McKean–Vlasov limit for interacting systems with simultaneous jumps

Luisa Andreis, Paolo Dai Pra & Markus Fischer

Tesla & Francis
Taylor & Francis Group

Check for updates

McKean–Vlasov limit for interacting systems with simultaneous jumps

Luisa Andreis[a], Paolo Dai Pra[b], and Markus Fischer[b]

[a]Weierstrass Institute, Berlin, Germany; [b]Department of Mathematics "Tullio Levi-Civita", University of Padua, Padova, Italy

ABSTRACT
Motivated by several applications, including neuronal models, we consider the McKean–Vlasov limit for a general class of mean-field systems of interacting diffusions characterized by an interaction via simultaneous jumps. We focus our interest on systems where the rate of the jumps is unbounded, which are rarely treated in the mean-field literature, and we prove well-posedness of the McKean–Vlasov limit together with propagation of chaos via a coupling technique. To highlight the role of simultaneous jumps, we introduce an intermediate process which is close to the original particle system but does not display simultaneous jumps. This shows in particular that the simultaneous jumps contribute to the overall rate of convergence of the N-particle empirical measures by a term of order $1/\sqrt{N}$.

1. Introduction

Treatable modeling for complex systems often involves the *mean-field* assumption: the system is comprised by several interacting components, whose distribution is permutation invariant. This assumption allows in several cases the derivation of *macroscopic* equations for the dynamics (*McKean–Vlasov* equations), in the limit as the number of components tends to infinity. Macroscopic behavior is also related to the phenomenon of *propagation of chaos*, which states that a fixed number of components become stochastically independent as the total number of components increases to infinity.

Since the introduction of this topic in the study of fluid dynamics [1, 2], dynamic mean-field models have been considered both in general [3–6] and for special models, motivated by life sciences [7–10] and social sciences [11–13]. General results on propagation of chaos include diffusions with jumps [4, 5], under suitable Lipschitz conditions on the coefficients. In recent years, neuronal networks have motivated the introduction of models whose components are allowed to jump simultaneously. Propagation of chaos is not obvious in these models, since simultaneous jumps could in principle interfere with asymptotic independence. In [14], this problem is solved via a rather involved approximation techniques, while existing techniques have been adapted to this context in [10, 15].

CONTACT Markus Fischer ✉ fischer@math.unipd.it 🖃 Department of Mathematics "Tullio Levi-Civita", University of Padua, via Trieste 63, 35121 Padova, Italy.

The purpose of this paper is to prove propagation of chaos results for general models with simultaneous jumps, which include the great majority of those cited above, by applying coupling techniques mainly borrowed from [4]. The peculiarity of this approach is the L^1 framework, reflected in the use of the Wasserstein-one distance. The L^1 framework, as opposed to the more common L^2 framework, allows to prescribe, in a natural way, the transition rates of the jumps. Besides simultaneous jumps, the models we treat may have drift or jump coefficients which are, though only locally Lipschitz as functions of the state variable, sufficiently stabilizing to yield strong existence and uniqueness of solutions. Notice that, when dealing with nonlinear Markov processes, the localization procedure usually employed in the proof of existence and uniqueness for SDEs with non-globally Lipschitz coefficients does not work in general [16].

We aim to highlight the role of the simultaneous jumps in this general class of systems. For this reason, we combine coupling arguments with the introduction of an "intermediate process" that will be convenient in handling the jump terms. The coupling approach for the proof of propagation of chaos, in particular in the presence of jump terms, is rarely chosen when the coefficients of the model are not globally Lipschitz continuous but it is of great interest since it provides a rate of convergence. Indeed, from a numerical or statistical points of view, it is desirable to know explicitly the rate at which the empirical measure of a system of interacting particles converges to the law of its McKean–Vlasov limit. Our work aims to underline that, for many systems considered in the literature, no ad hoc technique is necessary to obtain propagation of chaos results in the presence of jumps and only locally Lipschitz terms. It also provides a framework that can be applied to generalizations of the models recently proposed in neuroscience and in other applications.

The paper is organized as follows. In Section 2, we introduce at an informal level the framework, describing the main characteristics of both the particle system and the nonlinear process that we are interested in. In the following sections, we prove our results under precise assumptions, for three classes of systems. In Section 3, we work under Lipschitz conditions, where rather straightforward adaptations of standard techniques are applied. In Section 4, we introduce a class of nonlinear diffusions with jumps where the drift term comes from a convex potential. In Section 5, we take inspiration from the neuroscience models mentioned above and adapt our techniques to a class of piecewise deterministic processes, where the jump rate is superlinear.

2. Interacting particle systems and macroscopic limits

In this section, we introduce both the microscopic and the macroscopic dynamics at an informal level, and illustrate the phenomenon of propagation of chaos. In the remaining part of the paper well-posedness and convergence will be shown under various assumptions.

2.1. The microscopic dynamics

Let $X^N = (X_1^N, ..., X_N^N) \in \mathbb{R}^{d \times N}$ be the spatial positions of N different particles moving in $\mathbb{R}^d$. We introduce the corresponding *empirical measure*

$$\mu_X^N \doteq \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i^N}.$$

When the time variable appears explicitly in $X^N(t)$, we write $\mu_X^N(t)$ to indicate the time dependence of the empirical measure. Note that $\mu_X^N(t)$ is an element of $\mathcal{M}(\mathbb{R}^d)$, the set of probability measures on the Borel subsets of $\mathbb{R}^d$.

The particle positions $X^N(t)$ evolve as a jump diffusion process with the following specifications for the ith particle:

- a drift coefficient of the form $F(X_i^N(t), \mu_X^N(t))$ for some function $F : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \to \mathbb{R}^d$ common to all particles;
- a diffusion coefficient of the form $\sigma(X_i^N(t), \mu_X^N(t))$ for some function $\sigma : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \to \mathbb{R}^{d \times d_1}$, again the same for all particles;
- the jump amplitude and rate: particle i jumps by a random amplitude $\psi(X_i^N(t), \mu_X^N(t), h_i^N)$ with rate $\lambda(X_i^N(t), \mu_X^N(t))$; this *main* jump induces simultaneous *collateral* jumps of *all* other particles: the jth particle jumps by a random amplitude $\frac{\Theta(X_i^N(t), X_j^N(t), \mu_X^N(t), h_i^N, h_j^N)}{N}$, where randomness of the jumps is given by the random parameter $h^N = (h_i^N)_{i=1,\dots,n}$ which is distributed according to a symmetric probability measure ν_N on $[0,1]^N$, that is, ν_N is invariant under permutation of indices. Here λ, Ψ, and Θ are functions $\mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \to [0, \infty)$, $\mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \times [0,1] \to \mathbb{R}^d$, and $\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \times [0,1]^2 \to \mathbb{R}^d$, respectively.

