} \left(H(m,x) - H(m_0,x)\right) dm_0^{\otimes 2}(x) \\
&= \int_{\mathbb{R}^2} H(m_0,x_1,x_2)dm(x_1)d(m-m_0)(x_2) \\
&\quad + \int_{\mathbb{R}^2} H(m_0,x_1,x_2)d(m-m_0)(x_1)dm_0(x_2) \\
&\quad + \int_{\mathbb{R}^2} \left(H(m,x) - H(m_0,x)\right) dm_0^{\otimes 2}(x) \\
&\quad + \int_{\mathbb{R}^2} \left(H(m,x) - H(m_0,x)\right) d\left(m^{\otimes 2} - m_0^{\otimes 2}\right)(x).
\end{aligned}$$

Then,

$$\begin{aligned}
&\left|F(m) - F(m_0) - \int_{\mathbb{R}}\int_{\mathbb{R}^2} \delta(H_x)(m_0,y)dm_0^{\otimes 2}(x)d(m-m_0)(y)\right.\\
&\quad\left. - \int_{\mathbb{R}}\left(\int_{\mathbb{R}} H(m_0,x_1,y)dm_0(x_1) + \int_{\mathbb{R}} H(m_0,y,x_2)dm_0(x_2)\right)d(m-m_0)(y)\right|\\
&\leq \left|\int_{\mathbb{R}^2} H(m_0,x_1,x_2)d(m-m_0)^{\otimes 2}(x_1,x_2)\right| \qquad (18)\\
&\quad + \left|\int_{\mathbb{R}^2}\left(H(m,x) - H(m_0,x) - \int_{\mathbb{R}}\delta(H_x)(m_0,y)d(m-m_0)(y)\right)dm_0^{\otimes 2}(x)\right| \qquad (19)\\
&\quad + \left|\int_{\mathbb{R}^2}\left(H(m,x) - H(m_0,x)\right)d\left(m^{\otimes 2} - m_0^{\otimes 2}\right)(x)\right|. \qquad (20)
\end{aligned}$$

By Lemma 2.23 and thanks to the hypothesis (iii) of the corollary, the term at (18) is bounded by

$$D_{KR}(m,m_0)^2 \sup_{\mu\in\mathcal{P}_1(\mathbb{R}),x_1,x_2\in\mathbb{R}} \left|\partial^2_{x_1x_2}H(\mu,x)\right|.$$

According to Taylor-Lagrange's inequality (i.e. Theorem 2.24) and using Hypothesis (i), we know that the term at (19) is non-greater than

$$C\cdot D_{KR}(m,m_0)^2\left(1+\left(\int_{\mathbb{R}}|x|dm_0(x)\right)^2\right),$$

for some $C>0$ independent of (m,m_0).

For the last term (i.e. (20)), by definition of Kantorovich-Rubinstein's metric, is bounded by

$$D_{KR}\left(m^{\otimes 2},m_0^{\otimes 2}\right)\cdot \sup_{x_1,x_2\in\mathbb{R}} |\partial_{x_1}H_x(m) - \partial_{x_1}H_x(m_0)| + |\partial_{x_2}H_x(m) - \partial_{x_2}H_x(m_0)|.$$

As in the end of the proof of the previous corollary, we have

$$\sup_{x_1,x_2\in\mathbb{R}} |\partial_{x_1}H_x(m) - \partial_{x_1}H_x(m_0)| + |\partial_{x_2}H_x(m) - \partial_{x_2}H_x(m_0)| \leq C\cdot D_{KR}(m,m_0).$$

On the other hand, by Lemma 2.28, we have that

$$D_{KR}\left(m^{\otimes 2},m_0^{\otimes 2}\right) \leq 2D_{KR}(m,m_0).$$

So the term at (20) is bounded by $C\cdot D_{KR}(m,m_0)^2$. This proves finally that the function F is differentiable on $\mathcal{P}_1(\mathbb{R})$ and that the function

$$(m,y)\in\mathcal{P}_1(\mathbb{R})\times\mathbb{R} \longmapsto \int_{\mathbb{R}^2}\delta(H_x)(m,y)dm^{\otimes 2}(x) + \int_{\mathbb{R}} H(m,x_1,y)dm(x_1) + \int_{\mathbb{R}} H(m,y,x_2)dm(x_2)$$

is one version of the derivative of F. It is then sufficient to subtract $2F(m)$ to the quantity above to obtain the canonical derivative of F. □

Remark 2.29. *Note that, an idea to prove Corollary 2.27 could have been to write*

$$F(m) = \int_{\mathbb{R}^2} H(m,x)dm^{\otimes 2}(x) = \int_{\mathbb{R}} G(m,x_2)dm(x_2),$$

with

$$G(m,x_2) = \int_{\mathbb{R}} H(m,x_1,x_2)dm(x_1),$$

and to apply twice Corollary 2.26. The problem is that it would require more technical and stronger assumptions to be true than the assumptions of Corollary 2.27.

2.4. Measure-variable polynomials

A set of functions of interest is the following one

$$\left\{ F : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}} g(x)dm(x) \ : \ g \text{ smooth enough} \right\}.$$

Indeed the computation (and the stochastic calculus properties) on this kind of functions defined on $\mathcal{P}_1(\mathbb{R})$ rely directly on the computation for functions defined on $\mathbb{R}$. Since this set of functions is not a separating class for $\mathcal{P}_1(\mathbb{R})$, we need to introduce the richer class of functions composed of the *polynomials* (similar functions are also used in Dawson (1993), Cuchiero, Larsson and Svaluto-Ferro (2019) and Guo, Pham and Wei (2023)).

Definition 2.30 (Polynomials)**.** *For $p \in \mathbb{N} \cup \{\infty\}$, let $\mathscr{P}^p$ be the set of polynomials of order p, i.e. the functions F of the form*

$$F : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}^n} h(x)dm^{\otimes n}(x),$$

with $n \in \mathbb{N}$ and $h \in C_b^p(\mathbb{R}^n)$.

There exists another useful class of measure-variable functions called the *cylinder functions* (see e.g. Cox et al. (2021) and Guo, Pham and Wei (2023)) that can be defined as the set

$$\left\{ F : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto g\left(\int_{\mathbb{R}} h_1(x)dm(x), ..., \int_{\mathbb{R}} h_n(x)dm(x) \right) : g, h_1, ..., h_n \text{ smooth} \right\}.$$

Note that the terminology used by Guo, Pham and Wei (2023) slightly differs from the one of this paper (in Guo, Pham and Wei (2023), the functions $g, h_1, ..., h_n, h$ above are required to be polynomials). Consequently, in this paper the cylinder functions and the polynomials form two different sets of functions. In this paper, the class of polynomials satisfies an important property: if $(P_t)_t$ is the semigroup of a particular measure-valued Markov process, then, for any $G \in \mathscr{P}^4$ and $t \geq 0$, P_tG belongs to $\mathscr{P}^4$ (see Proposition 3.9 for a formal statement). The previous property does not seem to be true for other classes of measure-variable functions like the cylinder functions.

As mentioned above, the following proposition states that the set of the polynomials is a separating class for $\mathcal{P}_1(\mathbb{R})$.

Proposition 2.31. *Let $p \geq 1$ and μ_1, μ_2 be $\mathcal{P}_p(\mathbb{R})$-valued random variables. If for all $F \in \mathscr{P}^\infty$,*

$$\mathbb{E}\left[F(\mu_1)\right] = \mathbb{E}\left[F(\mu_2)\right],$$

then μ_1 and μ_2 have the same law.

The proof of this statement being quite both classical and technical, it is postponed to the end of Appendix A.

The next result concerns the regularity of the polynomials.

Proposition 2.32. *For all $p \geq 2$, the set $\mathscr{P}^p$ is included in both $C^{\infty,p}(\mathcal{P}_1(\mathbb{R}))$ and $C_b^p(\mathcal{P}_1(\mathbb{R}))$. The mixed derivatives of any $G \in \mathscr{P}^p$ up to order p exist and are bounded. In addition, for any $n \in \mathbb{N}^*$, for all $\varphi \in C_b^p(\mathbb{R}^n)$, defining $G \in \mathscr{P}^p$ as*

$$G : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}^n} \varphi(x) dm^{\otimes n}(x),$$

we have, for all $m \in \mathcal{P}_1(\mathbb{R})$ and $y \in \mathbb{R}$,

$$\delta G(m, y) = \sum_{k=1}^{n} \int_{\mathbb{R}^{n-1}} \varphi(x\backslash_k y) dm^{\otimes n-1}(x\backslash_k) - n \cdot G(m),$$

and, for all $m \in \mathcal{P}_1(\mathbb{R})$ and $y_1, y_2 \in \mathbb{R}$,

$$\delta^2 G(m, y_1, y_2) = \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} \int_{\mathbb{R}^{n-2}} \varphi(x\backslash_{(k,l)}(y_1, y_2)) dm^{\otimes n-2}(x\backslash_{(k,l)}) + C_1(m, y_1) + C_2(m, y_2),$$

with $C_1(m, y_1)$ (resp. $C_2(m, y_2)$) independent of y_2 (resp. y_1).

Before proving Proposition 2.32, let us remark that it is possible to have explicit and smooth expressions for the quantities denoted by $C_1(m, y_1)$ and $C_2(m, y_2)$: with the notation of Proposition 2.32,

$$\begin{aligned} C_1(m, y_1) &= -(n-1) \int_{\mathbb{R}^{n-1}} \varphi(x\backslash_k y_1) dm^{\otimes n-1}(x\backslash_k), \\ C_2(m, y_2) &= -n \cdot \delta G(m, y_2). \end{aligned}$$

We have omitted these expressions in the statement of the proposition since we do not use them in the proofs of the paper. It can still be of interest to note that the functions $(m, y) \mapsto C_1(m, y)$ and $(m, y) \mapsto C_2(m, y)$ are polynomials of order p w.r.t. m, and $C_b^p(\mathbb{R})$ w.r.t. y.

Remark 2.33. *As a consequence of Proposition 2.32, it can be noticed that all the polynomials are continuous on the space $\mathcal{P}_1(\mathbb{R})$, and so, also on each space $\mathcal{P}_p(\mathbb{R})$, since p-th order Wasserstein metric is finer than D_{KR} (for any $p \geq 1$).*

Proof of Proposition 2.32. Let us fix $p \geq 2$ in all the proof. We prove, by induction on $n \in \mathbb{N}^*$ that, for any $\varphi \in C_b^p(\mathbb{R}^n)$, and G defined as

$$G : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}^n} \varphi(x) dm^{\otimes n}(x), \tag{21}$$

the function G is differentiable on $\mathcal{P}_1(\mathbb{R})$ such that: for all $m \in \mathcal{P}_1(\mathbb{R})$ and $y \in \mathbb{R}$,

$$\delta G(m,y) = \sum_{k=1}^{n} \int_{\mathbb{R}^{n-1}} \varphi(x\backslash_k y) dm^{\otimes n-1}(x\backslash_k) + C(m). \tag{22}$$

Let us begin with the case $n = 1$. So, let us consider some $\varphi \in C_b^p(\mathbb{R})$ and define

$$G : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}} \varphi(x) dm(x).$$

Since, for any $m, m_0 \in \mathcal{P}_1(\mathbb{R})$,

$$G(m) - G(m_0) = \int_{\mathbb{R}} \varphi(x) d(m - m_0)(x),$$

we know that, the function $(m,x) \mapsto \varphi(x)$ is one version of the derivative of F. Its canonical derivative therefore satisfies: for all $m \in \mathcal{P}_1(\mathbb{R}), x \in \mathbb{R}$,

$$\delta G(m,x) = \varphi(x) - G(m).$$

Now, let us fix some $n \in \mathbb{N}^*$, and assume that (22) holds true for any function G of the form (21) (for any choice of $\varphi \in C_b^p(\mathbb{R}^n)$). Let $\varphi \in C_b^p(\mathbb{R}^{n+1})$ and

$$G : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}^{n+1}} \varphi(x) dm^{\otimes n+1}(x).$$

Let us introduce

$$H : (m,y) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \longmapsto \int_{\mathbb{R}^n} \varphi(x,y) dm^{\otimes n}(x),$$

such that, for all $m \in \mathcal{P}_1(\mathbb{R})$,

$$G(m) = \int_{\mathbb{R}} H(m,y) dm(y).$$

Now, we prove that G is differentiable using Corollary 2.26. Firstly, by induction hypothesis, applied twice successively, for all $y \in \mathbb{R}$, the function H_y is twice differentiable on $\mathcal{P}_1(\mathbb{R})$, with: for all $m \in \mathcal{P}_1(\mathbb{R}), z_1, z_2 \in \mathbb{R}$,

$$\delta^2 H_y(m,z_1,z_2) = \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} \int_{\mathbb{R}^{n-2}} \varphi(x\backslash_{(k,l)}(z_1,z_2), y) dm^{\otimes n-2}(x\backslash_{(k,l)}) + C_y(m,z_1) + C_y(m,z_2).$$

Then,

$$\partial^2_{z_1 z_2} \delta^2 H_y(m,z_1,z_2) = \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} \int_{\mathbb{R}^{n-2}} \partial^2_{kl} \varphi(x\backslash_{(k,l)}(z_1,z_2), y) dm^{\otimes n-2}(x\backslash_{(k,l)}),$$

which is bounded uniformly w.r.t. (y,z_1,z_2,m) (recalling that $\varphi \in C_b^p(\mathbb{R}^{n+1})$ with $p \geq 2$). Hence, the function H satisfies Condition (i) of Corollary 2.26.

Besides, by differentiation under the integral sign, we have, for all $m \in \mathcal{P}_1(\mathbb{R}), y \in \mathbb{R}$,

$$\partial_y H_y(m) = \int_{\mathbb{R}^n} \partial_n \varphi(x, y) dm^{\otimes n}(x).$$

And, by induction hypothesis and differentiating again under the integral sign, for all $m \in \mathcal{P}_1(\mathbb{R})$ and $y, z \in \mathbb{R}$,

$$\partial_z \delta \left(\partial_y H_y\right)(m, z) = \sum_{k=1}^{n} \int_{\mathbb{R}^{n-1}} \partial^2_{kn} \varphi \left(x \backslash_k z, y\right) dm^{\otimes n-1}(x),$$

which is bounded uniformly w.r.t. (y, z, m). So H also satisfies Condition (ii) of Corollary 2.26.

Then, by Corollary 2.26, the function G is differentiable, and for all $m \in \mathcal{P}_1(\mathbb{R}), y \in \mathbb{R}$,

$$\begin{aligned}
\delta G(m, y) &= H(m, y) + \int_{\mathbb{R}} \delta H_z(m, y) dm(z) - G(m) \\
&= \int_{\mathbb{R}^n} \varphi(x, y) dm^{\otimes n}(x) + \sum_{k=1}^{n} \int_{\mathbb{R}} \int_{\mathbb{R}^{n-1}} \varphi(x \backslash_k y, z) dm^{\otimes n-1}(x \backslash_k) dm(z) + C(m) \\
&= \int_{\mathbb{R}^n} \varphi(x, y) dm^{\otimes n}(x) + \sum_{k=1}^{n} \int_{\mathbb{R}^n} \varphi(x \backslash_k y) dm^{\otimes n}(x \backslash_k) + C(m),
\end{aligned}$$

which is exactly (22).

Now, one can note that, thanks to (22), for any $G \in \mathscr{P}^p$ (for $p \geq 2$), for any $x \in \mathbb{R}$, the function $(\delta G)_x$ still belongs to $\mathscr{P}^p$ and the function $\partial_x (\delta G)_x$ belongs to $\mathscr{P}^{p-1}$. This implies that $\mathscr{P}^p$ is included in $C^{\infty,0}(\mathcal{P}_1(\mathbb{R}))$. All the other statements of the proposition are a direct consequence of this result and the fact that, by definition, any $G \in \mathscr{P}^p$ can be written as in (21) with $\varphi \in C_b^p(\mathbb{R}^n)$. □

Remark 2.34. *Similarly as it was noted in Remark 2.29, it is possible to differentiate measure-variable functions that include all the functions of the forms given at Corollaries 2.26, 2.27 and Proposition 2.32, considering*

$$G : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}^d} H(m, x) dm^{\otimes d}(x).$$

However, such a result would need particularly technical and strong hypothesis that what is actually needed in this paper.

3. Markov theory of measure-valued processes

The aim of this section is to study the Markov properties of two kind of measure-valued Markov processes: the conditional laws of McKean-Vlasov processes (cf Section 3.2) and the empirical measures of some McKean-Vlasov particle systems (cf Section 3.4). Before studying these two kinds of processes, we need to introduce some definitions and some useful lemmas.

3.1. General results

Let us begin by introducing the notion of semigroups and generators that we use in the paper. This section is written for E-valued Markov processes, and is applied for $E = \mathcal{P}_p(\mathbb{R})$. It is assumed in all the section that E is a Polish space (the fact that $\mathcal{P}_p(\mathbb{R})$ is Polish is guaranteed by Theorem 6.18 of Villani (2009)). The notion of semigroup is the usual one, and the notion of infinitesimal generator has to be understood in a sense of martingale problem. Notice that these are the definitions used in Meyn and Tweedie (1993), and also in Erny, Löcherbach and Loukianova (2022b) only in the case $E = \mathbb{R}$. The results of this section generalize the ones of Appendix A and B of Erny, Löcherbach and Loukianova (2022b).

Definition 3.1. *Let E be a Polish space, and $(X_t)_t$ be some time homogeneous E-valued Markov process w.r.t. some filtration $(\mathcal{F}_t)_t$. Let $\mathcal{D}_{\mathcal{S}}(X)$ denote the domain of the semigroup of X:*

$$\mathcal{D}_{\mathcal{S}}(X) = \{g : E \to \mathbb{R} \text{ measurable } : \ \forall t > 0, \mathbb{E}_x\left[|g(X_t)|\right] < \infty\}.$$

The semigroup $(P_t)_t$ of $(X_t)_t$ is a family of operators defined by: for any $g \in \mathcal{D}_{\mathcal{S}}(X)$, for all $t \geq 0$,

$$P_t g : x \in E \longmapsto \mathbb{E}_x\left[g(X_t)\right],$$

with $\mathbb{E}_x$ the expectation related to the probability measure $\mathbb{P}_x$ under which $X_0 = x$.

Let $\mathcal{D}_{\mathcal{G}}(X)$ denote the domain of the generator of X: $g \in \mathcal{D}_{\mathcal{G}}(X)$ if and only if $g \in \mathcal{D}_{\mathcal{S}}(X)$ and there exists $h_g \in \mathcal{D}_{\mathcal{S}}(X)$ such that, for all $t \geq 0$,

$$\begin{cases} \mathbb{E}_x\left[\int_0^t |h_g(X_s)|ds\right] < \infty, \\ P_t g(x) - g(x) - \int_0^t P_s h_g(x) ds = 0. \end{cases}$$

In this case, the generator A of $(X_t)_t$ is the operator defined on $\mathcal{D}_{\mathcal{G}}(X)$ as $Ag = h_g$.

Note that, thanks to Remark 2.33, any polynomial is continuous (hence measurable) w.r.t. the topology of $\mathcal{P}_p(\mathbb{R})$, for all $p \geq 1$. In this paper, these are the only functions that are considered in the domains of the generators of the Markov processes.

Remark 3.2. *In this paper, we give the expressions of the generators for two classes of measure-valued Markov processes. In Theorem 3.6, we study the generator of processes defined as conditional laws of solutions of McKean-Vlasov Ito-SDEs, conditionally on a Brownian motion. In Theorem 3.10, we give a similar result for processes defined as empirical measures of particle systems whose dynamics are driven by McKean-Vlasov Ito-SDEs.*

We can now state a criterion for a $\mathcal{P}_p(\mathbb{R})$-valued process to be Markov. Since it is a direct consequence of the fact that the polynomials form a separating class for $\mathcal{P}_p(\mathbb{R})$ (i.e. Proposition 2.31), the proof is omitted.

Proposition 3.3. *Let $p \geq 1$ and $(\mu_t)_t$ be a $\mathcal{P}_p(\mathbb{R})$-valued process adapted to some filtration $(\mathcal{F}_t)_t$. Assume that, for any polynomial $F \in \mathscr{P}^\infty$, for all $0 \leq s \leq t$,*

$$\mathbb{E}\left[F(\mu_t)|\,\mathcal{F}_s\right] = \mathbb{E}\left[F(\mu_t)|\,\mu_s\right].$$

Then $(\mu_t)_t$ is a Markov process w.r.t. the filtration $(\mathcal{F}_t)_t$.

The exhaustive study of Markov theory for measure-valued processes is beyond the scope of this section. The aim is only to prove the following result stating a classical Trotter-Kato formula under some ad hoc assumptions.

Proposition 3.4. *Let $(\bar{P}_t)_t$ and $(P^N_t)_t$ (resp. $\bar{A}$ and A^N) be the semigroups (resp. generators) of two E-valued Markov processes $(\bar{X}_t)_t$ and $(X^N_t)_t$, let $\mathcal{A}$ be a subset of $\mathcal{C}(E)$ and $x \in E$. Assume that:*

1. *for all $g \in \mathcal{A}$ and $t \geq 0$, the function $y \in E \mapsto \bar{P}_t g(y)$ still belongs to $\mathcal{A}$,*
2. *the set $\mathcal{A}$ is included in the domains of the generators $\bar{A}$ and A^N,*
3. *for any $g \in \mathcal{A}$, the three functions*

$$t \mapsto \bar{P}_t \bar{A} g(x) \ , \quad t \mapsto P^N_t A^N g(x) \text{ and } t \mapsto P^N_t \bar{A} g(x)$$

   are continuous,
4. *and, for all $T \geq 0$, $g \in \mathcal{A}$ and $t \in [0,T]$,*

$$\sup_{r \leq T} \left| P^N_r \bar{A} \left(\bar{P}_t - \bar{P}_s \right) g(x) \right| + \sup_{r \leq T} \left| P^N_r A^N \left(\bar{P}_t - \bar{P}_s \right) g(x) \right| \underset{s \to t}{\longrightarrow} 0.$$

Then, for all $t \geq 0$, $g \in \mathcal{A}$,

$$\left(\bar{P}_t g - P^N_t g \right)(x) = \int_0^t P^N_{t-s} \left(\bar{A} - A^N \right) \bar{P}_s g(x) ds. \tag{23}$$

The proof of Proposition 3.4 requires the following result. Its classical, so its proof is omitted (see e.g. Remark A.2 and Proposition A.3 of Erny, Löcherbach and Loukianova (2022b)).

Lemma 3.5. *Let $(X_t)_t$ be some E-valued Markov process with semigroup $(P_t)_t$ and generator A (in the sense of Definition 3.1). For any $g \in \mathcal{D}_{\mathcal{G}}(X)$ and $x \in E$, if the function $t \mapsto P_t A g(x)$ is continuous on $\mathbb{R}_+$, then, for all $t \geq 0$,*

$$\partial_t \left(P_t g(x) \right) = P_t A g(x).$$

In addition, if for all $t \geq 0$, $P_t g \in \mathcal{D}_{\mathcal{G}}(X)$, then

$$\partial_t \left(P_t g(x) \right) = A P_t g(x) = P_t A g(x).$$

Proof of Proposition 3.4. Let us fix in all the proof some $g \in \mathcal{A}$, $t > 0$ and define

$$\Phi : s \in [0,t] \longmapsto P^N_{t-s} \bar{P}_s G(x).$$

By Lemma 3.5 and conditions 1-3 of Proposition 3.4, the function Φ is differentiable with, for all $s \in [0,t]$,

$$\begin{aligned} \Phi'(s) =& - \partial_u \left(P^N_u \bar{P}_s G(x) \right)_{|u=t-s} + \partial_v \left(P^N_{t-s} \bar{P}_v G(x) \right)_{|v=s} \\ =& - P^N_{t-s} A^N \bar{P}_s G(x) + P^N_{t-s} \bar{P}_s \bar{A} G(x) \\ =& P^N_{t-s} \left(\bar{A} - A^N \right) \bar{P}_s G(x). \end{aligned}$$

The only property required to prove the result of Proposition 3.4 is that Φ' is continuous on $[0,t]$. Indeed, it would imply (by the second fundamental theorem of calculus),

$$\Phi(t) - \Phi(0) = \int_0^t \Phi'(s)ds, \tag{24}$$

which is exactly (23).

To prove the continuity of Φ', let us fix, for the remainder of the proof, some $s \in [0,t]$ and some sequence $(s_k)_k$ converging to s. Then

$$\begin{aligned} |\Phi'(s) - \Phi'(s_k)| \leq & \left|P^N_{t-s}\left(\bar{A} - A^N\right)\bar{P}_s G(x) - P^N_{t-s_k}\left(\bar{A} - A^N\right)\bar{P}_s G(x)\right| \\ & + \left|P^N_{t-s_k}\left(\bar{A} - A^N\right)\bar{P}_s G(x) - P^N_{t-s_k}\left(\bar{A} - A^N\right)\bar{P}_{s_k} G(x)\right| \\ =: & A_k + B_k. \end{aligned} \tag{25}$$

Since

$$A_k \leq \left|\left(P^N_{t-s} - P^N_{t-s_k}\right)\bar{A}\bar{P}_s g(x)\right| + \left|\left(P^N_{t-s} - P^N_{t-s_k}\right)A^N\bar{P}_s g(x)\right|,$$

we know, by conditions 1-3 that A_k vanishes as k goes to infinity.

On the other hand,

$$B_k \leq \left|P^N_{t-s_k}\bar{A}\left(\bar{P}_s - \bar{P}_{s_k}\right)g(x)\right| + \left|P^N_{t-s_k}A^N\left(\bar{P}_s - \bar{P}_{s_k}\right)g(x)\right|,$$

also vanishes as k goes to infinity by condition 4. Consequently (25) implies that Φ' is continuous on $[0,t]$, implying that (24) holds true, which proves the proposition. □

3.2. Conditional law of diffusion

Let $(\bar{X}_t)_t$ be defined as the solution of the following SDE

$$\begin{aligned} d\bar{X}_t = & b\left(\bar{\mu}_t, \bar{X}_t\right)dt + \sigma\left(\bar{\mu}_t, \bar{X}_t\right)dB_t + \varsigma\left(\bar{\mu}_t, \bar{X}_t\right)dW_t \\ & + h\left(\bar{\mu}_{t-}, \bar{X}_{t-}\right)\int_{\mathbb{R}_+} \mathbb{1}_{\left\{z \leq f\left(\bar{\mu}_{t-}, \bar{X}_{t-}\right)\right\}} d\pi(t,z), \end{aligned} \tag{26}$$

where B, W are standard Brownian motions of dimension one, π a Poisson measure on $\mathbb{R}^2_+$ with Lebesgue intensity such that B, W and π are mutually independent, and

$$\bar{\mu}_t = \mathcal{L}\left(\bar{X}_t \mid \mathcal{W}_t\right),$$

with $\mathcal{W}_t = \sigma(W_s \ : \ s \leq t)$ the filtration of the Brownian motion W at time t.

Theorem 3.6. *Grant Assumption 1. The process $(\bar{\mu}_t)_t$ is a $\mathcal{P}_2(\mathbb{R})$-valued Markov process. The domain $\mathcal{D}_{\mathcal{G}}(\bar{\mu})$ contains the set $\mathscr{P}^2$, and the generator $\bar{A}$ of $\bar{\mu}$ satisfies: for any $G \in \mathscr{P}^2$ and $m \in \mathcal{P}_2(\mathbb{R})$,*

$$\begin{aligned} \bar{A}G(m) = & \int_{\mathbb{R}} \bar{L}(\delta G(m, \bullet))(m,x)dm(x) \\ & + \frac{1}{2}\int_{\mathbb{R}^2}\left(\partial^2_{xy}\delta^2 G(m,x,y)\right)\varsigma(m,x)\varsigma(m,y)dm^{\otimes 2}(x,y), \end{aligned} \tag{27}$$

where $\bar{L}$ is the operator defined as: for $g \in C^2(\mathbb{R}), x \in \mathbb{R}, m \in \mathcal{P}_1(\mathbb{R})$,

$$\begin{aligned}\bar{L}g(m,x) =& b(m,x)g'(x) + \frac{1}{2}\left(\sigma(m,x)^2 + \varsigma(m,x)^2\right)g''(x)\\ &+ f(m,x)\left[g\left(x + h(m,x)\right) - g(x)\right].\end{aligned}$$

It can be noticed that the operator $\bar{L}$ is the generator of the process $\bar{X}$ in the "martingale problem" sense. The process $\bar{X}$ itself being a non-Markov semimartingale we prefer not to call $\bar{L}$ a generator to avoid confusion.

Remark 3.7. *The expression of the generator $\bar{A}$ is a reminiscent of classical expressions for generators of $\mathbb{R}$-valued Markov processes. Indeed, the drift terms of the generator corresponds to the first order derivative of the test-function G (where it can be noted that, the Brownian and jump terms of the $\mathbb{R}$-valued process $\bar{X}$ creates some drift terms for the measure-valued process $\bar{\mu}$ since they are "averaged" by $\bar{\mu}$). Besides, the Brownian motion W being still a noise source for $\bar{\mu}$, it creates a Brownian term even in the dynamics of $\bar{\mu}$ which corresponds to the second order derivative of G. The expression at (27) is compact because the first term is written as an operator acting on the first order derivative of the test-function. This term would be harder to write with the notion of Lions-derivative instead of the linear derivative that we use. However, one can note that the second term would be a bit more compact with Lions-derivative instead of the linear derivative (by Corollary 2.25).*

The proof of Theorem 3.6 uses the same ideas as in Guo, Pham and Wei (2023) and Cox et al. (2021): using Ito's formula for $\mathbb{R}^n$-valued processes, and then integrating the result. In both references, the authors extend the expressions of their measure-valued Ito's formulas to sufficiently smooth test-functions. This is done by approximating smooth measure-variable functions with cylinder functions. It is not clear that the approximation scheme of Cox et al. (2021) can be adapted to our framework, whereas the one of Guo, Pham and Wei (2023) seems to be adaptable. This paper being sufficiently long and technical, we prefer to omit the use of approximation schemes, and restrict the result of Theorem 3.6 (and Theorem 3.10) for measure-variable polynomial test-functions.

Proof of Theorem 3.6. Let

$$G : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}^n} \varphi(x) dm^{\otimes n}(x),$$

with $n \in \mathbb{N}^*$ and $\varphi \in C_b^2(\mathbb{R}^n)$.

For a fixed $m \in \mathcal{P}_2(\mathbb{R})$, let us introduce, for all $1 \leq k \leq n$, the process $\bar{X}^k$ solution to

$$\begin{aligned}\bar{X}_t^k =& \bar{X}_0^k + \int_0^t b(\bar{\mu}_s, \bar{X}_s^k)ds + \int_0^t \sigma(\bar{\mu}_s, \bar{X}_s^k)dB_s^k + \int_0^t \varsigma(\bar{\mu}_s, \bar{X}_s^k)dW_s\\ &+ \int_{[0,t]\times\mathbb{R}_+} h(\bar{\mu}_{s-}, \bar{X}_{s-}^k)\mathbb{1}_{\left\{z \leq f(\bar{\mu}_{s-}, \bar{X}_{s-}^k)\right\}} d\pi^k(s,z),\end{aligned} \tag{28}$$

with $\bar{X}_0^1, ..., \bar{X}_0^n$ i.i.d. m-distributed, $B^1, ..., B^n, W$ independent Brownian motions, and $\pi^1, ..., \pi^n$ independent Poisson measures on $\mathbb{R}_+^2$ with Lebesque intensity, such that all these objects are mutually independent. The random measure $\bar{\mu}_t$ is the conditional law of any of the process given the filtration of W up to time t: $\bar{\mu}_t = \mathcal{L}(\bar{X}_t^1 | \mathcal{W}_t)$. The interest of this construction is to guarantee the following property: for all $t \geq 0$,

$$\text{conditionally on } \bar{\mu}_t, \quad \bar{X}_t^k \ (1 \leq k \leq n) \text{ are i.i.d. } \bar{\mu}_t\text{-distributed.} \tag{29}$$