In more analytic terms, we are considering a Markov process $X^N = \{X^N(t)\}_{t \in [0,T]}$ with values in $\mathbb{R}^{d \times N}$ whose infinitesimal generator takes the following form on a suitable family of test functions f:

$$\mathcal{L}^N f(x) = \sum_{i=1}^{N} \left[F(x_i, \mu_x^N) \cdot \partial_i f(x) + \frac{1}{2} \sum_{j,k=1}^{d} a(x_i, \mu_x^N)_{jk} \cdot \partial_i^2 f(x)_{jk} \right.$$
$$\left. + \lambda(x_i, \mu_x^N) \int_{[0,1]^N} \left(f(x + \Delta_i^N(x, \mu_x^N, h^N)) - f(x) \right) \nu_N(dh^N) \right],$$

where $\partial_i f(x)$ indicates the vector of first order derivatives w.r.t. x_i, $\partial_i^2 f(x)$ indicates the Hessian matrix of the second order derivatives w.r.t. x_i, $a(x_i, \mu_x^N) \doteq \sigma(x_i, \mu_x^N) \sigma(x_i, \mu_x^N)^*$ and

$$\Delta_i^N(x, \mu_x^N, h^N)_j \doteq \begin{cases} \dfrac{\Theta(x_i, x_j, \mu_x^N, h_i^N, h_j^N)}{N} & \text{for } j \neq i, \\ \psi(x_i, \mu_x^N, h_i^N) & \text{for } j = i. \end{cases}$$

Towards a rigorous construction, allowing the limit as $N \to +\infty$, let us consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$ satisfying the usual hypotheses, rich enough to carry an independent family $(B_i, \mathcal{N}^i)_{i \in \mathbb{N}}$ of d-dimensional Brownian motions B_i and Poisson random measures $\mathcal{N}^i$ with characteristic measure $l \times l \times \nu$. Here l is the Lebesgue measure restricted to $[0, \infty)$ and ν is a symmetric probability measure on $[0,1]^{\bar{N}}$ such that, for every $N \geq 1$, ν_N coincides with the projection of ν on the first N

coordinates, for a consistent family of symmetric probability distributions $\{\nu_N\}_{n\in\mathbb{N}}$. We will construct X^N as the solution of the following SDE

$$
dX_i^N(t) = F\big(X_i^N(t), \mu_X^N(t)\big)dt + \sigma\big(X_i^N(t), \mu_X^N(t)\big)dB_t^i
$$
$$
+\frac{1}{N}\sum_{j\neq i}\int_{[0,\infty)\times[0,1]^{\mathbb{N}}}\Theta\big(X_j^N(t^-), X_i^N(t^-), \mu_X^N(t^-), h_j, h_i\big)\mathbb{1}_{\left(0,\lambda\big(X_j^N(t^-),\mu_X^N(t^-)\big)\right]}(u)\mathcal{N}^j(dt, du, dh)
$$
$$
+\int_{[0,\infty)\times[0,1]^{\mathbb{N}}}\psi\big(X_i^N(t^-), \mu_X^N(t^-), h_i\big)\mathbb{1}_{\left(0,\lambda\big(X_i^N(t^-),\mu_X^N(t^-)\big)\right]}(u)\mathcal{N}^i(dt, du, dh),
$$

$$(1)$$

$i = 1, ..., N$. The existence and uniqueness of a solution starting from a vector of initial conditions $(X_1^N(0), ..., X_N^N(0))$ depends obviously on the assumptions on the coefficients, and we will specify sufficient conditions in the following sections.

Remark 2.1. Notice that we made the choice of considering separately the jump's rate and amplitude, a similar representation appears in [17]. This is motivated by the fact that we are mainly interested in the jumps and we want to state a clear framework, useful for applications. The non-compensated jump component is often represented by a measure that does not directly describe the behavior of the system. Here, we want to highlight the role of the jumps, therefore we describe a diffusion process that at each position has a certain jump rate and a set of possible jumps, represented by the functions λ and Δ^N, respectively. Our assumptions will allow for unbounded jump rates. In the next sections, we will see that the first natural assumption is to have globally Lipschitz conditions on the functions λ and Δ^N. This is the reason why we need to perform all our proofs in an L^1 framework, instead of the classical L^2 approach for stochastic calculus. Indeed, when dealing with the well-posedness of the nonlinear Markov process, we will need to bound expectations of the supremum over a time interval of an integral w.r.t. the Poisson random measure $\mathcal{N}$. In an L^2 framework, this involves the corresponding compensated martingale $\widetilde{\mathcal{N}}$ and needs bounds of the type, for X, Y $\in \mathbb{R}^{Nd}$,

$$
\int_0^{\infty}\int_{[0,1]^{\mathbb{N}}}\big\|\Delta^N(X, h)\mathbb{1}_{(0,\lambda(X)]}(u) - \Delta^N(Y, h)\mathbb{1}_{(0,\lambda(Y)]}(u)\big\|^p du\nu(dh) \leqslant C\big\|X-Y\big\|^p, \quad (2)
$$

for $p = 2$. However, sometimes (2) may hold for $p = 1$, but not for $p = 2$, which justifies the choice of getting the L^1 framework, where we do not need to compensate the process N. For instance, if Δ is constant and λ is globally Lipschitz, (2) holds for $p = 1$ and not $p = 2$.

2.2. Macroscopic process

Suppose the solution X^N of (1) exists, and that its initial condition has a permutation invariant distribution. Fix an arbitrary component i, and assume the process X_i^N has a limit in distribution; by symmetry, the law of the limit does not depend on i, so we denote by X the limit process. To identify, at a heuristic level, its law, we make the further assumption that a law of large numbers holds, i.e. $\mu_X^N(t)$ converges, as $N \to +\infty$, to the law μ_t of X(t). Letting $N \to +\infty$ in (1) we deduce, at a purely formal level, that the limit process X(t) has the law of the solution of the *McKean–Vlasov* SDE:

$$dX(t) = \left(F\big(X(t), \mu_t\big) + \left\langle \mu_t, \lambda(\cdot, \mu_t) \int_{[0,1]^2} \Theta\big(\cdot, X(t), \mu_t, h_1, h_2\big)\nu_2(dh_1, dh_2) \right\rangle \right) dt$$

$$+ \sigma\big(X(t), \mu_t\big)dB_t + \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \psi\big(X(t^-), \mu_t, h_1\big) 1_{\left(0, \lambda\big(X(t^-), \mu_t\big)\right]}(u) N(dt, du, dh). \tag{3}$$

Here, B is a d_1-dimensional Brownian motion and $\mathcal{N}$ an independent Poisson random measure with characteristic measure $dtdu\nu(dh)$ on $[0,\infty)^2 \times [0,1]^{\mathbb{N}}$ as above. By $\langle \cdot, \cdot \rangle$ we indicate the integral of a function on its domain with respect to a certain measure; thus, $\langle \mu, \phi \rangle = \int_{\mathbb{R}^d} \phi(y)\mu(dy)$.

Remark 2.2. The Poisson random measures appearing in Equations (1) and (3), respectively, have characteristic measure defined on $[0,\infty)^2 \times [0,1]^{\mathbb{N}}$. The two equations could equivalently be stated in terms of Poisson random measures with characteristic measures defined on $[0,\infty)^2 \times [0,1]^{\mathbb{N}}$ (namely, $l \times l \times \nu_N$) and on $[0,\infty)^2 \times [0,1]$ (namely, $l \times l \times \nu_1$). The reason for our seemingly unnatural choice is that it prepares for the coupling argument we will use below to establish propagation of chaos. We will need, for each N, a coupling of the N-particle system with N independent copies of the limit system.

Existence and uniqueness of solutions to (3) starting from a given initial condition X(0) will be discussed in the following sections. Note that (3) is not a standard SDE since the law μ_t of the solution appears as an argument of its coefficients. It is often referred to as McKean–Vlasov SDE, as it is customary to call McKean–Vlasov equation the partial differential equation solved by the law μ_t, namely, in the weak form,

$$\langle \mu_t, \phi \rangle - \langle \mu_0, \phi \rangle = \int_0^t \langle \mu_s, \mathcal{L}(\mu_s)\phi \rangle ds,$$

where

$$\mathcal{L}(\mu_t)\phi(x) \doteq F(x, \mu_t)\partial\phi(x) + \frac{1}{2}\sum_{j,k=1}^d a(x, \mu_t)_{jk}\partial^2\phi(x)_{jk}$$

$$+ \left\langle \mu_t, \lambda(\cdot, \mu_t)\int_{[0,1]^2} \Theta(\cdot, x, \mu_t, h_1, h_2)\nu_2(dh_1, dh_2) \right\rangle \partial\phi(x)$$

$$+ \lambda(x, \mu_t)\int_{[0,1]} \big(\phi(x + \psi(x, \mu_t, h_1)) - \phi(x)\big)\nu_1(dh_1).$$

2.3. Propagation of chaos

The link between the microscopic dynamics (1) and the macroscopic limit (3) is explained by the phenomenon of propagation of chaos. Let μ be a probability measure on $\mathbb{R}^d$. We assume that the sequence of the distributions of $X^N(0)$ is μ-*chaotic*: for every $k \in \mathbb{N}$, the vector $(X_1^N(0), X_2^N(0), ..., X_k^N(0))$ converges in distribution to the product measure $\mu^{\otimes k}$. Fix an arbitrary time horizon T>0, and denote by $X^N[0, T] = (X^N(t))_{t\in[0,T]}$ the random path of the microscopic process up to time T. We say that *propagation of chaos* holds if the distribution of $X^N[0, T]$ is itself Q-chaotic for some probability measure Q on the Skorohod space of *càdlàg* functions $\mathbb{D}([0, T], \mathbb{R}^d)$, that is,

for each fixed $k \in \mathbb{N}$, the vector of random paths $(X_1^N[0, T], X_2^N[0, T], ..., X_k^N[0, T])$ converges in distribution to $Q^{\otimes k}$. For a comprehensive introduction to the notion of propagation of chaos see [6].

To illustrate the general strategy of proof, it is useful to introduce an *intermediate process* $Y^N = (Y^N(t))_{t \in [0, T]}$ with values in $\mathbb{R}^{d \times N}$. This Markov process Y^N can be given as the solution of the SDE

$$
dY_i^N(t) = F\left(Y_i^N(t), \mu_Y^N(t)\right)dt + \sigma\left(Y_i^N(t), \mu_Y^N(t)\right)dB_t^i
$$
$$
+ \frac{1}{N}\sum_{j=1}^N \lambda\left(Y_i^N(t^-), \mu_Y^N(t^-)\right)\int_{[0,1]^2} \Theta\left(Y_j^N(t^-), Y_i^N(t^-), \mu_Y^N(t^-), h_1, h_2\right)\nu_2(dh_1, dh_2)dt \tag{4}
$$
$$
+ \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \psi\left(Y_i^N(t^-), \mu_Y^N(t^-), h\right)\mathbb{1}_{\left(0, \lambda\left(Y_i^N(t^-), \mu_i^N(t^-)\right)\right)}(u)N^i(dt, du, dh),
$$

$i = 1, ..., N$, where again B^i are independent d-dimensional Brownian motions and N^i are independent Poisson random measures with characteristic measure $l \times l \times \nu$. It is immediate to see that the process Y^N differs from the original process X^N in the jump terms; indeed, here the *collateral jumps* have been absorbed by a new drift term, while the amplitude of the remaining jumps affects only one component a time. By using the *same* Brownian motions and the *same* Poisson random measures in (1) and in (4), the processes X^N and Y^N are *coupled*, i.e. are realized on the same probability space: it will not be hard to give conditions for the L^1-convergence to zero of $X_1^N[0, T] - Y_1^N[0, T]$. Thus, the fact that the law of X^N is Q-chaotic will follow if one shows that the law of Y^N is Q-chaotic. Since Y^N has no simultaneous jumps, this can be obtained along the lines of the classical approach. The intermediate process has the nice feature of highlighting the role of simultaneous jumps in the rate of convergence in W_1 Wasserstein distance of the empirical measure. Indeed by comparing the empirical measures of X^N and Y^N, we obtain that the rate of convergence due to the simultaneous jumps is of the order $1/\sqrt{N}$, while the final rate obviously depends on the moments of initial conditions and of the process itself, see [18].

3. Globally Lipschitz conditions on all coefficients

In this section, we give Lipschitz conditions under which we can prove rigorously the results informally stated in the previous section. To state these conditions and the corresponding theorems, we need a suitable metric on spaces of probability measures.

Let $\mathcal{M}^1(\mathbb{R}^d)$ be the space of probability measures on $\mathbb{R}^d$ with finite first moment:

$$
\mathcal{M}^1(\mathbb{R}^d) = \left\{\mu \in \mathcal{M}(\mathbb{R}^d) : \int \|x\| \mu(dx) < +\infty\right\}.
$$

This space is equipped with the W_1 Wasserstein metric:

$$
\rho(\mu, \nu) \doteq \inf\left\{\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\| \pi(dx, dy); \pi \text{ has marginals } \mu \text{ and } \nu\right\}
$$
$$
= \sup\left\{\langle g, \mu\rangle - \langle g, \nu\rangle : g : \mathbb{R}^d \to \mathbb{R}, \|g(x) - g(y)\| \leqslant \|x - y\|\right\}.
$$

We also consider the following subset of $\mathcal{M}(\mathbf{D}([0, T], \mathbb{R}^d))$, the set of the probability measures on $\mathbf{D}([0, T], \mathbb{R}^d)$:

$$\mathcal{M}^1\big(\mathbf{D}\big([0,T],\mathbb{R}^d\big)\big)\doteq\left\{\alpha\in\mathcal{M}\big(\mathbf{D}\big([0,T],\mathbb{R}^d\big)\big):\int_{\mathbf{D}}\sup_{t\in[0,T]}\parallel\mathbf{x}(t)\parallel\alpha(dx)<+\infty\right\},$$

and provide it with the metric

$$\rho_T(\alpha,\beta)\doteq\inf\left\{\int_{\mathbf{D}\times\mathbf{D}}\sup_{t\in[0,T]}\parallel\mathbf{x}(t)-\mathbf{y}(t)\parallel P(dx,dy);\quad\text{where P has marginals }\alpha\text{ and }\beta\right\}.$$

The metric ρ_T is not complete on $\mathcal{M}^1(\mathbf{D}([0,T_0],\mathbb{R}^d))$ as it is based on the supremum norm. In our proofs all the estimates that we make will clearly hold also for the weaker Wasserstein metric

$$\bar{\rho}_T(\alpha,\beta)\doteq\inf\left\{\int_{\mathbf{D}\times\mathbf{D}}d_T(x,y)P(dx,dy);\quad\text{where P has marginals }\alpha\text{ and }\beta\right\},$$

where d_T is the classical Skorohod metric, see [19]. Since d_T is complete on $\mathbf{D}([0,T_0],\mathbb{R}^d)$, $\bar{\rho}_T$ is complete on $\mathcal{M}^1(\mathbf{D}([0,T_0],\mathbb{R}^d))$ (see, for instance, [20]), and it is weaker than ρ_T because the Skorohod metric d_T is bounded from above by the distance in supremum norm. Therefore, we choose to give our estimates with respect to ρ_T because of its simplicity.

In what follows, we shall adopt a notion of chaoticity that is stronger than the one we stated above.

Definition 3.1. Let $X^N=(X_1^N,X_2^N,...,X_N^N)$ be a sequence of random vectors with components $X_i^N\in\mathbb{R}^d$ (resp. $X_i^N\in\mathbf{D}([0,T],\mathbb{R}^d)$). For $\mu\in\mathcal{M}^1(\mathbb{R}^d)$ (resp. $\mu\in\mathcal{M}^1(\mathbf{D}([0,T],\mathbb{R}^d))$), we say that X^N is μ-chaotic in W_1 if its distribution is permutation invariant and, for each $k\in\mathbb{N}$, the law of the vector $(X_1^N,X_2^N,...,X_k^N)$ converges to $\mu^{\otimes k}$ with respect to the metric ρ (resp. ρ_T).