Let us denote $\bar{\mathbf{X}}_t = (\bar{X}_t^1, ..., \bar{X}_t^n)$, and, for $1 \leq k \leq n$, $\bar{X}^{k,c}$ the continuous part of the semi-martingale $\bar{X}^k$. By Ito's formula, for all $t \geq r \geq 0$,

$$\begin{aligned}
\varphi(\bar{\mathbf{X}}_t) =& \varphi(\bar{\mathbf{X}}_r) + \sum_{k=1}^n \int_r^t \partial_k \varphi(\bar{\mathbf{X}}_s) d\bar{X}_s^{k,c} + \frac{1}{2} \sum_{k,l=1}^n \int_r^t \partial_{kl}^2 \varphi(\bar{\mathbf{X}}_s) d \left\langle \bar{X}^{k,c}, \bar{X}^{l,c} \right\rangle_s \\
&+ \sum_{r<s\leq t} \left(\varphi(\bar{\mathbf{X}}_t) - \varphi(\bar{\mathbf{X}}_{t-}) \right) \\
=& \varphi(\bar{\mathbf{X}}_r) + \sum_{k=1}^n \int_r^t \partial_k \varphi(\bar{\mathbf{X}}_s) b(\bar{\mu}_s, \bar{X}_s^k) ds + \frac{1}{2} \sum_{k=1}^n \int_r^t \sigma(\bar{\mu}_s, \bar{X}_s^k)^2 \partial_{kk}^2 \varphi(\bar{\mathbf{X}}_s) ds \\
&+ \frac{1}{2} \sum_{k,l=1}^n \int_r^t \varsigma(\bar{\mu}_s, \bar{X}_s^k) \varsigma(\bar{\mu}_s, \bar{X}_s^l) \partial_{kl}^2 \varphi(\bar{\mathbf{X}}_s) ds \\
&+ \sum_{k=1}^n \int_r^t \partial_k \varphi(\bar{\mathbf{X}}_s) \sigma(\bar{\mu}_s, \bar{X}_s^k) dB_s^k + \sum_{k=1}^n \int_r^t \partial_k \varphi(\bar{\mathbf{X}}_s) \varsigma(\bar{\mu}_s, \bar{X}_s^k) dW_s \\
&+ \sum_{k=1}^n \int_{]r,t]\times\mathbb{R}_+} \mathbb{1}_{\left\{z \leq f(\bar{\mu}_{s-}, \bar{X}_{s-}^k)\right\}} \\
&\qquad \left[\varphi\left(\bar{\mathbf{X}}_{s-} + h(\bar{\mu}_{s-}, \bar{X}_{s-}^k) \cdot e_k \right) - \varphi\left(\bar{\mathbf{X}}_{s-} \right) \right] d\pi^k(s,z),
\end{aligned}$$

with $e_k = (\mathbb{1}_{\{k=l\}})_{1\leq l\leq n} \in \mathbb{R}^n$. Then, thanks to the calculation above, the property (29), and Lemmas D.1 and D.2,

$$\begin{aligned}
G(\bar{\mu}_t) =& G(\bar{\mu}_r) + \sum_{k=1}^n \int_r^t \int_{\mathbb{R}^n} \partial_k \varphi(x) b(\bar{\mu}_s, x_k) d\bar{\mu}_s^{\otimes n}(x) ds \\
&+ \frac{1}{2} \sum_{k=1}^n \int_r^t \int_{\mathbb{R}^n} \sigma(\bar{\mu}_s, x_k)^2 \partial_{kk}^2(x) d\bar{\mu}_s^{\otimes n}(x) ds \\
&+ \frac{1}{2} \sum_{k,l=1}^n \int_r^t \int_{\mathbb{R}^n} \varsigma(\bar{\mu}_s, x_k) \varsigma(\bar{\mu}_s, x_l) \partial_{kl}^2 \varphi(x) d\bar{\mu}_s^{\otimes n}(x) ds \\
&+ \sum_{k=1}^n \int_r^t \int_{\mathbb{R}^n} \partial_k \varphi(x) \varsigma(\bar{\mu}_s, x_k) d\bar{\mu}_s^{\otimes n}(x) dW_s \\
&+ \sum_{k=1}^n \int_r^t \int_{\mathbb{R}^n} f(\bar{\mu}_s, x_k) \left[\varphi\left(x + h(\bar{\mu}_s, x_k) \cdot e_k \right) - \varphi(x) \right] d\bar{\mu}_s^{\otimes n}(x) ds \\
&+ \sum_{k=1}^n \int_{]r,t]\times\mathbb{R}_+} \int_{\mathbb{R}^n} \mathbb{1}_{\{z \leq f(\bar{\mu}_{s-}, x_k)\}} \\
&\qquad \left[\varphi\left(x + h(\bar{\mu}_{s-}, x_k) \cdot e_k \right) - \varphi\left(x \right) \right] d\bar{\mu}_{s-}^{\otimes n}(x) d\tilde{\pi}^k(s,z),
\end{aligned}$$

with $\tilde{\pi}^k$ the compensated version of π^k:

$$d\tilde{\pi}^k(t,z) = d\pi^k(t,z) - dt \cdot dz.$$

Then, using the computation above and Proposition 3.3, we know that the process $(\bar{\mu}_t)_t$ is a Markov process. In addition, by (3) and the fact that ς and f are sublinear, we know that the stochastic integrals w.r.t. W and $\tilde{\pi}^k$ above are real martingales. So the function G belongs to $\mathcal{D}_{\mathcal{G}}(\bar{\mu})$, and

$$\begin{aligned}\bar{A}G(m) =& \sum_{k=1}^{n} \int_{\mathbb{R}^n} \partial_k \varphi(x) b(m, x_k) dm^{\otimes n}(x) + \frac{1}{2} \sum_{k=1}^{n} \int_{\mathbb{R}^n} \sigma(m, x_k)^2 \partial^2_{kk} \varphi(x) dm^{\otimes n}(x) \\ &+ \frac{1}{2} \sum_{k,l=1}^{n} \int_{\mathbb{R}^n} \varsigma(m, x_k) \varsigma(m, x_l) \partial^2_{kl} \varphi(x) dm^{\otimes n}(x) \\ &+ \sum_{k=1}^{n} \int_{\mathbb{R}^n} f(m, x_k) \left[\varphi\left(x + h(m, x_k) \cdot e_k\right) - \varphi(x) \right] dm^{\otimes n}(x).\end{aligned} \tag{30}$$

Then, by Proposition 2.32, for all $x, y \in \mathbb{R}, m \in \mathcal{P}_1(\mathbb{R})$,

$$\delta G(m, x) = \sum_{k=1}^{n} \int_{\mathbb{R}^{n-1}} \varphi\left(y\backslash_k x\right) dm^{\otimes n-1}(y\backslash_k) + C_0(m), \tag{31}$$

$$\delta^2 G(m, x, y) = \sum_{\substack{k,l=1 \\ k \neq l}}^{n} \int_{\mathbb{R}^{n-2}} \varphi(z\backslash_{(k,l)}(x, y)) dm^{\otimes n-2}(z\backslash_{(k,l)}) + C_1(m, x) + C_2(m, y), \tag{32}$$

with $C_2(m, y)$ independent of x, $C_1(m, x)$ independent of y, and $C_0(m)$ independent of both x and y.

Consequently,

$$\begin{aligned}\partial_x \delta G(m, x) =& \sum_{k=1}^{n} \int_{\mathbb{R}^{n-1}} \partial_k \varphi\left(z\backslash_k x\right) dm^{\otimes n-1}(z\backslash_k), \\ \int_{\mathbb{R}} b(m, x) \partial_x \delta G(m, x) dm(x) =& \sum_{k=1}^{n} \int_{\mathbb{R}} \int_{\mathbb{R}^{n-1}} b(m, x) \partial_k \varphi\left(y\backslash_k x\right) dm^{\otimes n-1}(y\backslash_k) dm(x) \\ =& \sum_{k=1}^{n} \int_{\mathbb{R}^n} b(m, y_k) \partial_k \varphi(y) dm^{\otimes n}(y).\end{aligned}$$

Using the same trick on the other terms of (30), we have

$$\begin{aligned}\int_{\mathbb{R}} \sigma(m, x)^2 \partial^2_{xx} \delta G(m, x) dm(x) =& \sum_{k=1}^{n} \int_{\mathbb{R}^n} \sigma(m, y_k)^2 \partial^2_{kk} \varphi(y) dm^{\otimes n}(y), \\ \int_{\mathbb{R}} \varsigma(m, x)^2 \partial^2_{xx} \delta G(m, x) dm(x) =& \sum_{k=1}^{n} \int_{\mathbb{R}^n} \varsigma(m, y_k)^2 \partial^2_{kk} \varphi(y) dm^{\otimes n}(y), \\ \int_{\mathbb{R}^2} \varsigma(m, x) \varsigma(m, y) \partial^2_{xy} \delta^2 G(m, x, y) dm^{\otimes 2}(x, y) =& \sum_{\substack{k,l=1 \\ k \neq l}}^{n} \int_{\mathbb{R}^n} \varsigma(m, z_k) \varsigma(m, z_l) \partial^2_{kl} \varphi(z) dm^{\otimes n}(z).\end{aligned}$$

As a consequence, it is possible to rewrite (30) exactly as (27). □

3.3. Regularity of semigroups of conditional laws of diffusions

In this section, we only consider a particular case of the SDE (26) without the jump term:

$$d\bar{X}_t = b(\bar{\mu}_t, \bar{X}_t)dt + \sigma(\bar{\mu}_t, \bar{X}_t)dB_t + \varsigma(\bar{\mu}_t, \bar{X}_t)dW_t, \tag{33}$$

where B, W are still independent Brownian motions, and $\bar{\mu}_t = \mathcal{L}(\bar{X}_t|\mathcal{W}_t)$. In addition, for any $x \in \mathbb{R}$, let $(\bar{X}_t^{(x)})_t$ be the (only) strong solution of (33) with initial condition $\bar{X}_0^{(x)} = x$.

Lemma 3.8. *Assume that the functions b, σ, ς admit fifth order mixed derivatives such that the mixed derivatives of orders from one to five are bounded. Then, there exists some $T > 0$ such that, for all $0 \leq t \leq T$, the function*

$$x \in \mathbb{R} \longmapsto \bar{X}_t^{(x)}$$

belongs to $C_b^4(\mathbb{R})$.

In addition, for all $T > 0$, $p \in \mathbb{N}^*$, and $1 \leq k \leq 4$,

$$\sup_{x\in\mathbb{R}} \mathbb{E}\left[\sup_{t\leq T} \left|\partial_{x^k}^k \bar{X}_t^{(x)}\right|^p\right] < \infty.$$

The proof of the previous lemma uses classical technics (see e.g. the proof of Lemma 4.17 of Chassagneux, Crisan and Delarue (2022)), but due to the particularity of our framework, we prefer to write an explicit proof at Appendix B.

Proposition 3.9. *Let $\bar{P}$ be the semigroup of the process $\bar{X}$ defined at (33). Under the assumption of Lemma 3.8, for any $G \in \mathscr{P}^4$ and $t \geq 0$,*

(a) *the function*

$$m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \bar{P}_t G(m)$$

belongs to $\mathscr{P}^4$,

(b) *there exists $C_t > 0$ independent of G such that, for all $0 \leq k \leq 4$,*

$$\sup_{s\leq t} ||\bar{P}_s G||_k \leq C_t ||G||_k,$$

where $||\bullet||_k$ is introduced in Definition 2.14,

(c) *for all $m \in \mathcal{P}_1(\mathbb{R})$ and $x, y \in \mathbb{R}$, the following functions are continuous*

$$\begin{aligned}
&t \in \mathbb{R}_+ \longmapsto \bar{P}_t G(m),\\
&t \in \mathbb{R}_+ \longmapsto \delta\left(\bar{P}_t G\right)(m, x),\\
&t \in \mathbb{R}_+ \longmapsto \partial_x \delta\left(\bar{P}_t G\right)(m, x),\\
&t \in \mathbb{R}_+ \longmapsto \partial_{xx}^2 \delta\left(\bar{P}_t G\right)(m, x),\\
&t \in \mathbb{R}_+ \longmapsto \partial_{xy}^2 \delta^2\left(\bar{P}_t G\right)(m, x, y).
\end{aligned}$$

Proof. Let $G \in \mathscr{P}^4$ satisfying for all $m \in \mathcal{P}_1(\mathbb{R})$,

$$G(m) = \int_{\mathbb{R}^n} \varphi(x) dm^{\otimes n}(x),$$

with $n \in \mathbb{N}^*$ and $\varphi \in C_b^4(\mathbb{R}^n)$.

Let us prove Item (a). For all $t \geq 0$,

$$G(\bar{\mu}_t) = \mathbb{E}\left[\varphi(\bar{X}_t^1, ..., \bar{X}_t^n) \middle| \mathcal{W}_t\right],$$

where the processes $\bar{X}^k$ $(1 \leq k \leq n)$ are solutions to (33) w.r.t. the same W, but independent B^k and $\bar{X}_0^k$ (similarly as in (28)). Since the variables $\bar{X}_0^1, ..., \bar{X}_0^n$ are i.i.d. $\bar{\mu}_0$-distributed, we have, for any $m \in \mathcal{P}_1(\mathbb{R})$,

$$\bar{P}_t G(m) = \mathbb{E}_m\left[G(\bar{\mu}_t)\right] = \int_{\mathbb{R}^n} \mathbb{E}\left[\varphi\left(\bar{X}_t^{1,(x_1)}, ..., \bar{X}_t^{n,(x_n)}\right)\right] dm^{\otimes n}(x) = \int_{\mathbb{R}^n} \psi_t(x) dm^{\otimes n}(x), \tag{34}$$

with

$$\psi_t : x \in \mathbb{R}^n \longmapsto \mathbb{E}\left[\varphi\left(\bar{X}_t^{1,(x_1)}, ..., \bar{X}_t^{n,(x_n)}\right)\right],$$

and $\bar{X}^{k,(x_k)}$ the process $\bar{X}^k$ starting from initial condition $\bar{X}_0^{k,(x_k)} = x_k$ $(1 \leq k \leq n)$. Then, by Lemma 3.8 and by differentiation under the integral sign (using uniform integrability condition, like, for example, in Lemma 6.1 of Eldredge (2018)), there exists some $T > 0$ such that, for all $t \in [0, T]$, the function ψ_t belongs to $C_b^4(\mathbb{R}^n)$. This proves that, for all $0 \leq t \leq T$, $\bar{P}_t G$ belongs to $\mathscr{P}^4$.

Then, since the process $(\bar{\mu}_t)_t$ is Markov, we have: for all $t \in [T, 2T]$, for all $m \in \mathcal{P}_1(\mathbb{R})$,

$$\bar{P}_t G(m) = \bar{P}_{t-T}\left(\bar{P}_T G\right)(m).$$

Whence, since $\bar{P}_T G \in \mathscr{P}^4$ and $t - T \in [0, T]$, the above equality implies that, for all $T \leq t \leq 2T$, the function $\bar{P}_t G$ also belongs to $\mathscr{P}^4$. Iterating this argument concludes the proof.

The proof of Item (b) follows from the fact that we have the explicit expression of $\bar{P}_t G$ at (34) and the control given at Lemma 3.8 (recalling Proposition 2.32). The proof of Item (c) relies on the same arguments as for Item (b) with the additional use of the Dominated Convergence Theorem. □

3.4. Empirical measure of particle system

Let $N \in \mathbb{N}^*$, and $X^{N,k}$ $(1 \leq k \leq N)$ be solutions to the following SDEs:

$$\begin{aligned} dX_t^{N,k} =& b\left(\mu_t^N, X_t^{N,k}\right) dt + \sigma\left(\mu_t^N, X_t^{N,k}\right) dB_t^k + \varsigma\left(\mu_t^N, X_t^{N,k}\right) dW_t \\ &+ \sum_{l=1}^{N} \int_{\mathbb{R}_+ \times \mathbb{R}} h\left(\mu_{t-}^N, X_{t-}^{N,l}, u\right) \cdot \mathbb{1}_{\left\{z \leq f\left(\mu_{t-}^N, X_{t-}^{N,l}\right)\right\}} d\pi^l(t, z, u), \end{aligned} \tag{35}$$

where $B^1, ..., B^N, W$ are standard Brownian motions, $\pi^1, ..., \pi^N$ are Poisson measures on $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$ with intensity $dt \cdot dz \cdot d\nu(u)$ (with ν a probability measure on $\mathbb{R}$), such that they are all mutually independent, and

$$\mu_t^N = \frac{1}{N} \sum_{l=1}^{N} \delta_{X_t^{N,l}}.$$

Theorem 3.10. *Assume that the functions b, σ, ς are Lipschitz continuous, and that there exists $C > 0$ such that for all $x_1, x_2 \in \mathbb{R}$, and $m_1, m_2 \in \mathcal{P}_1(\mathbb{R})$,*

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} \left| h(m_1, x_1, u) \mathbb{1}_{\{z \leq f(m_1, x_1)\}} - h(m_2, x_2, u) \mathbb{1}_{\{z \leq f(m_2, x_2)\}} \right| dz\, d\nu(u) \\ \leq C\left(|x_1 - x_2| + D_{KR}(m_1, m_2)\right). \end{aligned}$$

Then, the process $(\mu_t^N)_t$ is a $\mathcal{P}_2(\mathbb{R})$-valued Markov Process. The domain $\mathcal{D}_{\mathcal{G}}(\mu^N)$ contains the set $\mathscr{P}^2$, and the generator A^N of μ^N satisfies: for any $G \in \mathscr{P}^2$, and $m \in \mathcal{P}_2(\mathbb{R})$,

$$\begin{aligned} A^N G(m) = &\int_{\mathbb{R}} \left(b(m,x)\partial_x \delta G(m,x) + \frac{1}{2}\left[\sigma(m,x)^2 + \varsigma(m,x)^2\right] \partial^2_{xx}\delta G(m,x)\right) dm(x) \\ &+ N \int_{\mathbb{R}}\int_{\mathbb{R}} f(m,x)\left[G(\mathcal{S}(m,h(m,x,u))) - G(m)\right] d\nu(u) dm(x) \\ &+ \frac{1}{2}\int_{\mathbb{R}^2} \partial^2_{xy}\delta^2 G(m,x,y)\varsigma(m,x)\varsigma(m,y) dm^{\otimes 2}(x,y) \\ &+ \frac{1}{2N}\int_{\mathbb{R}} \left(\partial^2_{y_1 y_2}\delta^2 G(m,y_1,y_2)\right)_{|y_1=y_2=x} \sigma(m,x)^2 dm(x), \end{aligned}$$

where we recall that $\mathcal{S}$ is a shift operator defined as: for $m \in \mathcal{P}(\mathbb{R}), \lambda \in \mathbb{R}$,

$$\mathcal{S}(m,\lambda) : A \in \mathcal{B}(\mathbb{R}) \longmapsto m\left(\{x - \lambda \ : \ x \in A\}\right).$$

Remark 3.11. *The same phenomenon as the one described at Remark 3.7 can be observed for μ^N: the expression for A^N can be interpreted in a similar way as the expression of $\mathbb{R}$-valued processes. The terms of the generator that depend on the first order derivatives corresponds to drift term in the dynamics of μ^N, the terms that depend on the second order derivatives to Brownian terms (here all Brownian motions B^k, W give Brownian dynamics for μ^N, since the law μ^N, being an empirical measure instead of a (conditional) law, does not make disappear the Brownian dynamics of B^k, on the contrary of the situation with $\bar{\mu}$ for the Brownian motion B). Compared to the generator of $\bar{\mu}$, there is an additional term corresponding to the jump term of the measure-valued process μ^N (whereas the process $\bar{\mu}$ is continuous in time) that depends on the shift operator $\mathcal{S}$ and on the increasing of the test-function G.*

Proof of Theorem 3.10. The fact that μ^N is a Markov process follows from Proposition 2.3.3 of Dawson (1993). It can also be proved as in Theorem 3.6 using Proposition 3.3 and the following computation that are in any case required to prove the remains of the statement of the theorem.