3.1. Existence and uniqueness of solutions for the particle system and the McKean–Vlasov equation

The conditions on the coefficients of system (1) and the corresponding limit (3) are as follows:

Assumption 1. (Li) The classical global Lipschitz assumption on F and σ: $\exists\,\widetilde{L}>0$ such that, for all $x,y\in\mathbb{R}^d$, all $\alpha,\gamma\in\mathcal{M}^1(\mathbb{R}^d)$,

$$\parallel F(\mathbf{x},\alpha)-F(\mathbf{y},\gamma)\parallel\vee\parallel\sigma(\mathbf{x},\alpha)-\sigma(\mathbf{y},\gamma)\parallel\leqslant\widetilde{L}\Big(\parallel\mathbf{x}-\mathbf{y}\parallel+\rho(\alpha,\gamma)\Big).$$

(I) The integrability condition: for all $N\in\mathbb{N}$, for all $\mathbf{x}\in\mathbb{R}^{d\times N}$ and all $\alpha\in\mathcal{M}^1(\mathbb{R}^d)$

$$\int_{[0,1]^N}\|\Delta_i^N\big(\mathbf{x},\alpha,h^N\big)\|\nu_N(dh^N)<\infty.$$

(L1) The L^1-Lipschitz assumption on the jump coefficients: $\exists\,\bar{L}>0$ such that, for all $x,y\in\mathbb{R}^d$, all $\alpha,\gamma\in\mathcal{M}^1(\mathbb{R}^d)$,

$$\int_{[0,\infty)\times[0,1]} \Big\| \psi(x,\alpha,h)\mathbb{1}_{\left(0,\lambda(x,\alpha)\right]}(u) - \psi(y,\gamma,h)\mathbb{1}_{\left(0,\lambda(y,\gamma)\right]}(u) \Big\| \, du\nu_1(dh) \leqslant \widetilde{L}\Big(\| x - y \| + \rho(\alpha,\gamma)\Big)$$

and

$$\Big\| \langle \alpha, \lambda(\cdot,\alpha)\int_{[0,1]^2}\Theta(\cdot,x,\alpha,h_1,h_2)\nu_2(dh_1,dh_2)\rangle - \langle \gamma, \lambda(\cdot,\gamma)\int_{[0,1]^2}\Theta(\cdot,y,\gamma,h_1,h_2)\nu_2(dh_1,dh_2)\rangle \Big\|$$
$$\leqslant \bar{L}\Big(\| x - y \| + \rho(\alpha,\gamma)\Big).$$

In the following, we set $L \doteq \widetilde{L}\vee\bar{L}$.

Existence and uniqueness of a square integrable strong solution of (1) and (4) starting from a vector of square integrable initial conditions $(X_1^N(0),...,X_N^N(0))$, independent of the family $(B_i, N^i)_{i\in\mathbb{N}}$, are ensured by Assumption 1; see Theorem 1.2 in [4]. The same assumptions also guarantee existence and uniqueness of a strong solution of (3) starting from any square integrable initial condition $X(0)$; see Theorem 2.1 in [4].

3.2. Propagation of chaos

In addition to the aforementioned assumptions, for the proof of propagation of chaos, we will need the following square integrability condition on the amplitude of the collateral jumps:

Assumption 2. (I2) There exists a constant $C > 0$ such that, for all $x, y \in \mathbb{R}^d$ and all $\alpha \in \mathcal{M}^1(\mathbb{R}^d)$

$$\int_{[0,\infty)\times[0,1]^\mathbb{N}} \Big\| \Theta(x,y,\alpha,h_1,h_2)\mathbb{1}_{\left(0,\lambda(x,\alpha)\right]}(u) \Big\|^2 du\nu_2(dh) \leqslant C.$$

We begin by establishing the closeness between the original particle system X^N and the intermediate process Y^N.

Proposition 3.1. *Grant Assumptions 1 and 2. Let X^N and Y^N be the solutions of (1) and (4), respectively. We assume the two processes are driven by the same Brownian motions and Poisson random measures, and start from the same square integrable and permutation invariant initial condition. Then there exists a constant $C_T > 0$ such that, for each fixed $i \in \mathbb{N}$, for all $N \geqslant 1$*

$$\mathbf{E}\left[\sup_{t\in[0,T]} \big\| X_i^N(t) - Y_i^N(t) \big\| \right] \leqslant \frac{C_T}{\sqrt{N}}. \tag{5}$$

Proof. To simplify notation, we adopt the following abbreviations:

$$\Theta_{i,j}\big(X^N(s^-),h\big) \doteq \Theta\Big(X_i^N(s^-), X_j^N(s^-), \mu_X^N(s^-), h_i, h_j\Big),$$
$$\lambda_i\big(X^N(s^-)\big) \doteq \lambda\big(X_i^N(s^-), \mu_X^N(s^-)\big),$$
$$\psi_i\big(X^N(s^-),h\big) \doteq \psi\big(X_i^N(s^-), \mu_X^N(s^-), h_i\big),$$
$$U \doteq [0,\infty) \times [0,1]^\mathbb{N}.$$

By permutation invariance of both the initial condition and the dynamics, we have, for every $t \in [0,T]$,

$$\mathbf{E}\left[\sup_{s\in[0,t]}\|X_i^N(s)-Y_i^N(s)\|\right]=\frac{1}{N}\sum_{j=1}^{N}\mathbf{E}\left[\sup_{s\in[0,t]}\|X_j^N(s)-Y_j^N(s)\|\right].$$

Moreover this, together with the coupling bound for the W_1 Wasserstein metric, implies that, for all $t\geqslant0$

$$\mathbf{E}\left[\rho\big(\mu_X^N(t),\mu_Y^N(t)\big)\right]\leqslant\frac{1}{N}\sum_{j=1}^{N}\mathbf{E}\left[\|X_j^N(t)-Y_j^N(t)\|\right]=\mathbf{E}\left[\|X_i^N(t)-Y_i^N(t)\|\right].$$

Fix $t\in[0,T]$, and set

$$F_i\doteq\mathbf{E}\left[\int_0^t\|F\big(X_i^N(s),\mu_X^N(s)\big)-F(Y_i^N(s),\mu_Y^N(s))\|\,ds\right],$$

$$\sigma_i\doteq\mathbf{E}\left[\sup_{r\in[0,t]}\|\int_0^r(\sigma(X_i^N(s),\mu_X^N(s))-\sigma(Y_i^N(s),\mu_Y^N(s)))dB_s^i\|\right],$$

$$\Theta_i\doteq\mathbf{E}[\sup_{r\in[0,t]}\|\frac{1}{N}\sum_{j\neq i}\int_{[0,r]\times U}\Theta_{j,i}(X^N(s^-),h)\mathbb{1}_{(0,\lambda_j(X^N(s^-)))}(u)N^j(ds,du,dh)$$

$$-\frac{1}{N}\sum_{j=0}^{N}\int_{[0,r]\times U}\Theta_{j,i}(Y^N(s^-),h)\mathbb{1}_{(0,\lambda_j(Y^N(s^-)))}(u)ds\;du\nu(dh)\|],$$

$$\psi_i\doteq\mathbf{E}[\sup_{r\in[0,t]}\|\int_{[0,r]\times U}\psi_i(X^N(s^-),h)\mathbb{1}_{(0,\lambda_i(X^N(s^-)))}(u)N^i(ds,du,dh)$$

$$-\int_{[0,r]\times U}\psi_i(Y^N(s^-),h)\mathbb{1}_{(0,\lambda_i(Y^N(s^-)))}(u)N^i(ds,du,dh)\|].$$