So, as in the proof of Theorem 3.6, let $G \in \mathscr{P}^2$ such that: for all $m \in \mathcal{P}_1(\mathbb{R})$,

$$G(m) = \int_{\mathbb{R}^n} \varphi(x) dm^{\otimes n}(x),$$

with $n \in \mathbb{N}^*$ and $\varphi \in C_b^2(\mathbb{R}^n)$.

Then,

$$G\left(\mu_t^N\right) = \frac{1}{N^n} \sum_{k_1,\dots,k_n=1}^{N} \varphi\left(X_t^{N,k_1},\dots,X_t^{N,k_n}\right).$$

For a given tuple $\mathbf{k} = (k_1,\dots,k_n) \in [\![1,N]\!]^n$, let us denote

$$\mathbf{X}_t^{N,\mathbf{k}} = \left(X_t^{N,k_1},\dots,X_t^{N,k_n}\right),$$

and $X^{N,k_i,c}$ the continuous part of the semimartingale X^{N,k_i}.

By Ito's formula, for all $t \geq r \geq 0$,

$$\begin{aligned}
\varphi\left(\mathbf{X}_t^{N,\mathbf{k}}\right) =& \varphi\left(\mathbf{X}_r^{N,\mathbf{k}}\right) + \sum_{i=1}^{n} \int_r^t \partial_i \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) dX_s^{N,k_i,c} \\
&+ \frac{1}{2} \sum_{i,j=1}^{n} \int_r^t \partial_{ij}^2 \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) d\left\langle X^{N,k_i,c}, X^{N,k_j,c} \right\rangle_s \\
&+ \sum_{r<s\leq t} \left(\varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) - \varphi\left(\mathbf{X}_{s-}^{N,\mathbf{k}}\right)\right) \\
=& \varphi\left(\mathbf{X}_r^{N,\mathbf{k}}\right) + \sum_{i=1}^{n} \int_r^t \partial_i \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) b\left(\mu_s^N, X_s^{N,k_i}\right) ds \\
&+ \sum_{i=1}^{n} \int_r^t \partial_i \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) \sigma\left(\mu_s^N, X_s^{N,k_i}\right) dB_s^{k_i} \\
&+ \sum_{i=1}^{n} \int_r^t \partial_i \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) \varsigma\left(\mu_s^N, X_s^{N,k_i}\right) dW_s \\
&+ \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij}^2 \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) \sigma\left(\mu_s^N, X_s^{N,k_i}\right) \sigma\left(\mu_s^N, X_s^{N,k_j}\right) \mathbb{1}_{\{k_i=k_j\}} ds \\
&+ \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij}^2 \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) \varsigma\left(\mu_s^N, X_s^{N,k_i}\right) \varsigma\left(\mu_s^N, X_s^{N,k_j}\right) ds \\
&+ \sum_{l=1}^{N} \int_{]r,t]\times\mathbb{R}_+\times\mathbb{R}} \mathbb{1}_{\left\{z\leq f\left(\mu_{s-}^N, X_{s-}^{N,l}\right)\right\}} \\
&\qquad \left[\varphi\left(\mathbf{X}_{s-}^{N,\mathbf{k}} + h\left(\mu_{s-}^N, X_{s-}^{N,l}, u\right)\cdot \mathbf{1}\right) - \varphi\left(\mathbf{X}_{s-}^{N,\mathbf{k}}\right)\right] d\pi^l(s,z,u),
\end{aligned}$$

with $\mathbf{1} = (1, 1, ..., 1) \in \mathbb{R}^n$. Consequently,

$$\begin{aligned} G(\mu_t^N) =& \frac{1}{N^n} \sum_{k_1,\dots,k_n=1}^{N} \varphi\left(\mathbf{X}_t^{N,\mathbf{k}}\right) \\ =& G(\mu_r^N) + M_t^{N,r} + \sum_{i=1}^{n} \int_r^t \int_{\mathbb{R}^n} \partial_i \varphi(x) b\left(\mu_s^N, x_i\right) d\left(\mu_s^N\right)^{\otimes n}(x) ds \\ &+ \frac{1}{N^n} \sum_{\mathbf{k} \in [\![1,N]\!]^n} \frac{1}{2} \sum_{i,j=1}^{n} \qquad (36) \\ &\qquad \int_r^t \partial_{ij}^2 \varphi\left(\mathbf{X}_s^{N,\mathbf{k}}\right) \sigma\left(\mu_s^N, X_s^{N,k_i}\right) \sigma\left(\mu_s^N, X_s^{N,k_j}\right) \mathbb{1}_{\{k_i = k_j\}} ds \\ &+ \frac{1}{2} \sum_{i,j=1}^{n} \int_r^t \int_{\mathbb{R}^n} \partial_{ij}^2 \varphi(x) \varsigma(\mu_s^N, x_i) \varsigma(\mu_s^N, x_j) d\left(\mu_s^N\right)^{\otimes n}(x) ds \\ &+ \sum_{l=1}^{N} \int_r^t \int_{\mathbb{R}} \int_{\mathbb{R}^n} f\left(\mu_s^N, X_s^{N,l}\right) \qquad (37) \\ &\qquad \left[\varphi\left(x + h\left(\mu_s^N, X_s^{N,l}, u\right) \cdot \mathbf{1}\right) - \varphi(x)\right] d\left(\mu_s^N\right)^{\otimes n}(x) d\nu(u) ds, \end{aligned}$$

with $(M_t^{N,r})_{t \geq r}$ some martingale (the fact that $M^{N,r}$ is not only a local martingale holds true thanks to controls like (3) and the assumption that f, σ, ς are sublinear).

Then, noticing that, for any $x_1, ..., x_N \in \mathbb{R}$ and $\lambda \in \mathbb{R}$,

$$\mathcal{S}\left(\frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, \lambda\right) = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k + \lambda},$$

it is possible to rewrite the expression at (37) as

$$\begin{aligned} N \int_r^t \int_{\mathbb{R}} \int_{\mathbb{R}^n} \int_{\mathbb{R}} & f\left(\mu_s^N, y\right) \left[\varphi\left(x + h(\mu_s^N, y, u) \cdot \mathbf{1}\right) - \varphi(x)\right] d\mu_s^N(y) d\left(\mu_s^N\right)^{\otimes n}(x) d\nu(u) ds \\ &= N \int_r^t \int_{\mathbb{R}} \int_{\mathbb{R}} f\left(\mu_s^N, y\right) \left[G\left(\mathcal{S}\left(\mu_s^N, h(\mu_s^N, y, u)\right)\right) - G\left(\mu_s^N\right)\right] d\mu_s^N(y) d\nu(u) ds, \end{aligned}$$

Besides, separating the terms $i = j$ and $i \neq j$ in the double-sum of the expression at (36), this

expression can be rewritten as

$$\begin{aligned}
&\frac{1}{2}\int_r^t \sum_{i=1}^n \frac{1}{N^n} \sum_{\mathbf{k}\in[\![1,N]\!]^n} \partial^2_{ii}\varphi(\mathbf{X}_s^{N,\mathbf{k}})\sigma\left(\mu_s^N, X_s^{N,k_i}\right)^2 ds \\
&+\frac{1}{2}\int_r^t \sum_{\substack{i,j=1\\ i\neq j}}^n \frac{1}{N}\cdot\frac{1}{N^{n-1}} \sum_{\mathbf{k}\backslash_j\in[\![1,N]\!]^{n-1}} \partial^2_{ij}\varphi(\mathbf{X}_s^{N,(\mathbf{k}\backslash_j k_i)})\sigma\left(\mu_s^N, X_s^{N,k_i}\right)^2 ds \\
&\quad=\frac{1}{2}\int_r^t \sum_{i=1}^n \int_{\mathbb{R}^n} \partial^2_{ii}\varphi(x)\sigma\left(\mu_s^N, x_i\right)^2 d\left(\mu_s^N\right)^{\otimes n}(x) ds \\
&\qquad+\frac{1}{2N}\int_r^t \sum_{\substack{i,j=1\\ i\neq j}}^n \int_{\mathbb{R}^{n-1}} \partial^2_{ij}\varphi(x\backslash_j x_i)\sigma\left(\mu_s^N, x_i\right)^2 d\left(\mu_s^N\right)^{\otimes n-1}(x\backslash_j) ds \\
&\quad=\frac{1}{2}\int_r^t \int_{\mathbb{R}} \sigma\left(\mu_s^N, x\right)^2 \partial^2_{xx}\delta G\left(\mu_s^N, x\right) d\mu_s^N(x) ds \\
&\qquad+\frac{1}{2N}\int_r^t \int_{\mathbb{R}} \sigma\left(\mu_s^N, x\right)^2 \left(\partial^2_{y_1y_2}\delta^2 G\left(\mu_s^N, y_1, y_2\right)\right)_{|y_1=y_2=x} d\mu_s^N(x) ds,
\end{aligned}$$

where the last equality uses (31) and (32), and is mostly a notation matter.

In particular, doing the same computation trick as the one done at the end of the proof of Theorem 3.6, it is possible to write the dynamics of $G(\mu_t^N)$ as follows

$$\begin{aligned}
G(\mu_t^N) =& G(\mu_r^N) + M_t^{N,r} + \int_r^t \int_{\mathbb{R}} b\left(\mu_s^N, x\right)\partial_x\delta G(m,x) d\mu_s^N(x) ds \\
&+\frac{1}{2}\int_r^t \int_{\mathbb{R}} \sigma\left(\mu_s^N, x\right)^2 \partial^2_{xx}\delta G\left(\mu_s^N, x\right) d\mu_s^N(x) ds \\
&+\frac{1}{2N}\int_r^t \int_{\mathbb{R}} \sigma\left(\mu_s^N, x\right)^2 \left(\partial^2_{y_1y_2}\delta^2 G\left(\mu_s^N, y_1, y_2\right)\right)_{|y_1=y_2=x} d\left(\mu_s^N(x)\right)^{\otimes 2}(x) ds \\
&+\frac{1}{2}\int_r^t \int_{\mathbb{R}^2} \varsigma(\mu_s^N, x)\varsigma(\mu_s^N, y)\partial^2_{xy}\delta^2 G(\mu_s^N, x, y) d\left(\mu_s^N\right)^{\otimes 2}(x,y) ds \\
&+N\int_r^t \int_{\mathbb{R}}\int_{\mathbb{R}} f\left(\mu_s^N, y\right)\left[G\left(\mathcal{S}\left(\mu_s^N, h(\mu_s^N, y, u)\right)\right) - G\left(\mu_s^N\right)\right] d\mu_s^N(y) d\nu(u) ds,
\end{aligned}$$

with $(M_t^{N,r})_t$ a martingale that depends only on the process μ^N, on the Brownian motions W, B^k $(1 \leq k \leq N)$ and on the Poisson measures π^l $(1 \leq l \leq N)$ such that it adapted to the union of the filtrations of these objects. The equality above being true for all $G \in \mathscr{P}^2$, Proposition 3.3 entails that the measure-valued process $(\mu_t^N)_t$ is a Markov process (which was already guaranteed by Proposition 2.3.3 of Dawson (1993)), but also that the domain of the generator of μ^N contains the set $\mathscr{P}^2$, and that the expression of the generator on any $G \in \mathscr{P}^2$ is the one given in the statement of Theorem 3.10. □

4. Proof of the main results

4.1. Proof of Theorem 1.3

Let us recall that, for each $N \in \mathbb{N}^*$, the processes $X^{N,k}$ $(1 \leq k \leq N)$ are defined as the (strong) solutions of (1), and $\bar{X}$ as the (strong) solution of (2). And, by definition, for all $t \geq 0$,

$$\mu_t^N = \frac{1}{N}\sum_{k=1}^N \delta_{X_t^{N,k}} \quad \text{and} \quad \bar{\mu}_t = \mathcal{L}\left(\bar{X}_t | \mathcal{W}_t\right).$$

Let us denote

$$\mathcal{E}_N = \left\{ m_w = \frac{1}{N}\sum_{k=1}^N \delta_{w_k} \ : \ w_1, ..., w_N \in \mathbb{R} \right\} \subseteq \mathcal{P}_\infty(\mathbb{R}).$$

It is classical that for any $t \geq 0$, there exists $C_t > 0$ such that for all $p \in \mathbb{N}^*$, $m \in \mathcal{P}_p(\mathbb{R})$ and $w_1, ..., w_N \in \mathbb{R}$

$$\mathbb{E}_m\left[\sup_{s\leq t} \left|\bar{X}_s\right|^p\right] \leq C_t\left(1 + \int_{\mathbb{R}} |x|^p dm(x)\right), \tag{38}$$

$$\frac{1}{N}\sum_{k=1}^N \mathbb{E}_{m_w}\left[\sup_{s\leq t} \left|X_s^{N,k}\right|^p\right] \leq C_t\left(1 + \frac{1}{N}\sum_{k=1}^N |w_k|^p\right), \tag{39}$$

where $m_w = N^{-1}\sum_{k=1}^N w_k$, $\mathbb{E}_m$ (resp. $\mathbb{E}_{m_w}$) denotes the expectation w.r.t. the probability measure under which the Markov process $\bar{\mu}$ (resp. μ^N) has m (resp. m_w) as initial condition.

Step 1. The first step of the proof consists in showing that, for any fixed $N \in \mathbb{N}^*$, the $\mathcal{P}_1(\mathbb{R})$-valued Markov processes μ^N and $\bar{\mu}$ satisfy the following Trotter-Kato formula: for all $G \in \mathscr{P}^4$, $m \in \mathcal{E}_N$ and $t \geq 0$,

$$\left(\bar{P}_t - P_t^N\right) G(m) = \int_0^t P_{t-s}^N \left(\bar{A} - A^N\right) \bar{P}_s G(m) ds,$$

with $P^N, \bar{P}$ (resp. $A^N, \bar{A}$) the semigroups (resp. generators) of $\mu^N, \bar{\mu}$, in the sense of Definition 3.1.

To prove this, it is sufficient to check that $\mu^N, \bar{\mu}$ satisfy the four conditions of Proposition 3.4. Condition 1 holds true for $\mathcal{A} = \mathscr{P}^4$ by Proposition 3.9. Besides, Condition 2 is a direct consequence of Theorems 3.6 and 3.10.

Now, let us verify Condition 3 of Proposition 3.4. Let $G \in \mathscr{P}^4$, by Theorem 3.6, the generator $\bar{A}$ of $\bar{\mu}$ satisfy: for all $m \in \mathcal{P}_1(\mathbb{R})$,

$$\bar{A}G(m) = \int_{\mathbb{R}} I(m,x) dm(x) + \int_{\mathbb{R}^2} J(m,x,y) dm^{\otimes 2}(x,y),$$

with

$$\begin{aligned} I(m,x) = &\left(b(m,x) + \int_{\mathbb{R}} h(m,y) f(m,y) dm(y)\right) \partial_x \delta G(m,x) \\ &+ \frac{1}{2}\left(\sigma(m,x)^2 + \varsigma(m,x)^2\right) \partial_{xx}^2 \delta G(m,x), \end{aligned} \tag{40}$$

$$J(m,x,y) = \frac{1}{2}\varsigma(m,x)\varsigma(m,y)\partial_{xy}^2 \delta^2 G(m,x,y). \tag{41}$$

In particular thanks to Proposition 2.32, we have the following control: there exists $C > 0$ such that for all $G \in \mathscr{P}^4$ and $m \in \mathcal{P}_2(\mathbb{R})$,

$$\left|\bar{A}G(m)\right| \leq C\left(1 + \int_{\mathbb{R}} |x|^2 dm(x)\right) ||G||_2. \tag{42}$$

And, for all $t_0, t \leq T$,

$$\begin{aligned}
\left|\bar{A}G(\bar{\mu}_{t_0}) - \bar{A}G(\bar{\mu}_t)\right| \leq & \left|\int_{\mathbb{R}} I(\bar{\mu}_{t_0}, x) d\left(\bar{\mu}_{t_0} - \bar{\mu}_t\right)(x)\right| + \int_{\mathbb{R}} \left|I(\bar{\mu}_{t_0}, x) - I(\bar{\mu}_t, x)\right| d\bar{\mu}_t(x) \\
& + \left|\int_{\mathbb{R}^2} J(\bar{\mu}_{t_0}, x, y) d\left(\bar{\mu}_{t_0}^{\otimes 2} - \bar{\mu}_t^{\otimes 2}\right)(x, y)\right| \\
& + \int_{\mathbb{R}^2} \left|J(\bar{\mu}_{t_0}, x, y) - J(\bar{\mu}_t, x, y)\right| d\bar{\mu}_t^{\otimes 2}(x, y) \\
\leq & \mathbb{E}\left[\left|I(\bar{\mu}_{t_0}, \bar{X}_{t_0}) - I(\bar{\mu}_{t_0}, \bar{X}_t)\right| \middle| \mathcal{W}_T\right] \\
& + \mathbb{E}\left[\left|I(\bar{\mu}_{t_0}, \bar{X}_t) - I(\bar{\mu}_t, \bar{X}_t)\right| \middle| \mathcal{W}_T\right] \\
& + \mathbb{E}\left[\left|J(\bar{\mu}_{t_0}, \bar{X}^1_{t_0}, \bar{X}^2_{t_0}) - J(\bar{\mu}_{t_0}, \bar{X}^1_t, \bar{X}^2_t)\right| \middle| \mathcal{W}_T\right] \\
& + \mathbb{E}\left[\left|J(\bar{\mu}_{t_0}, \bar{X}^1_t, \bar{X}^2_t) - J(\bar{\mu}_t, \bar{X}^1_t, \bar{X}^2_t)\right| \middle| \mathcal{W}_T\right],
\end{aligned}$$

with $\bar{X}^1, \bar{X}^2$ solutions to the SDE (2) w.r.t. the same Brownian motion W, but w.r.t. two independent Brownian motions B^1, B^2 instead of B. Let us note that, the functions I, J are locally Lipschitz continuous in the following sense: there exists $C_G > 0$ such that, for all $m_1, m_2 \in \mathcal{P}_2(\mathbb{R})$, and $x_1, x_2, y_1, y_2 \in \mathbb{R}$,

$$\begin{aligned}
& |I(m_1, x_1) - I(m_2, x_2)| + |J(m_1, x_1, y_1) - J(m_2, x_2, y_2)| \\
& \qquad \leq C_G\left(1 + |x_1|^2 + |x_2|^2 + |y_1|^2 + |y_2|^2 + \int_{\mathbb{R}} |z|^2 dm_1(z) + \int_{\mathbb{R}} |z|^2 dm_2(z)\right) \\
& \qquad\qquad \left(|x_1 - x_2| + D_{KR}(m_1, m_2)\right).
\end{aligned}$$

In particular, by Cauchy-Schwarz' inequality and Jensen's inequality and (3), denoting $\mathbb{E}_m$ the expectation w.r.t. the probability measure under which $\bar{\mu}_0 = m$,

$$\begin{aligned}
\left|\bar{P}_{t_0}\bar{A}G(m) - \bar{P}_t\bar{A}G(m)\right| &\leq \mathbb{E}_m\left[\left|\bar{A}G(\bar{\mu}_{t_0}) - \bar{A}G(\bar{\mu}_t)\right|\right] \\
&\leq C_G\mathbb{E}_m\left[\left(1 + \sup_{s \leq T} \mathbb{E}\left[|\bar{X}_s|^2 \middle| \mathcal{W}_T\right]\right)\mathbb{E}\left[\left|\bar{X}_{t_0} - \bar{X}_t\right| \middle| \mathcal{W}_T\right]\right] \\
&\leq C_G\mathbb{E}_m\left[\left|\bar{X}_{t_0} - \bar{X}_t\right|^2\right]^{1/2}.
\end{aligned}$$

Recalling that $\bar{X}$ is solution to (2), it is classical (using for example Burkholder-Davis-Gundy's inequality with (3)) that the RHS of the above inequality vanishes as t goes to t_0. This implies that, for any $m \in \mathcal{P}_2(\mathbb{R})$ and $G \in \mathscr{P}^4$, the function

$$t \mapsto \bar{P}_t\bar{A}G(m)$$

is continuous.

The continuity of the function

$$t \mapsto P_t^N A^N G(m),$$

for $m \in \mathcal{E}_N$ follows from the same reasoning provided Lemma 4.1 below to deal with the jump term that depends on a shift operator $\mathcal{S}$. Note that it is important to work with an initial condition belonging to $\mathcal{E}_N$ in order to interpret μ_t^N as the empirical measure of some N-particle system satisfying the SDEs (1).

Lemma 4.1. *Let $m_1, m_2 \in \mathcal{P}_1(\mathbb{R})$ and $h_1, h_2 \in \mathbb{R}$. Then*

$$D_{KR}\left(\mathcal{S}(m_1,h_1), \mathcal{S}(m_2,h_2)\right) \leq D_{KR}(m_1,m_2) + |h_1 - h_2|.$$

Proof. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be any Lipschitz continuous function with Lipschitz constant non-greater than one.

$$\begin{aligned}
&\left| \int_{\mathbb{R}} \varphi(x) d\left(\mathcal{S}(m_1,h_1)\right)(x) - \int_{\mathbb{R}} \varphi(x) d\left(\mathcal{S}(m_2,h_2)\right)(x) \right| \\
&= \left| \int_{\mathbb{R}} \varphi(x+h_1) dm_1(x) - \int_{\mathbb{R}} \varphi(x+h_2) dm_2(x) \right| \\
&\leq \left| \int_{\mathbb{R}} \varphi(x+h_1) d(m_1 - m_2)(x) \right| + \int_{\mathbb{R}} |\varphi(x+h_1) - \varphi(x+h_2)| \, dm_2(x) \\
&\leq D_{KR}(m_1,m_2) + |h_1 - h_2|.
\end{aligned}$$

Taking the supremum over all such functions φ proves the result. □

Indeed, with the same computation, we obtain

$$\begin{aligned}
\left|P_{t_0}^N A^N G(m) - P_t^N A^N G(m)\right| &\leq \mathbb{E}_m\left[\left|A^N G(\mu_{t_0}^N) - A^N G(\mu_t^N)\right|\right] \\
&\leq C_G \mathbb{E}_m \left[\left(1 + \sup_{s \leq T} \frac{1}{N} \sum_{k=1}^N \left|X_s^{N,k}\right|^2 \right) \frac{1}{N} \sum_{k=1}^N \left| X_{t_0}^{N,k} - X_t^{N,k} \right| \right] \\
&\leq C_G \left(\frac{1}{N} \sum_{k=1}^N \mathbb{E}_m \left[\left| X_{t_0}^{N,k} - X_t^{N,k} \right|^2 \right] \right)^{1/2}.
\end{aligned}$$

Once again, this proves that for any $m \in \mathcal{E}_N$ and $G \in \mathscr{P}^4$, the function

$$t \mapsto P_t^N A^N G(m)$$

is continuous. The continuity of $t \mapsto P_t^N \bar{A} G(m)$ is guaranteed by the same arguments. So Condition 3 of Proposition 3.4 is verified.

To end the *Step 1* of the proof, let us now check Condition 4. Let us fix in the rest of *Step 1* some $G \in \mathscr{P}^4$, $0 \leq t < T$ and some sequence $(s_n)_n$ converging to t such that $s_n \leq T$ for all $n \in \mathbb{N}$. For $M > 0$, let us introduce

$$\mathcal{K}_M = \{m \in \mathcal{P}_1(\mathbb{R}) \ : \ \mathrm{Supp}(m) \subseteq [-M, M]\}.$$

For any $M > 0$, $\mathcal{K}_M$ is a compact set of $\mathcal{P}_p(\mathbb{R})$, for all $p \geq 1$ (see Remark 6.19 of Villani (2009)). Let us denote, for every $n \in \mathbb{N}$,

$$G_n = \bar{P}_t G - \bar{P}_{s_n} G.$$

Thanks to Proposition 3.9, we know that each G_n belongs to $\mathscr{P}^4$. Hence, by Proposition 2.32, we can use Corollaries 2.26 and 2.27 (and Lemma 2.18 to differentiate the jump term from the generator A^N) to compute $\partial(\bar{A}G_n)(m,x)$ and $\partial(A^N g_n)(m,x)$ (recalling the expressions of $\bar{A}$ and A^N given in Theorems 3.6 and 3.10) to prove that, for all $M>0$,

$$\sup_{\substack{n\in\mathbb{N}\\ m\in\mathcal{K}_M\\ |x|\leq M}} \left|\partial_x\delta\left(\bar{A}G_n\right)(m,x)\right| + \sup_{\substack{n\in\mathbb{N}\\ m\in\mathcal{K}_M\\ |x|\leq M}} \left|\partial_x\delta\left(A^N G_n\right)(m,x)\right| \leq C_G\left(1+M^2\right).$$

Note that the fact that the control above is uniform in n is guaranteed by Proposition 3.9.(b).

Then, by the Mean Value Theorem for measure-variable functions (i.e. Proposition 2.21) and since $\mathcal{K}_M$ $(M>0)$ are convex sets, we know that the families of functions $(\bar{A}G_n)_n$ and $(A^N G_n)_n$ are both uniformly equicontinuous on each compact set $\mathcal{K}_M$ (for any $M>0$). In addition, let us recall that it is possible to write the generator $\bar{A}$ and A^N in the following form: for any $m\in\mathcal{P}_2(\mathbb{R})$,

$$\begin{aligned}\bar{A}G_n(m) &= \int_{\mathbb{R}} I_n(m,x)dm(x) + \int_{\mathbb{R}^2} J_n(m,x,y)dm^{\otimes 2}(x,y)\\ A^N G_n(m) &= \int_{\mathbb{R}} I_n^N(m,x)dm(x) + \int_{\mathbb{R}^2} J_n^N(m,x,y)dm^{\otimes 2}(x,y),\end{aligned}$$

with I_n, J_n, I_n^N and J_n^N defined similarly as I, J in (40) and (41), for the function G_n instead of G.

And, by Proposition 3.9.(c), for any $x,y\in\mathbb{R}$ and $m\in\mathcal{P}_2(\mathbb{R})$, the integrands above vanish as n goes to infinity, and are bounded by

$$C_G\left(1+x^2+y^2+\int|z|^2dm(z)\right),$$

where $C_G>0$ does not depend on (m,x,y,n). Then, by the Dominated Convergence Theorem, for any $m\in\mathcal{K}_M$ (for any $M>0$), both $\bar{A}G_n(m)$ and $A^N G_n(m)$ vanish as n goes to infinity. Since these two sequences are uniformly equicontinuous on each compact set $\mathcal{K}_M$, this implies that, for all $M>0$,

$$\sup_{m\in\mathcal{K}_M}\left|\bar{A}G_n(m)\right| + \sup_{m\in\mathcal{K}_M}\left|A^N G_n(m)\right| \underset{n\to\infty}{\longrightarrow} 0. \tag{43}$$

Then, let us introduce, for $M>0$ the following event

$$D_M = \left\{\sup_{\substack{r\leq T\\ 1\leq k\leq N}}\left|X_r^{N,k}\right|\leq M\right\},$$

and D_M^c its complementary. We have, by Cauchy-Schwarz' inequality and Markov's inequality, for any $m\in\mathcal{E}_N$,

$$\begin{aligned}\sup_{r\leq T}\left|P_r^N\bar{A}G_n(m)\right| \leq& \mathbb{E}_m\left[\sup_{r\leq T}\left|\bar{A}G_n(\mu_r^N)\right|\right]\\ \leq& \mathbb{E}_m\left[\sup_{r\leq T}\left|\bar{A}G_n(\mu_r^N)\right|\mathbb{1}_{D_M}\right] + \mathbb{E}_m\left[\sup_{r\leq T}\left|\bar{A}G_n(\mu_r^N)\mathbb{1}_{D_M^c}\right|\right]\\ \leq& \mathbb{E}_m\left[\sup_{r\leq T}\left|\bar{A}G_n(\mu_r^N)\right|\mathbb{1}_{D_M}\right] + C_{T,G,m}M^{-1/2},\end{aligned}$$

where we have used the controls given at (39), (42) and Proposition 3.9.(b).

In particular, for any $\varepsilon > 0$ it is possible to fix some $M_\varepsilon > 0$ (whose also depends on T, G, m, but not on n) such that

$$\sup_{r\leq T} \left|P_r^N \bar{A} G_n(m)\right| \leq \mathbb{E}_m \left[\sup_{r\leq T} \left|\bar{A} G_n(\mu_r^N)\right| \mathbb{1}_{D_{M_\varepsilon}}\right] + \varepsilon.$$

Noticing that,

$$D_{M_\varepsilon} \subseteq \left\{\forall r \leq T, \mu_r^N \in \mathcal{K}_{M_\varepsilon}\right\},$$

we know, by (43), that, almost surely,

$$\sup_{r\leq T} \left|\bar{A} G_n(\mu_r^N)\right| \mathbb{1}_{D_{M_\varepsilon}} \underset{n\to\infty}{\longrightarrow} 0,$$

and, by the Dominated Convergence Theorem (and (39), (42) and Proposition 3.9.(b)), this entails that, for $m \in \mathcal{E}_N$,

$$\sup_{r\leq T} \left|P_r^N \bar{A} G_n(m)\right| \underset{n\to\infty}{\longrightarrow} 0.$$

With exactly the same reasoning, we also have

$$\sup_{r\leq T} \left|P_r^N A^N G_n(m)\right| \underset{n\to\infty}{\longrightarrow} 0.$$

So Condition 4 of Proposition 3.4 is satisfied.

Step 2. We have proved that for all $G \in \mathscr{P}^4$, $m \in \mathcal{E}_N$ and $t \geq 0$,

$$\left(\bar{P}_t - P_t^N\right) G(m) = \int_0^t P_{t-s}^N \left(\bar{A} - A^N\right) \bar{P}_s G(m) ds.$$

This formula allows to obtain a convergence speed for the semigroups provided a convergence speed of the generators. So, the goal of *Step 2* is to prove that, for all $G \in \mathscr{P}^2$ and $m \in \mathcal{P}_3(\mathbb{R})$,

$$\left|\left(A^N - \bar{A}\right) G(m)\right| \leq C||G||_2 \frac{1}{N} \left(1 + \int_{\mathbb{R}} |x|^3 dm(x)\right), \tag{44}$$

where $C > 0$ is independent of G, m and N.

To begin with, by Theorems 3.6 and 3.10, for all $G \in \mathscr{P}^2$ and $m \in \mathcal{P}_2(\mathbb{R})$,

$$\begin{aligned}
\bar{A}G(m) = &\int_{\mathbb{R}} \left[b(m,x) + \left(\int_{\mathbb{R}} h(m,y) f(m,y) dm(y) \right) \right] \partial_x \delta G(m,x) dm(x) \\
&+ \frac{1}{2} \int_{\mathbb{R}} \left[\sigma(m,x)^2 + \varsigma(m,x)^2 \right] \partial^2_{xx} \delta G(m,x) dm(x) \\
&+ \frac{1}{2} \int_{\mathbb{R}^2} \varsigma(m,x)\varsigma(m,y) \partial^2_{xy} \delta^2 G(m,x,y) dm^{\otimes 2}(x,y), \\
A^N G(m) = &\int_{\mathbb{R}} \left[b(m,x) \partial_x \delta G(m,x) + \frac{1}{2} \int_{\mathbb{R}} \left[\sigma(m,x)^2 + \varsigma(m,x)^2 \right] \partial^2_{xx} \delta G(m,x) \right] dm(x) \\
&+ \frac{1}{2N} \int_{\mathbb{R}} \sigma(m,x)^2 \left(\partial^2_{y_1 y_2} \delta^2 G(m,y_1,y_2) \right)_{|y_1 = y_2 = x} dm(x) \\
&+ \frac{1}{2} \int_{\mathbb{R}^2} \varsigma(m,x)\varsigma(m,y) \partial^2_{xy} \delta^2 G(m,x,y) dm^{\otimes 2}(x,y) \\
&+ N \int_{\mathbb{R}} f(m,x) \left[G\left(\mathcal{S}\left(m, \frac{1}{N} h(m,x) \right) \right) - G(m) \right] dm(x).
\end{aligned}$$

So

$$\begin{aligned}
\left| \left(A^N - \bar{A} \right) G(m) \right| \leq &\frac{1}{2N} \int_{\mathbb{R}} \left| \sigma(m,x)^2 \left(\partial^2_{y_1 y_2} \delta^2 G(m,y_1,y_2) \right)_{|y_1 = y_2 = x} \right| dm(x) \qquad (45) \\
&+ N \int_{\mathbb{R}} f(m,x) \left| G\left(\mathcal{S}\left(m, \frac{1}{N} h(m,x) \right) \right) - G(m) \right. \qquad (46) \\
&\left. - \frac{h(m,x)}{N} \int_{\mathbb{R}} \partial_y \delta G(m,y) dm(y) \right| dm(x).
\end{aligned}$$

Firstly, the term (45) is bounded by

$$C ||G||_2 \frac{1}{N} \left(1 + \int_{\mathbb{R}} |x|^2 dm(x) \right).$$

To control the other term (46), let us define

$$H : (m,x) \in \mathcal{P}_1(\mathbb{R}) \times \mathbb{R} \longmapsto \int_{\mathbb{R}} \delta G(m,y) \, d\left(\mathcal{S}\left(m, \frac{1}{N} h(m,x) \right) - m \right)(y),$$

and notice that, for all $m \in \mathcal{P}_1(\mathbb{R}), x \in \mathbb{R}$,

$$H(m,x) = \int_{\mathbb{R}} \left[\delta G\left(m, y + \frac{1}{N} h(m,x) \right) - \delta G(m,y) \right] dm(y).$$

Consequently, by subtracting $H(m,x)$ in the first line of the integrand of (46), and adding it to

the second line, we deduce that the term at (46) is non-greater than

$$\begin{aligned}
&N\int_{\mathbb{R}} f(m,x)\left|G\left(\mathcal{S}\left(m,\frac{1}{N}h(m,x)\right)\right)-G(m)-H(m,x)\right|dm(x)\\
&\quad+N\int_{\mathbb{R}} f(m,x)\int_{\mathbb{R}}\left|\delta G\left(m,y+\frac{1}{N}h(m,x)\right)-\delta G(m,y)\right.\\
&\qquad\qquad\left.-\frac{1}{N}h(m,x)\partial_y\delta G(m,y)\right|dm(y)dm(x)\\
&\leq\frac{1}{2}N||G||_2\int_{\mathbb{R}} f(m,x)\left(D_{KR}\left(\mathcal{S}\left(m,\frac{1}{N}h(m,x)\right),m\right)^2+\frac{1}{N^2}h(m,x)^2\right)dm(x),
\end{aligned}$$

where we have used Taylor-Lagrange's inequality for measure-variable functions (i.e. Theorem 2.24) to control the first term of the sum above, and the standard Taylor-Lagrange's inequality for the second one. Besides, by Lemma 2.18, for any $m\in\mathcal{P}_1(\mathbb{R})$ and $\lambda\in\mathbb{R}$,

$$D_{KR}(\mathcal{S}(m,\lambda),m)\leq|\lambda|,$$

such that (46) is bounded by

$$\frac{1}{N}||G||_2\int_{\mathbb{R}} f(m,x)h(m,x)^2dm(x)\leq C||G||_2\frac{1}{N}\left(1+\int_{\mathbb{R}}|x|^3dm(x)\right),$$

where we have used the sublinearity of f,h. So we have proved (44).