Note that all these quantities do not depend on the choice of i, which is therefore omitted in what follows. Then

$$\mathbf{E}\left[\sup_{s\in[0,t]}\|X_i^N(s)-Y_i^N(s)\|\right]\leqslant F+\sigma+\Theta+\psi.\qquad(6)$$

The term F can be easily bounded thanks to the Lipschitz condition (Li) and the coupling bound for the W_1 Wasserstein metric, and we obtain

$$F\leqslant2L\int_0^t\mathbf{E}\left[\|X_i^N(s)-Y_i^N(s)\|\right]ds.$$

The bound on σ involves the Burkholder–Davis–Gundy inequality, and we get, for some constant M not depending on N nor t,

$$\sigma\leqslant M\mathbf{E}\left[\left(\int_0^t\left(\|X_i^N(s)-Y_i^N(s)\|+\frac{1}{N}\sum_{j=1}^{N}\|X_j^N(s)-Y_j^N(s)\|\right)^2ds\right)^{1/2}\right]$$

$$\leqslant2M\sqrt{t}\mathbf{E}\left[\sup_{s\in[0,t]}\|X_i^N(s)-Y_i^N(s)\|\right].$$

The term Θ needs to be treated again with the Burkholder–Davis–Gundy inequality. In what follows, we denote by $\widetilde{\mathcal{N}^i}$ the compensated Poisson measure associated to $\mathcal{N}^i$. Notice the fact that $\{\widetilde{\mathcal{N}^i}\}_{i=1,\ldots,N}$ is a family of orthogonal martingales. Therefore, for a certain constant $K>0$ coming from the Burkholder–Davis–Gundy inequality, the constant $L>0$ coming from condition (L1) and a constant $C>0$ not depending on N nor t, we have

$$\Theta \leqslant \mathbf{E}\left[\sup_{r\in[0,t]} \| \frac{1}{N}\sum_{j\neq i}\int_{[0,r]\times U}\Theta_{j,i}(X^N(s^-),h)\mathbb{1}_{[0,\lambda_j(X^N(s^-)))}\widetilde{N}^j(ds,du,dh) \| \right]$$

$$+ \mathbf{E}\left[\sup_{r\in[0,t]} \| \frac{1}{N}\sum_{j=1}^{N}\int_{[0,r]\times U}(\Theta_{j,i}(X^N(s^-),h)\mathbb{1}_{(0,\lambda_j(X^N(s^-))]} - \Theta_{j,i}(Y^N(s^-),h)1_{(0,\lambda_j(Y^N(s^-)))})ds\ du\nu(dh) \| \right]$$

$$+\frac{1}{N}\mathbf{E}\left[\sup_{r\in[0,t]} \| \int_{[0,r]\times U}\Theta_{i,i}(X^N(s^-),h)\mathbb{1}_{(0,\lambda_j(X^N(s^-)))}ds\ du\ \nu(dh) \| \right]$$

$$\leqslant\frac{K}{N}\mathbf{E}\left[\left(\sum_{j\neq i}\int_0^t\int_U \| \Theta_{j,i}(X^N(s^-),h)\mathbb{1}_{(0,\lambda_j(X^N(s^-)))}(u) \|^2 ds\ du\nu(dh)\right)^{1/2}\right]$$

$$+\int_0^t\mathbf{E}\left[\| \left\langle \mu_s^N,\int_U\Theta_{\cdot,i}(X^N(s^-),h)\mathbb{1}_{[0,\lambda.(X^N(s^-)))}(u)\ du\nu(dh)\right\rangle\right.$$

$$\left.-\left\langle \bar{\mu}_s^N,\int_U\Theta_{\cdot,i}(Y^N(s^-),h)\mathbb{1}_{[0,\lambda.(Y^N(s^-)))}(u)\ du\nu(dh)\right\rangle \| \right]ds$$

$$+\frac{1}{N}\mathbf{E}\left[\int_0^t\int_U \| \Theta_{i,i}(X^N(s^-),h)\mathbb{1}_{[0,\lambda_i(X^N(s^-)))}(u) \| du\nu(dh)ds\right]$$

$$\leqslant\frac{C}{\sqrt{N}}+2L\int_0^t\mathbf{E}[\| X_i^N(s)-Y_i^N(s) \|]ds+\frac{C}{N}.$$

The term ψ concerns the main jumps of the particle system and is bounded by the positivity property of Poisson processes and the Lipschitz condition (L1):

$$\psi \leqslant \mathbf{E}\left[\int_{[0,t]\times U} \| \psi_i(X^N(s^-),h)\mathbb{1}_{(0,\lambda_i(X^N(s^-)))}(u)-\psi_i(Y^N(s^-),h)\mathbb{1}_{(0,\lambda_i(Y^N(s^-)))}(u) \| N^i(ds,du,dh)\right]$$

$$= \mathbf{E}\left[\int_{[0,t]\times U} \| \psi_i(X^N(s^-),h)\mathbb{1}_{(0,\lambda_i(X^N(s^-)))}(u)-\psi_i(Y^N(s^-),h)\mathbb{1}_{(0,\lambda_i(Y^N(s^-)))}(u) \| ds\ du\nu(dh)\right]$$

$$\leqslant 2L\int_0^t\mathbf{E}\left[\| X_i^N(s)-Y_i^N(s) \| \right]ds.$$

Therefore, recalling (6), we find that, for every $t\in[0,T]$,

$$\mathbf{E}[\sup_{s\in[0,t]} \| X_i^N(s)-Y_i^N(s) \|]\leqslant 2M\sqrt{t}\mathbf{E}[\sup_{s\in[0,t]} \| X_i^N(s)-Y_i^N(s) \|]+6L$$

$$\int_0^t\mathbf{E}[\| X_i^N(s)-Y_i^N(s) \|]ds+\frac{C}{N}+\frac{C}{\sqrt{N}}.$$

Choose $T_0>0$ small enough so that $(1-2M\sqrt{T_0})>0$. By summing over the index i in the above inequality and dividing both sides by N, we can move the first two terms on the right-hand side to the left, obtaining, for every $t\in[0,T_0]$,

$$\mathbf{E}\left[\sup_{s\in[0,t]} \| X_i^N(s)-Y_i^N(s) \| \right]\leqslant\frac{6K}{1-2M\sqrt{t}}\int_0^t\mathbf{E}\left[\sup_{s\in[0,r]} \| X_i^N(s)-Y_i^N(s) \| \right]dr$$

$$+\frac{C}{N(1-2M\sqrt{t})}+\frac{C}{\sqrt{N}(1-2M\sqrt{t})}.$$

An application of Gronwall's lemma yields

$$\mathbf{E}\left[\sup_{t\in[0,T_0]} \| X_i^N(t)-Y_i^N(t) \| \right]\leqslant\frac{C_{T_0}}{\sqrt{N}} \tag{7}$$

for some finite constant C_{T_0} not depending on N. Recall that (7) holds on a time interval $[0,T_0]$ for T_0 sufficiently small. If T_0 is smaller than T, then we can repeat the

procedure of estimates on the interval $[T_0, (2T_0) \wedge T]$. In this case, we find that, for every $t \in [T_0, (2T_0) \wedge T]$,

$$\mathbf{E}\left[\sup_{s \in [T_0, t]} \| X_i^N(s) - Y_i^N(s) \|\right] \leqslant \frac{1}{1 - 2M\sqrt{t - T_0}} \left(\mathbf{E}\left[\sup_{s \in [0, T_0]} \| X_i^N(s) - Y_i^N(s) \|\right]\right)$$
$$+ \frac{6K}{1 - 2M\sqrt{t - T_0}} \int_{T_0}^t \mathbf{E}\left[\sup_{s \in [T_0, r]} \| X_i^N(s) - Y_i^N(s) \|\right] dr$$
$$+ \frac{C}{N(1 - 2M\sqrt{t - T_0})} + \frac{C}{\sqrt{N}(1 - 2M\sqrt{t - T_0})},$$

where the first term comes from a bound on the initial condition $\mathbf{E}[\|X_i^N(T_0) - Y_i^N(T_0)\|]$. Hence, again thanks to Gronwall's lemma, for some constant C_{2, T_0},

$$\mathbf{E}\left[\sup_{s \in [0, (2T_0) \wedge T]} \| X_i^N(s) - Y_i^N(s) \|\right] \leqslant \left(\frac{C_{T_0} + C}{\sqrt{N}(1 - 2M\sqrt{t - T_0})} + \frac{C}{N(1 - 2M\sqrt{t - T_0})}\right) e^{\frac{6KT_0}{1 - 2M\sqrt{t - T_0}}}$$
$$\leqslant \frac{C_{2, T_0}}{\sqrt{N}}.$$