Step 3. From (44), we deduce a control between the semigroups of $\mu^N,\bar{\mu}$. Let $w_1,...,w_N\in\mathbb{R}$, and

$$m_w=\frac{1}{N}\sum_{k=1}^N w_k.$$

Then, thanks to *Step 1*,

$$\begin{aligned}
\left|\left(\bar{P}_t-P_t^N\right)G(m_w)\right|\leq&\int_0^t\mathbb{E}_{m_w}\left[\left|\left(\bar{A}-A^N\right)\bar{P}_sG(\mu^N_{t-s})\right|\right]ds\\
\leq&C\frac{1}{N}\int_0^t||\bar{P}_sG||_2\left(1+\frac{1}{N}\sum_{k=1}^N\mathbb{E}_{m_w}\left[|X^{N,k}_{t-s}|^3\right]\right)ds\\
\leq&C_t\frac{1}{N}||G||_2\left(1+\frac{1}{N}\sum_{k=1}^N|w_k|^3\right),
\end{aligned}\tag{47}$$

where we have used Proposition 3.9.

Step 4. Then end of the proof is now quite classical. For $G\in\mathscr{P}^4$,

$$\begin{aligned}
\left|\mathbb{E}\left[G(\bar{\mu}_t)\right]-\mathbb{E}\left[G(\mu_t^N)\right]\right|\leq&\left|\int_{\mathcal{P}_1(\mathbb{R})}\bar{P}_tG(m)\,d\left(\mathcal{L}\left(\bar{\mu}_0\right)-\mathcal{L}\left(\mu_0^N\right)\right)(m)\right|\\
&+\int_{\mathcal{P}_1(\mathbb{R})}\left|\left(\bar{P}_t-P_t^N\right)G(m)\right|d\left(\mathcal{L}\left(\mu_0^N\right)\right)(m)\\
\leq&C_t||G||_1\mathcal{D}_{KR}\left(\mathcal{L}\left(\bar{\mu}_0\right),\mathcal{L}\left(\mu_0^N\right)\right)+C_t||G||_2\frac{1}{N},
\end{aligned}$$

where the last line has been obtained using Proposition 3.9 (and the definition of the Kantorovich-Rubinstein metric) for the first term of the sum, and (47) for the second one.

This ends the proof of Theorem 1.3.

4.2. Proof of Corollary 1.5

The proof of Corollary 1.5 consists in remarking that the SDEs (4) and (5) are particular cases of (1) and (2) with: for $m \in \mathcal{P}_1(\mathbb{R})$ and $x \in \mathbb{R}$,

$$b(m,x) = \tilde{b}(x); \quad \sigma(m,x) = \tilde{\sigma}(x); \quad \varsigma(m,x) = 0; \quad h(m,x) = 1; \quad f(m,x) = \tilde{f}(x).$$

In addition, since the processes $Y^{N,k}$ $(1 \leq k \leq N)$ and $\bar{Y}$ are assumed to have the same deterministic initial condition $x_0 \in \mathbb{R}$, the measures μ^N $(N \in \mathbb{N}^*)$ and $\bar{\mu}_0$ are deterministic and equal to the Dirac mass at x_0. Hence

$$\mathcal{D}_{KR}\left(\mathcal{L}\left(\bar{\mu}_0\right), \mathcal{L}\left(\mu_0^N\right)\right) = D_{KR}\left(\delta_{x_0}, \delta_{x_0}\right) = 0.$$

Finally, the result of Corollary 1.5 comes from the fact that, for any $\varphi \in C_b^4(\mathbb{R})$, the function

$$G : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}} \varphi(x) dm(x)$$

belongs to $\mathscr{P}^4$, and that

$$\mathbb{E}\left[G(\bar{\mu}_t)\right] = G(\bar{\mu}_t) = \mathbb{E}\left[\varphi(\bar{Y}_t)\right] \quad \text{and} \quad \mathbb{E}\left[G(\mu_t^N)\right] = \mathbb{E}\left[\frac{1}{N}\sum_{k=1}^{N} \varphi(Y_t^{N,k})\right] = \mathbb{E}\left[\varphi(Y_t^{N,1})\right].$$

4.3. Proofs of Theorem 1.7 and Corollary 1.9

The proof of Theorem 1.7 is essentially the same as the one of Theorem 1.3. The only difference is the *Step 2* of the proof concerning the control of the difference of the generators A^N and $\bar{A}$, where we need Taylor-Lagrange's inequalities at order three instead of two. So concerning the proof of Theorem 1.7, let us only write the proof of the equivalent of (44) for the diffusive model: for $G \in \mathscr{P}^3$ and $m \in \mathcal{P}_4(\mathbb{R})$,

$$\left|\left(A^N - \bar{A}\right) G(m)\right| \leq C||G||_3 \frac{1}{\sqrt{N}} \left(1 + \int_{\mathbb{R}} |x|^4 dm(x)\right). \tag{48}$$

To this end, let us admit the following lemma, whose proof consists in using three times the Taylor-Lagrange's inequality. This lemma allows to simplify computation in our proof.

Lemma 4.2. *Let $g \in C_b^3(\mathbb{R}^2)$. Then, for all $x, y, \lambda \in \mathbb{R}$,*

$$\left|g(x+\lambda, y+\lambda) - g(x+\lambda, y) - g(x, y+\lambda) + g(x,y) - \frac{\lambda^2}{2}\partial_{xy}^2 g(x,y)\right| \leq \frac{|\lambda|^3}{2} \sum_{|\alpha|=3} ||\partial_\alpha g||_\infty.$$

Let us denote $\zeta > 0$ the standard deviation of the probability measure ν:

$$\zeta = \left(\int_{\mathbb{R}} u^2 d\nu(u) \right)^{1/2}.$$

Then, by Theorems 3.6 and 3.10, for all $G \in \mathscr{P}^2$ and $m \in \mathcal{P}_2(\mathbb{R})$,

$$\begin{aligned}
\bar{A}G(m) = &\int_{\mathbb{R}} \left[b(m,x)\partial_x \delta G(m,x) + \frac{1}{2}\sigma(m,x)^2 \partial^2_{xx} \delta G(m,x) \right] dm(x) \\
&+ \frac{1}{2}\zeta^2 \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(m,y)^2 f(m,y) dm(y) \right) \partial^2_{xx} \delta G(m,x) dm(x) \\
&+ \frac{1}{2}\zeta^2 \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}} h(m,z)^2 f(m,z) dm(z) \right) \partial^2_{xy} \delta^2 G(m,x,y) dm^{\otimes 2}(x,y), \\
A^N G(m) = &\int_{\mathbb{R}} \left[b(m,x)\partial_x \delta G(m,x) + \frac{1}{2}\sigma(m,x)^2 \partial^2_{xx} \delta G(m,x) \right] dm(x) \\
&+ \frac{1}{2N} \int_{\mathbb{R}} \sigma(m,x)^2 \left(\partial^2_{y_1 y_2} \delta^2 G(m,y_1,y_2) \right)_{|y_1=y_2=x} dm(x) \\
&+ N \int_{\mathbb{R}} \int_{\mathbb{R}} f(m,x) \left[G\left(\mathcal{S}\left(m, \frac{u}{\sqrt{N}} h(m,x) \right) \right) - G(m) \right] d\nu(u) dm(x).
\end{aligned}$$

So

$$\begin{aligned}
\left(A^N - \bar{A} \right) G(m) = &\frac{1}{2N} \int_{\mathbb{R}} \sigma(m,x)^2 \left(\partial^2_{y_1 y_2} \delta^2 G(m,y_1,y_2) \right)_{|y_1=y_2=x} dm(x) \qquad (49) \\
&+ N \int_{\mathbb{R}} f(m,z) \int_{\mathbb{R}} \left[G\left(\mathcal{S}\left(m, \frac{u}{\sqrt{N}} h(m,z) \right) \right) - G(m) \right. \qquad (50) \\
&\quad - \frac{u^2}{2N} h(m,z)^2 \int_{\mathbb{R}} \partial^2_{xx} \delta G(m,x) dm(x) \\
&\quad \left. - \frac{u^2}{2N} h(m,z)^2 \int_{\mathbb{R}^2} \partial^2_{xy} \delta^2 G(m,x,y) dm^{\otimes 2}(x,y) \right] d\nu(u) dm(z).
\end{aligned}$$

The absolute value of the term at (49) is bounded by

$$C \frac{1}{N} ||G||_2 \left(1 + \int_{\mathbb{R}} x^2 dm(x) \right).$$

For the rest of the proof, let us introduce, for $u, z \in \mathbb{R}$, the signed measure

$$m^N_{z,u} = \mathcal{S}\left(m, \frac{u}{\sqrt{N}} h(m,z) \right) - m. \qquad (51)$$

To control the term at (50), by using Theorem 2.24, we need to compute the integrals of the k-th order derivative of G w.r.t. the power of $m^N_{z,u}$ to k, for $k \in \{1,2\}$.

$$\begin{aligned}
\int_{\mathbb{R}} \int_{\mathbb{R}} \delta G(m,x) dm^N_{z,u}(x) d\nu(u) = &\int_{\mathbb{R}} \int_{\mathbb{R}} \left[\delta G\left(m, x + \frac{u}{\sqrt{N}} h(m,z) \right) - \delta G(m,x) \right. \\
&\left. - \frac{u}{\sqrt{N}} h(m,z) \partial_x \delta G(m,x) \right] d\nu(u) dm(x),
\end{aligned}$$

where the term at the second line above has been added artificially since ν is a centered measure. In addition,

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}^2} \delta^2 G(m,x,y) d\left(m^N_{z,u}\right)^{\otimes 2}(x,y) d\nu(u) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} \Big[\delta^2 G\left(m, x+\frac{u}{\sqrt{N}}h(m,z), y+\frac{u}{\sqrt{N}}h(m,z)\right) \\
&\qquad -\delta^2 G\left(m, x+\frac{u}{\sqrt{N}}h(m,z), y\right) - \delta^2 G\left(m, x, y\frac{u}{\sqrt{N}}h(m,z)\right) \\
&\qquad + \delta^2 G(m,x,y)\Big] d\nu(u) dm(x) dm(y).
\end{aligned}$$

Whence, the absolute value of the term within the brackets at (50) is non-greater than the sum of the three following quantities:

$$\begin{aligned}
&\left| G\left(\mathcal{S}\left(m, \frac{u}{\sqrt{N}}h(m,z)\right)\right) - G(m) - \int_{\mathbb{R}} \delta G(m,x) dm^N_{z,u}(x) dm(x) \right. \\
&\qquad \left. -\frac{1}{2}\int_{\mathbb{R}^2} \delta^2 G(m,x,y) d\left(m^N_{z,u}\right)^{\otimes 2}(x,y)\right|, \qquad (52)
\end{aligned}$$

$$\begin{aligned}
&\int_{\mathbb{R}} \left| \delta G\left(m, x+\frac{u}{\sqrt{N}}h(m,z)\right) - \delta G(m,x) - \frac{u}{\sqrt{N}}h(m,z)\partial_x \delta G(m,x) \right. \\
&\qquad \left. -\frac{u^2}{2N}h(m,z)^2 \partial^2_{xx}\delta G(m,x)\right| dm(x), \qquad (53)
\end{aligned}$$

$$\begin{aligned}
&\int_{\mathbb{R}^2} \left| \delta^2 G\left(m, x+\frac{u}{\sqrt{N}}h(m,z), y+\frac{u}{\sqrt{N}}h(m,z)\right) - \delta^2 G\left(m, x+\frac{u}{\sqrt{N}}h(m,z), y\right) \right. \\
&\left. -\delta^2 G\left(m, x, y\frac{u}{\sqrt{N}}h(m,z)\right) + \delta^2 G(m,x,y) - \frac{u^2}{2N}h(m,z)^2 \partial^2_{xy}\delta^2 G(m,x,y)\right| dm^{\otimes 2}(x,y). \qquad (54)
\end{aligned}$$

Each of the three terms above is bounded by

$$C||G||_3 \frac{1}{N\sqrt{N}}\left(1+|z|^3+\int_{\mathbb{R}}|x|^3 dm(x)\right)\left(1+|u|^3\right).$$

Indeed, for (52), it is a consequence of the Taylor-Lagrange's inequality for measure-variable functions (i.e. Theorem 2.24), and the fact that (by Lemma 2.18)

$$D_{KR}\left(\mathcal{S}\left(m, \frac{u}{\sqrt{N}}h(m,z)\right), m\right) \leq \frac{|u|}{\sqrt{N}}|h(m,z)|,$$

for (53), it comes from the Taylor-Lagrange's inequality for real-variable functions, and the control of (54) is obtained by Lemma 4.2.

Finally using this last bound to control (50) is sufficient to prove (48). As explained previously, the rest of the proof of Theorem 1.7 being almost unchanged compared to the proof of Theorem 1.3, it is omitted.

The proof of Corollary 1.9 follows the same idea as the one of Corollary 1.5: it consists in remarking that Corollary 1.9 is a particular case of Theorem 1.7. The only difference is that we

need to prove that

$$\mathcal{D}_{KR}\left(\mathcal{L}\left(\frac{1}{N}\sum_{k=1}\delta_{Y_0^{N,k}}\right),\delta_{\mathcal{L}(\bar{Y}_0)}\right)=D_{KR}\left(\frac{1}{N}\sum_{k=1}\delta_{Y_0^{N,k}},\mathcal{L}(\bar{Y}_0)\right)\leq C\frac{1}{\sqrt{N}}.$$

Let us recall that, by hypothesis, the variables $Y_0^{N,k}$ $(1\leq k\leq N)$ are i.i.d. with the same law as $\bar{Y}_0$ with finite fourth order moment. Then, the inequality above is a direct consequence of Theorem 3.2 of Bobkov and Ledoux (2019) and the discussion thereafter, or of Theorem 1 of Fournier and Guillin (2015).

Appendix A: Separating class on spaces of probability measures

The goal of this section is to prove Proposition 2.31. It means that the set $\mathscr{P}^\infty$ is a separating class for $\mathcal{P}_p(\mathbb{R})$ (for any $p\geq 1$). Before proving this statement, let us introduce some definitions and useful results about separating classes. This section relies strongly on the content of Section 3.4 of Ethier and Kurtz (2005).

In all this section, E denotes some Polish space, and $\mathcal{P}(E)$ the space of probability measures on E endowed with Prohorov metric (i.e. the topology of the weak convergence).

Definition A.1. *Some set $\mathcal{A}$ of measurable functions $f:E\to\mathbb{R}$ is said to separate E if, for all $x,y\in E$,*

$$(\forall f\in\mathcal{A}, f(x)=f(y))\Longrightarrow x=y.$$

A set $\mathcal{A}$ of measurable functions $f:E\to\mathbb{R}$ is called a separating class for E if, for all $m,\mu\in\mathcal{P}(E)$,

$$\left(\forall f\in\mathcal{A},\int_E f(x)dm(x)=\int_E f(x)d\mu(x)\right)\Longrightarrow m=\mu.$$

In other words, $\mathcal{A}$ is a separating class for E means that

$$\left\{m\in\mathcal{P}(E)\longmapsto\int_E f(x)dm(x)\ :\ f\in\mathcal{A}\right\}$$

separates $\mathcal{P}(E)$. The interesting point of a separating class for E is that it allows to identify the laws on E in the following sense (it is a mere rephrasing of the definition of a separating class).