We proceed by induction until we cover, after finitely many steps, the entire interval $[0, T]$. By exchangeability of the laws of both the initial and the intermediate process, this yields, for $i = 1, ..., N$,

$$\mathbf{E}\left[\sup_{s \in [0, T]} \| X_i^N(s) - Y_i^N(s) \|\right] \leqslant \frac{C_T}{\sqrt{N}}$$

and (5) holds. □

In the next proposition, we use a similar coupling technique and we now show propagation of chaos for Y^N.

Proposition 3.2. *Grant Assumptions 1 and 2. Let μ_0 be a probability measure on $\mathbb{R}^d$ such that $\int \|x\|^2 \mu_0(dx) < +\infty$. For $N \in \mathbb{N}$, let Y^N be a solution of Equation (4) in $[0, T]$. Assume that $Y^N(0) = (Y_1^N(0), ..., Y_N^N(0))$, $N \in \mathbb{N}$, form a sequence of square integrable random vectors that is μ_0-chaotic in W_1. Let μ be the law of the solution of Equation (3) in $[0, T]$ with initial law $\mathbf{P} \circ X(0)^{-1} = \mu_0$. Then Y^N is μ-chaotic in W_1.*

Proof. Let the processes Y_i^N, $N \in \mathbb{N}$, $i \in \{1, ..., N\}$ be all defined on the filtered probability space $(\Omega, F, (F_t)_{t \geqslant 0}, \mathbf{P})$ with respect to the family $(B_i, N^i)_{i \in \mathbb{N}}$ of Brownian motions and Poisson random measures. Since $(Y^N(0))$ is μ_0-chaotic in W_1 by hypothesis, we assume, as we may, that our stochastic basis carries a triangular array $(\bar{X}_i^N(0))_{i \in \{1, ..., N\}, N \in \mathbb{N}}$ of identically distributed $\mathbb{R}^d$-valued random variables with common distribution μ_0 such that $(\bar{X}_i^N(0))_{i \in \{1, ..., N\}, N \in \mathbb{N}}$ and $(B_i, N^i)_{i \in \mathbb{N}}$ are independent, the sequence $(\bar{X}_i^N(0))_{i \in \{1, ..., N\}}$ is independent for each N, and

$$\phi^N \doteq \mathbf{E}\left[\left\|\bar{X}_i^N(0) - Y_i^N(0)\right\|\right]$$

tends to zero as $N \to \infty$.

For $N \in \mathbb{N}$, $i \in \{1, ..., N\}$, let $\bar{X}_i^N$ be the unique strong solution of Equation (3) in $[0, T]$ with initial condition $\bar{X}_i^N(0) = X_i^N(0)$ a.s., driving Brownian motion B_i and

Poisson random measure N^i. Notice that the processes $X_1^N, \ldots, X_N^N$ are independent and identically distributed for each N.

By definition of the metric ρ_T, the μ-chaoticity in W_1 of the sequence Y^N follows from

$$\lim_{N \to \infty} \mathbf{E}\left[\sup_{t \in [0,T]} \left\| \bar{X}_i^N(t) - Y_i^N(t) \right\| \right] = 0, \tag{8}$$

for every fixed $i \in \mathbb{N}$. However, the limit is the same by exchangeability of components. The term in (8) is bounded by

$$\mathbf{E}\left[\sup_{t \in [0,T]} \left\| Y_i^N(t) - \bar{X}_i^N(t) \right\| \right] \leq \phi^N + \bar{F} + \bar{\sigma} + \bar{\Theta} + \bar{\psi},$$

where

$$\bar{F} \doteq \mathbf{E}\left[\int_0^T \left\| F(Y_i^N(s), \mu_Y^N(s)) - F(\bar{X}_i^N(s), \mu_s) \right\| ds \right],$$

$$\bar{\sigma} \doteq \mathbf{E}\left[\sup_{t \in [0,T]} \left\| \int_0^t \sigma(Y_i^N(s), \mu_Y^N(s)) - \sigma(\bar{X}_i^N(s), \mu_s) dB_s^i \right\| \right],$$

$$\bar{\Theta} \doteq \mathbf{E}\left[\sup_{t \in [0,T]} \left\| \int_0^t \langle \mu_Y^N(s), \int_U \Theta(\cdot, Y_i^N(s), \mu_Y^N(s), h) \mathbb{1}_{(0, \lambda_j(\cdot, \mu_Y^N(s))]}(u) \; du\nu(dh) \rangle ds \right. \right.$$
$$\left. \left. - \int_0^t \int_U \langle \mu_s, \Theta(\cdot, \bar{X}_i^N(s), \mu_s, h) \mathbb{1}_{(0, \lambda_j(\cdot, \mu_s)]}(u) \; du\nu(dh) \rangle ds \right\| \right],$$

$$\bar{\psi} \doteq \mathbf{E}\left[\sup_{t \in [0,T]} \left\| \int_{[0,t] \times U} \psi(Y_i^N(s^-), \mu_Y^N(s), h) \mathbb{1}_{(0, \lambda(Y_i^N(s^-), \mu_Y^N(s^-))]}(u) \right. \right.$$
$$\left. \left. - \psi(\bar{X}_i^N(s^-), \mu_{s^-}, h) \mathbb{1}_{(0, \lambda_i(\bar{X}_i^N(s^-), \mu_{s^-})]}(u) N^i(ds, du, dh) \right\| \right].$$

The terms $\bar{F}$, $\bar{\sigma}$, and $\bar{\psi}$ are treated exactly as in Proposition 3.1, whereas the term $\bar{\Theta}$ only requires the application of the Lipschitz condition (L1). By mimicking the steps in Proposition 3.1, there exists a $T_0 > 0$ small enough and a constant $C_{T_0} > 0$, independent of N, such that we can apply Gronwall's lemma and obtain

$$\mathbf{E}\left[\sup_{t \in [0,T_0]} \left\| Y_i^N(t) - \bar{X}_i^N(t) \right\| \right] \leqslant C_{T_0} \left(\int_0^{T_0} \mathbf{E}[\rho(\mu_Y^N(t), \mu_t)] dt + \sqrt{T_0} \mathbf{E}\left[\sup_{t \in [0,T_0]} \rho(\mu_Y^N(t), \mu_t) \right] + \phi^N \right). \tag{9}$$

By the triangle inequality, for every fixed $t \in [0, T_0]$,

$$\mathbf{E}[\rho(\mu_Y^N(t), \mu_t)] \leqslant \mathbf{E}\left[\rho\left(\mu_Y^N(t), \mu\bar{X}^N(t) \right) \right] + \mathbf{E}\left[\sup_{t \in [0,T_0]} \rho\left(\mu\bar{X}^N(t), \mu_t \right) \right]$$

$$\leqslant \mathbf{E}\left[\sup_{t \in [0,T_0]} \left\| Y_i^N(t) - \bar{X}_i^N(t) \right\| \right] + \mathbf{E}\left[\rho_{T_0}\left(\mu\bar{X}^N, \mu \right) \right]$$