Lemma A.2. *If $\mathcal{A}$ is a separating class for E, and X,Y are two E-valued random variables such that: for all $f\in\mathcal{A}$,*

$$\mathbb{E}\left[f(X)\right]=\mathbb{E}\left[f(Y)\right].$$

Then X and Y have the same law.

The lemma below should be classical, but since we have not found a proof in the literature, we provide one for self-completeness.

Lemma A.3. *The set $C_c^\infty(\mathbb{R})$ is a separating class for $\mathbb{R}$.*

Proof. Let $m,\mu\in\mathcal{P}(\mathbb{R})$ such that, for all $g\in C_c^\infty(\mathbb{R})$,

$$\int_{\mathbb{R}} g(x)dm(x)=\int_{\mathbb{R}} g(x)d\mu(x). \tag{55}$$

Let $a < b$. Let $(g_n)_n$ be a $C_c^\infty(\mathbb{R})-$valued sequence such that

$$\begin{cases} (g_n)_n \text{ converges pointwise to } \mathbb{1}_{]a,b[}, \\ \forall (x,n) \in \mathbb{R} \times \mathbb{N}, \quad 0 \leq g_n(x) \leq 1. \end{cases}$$

Then, rewriting (55) with g_n instead of g (for all $n \in \mathbb{N}^*$), and using the Dominated Convergence Theorem,

$$m(]a,b[) = \mu(]a,b[).$$

The previous equality being true for any $a < b$, this proves that $m = \mu$. □

The proof that the polynomials form a separating class for $\mathcal{P}_p(\mathbb{R})$ relies on the following criterion.

Theorem A.4 (Theorem 3.4.5.(a) of Ethier and Kurtz (2005))**.** *Any algebra that separates E is a separating class for E.*

Besides, since $\mathcal{P}_p(\mathbb{R})$ is Polish (see Theorem 6.18 of Villani (2009)), Theorem A.4 above can be used with $E = \mathcal{P}_p(\mathbb{R})$. Now we can prove that the polynomials form a separating class of $\mathcal{P}_p(\mathbb{R})$.

Proof of Proposition 2.31. By Lemma A.3, the set

$$\left\{ m \in \mathcal{P}(\mathbb{R}) \longmapsto \int_{\mathbb{R}} h(x) dm(x) \ : \ h \in C_c^\infty(\mathbb{R}) \right\}$$

separates $\mathcal{P}(\mathbb{R})$, hence it also separates $\mathcal{P}_p(\mathbb{R})$. The set $\mathscr{P}^\infty$ being an algebra containing the set above, it is a separating class by Theorem A.4. □

Appendix B: Proof of Lemma 3.8

The proof of Lemma 3.8 uses the following classical lemma whose proof is omitted.

Lemma B.1. *Let $(f_n)_n$ be a sequence of $C^1(\mathbb{R})$ such that:*

- *$(f_n)_n$ converges point-wisely to some function f,*
- *$(f_n')_n$ converges uniformly on every compact set to some function g.*

Then, $f \in C^1(\mathbb{R})$ and $f' = g$.

Let us recall that, by definition, for all $x \in \mathbb{R}, t \geq 0$,

$$\bar{X}_t^{(x)} = x + \int_0^t b\left(\bar{\mu}_s^{(x)}, \bar{X}_s^{(x)}\right) ds + \int_0^t \sigma\left(\bar{\mu}_s^{(x)}, \bar{X}_s^{(x)}\right) dB_s + \int_0^t \varsigma\left(\bar{\mu}_s^{(x)}, \bar{X}_s^{(x)}\right) dW_s,$$

where $\bar{\mu}_s^{(x)} = \mathcal{L}(\bar{X}_s^{(x)} | \mathcal{W}_s)$.

In order to prove Lemma 3.8, we use the Banach-Picard iteration scheme related to the above equation. Namely, for all $x \in \mathbb{R}, t \geq 0, n \in \mathbb{N}$,

$$\begin{aligned}
\bar{X}_t^{(x),[0]} =& x, \\
\bar{\mu}_t^{(x),[0]} =& \delta_x, \\
\bar{X}_t^{(x),[n+1]} =& x + \int_0^t b\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right) ds + \int_0^t \sigma\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right) dB_s \\
&+ \int_0^t \varsigma\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right) dW_s, \\
\bar{\mu}_t^{(x),[n+1]} =& \mathcal{L}\left(\bar{X}_t^{(x),[n+1]} | \mathcal{W}_t\right).
\end{aligned}$$

Step 1. In this first step, we prove the almost sure convergence of $\bar{X}_t^{(x),[n]}$ to $\bar{X}_t^{(x)}$ as n goes to infinity, for any t belonging to some sufficiently small interval, and locally uniformly w.r.t. x.

For $n \in \mathbb{N}, p \in \mathbb{N}^*, t \geq 0$ and $M > 0$,

$$u_t^{[n],0}(M,p) = \mathbb{E}\left[\sup_{s \leq t, |x| \leq M} \left|\bar{X}_s^{(x),[n+1]} - \bar{X}_s^{(x),[n]}\right|^p\right].$$

With classical computation (using Burkholder-Davis-Gundy's inequality and the assumption that b, σ, ς are Lipschitz continuous), we have that, for all $T > 0$, $p \in \mathbb{N}^*$, $n \in \mathbb{N}$,

$$u_T^{[n+1],0}(M,p) \leq C_p\left(T^p + T^{p/2}\right) u_T^{[n],0}(M,p),$$

with $C_p > 0$ independent of T, M and n. In particular, let us fix some small enough $T_p > 0$ such that, for all $n \in \mathbb{N}, p \in \mathbb{N}, M > 0$,

$$u_{T_p}^{[n+1],0}(M,p) \leq \frac{1}{2} u_{T_p}^{[n],0}(M,p),$$

whence

$$u_{T_p}^{[n],0}(M,p) \leq C_p \frac{1}{2^n}\left(1 + M^p\right). \tag{56}$$

and

$$\mathbb{E}\left[\sup_{s \leq T_p, |x| \leq M} \left|\bar{X}_s^{(x),[n+1]} - \bar{X}_s^{(x),[n]}\right|\right] \leq u_{T_p}^{[n],0}(M,p)^{1/p} \leq C_p \frac{1}{2^{n/p}}\left(1 + M\right).$$

As a consequence, for all $M > 0, p \in \mathbb{N}^*$,

$$\mathbb{E}\left[\sup_{t \leq T_p, |x| \leq M} \left|\sum_{n=0}^{+\infty}\left(\bar{X}_t^{(x),[n+1]} - \bar{X}_t^{(x),[n]}\right)\right|\right] < \infty,$$

and, almost surely,

$$\sup_{t \leq T_p, |x| \leq M} \left|\sum_{n=0}^{+\infty}\left(\bar{X}_t^{(x),[n+1]} - \bar{X}_t^{(x),[n]}\right)\right| < \infty.$$

This proves that, almost surely, for all $t \in [0, T_p]$, $\bar{X}_t^{(x),[n]}$ converges as n goes to infinity to $\bar{X}_t^{(x)}$ (recalling that $\bar{X}^{(x)}$ is solution to (33)) locally uniformly w.r.t. x.

In addition, with similar technics and Fatou's lemma, we can prove that, for all $p \in \mathbb{N}^*$, $M > 0$, $T > 0$,

$$\mathbb{E}\left[\sup_{t \leq T, |x| \leq M} \left|\bar{X}_t^{(x)}\right|^p\right] < \infty. \tag{57}$$

Step 2. Now we study the regularity of the functions

$$x \longmapsto \bar{X}_t^{(x),[n]}.$$

Let us prove by induction on $n \in \mathbb{N}$ that there exists $T > 0$ (independent of n) such that, for all $t \in [0, T]$ the function $x \mapsto \bar{X}_t^{(x),[n]}$ C^1 and that, for all $M > 0$,

$$\mathbb{E}\left[\sup_{t \leq T, |x| \leq M} \left|\partial_x \bar{X}_t^{(x),[n]}\right|\right] < \infty \quad \text{and} \quad \sup_{x \in \mathbb{R}} \mathbb{E}\left[\sup_{t \leq T} \left|\partial_x \bar{X}_t^{(x),[n]}\right|\right] < \infty.$$

For $n = 0$,

$$\partial_x \bar{X}_t^{(x),[0]} = 1.$$

Then fix some $n \in \mathbb{N}$ and assume that the induction hypothesis holds true for this n. Then, by Lemma 2.19 (and Remark 2.20), and assuming that b, σ, ς admits bounded second order mixed derivatives (what implies that the first order mixed derivatives are continuous w.r.t. all their variables, by Remark 2.13),

$$\begin{aligned}
\partial_x \bar{X}_t^{(x),[n+1]} =& 1 + \int_0^t \left(\partial_x \bar{X}_s^{(x),[n]}\right) \partial_{(0,1)} b\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right) ds \\
&+ \int_0^t \tilde{\mathbb{E}}\left[\left(\partial_x \tilde{X}_s^{(x),[n]}\right) \partial\left(b_{\bar{X}_s^{(x),[n]}}\right)\left(\bar{\mu}_s^{(x),[n]}, \tilde{X}_s^{(x),[n]}\right) \middle| \mathcal{W}_s\right] ds \\
&+ \int_0^t \left(\partial_x \bar{X}_s^{(x),[n]}\right) \partial_{(0,1)} \sigma\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right) dB_s \\
&+ \int_0^t \tilde{\mathbb{E}}\left[\left(\partial_x \tilde{X}_s^{(x),[n]}\right) \partial\left(\sigma_{\bar{X}_s^{(x),[n]}}\right)\left(\bar{\mu}_s^{(x),[n]}, \tilde{X}_s^{(x),[n]}\right) \middle| \mathcal{W}_s\right] dB_s \\
&+ \int_0^t \left(\partial_x \bar{X}_s^{(x),[n]}\right) \partial_{(0,1)} \varsigma\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right) dW_s \\
&+ \int_0^t \tilde{\mathbb{E}}\left[\left(\partial_x \tilde{X}_s^{(x),[n]}\right) \partial\left(\varsigma_{\bar{X}_s^{(x),[n]}}\right)\left(\bar{\mu}_s^{(x),[n]}, \tilde{X}_s^{(x),[n]}\right) \middle| \mathcal{W}_s\right] dW_s,
\end{aligned} \tag{58}$$

where $\tilde{X}$ is defined in the same way as $\bar{X}$ w.r.t. the same Brownian motion W, but w.r.t. a Brownian motion $\tilde{B}$ independent of (B, W), and $\tilde{\mathbb{E}}$ is the expectation w.r.t. the law of $\tilde{B}$. One can note that, the expression above of $\partial_x \bar{X}_t^{(x),[n+1]}$ is closely related to the ones of Proposition 3.1 of Crisan and McMurray (2018) (we actually use their notation for $\tilde{\mathbb{E}}$ and $\tilde{X}$).

Let us remark an important point: the functions appearing as coefficients in the expression above are exactly the first order mixed derivatives of the functions b, σ, ς.

And, with similar computation as in *Step 1*, we can prove that, for all $p \in \mathbb{N}^*$, $M > 0$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\sup_{t\leq T_p,|x|\leq M}\left|\partial_x \bar{X}_t^{(x),[n]}\right|^p\right] < \infty \text{ and } \sup_{n\in\mathbb{N},x\in\mathbb{R}} \mathbb{E}\left[\sup_{t\leq T_p}\left|\partial_x \bar{X}_t^{(x),[n]}\right|\right] < \infty, \tag{59}$$

for some $T_p > 0$ independent of n, M.

Let

$$u_t^{[n],1}(M,p) = \mathbb{E}\left[\sup_{s\leq t,|x|\leq M}\left|\partial_x \bar{X}_s^{(x),[n+1]} - \partial_x \bar{X}_s^{(x),[n]}\right|^p\right].$$

To control the above quantity in a similar way as in *Step 1*, it is required to control two kind of terms. In order to simplify the reading we only handle the drift terms (the Brownian terms can be treated in the exact same way after using Burkholder-Davis-Gundy's inequality). The first term is of the following form: for $|x| \leq M$,

$$\begin{aligned}
&\left|\left(\partial_x \bar{X}_s^{(x),[n+1]}\right)\partial_{(0,1)}b\left(\bar{\mu}_s^{(x),[n+1]}, \bar{X}_s^{(x),[n+1]}\right) - \left(\partial_x \bar{X}_s^{(x),[n]}\right)\partial_{(0,1)}b\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right)\right| \\
&\leq \left(\sup_{m\in\mathcal{P}_1(\mathbb{R}),|y|\leq M}|\partial_{(0,1)}b(m,y)|\right)\left|\partial_x \bar{X}_s^{(x),[n+1]} - \partial_x \bar{X}_s^{(x),[n]}\right| \\
&\quad + C\left|\partial_x X_s^{(x),[n]}\right|\left(\left|\bar{X}_s^{(x),[n+1]} - \bar{X}_s^{(x),[n]}\right| + D_{KR}\left(\bar{\mu}_s^{(x),[n+1]}, \bar{\mu}_s^{(x),[n]}\right)\right).
\end{aligned}$$

In particular, thanks to (56), (59) and Cauchy-Schwarz' inequality, for any $p \in \mathbb{N}^*$, for $t \leq T_p$ (where T_p has possibly been reduced compared to the previous one),

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^t \left|\left(\partial_x \bar{X}_s^{(x),[n+1]}\right)\partial_{(0,1)}b\left(\bar{\mu}_s^{(x),[n+1]}, \bar{X}_s^{(x),[n+1]}\right)\right.\right. \\
&\qquad\qquad \left.\left. - \left(\partial_x \bar{X}_s^{(x),[n]}\right)\partial_{(0,1)}b\left(\bar{\mu}_s^{(x),[n]}, \bar{X}_s^{(x),[n]}\right)\right| ds\right)^p\Big] \\
&\leq C_{M,p}T^p u_t^{[n],1}(M,p) + C_{M,p}T^p\frac{1}{2^{n/2}}.
\end{aligned} \tag{60}$$

The second term involves the following term

$$\left(\partial_x \tilde{X}_s^{(x),[n]}\right)\partial\left(b_{\bar{X}_s^{(x),[n]}}\right)\left(\bar{\mu}_s^{(x),[n]}, \tilde{X}_s^{(x),[n]}\right),$$

and can be handled exactly as (60), since the function $(m,x,y) \mapsto \partial_y\delta(b_x)(m,y)$ is assumed to be Lipschitz continuous and bounded (recalling that this function belongs to the set of first order mixed derivatives of b). Whence, for $T \leq T_p$,

$$u_T^{[n+1],1}(M,p) \leq C_{M,p}\left(T^p + T^{p/2}\right)\left(u_T^{[n],1}(M,p) + \frac{1}{2^{n/2}}\right).$$

And, possibly by reducing T_p,

$$u_T^{[n+1],1}(M,p) \leq \frac{1}{2}u_T^{[n],1}(M,p) + \frac{1}{2^{1+n/2}}.$$

Consequently, for $T \leq T_p$,

$$u_T^{[n],1}(M,p) \leq C_{M,p}2^{-n/2}, \tag{61}$$

and

$$\mathbb{E}\left[\sup_{s\leq T_p,|x|\leq M}\left|\partial_x\bar{X}_s^{(x),[n+1]}-\partial_x\bar{X}_s^{(x),[n]}\right|\right]\leq u_T^{[n],1}(M,p)^{1/p}\leq C_{M,p}2^{-n/(2p)}.$$

The quantities above being the terms of a convergent series, we can conclude as in *Step 1* that, almost surely, for all $t\in[0,T_p]$, $\partial_x\bar{X}_t^{(x),[n]}$ converges as n goes to infinity to some function locally uniformly w.r.t. x. Then, by Lemma B.1, we know that, almost surely, for all $t\in[0,T_p]$, the function $x\mapsto\bar{X}_t^{(x)}$ is C^1, and that its derivative is the limit of $\partial_x\bar{X}_t^{(x),[n]}$. In particular, by Fatou's lemma and (59),

$$\mathbb{E}\left[\sup_{t\leq T,|x|\leq M}\left|\partial_x\bar{X}_t^{(x)}\right|^p\right]<\infty\quad\text{and}\quad\sup_{x\in\mathbb{R}}\mathbb{E}\left[\sup_{t\leq T}\left|\partial_x\bar{X}_t^{(x)}\right|^p\right]<\infty. \tag{62}$$

Step 3. The proof that the function $x\mapsto\bar{X}_t^{(x)}$ is C^2 uses the same arguments as the ones used in *Step 2*. In order to make it clear, we just write the terms of the dynamics of $\partial^2_{xx}\bar{X}_t^{(x),[n+1]}$ coming from the two first lines of (58).

$$\begin{aligned}
\partial^2_{xx}\bar{X}_t^{(x),[n+1]}&=\int_0^t\left(\partial^2_{xx}\bar{X}_s^{(x),[n]}\right)\partial_{(0,1)}b\left(\bar{\mu}_s^{(x),[n]},\bar{X}_s^{(x),[n]}\right)ds\\
&+\int_0^t\left(\partial_x\bar{X}_s^{(x),[n]}\right)^2\partial_{(0,2)}b\left(\bar{\mu}_s^{(x),[n]},\bar{X}_s^{(x),[n]}\right)ds\\
&+\int_0^t\left(\partial_x\bar{X}_s^{(x),[n]}\right)\tilde{\mathbb{E}}\left[\left(\partial_x\tilde{X}_s^{(x),[n]}\right)\partial\left(\left(\partial_{(0,1)}b\right)_{\bar{X}_s^{(x),[n]}}\right)\left(\bar{\mu}_s^{(x),[n]},\tilde{X}^{(x),[n]}\right)\middle|\mathcal{W}_s\right]ds\\
&+\int_0^t\tilde{\mathbb{E}}\left[\left(\partial^2_{xx}\tilde{X}_s^{(x),[n]}\right)\partial\left(b_{\bar{X}_s^{(x),[n]}}\right)\left(\bar{\mu}_s^{(x),[n]},\tilde{X}_s^{(x),[n]}\right)\middle|\mathcal{W}_s\right]ds\\
&+\int_0^t\left(\partial_x\bar{X}_s^{(x),[n]}\right)\tilde{\mathbb{E}}\left[\left(\partial_x\tilde{X}_s^{(x),[n]}\right)\left(\partial_{\bar{x}}\partial\left(b_{\bar{x}}\right)\left(\bar{\mu}_s^{(x),[n]},\tilde{X}_s^{(x),[n]}\right)\right)_{\left|\bar{x}=\bar{X}_s^{(x),[n]}\right.}\middle|\mathcal{W}_s\right]ds\\
&+\int_0^t\check{\mathbb{E}}\tilde{\mathbb{E}}\left[\left(\partial_x\tilde{X}_s^{(x),[n]}\right)\left(\partial_x\check{X}_s^{(x),[n]}\right)\partial^2\left(b_{\bar{X}_s^{(x),[n]}}\right)\left(\bar{\mu}_s^{(x),[n]},\tilde{X}_s^{(x),[n]},\check{X}_s^{(x),[n]}\right)\middle|\mathcal{W}_s\right]ds\\
&+\ldots
\end{aligned}$$

where the omitted terms (hidden in the ellipsis) are the same as the ones written replacing the function b respectively by σ and ς, and ds by dB_s and dW_s. As previously, $\tilde{X}$ and $\check{X}$ are defined as $\bar{X}$ w.r.t. the same Brownian motion W, and w.r.t. respective Brownian motions $\tilde{B}$ and $\check{B}$ such that $W,\hat{B}$ and $\check{B}$ are independent. The expectations $\tilde{\mathbb{E}}$ and $\check{\mathbb{E}}$ are the expectation w.r.t. the laws of $\tilde{B}$ and $\check{B}$.

Once again, it can be noted that the functions appearing in the expression of $\partial^2_{xx}\bar{X}_t^{(x),[n+1]}$ are the mixed-derivatives of b,σ,ς up to order two. Hence the same reasoning as in *Step 2* allows to conclude. Finally, the proof that $x\mapsto\bar{X}_t^{(x)}$ is C^3 (resp. C^4) is the same as previously remarking that, calculating the expression of $\partial^3_{xxx}\bar{X}_t^{(x),[n+1]}$ (resp. $\partial^4_{xxxx}\bar{X}_t^{(x),[n+1]}$) makes appear the mixed-derivatives of b,σ,ς up to order three (resp. four) . Since it is assumed that all the mixed-derivatives up to order five are bounded, we know by, Remark 2.20, that the mixed-derivatives up to order four are Lipschitz continuous. So the end of the proof follows the same arguments as *Step 2*, and similar controls as (62) can be proved for the third and fourth derivatives.

Appendix C: Proofs of some technical lemmas on spaces of probability measures

Proof of Lemma 2.3. Let us consider $m_0 \in \mathcal{P}_1(\mathbb{R})$ and $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ such that, for all $m \in \mathcal{P}_1(\mathbb{R})$,

$$\int_{\mathbb{R}} h_1(x) d(m - m_0)(x) = \int_{\mathbb{R}} h_2(x) d(m - m_0)(x) + \varepsilon_{m_0}(m),$$

where $\varepsilon_{m_0}(m)/D_{KR}(m, m_0)$ vanishes as m converges to m_0.

Let us define

$$\Phi : m \in \mathcal{P}_1(\mathbb{R}) \longmapsto \int_{\mathbb{R}} (h_1(x) - h_2(x))\, d(m - m_0)(x).$$

So, by hypothesis, $\Phi(m)/D_{KR}(m, m_0)$ vanishes as m goes to m_0.

Step 1. In a first time, we prove that Φ is the zero function. Let us fix $m \in \mathcal{P}_1(\mathbb{R})$ and introduce, for any $n \in \mathbb{N}^*$,

$$\mu_n = \frac{n-1}{n} m_0 + \frac{1}{n} m \in \mathcal{P}_1(\mathbb{R}).$$

We have that, for all $n \in \mathbb{N}^*$,

$$D_{KR}(m_0, \mu_n) \leq \frac{1}{n} D_{KR}(m_0, m),$$

hence, μ_n converges to m_0 as n goes to infinity.

On the other hand,

$$\Phi(m) = n\Phi(\mu_n) = n \frac{\Phi(\mu_n)}{D_{KR}(m_0, \mu_n)} D_{KR}(m_0, \mu_n) \leq \frac{\Phi(\mu_n)}{D_{KR}(m_0, \mu_n)} D_{KR}(m_0, m) \underset{n \to \infty}{\longrightarrow} 0.$$

This proves that, for all $m \in \mathcal{P}_1(\mathbb{R})$, $\Phi(m) = 0$.

Step 2. To conclude the proof, it is then sufficient to notice that, thanks to *Step 1*, for any $y \in \mathbb{R}$, $\Phi(\delta_y) = 0$. Whence, for any $y \in \mathbb{R}$,

$$h_1(y) - h_2(y) = \int_{\mathbb{R}} (h_1(x) - h_2(x))\, dm_0(x).$$

In particular, the function $h_1 - h_2$ is constant, and the lemma is proved. □

Proof of Lemma 2.28. For each $1 \leq k \leq d$, there exists a probability space $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$ and two random variables on this space X_k, Y_k of respective laws m_k, μ_k such that

$$D_{KR}(m_k, \mu_k) = \mathbb{E}_k \left[|X_k - Y_k| \right],$$

with $\mathbb{E}_k$ the expectation w.r.t. $\mathbb{P}_k$ (the existence of these random variables is guaranteed by Theorem 4.1 of Villani (2009)). Let us assume that all the probability spaces $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$ are disjoint, and consider the product probability space of these spaces, denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. In particular, the variables X_k (resp. Y_k) $(1 \leq k \leq d)$ are independent, whence

$$\mathcal{L}(X_1, ..., X_d) = \bigotimes_{k=1}^{d} m_k \quad \text{and} \quad \mathcal{L}(Y_1, ..., Y_d) = \bigotimes_{k=1}^{d} \mu_k.$$

Consequently,

$$D_{KR}\left(\bigotimes_{k=1}^{d} m_k, \bigotimes_{k=1}^{d} \mu_k \right) \leq \mathbb{E}\left[\sum_{k=1}^{d} |X_k - Y_k| \right] = \sum_{k=1}^{d} \mathbb{E}_k \left[|X_k - Y_k| \right] = \sum_{k=1}^{d} D_{KR}(m_k, \mu_k),$$

which proves the result. □

Appendix D: Some technical lemmas about Ito's integrals

Let $(X_t)_{t\geq 0}$ be some càdlàg $\mathbb{R}$-valued process that is locally L^2: for all $t \geq 0$,

$$\int_0^t \mathbb{E}\left[X_s^2\right] ds < \infty. \tag{63}$$

Let W, B be two standard Brownian motions of dimension one, and $(\mathcal{F}_t)_t$ be a filtration such that $\mathcal{F}$, B and W are independent. Let us denote $(\mathcal{W}_t)_t$ (resp. $(\mathcal{B}_t)_t$) the filtration of W (resp. B), meaning

$$\mathcal{W}_t = \sigma\left(W_s \ : \ s \leq t\right) \text{ and } \mathcal{B}_t = \sigma\left(B_s \ : \ s \leq t\right),$$

and define $(\mathcal{G}_t)_t$ the union (in the filtration sense) of $\mathcal{F}$, $\mathcal{W}$ and $\mathcal{B}$: for all $t \geq 0$,

$$\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{W}_t \vee \mathcal{B}_t.$$

Lemma D.1. *Assume that X is $\mathcal{G}$-adapted. Then, for all $t \geq 0$,*

$$\mathbb{E}\left[\int_0^t X_s dW_s \middle| \mathcal{W}_t\right] = \int_0^t \mathbb{E}\left[X_s \middle| \mathcal{W}_s\right] dW_s.$$

Proof. Let us fix $t \geq 0$. By definition of Ito's integral,

$$\int_0^t X_s dW_s = \lim_{n\to\infty} \sum_{k=0}^{n-1} X_{s_k}\left(W_{s_{k+1}} - W_{s_k}\right)$$

in probability, where, for any $0 \leq k \leq n$, $s_k = t \cdot k/n$.

Then, thanks to (63), the (conditional) Vitali's convergence theorem implies that, almost surely,

$$\begin{aligned}
\mathbb{E}\left[\int_0^t X_s dW_s \middle| \mathcal{W}_t\right] &= \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{E}\left[X_{s_k}\left(W_{s_{k+1}} - W_{s_k}\right) \middle| \mathcal{W}_t\right] \\
&= \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{E}\left[X_{s_k} \middle| \mathcal{W}_t\right]\left(W_{s_{k+1}} - W_{s_k}\right) \\
&= \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{E}\left[X_{s_k} \middle| \mathcal{W}_{s_k}\right]\left(W_{s_{k+1}} - W_{s_k}\right) \\
&= \int_0^t \mathbb{E}\left[X_s \middle| \mathcal{W}_s\right] dW_s,
\end{aligned}$$

where the before last equality above comes from the fact that X is $\mathcal{G}$-adapted and that $\mathcal{W}_t$ can be written as the union (in the filtration sense) of $\mathcal{W}_{s_k}$ and $\sigma(W_r - W_{s_k} : s_k < r \leq t)$ which are independent (whence, X_{s_k} is also necessarily independent of $\sigma(W_r - W_{s_k} : s_k < r \leq t)$). □

Lemma D.2. *Assume that X is $\mathcal{G}$-adapted. Then, for all $t \geq 0$,*

$$\mathbb{E}\left[\int_0^t X_s dW_s \middle| \mathcal{B}_t\right] = 0.$$

Proof. With the same reasoning (and using the same notation) as in the beginning of the proof of Lemma D.1, for all $t \geq 0$,

$$\mathbb{E}\left[\int_0^t X_s dB_s \middle| \mathcal{W}_t\right] = \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{E}\left[X_{s_k}\left(B_{s_{k+1}} - B_{s_k}\right) \middle| \mathcal{W}_t\right]. \tag{64}$$

Let us recall that X is $\mathcal{G}$-adapted and that B, W are independent. Let us consider any $\mathcal{W}_t$-measurable random variable Z. Then, both Z and X_{s_k} are independent of $B_{s_{k+1}} - B_{s_k}$. So,

$$\mathbb{E}\left[X_{s_k}\left(B_{s_{k+1}} - B_{s_k}\right) Z\right] = \mathbb{E}\left[X_{s_k} Z\right] \mathbb{E}\left[B_{s_{k+1}} - B_{s_k}\right] = 0,$$

implying

$$\mathbb{E}\left[X_{s_k}\left(B_{s_{k+1}} - B_{s_k}\right) \middle| \mathcal{W}_t\right] = 0.$$

Combining the equation above with (64) proves the result. □

References

Andreis, L., Dai Pra, P. and Fischer, M. (2018). McKean–Vlasov limit for interacting systems with simultaneous jumps. *Stochastic Analysis and Applications* **36** 960–995.

Bauwens, L. and Hautsch, N. (2009). Modelling Financial High Frequency Data Using Point Processes. In *Handbook of Financial Time Series* (T. Mikosch, J.-P. Kreiß, R. A. Davis and T. G. Andersen, eds.) 953–979. Springer, Berlin, Heidelberg.

Billiard, S., Leman, H., Rey, T. and Tran, V. (2022). Continuous limits of large plant-pollinator random networks and some applications.

Bobkov, S. and Ledoux, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Memoirs of the American Mathematical Society* **261**.

Cardaliaguet, P. (2013). Notes on Mean Field Games.

Carmona, R. and Delarue, F. (2018). *Probabilistic Theory of Mean Field Games with Applications I. Probability Theory and Stochastic Modelling* **83**. Springer International Publishing, Cham.

Chassagneux, J.-F., Crisan, D. and Delarue, F. (2022). A Probabilistic Approach to Classical Solutions of the Master Equation for Large Population Equilibria. *Memoirs of the American Mathematical Society* **280**.

Cox, A. M. G., Källblad, S., Larsson, M. and Svaluto-Ferro, S. (2021). Controlled Measure-Valued Martingales: a Viscosity Solution Approach. arXiv:2109.00064 [math].

Crisan, D. and McMurray, E. (2018). Smoothing properties of McKean–Vlasov SDEs. *Probability Theory and Related Fields* **171** 97–148.

Cuchiero, C., Larsson, M. and Svaluto-Ferro, S. (2019). Probability measure-valued polynomial diffusions. *Electronic Journal of Probability* **24** 1–32. Publisher: Institute of Mathematical Statistics and Bernoulli Society.

Dawson, D. (1993). Measure-valued Markov processes. In *Ecole d'Eté de Probabilités de Saint-Flour XXI - 1991* (D. A. Dawson, B. Maisonneuve, J. Spencer and P.-L. Hennequin, eds.). *Lecture Notes in Mathematics* 1–260. Springer, Berlin, Heidelberg.

Eldredge, N. (2018). Strong hypercontractivity and strong logarithmic Sobolev inequalities for log-subharmonic functions on stratified Lie groups. *Nonlinear Analysis* **168** 1–26.

Erny, X., Löcherbach, E. and Loukianova, D. (2021). Conditional propagation of chaos for mean field systems of interacting neurons. *Electronic Journal of Probability* **26** 1–25. Publisher: Institute of Mathematical Statistics and Bernoulli Society.

Erny, X., Löcherbach, E. and Loukianova, D. (2022a). White-noise driven conditional McKean–Vlasov limits for systems of particles with simultaneous and random jumps. *Probability Theory and Related Fields* **183** 1027–1073.

Erny, X., Löcherbach, E. and Loukianova, D. (2022b). Mean field limits for interacting Hawkes processes in a diffusive regime. *Bernoulli* **28** 125–149. Publisher: Bernoulli Society for Mathematical Statistics and Probability.

Erny, X., Löcherbach, E. and Loukianova, D. (2023). Strong error bounds for the convergence to its mean field limit for systems of interacting neurons in a diffusive scaling. *The Annals of Applied Probability* **33** 3563–3586. Publisher: Institute of Mathematical Statistics.

Ethier, S. N. and Kurtz, T. G. (2005). *Markov Processes: Characterization and Convergence*, John wiley & sons ed. Wiley-Interscience, Hoboken, New Jersey.

Fleming, W. H. and Viot, M. (1979). Some Measure-Valued Markov Processes in Population Genetics Theory. *Indiana University Mathematics Journal* **28** 817–843. Publisher: Indiana University Mathematics Department.

Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **162** 707–738.

Fournier, N. and Löcherbach, E. (2016). On a toy model of interacting neurons. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **52** 1844–1876. Publisher: Institut Henri Poincaré.

Graham, C. (1992). McKean-Vlasov Ito-Skorohod equations, and nonlinear diffusions with discrete jump sets. *Stochastic Processes and their Applications* **40** 69–82.

Guo, X., Pham, H. and Wei, X. (2023). Itô's formula for flows of measures on semimartingales. *Stochastic Processes and their Applications* **159** 350–390.

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **34** 33–58.

Lu, X. and Abergel, F. (2018). High-dimensional Hawkes processes for limit order books: modelling, empirical analysis and numerical calibration. *Quantitative Finance* **18** 249–264. Publisher: Routledge _eprint: https://doi.org/10.1080/14697688.2017.1403142.

Meyn, S. P. and Tweedie, R. L. (1993). Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability* **25** 518–548. Publisher: Applied Probability Trust.

Mitchell, L. and Cates, M. E. (2009). Hawkes process as a model of social interactions: a view on video dynamics. *Journal of Physics A: Mathematical and Theoretical* **43** 045101.

Reynaud-Bouret, P. and Schbath, S. (2010). Adaptive Estimation for Hawkes Processes; Application to Genome Analysis. *The Annals of Statistics* **38** 2781–2822. Publisher: Institute of Mathematical Statistics.

Sznitman, A.-S. (1991). Topics in Propagation of Chaos. In *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989* (D. L. Burkholder, E. Pardoux and A.-S. Sznitman, eds.). *Lecture Notes in Mathematics* **1464** 165–251. Springer, Berlin, Heidelberg.

Vedenyapin, V., Sinitsyn, A. and Dulov, E. (2011). Boltzmann Equation. In *Kinetic Boltzmann, Vlasov and Related Equations* 165–182. Elsevier, London.

Villani, C. (2009). *Optimal Transport. Grundlehren der mathematischen Wissenschaften* **338**.

Springer, Berlin, Heidelberg.