Then, for a T_0 sufficiently small, using again Gronwall's lemma, there exists a positive constant, depending on T_0, that by abuse of notation we will indicate again with $C_{T_0} > 0$, such that

$$\mathbf{E}\left[\sup_{t\in[0,T_0]}\left\|Y_i^N(t)-\bar{X}_i^N(t)\right\|\right]\leqslant C_{T_0}\left(\int_0^{T_0}\mathbf{E}\left[\rho_t\left(\mu\bar{X}^N,\mu\right)\right]dt+\sqrt{T_0}\mathbf{E}\left[\rho_{T_0}\left(\mu\bar{X}^N,\mu\right)\right]+\phi^N\right).$$

We see that the bound on (8) depends on the initial conditions and on $\mathbf{E}[\rho_{T_0}(\mu\bar{X}^N,\mu)]$, that is, the distance, on the time interval $[0,T_0]$, between the empirical measure of N i.i.d. copies of the solution of the process with law μ and the law μ_t itself. By the Glivenko–Cantelli theorem, setting

$$\beta^N:=\mathbf{E}\left[\rho_T\left(\mu\bar{X}^N(t),\mu_t\right)\right],$$

we have

$$\lim_{N\to\infty}\beta^N=0.$$

Therefore, we know that there exists a constant $C_{T_0}>0$ such that, for N going to infinity, we have

$$\mathbf{E}\left[\sup_{t\in[0,T_0]}\left\|Y_i^N(t)-\bar{X}_i^N(t)\right\|\right]\leqslant C_{T_0}\left(\beta^N+\phi^N\right).$$

Iterating this procedure as in Proposition 3.1, we extend the above result to $[0,T]$, i.e.

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left\|Y_i^N(t)-\bar{X}_i^N(t)\right\|\right]\leqslant C_T\left(\beta^N+\phi^N\right),$$

for a suitable constant C_T. This establishes μ-chaoticity of Y^N in W_1. □

Remark 3.1. Since we have an explicit rate of convergence in (5), it would be desirable to give also some bounds on β_N and ϕ_N in terms of N. However, the rate of convergence of empirical measures of i.i.d. random variables in Wasserstein distance is a delicate issue, see for example [18, 21]. By the results in [18], we have sharp bounds for measures in $\mathbb{R}^d$, depending on the moments and on the dimension d. If the diffusion coefficient does not depend on the measure, it is possible to exploit these results to get an explicit rate of convergence. Indeed, the second term in the rhs of (9) would not appear and all the subsequent estimates would only depend on distance of measures on $\mathbb{R}^d$ and not on the space of trajectories. In Proposition 3.1 we prove that, in any situation, the simultaneous jumps in the form presented here, do not worsen this rate of convergence, since they contribute with a term of order $1/\sqrt{N}$.

The propagation of chaos property for X^N is now an immediate consequence of Propositions 3.1 and 3.2.

Corollary 3.3. *Grant Assumptions 1 and 2. Let μ_0 be a probability measure on $\mathbb{R}^d$ such that $\int\|x\|^2\mu_0(dx)<+\infty$. For $N\in\mathbb{N}$, let X^N be a solution of Equation (1) in $[0,T]$. Assume that $X^N(0)=(X_1^N(0),...,X_N^N(0))$, $N\in\mathbb{N}$, form a sequence of square integrable random vectors that is μ_0-chaotic in W_1. Let μ be the law of the solution of Equation (3) in $[0,T]$ with initial law $\mathbf{P}\circ X(0)^{-1}=\mu_0$. Then X^N is μ chaotic in W_1.*

4. Non-globally Lipschitz drift

We are interested in enlarging the class of systems for which propagation of chaos holds. In this section, we relax the Lipschitz assumption on the drift, allowing gradients of general convex potentials. This includes relevant examples as those appeared in [3, 12].

4.1. Particle system

Consider the particle system (1) in Section 2. The coefficients are supposed to satisfy the following set of conditions.

Assumption 3. (U) The drift coefficient $F : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \to \mathbb{R}^d$ is of the form

$$F(x, \alpha) = -\nabla U(x) + b(x, \alpha),$$

for all $x \in \mathbb{R}^d$ and all $\alpha \in \mathcal{M}(\mathbb{R}^d)$, where U is convex and C^1. The function b is assumed to be globally Lipschitz in both variables, and for all $x \in \mathbb{R}^d$ we have $\sup_{\alpha \in \mathcal{M}^1(\mathbb{R}^d)} b(x, \alpha) < \infty$.

(LD) The diffusion coefficient $\sigma : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \to \mathbb{R}^{d \times d_1}$ satisfies the usual global Lipschitz condition, i.e., $\exists \ \widetilde{L} > 0$ such that for all $x, y \in \mathbb{R}^d$, all $\alpha, \gamma \in \mathcal{M}(\mathbb{R}^d)$,

$$\left\| \sigma(x, \alpha) - \sigma(y, \gamma) \right\| \leqslant \widetilde{L}\left(\left\| x - y \right\| + \rho(\alpha, \gamma) \right).$$

Moreover, for all $x \in \mathbb{R}^d \sup_{\alpha \in \mathcal{M}^1(\mathbb{R}^d)} \sigma(x, \alpha) < \infty$.
The jump coefficients satisfy conditions (I) and (L1) from Assumption 1.
As before, we set $L \doteq \bar{L} \vee \widetilde{L}$.

Remark 4.1. Condition (U) is a natural choice when one wants to relax globally-Lipschitz conditions on coefficients. It induces a process whose trajectories are strongly constrained by the convex potential. This attracting drift, even when combined with an unbounded jump rate, should prevent the process from exploding in finite time. We will see that this is what happens provided the jump rate is in some way "controllable," as it is under the Lipschitz assumption (L1).

4.2. McKean–Vlasov equation with non-Lipschitz drift

Here, we focus on the well-posedness of the McKean–Vlasov limit of the particle system introduced in Section 4.1, that is, Equation (1) under Assumptions 2 and 3. This limit will satisfy Equation (3), for which we prove existence and uniqueness of strong solutions in the following. Since the collateral jumps part in the limit gives rise to a globally Lipschitz drift term, in this subsection, by abuse of notation, we absorb it into the drift coefficient F, satisfying the weaker condition (U) in Assumption 3. Thus, consider the stochastic differential equation

$$\begin{aligned} dX(t) = F\big(X(t), \mu_t\big)dt + \sigma\big(X(t), \mu_t\big)dB_t \\ + \int_{[0,\infty) \times [0,1]^{\mathbb{N}}} \psi\big(X(t^-), \mu_{t^-}, h_1\big)\mathbb{1}_{\big(0, \lambda(X(t^-), \mu_{t^-})\big]}(u)N(dt, du, dh), \end{aligned} \tag{10}$$

where $\mu_t = \text{Law}(X(t))$, B is a d_1-dimensional Brownian motion and $\mathcal{N}$ a stationary Poisson random measure with characteristic measure $l \times l \times \nu$.

Theorem 4.1. *Let the coefficients of the nonlinear SDE (10) satisfy Assumption 3. Then for all square integrable initial conditions* $X(0) \in \mathbb{R}^d$, *Equation (10) admits a unique strong solution.*

Proof. Let P^1 and P^2 be two laws on $\mathbf{D}([0,T], \mathbb{R}^d)$, and suppose that X^1 and X^2 are two solutions of the following SDE, for $k = 1, 2$:

$$
\begin{aligned}
dX^k(t) &= F\left(X^k(t), P^k_t\right)dt + \sigma\left(X^k(t), P^k_t\right)dB_t \\
&+ \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \psi\left(X^k(t^-), P^k_{t-}, h_1\right)\mathbb{1}_{\left(0,\lambda\left(X^k(t^-),P^k_{t-}\right)\right]}(u)\mathcal{N}(dt, du, dh),
\end{aligned}
\tag{11}
$$

defined on the same probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbf{P})$ with the same $\mathcal{F}_t$-Brownian motion B, the same Poisson random measure N and with initial condition $X^1(0) = X^2(0) = \xi$ $\mathbf{P}$-almost surely. The well-posedness of Equation (11) is ensured by Lemma A.1. Let Q^1 and Q^2 be the laws of the solutions on $\mathbf{D}([0,T), \mathbb{R}^d)$, and let Γ be the map that associates Q^k to P^k. We are interested in proving that the map Γ is a contraction for the W_1 Wasserstein norm. Hence, we want to bound the distance

$$
\rho_T(Q^1, Q^2) \leqslant \mathbf{E}\left[\sup_{t\in[0,T]} \left\| X^1(t) - X^2(t) \right\|\right].
\tag{12}
$$

The idea here, in order to exploit the convexity of U, is to apply Ito's rule. A classical approach consists in applying Ito's rule to a quantity of the type $(X^1_t - X^2_t)^2$; this L^2 approach does not work in the presence of jump terms. For this reason, we rather use an L^1 approach. To this aim, for all $\epsilon > 0$ we define the following smooth approximation of the norm

$$
f^\epsilon(x) \doteq \sqrt{\left\| x \right\|^2 + \epsilon}.
$$

Then, by Ito's rule and Fatou's Lemma, we have

$$
\mathbf{E}\left[\sup_{t\in[t_0,t_1]} \left\| X^1(t) - X^2(t) \right\|\right] \leqslant \liminf_{\epsilon\downarrow 0} \mathbf{E}\left[\sup_{t\in[t_0,t_1]} f^\epsilon\left(X^1(t) - X^2(t)\right)\right]
$$

$$
\leqslant \liminf_{\epsilon\downarrow 0}\left(i_\epsilon[t_0, t_1] + u_\epsilon[t_0, t_1] + b_\epsilon[t_0, t_1] + \sigma_\epsilon[t_0, t_1] + \Sigma_\epsilon[t_0, t_1] + \Lambda_\epsilon[t_0, t_1]\right),
$$

where, for $t_1 \in [t_0, T]$, we set

$$
i_\epsilon[t_0, t_1] \doteq \mathbf{E}\left[f^\epsilon\left(X^1(t_0) - X^2(t_0)\right)\right],
$$

$$
u_\epsilon[t_0, t_1] \doteq \mathbf{E}\left[\sup_{t\in[t_0,t_1]} -\int_{t_0}^t \nabla f^\epsilon\left(X^1(s) - X^2(s)\right)\cdot\nabla\left(U(X^1(s)) - U(X^2(s))\right)ds\right],
$$

$$
b_\epsilon[t_0, t_1] \doteq \mathbf{E}\left[\sup_{t\in[t_0,t_1]} \int_{t_0}^t \nabla f^\epsilon\left(X^1(s) - X^2(s)\right)\cdot\left(b(X^1(s), P^1_s) - b(X^2(s), P^2_s)\right)ds\right],
$$

$$
\sigma_\epsilon[t_0, t_1] \doteq \frac{1}{2}\mathbf{E}\left[\sup_{t\in[t_0,t_1]} \int_{t_0}^t \mathrm{Tr}\left(\sigma(X^1(s), P^1_s) - \sigma(X^2(s), P^2_s)\right)^{\mathrm{T}} H_{f^\epsilon\left(X^1(s)-X^2(s)\right)}\left(\sigma(X^1(s), P^1_s) - \sigma(X^2(s), P^2_s)\right)ds\right],
$$

$$
\Sigma_\epsilon[t_0, t_1] \doteq \mathbf{E}\left[\sup_{t\in[t_0,t_1]} \int_{t_0}^t \nabla f^\epsilon\left(X^1(s) - X^2(s)\right)\cdot\left(\sigma(X^1(s), P^1_s) - \sigma(X^2(s), P^2_s)\right)dB_s\right],
$$

$$
\begin{aligned}
\Lambda_\epsilon[t_0, t_1] \doteq \mathbf{E}\Big[&\sup_{t\in[t_0,t_1]} \int_{t_0}^t \int_{[0,1]}\int_0^\infty f^\epsilon(X^1(s) + \psi(X^1(s), P^1_s, h)\mathbb{1}_{u\leqslant\lambda(X^1(s),P^1_s)} - X^2(s)\\
&-\psi(X^2(s), P^2_s, h)\mathbb{1}_{u\leqslant\lambda(X^2(s),P^2_s)}) - f^\epsilon(X^1(s) - X^2(s))du\nu_1(dh)ds\Big].
\end{aligned}
$$

Notice that, by the assumption of convexity of U, for all $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^d$, it holds

$$\nabla f^\epsilon(\mathbf{x}-\mathbf{y}) \cdot \nabla\big(U(\mathbf{x}) - U(\mathbf{y})\big) = \frac{1}{\sqrt{\|\mathbf{x}-\mathbf{y}\|^2 + \epsilon}}(\mathbf{x}-\mathbf{y}) \cdot \nabla\big(U(\mathbf{x}) - U(\mathbf{y})\big) \geqslant 0.$$

Therefore, the term $u_\epsilon[t_0, t_1]$ is easily bounded, since it is always non-positive, i.e.

$$\liminf_{\epsilon \downarrow 0} u_\epsilon[t_0, t_1] \leqslant 0.$$

For the term $b_\epsilon[t_0, t_1]$, we use the global Lipschitz condition on the function b, together with the properties of W_1 Wasserstein distance and inequality (12):

$$b_\epsilon[t_0, t_1] \leqslant \mathbf{E}\left[\int_{t_0}^{t_1} \big\| b\big(X^1(s), P_s^1\big) - b\big(X^2(s), P_s^2\big) \big\| \, ds\right]$$
$$\leqslant L\left(\int_{t_0}^{t_1} \mathbf{E}\left[\sup_{s \in [0,t]} \big\| X^1(s) - X^2(s) \big\|\right] dt + (t_1 - t_0)\rho_{[t_0,t_1]}\big(P^1, P^2\big)\right).$$

For estimating the term $\sigma_\epsilon[t_0, t_1]$, we observe that the Hessian matrix of f^ϵ has the following form:

$$H_{f^\epsilon(x)} = \left(\frac{1}{\sqrt{\|x\|^2 + \epsilon}} I - \frac{1}{\left(\|x\|^2 + \epsilon\right)^{3/2}} A\right),$$

where A is the $d \times d$ matrix such that, for all i, j, $A_{i,j} = x_i x_j$ and I is the identity $d \times d$ matrix. Therefore,

$$\sigma_\epsilon[t_0, t_1] \leqslant \int_{t_0}^{t_0} \mathbf{E}\left[\frac{1}{2\sqrt{\big\| X^1(s) - X^2(s) \big\|^2 + \epsilon}} \mathrm{Tr}(\sigma(X^1(s), P_s^1) - \sigma(X^2(s), P_s^2))^T\right.$$
$$\left. (\sigma(X^1(s), P_s^1) - \sigma(X^2(s), P_s^2))\right] ds$$
$$+ \int_{t_0}^{t_0} \mathbf{E}\left[\frac{1}{2\big\| X^1(s) - X^2(s) \big\|^2 + \epsilon)^{3/2}} \mathrm{Tr}(\sigma(X^1(s), P_s^1) - \sigma(X^2(s), P_s^2))^T\right.$$
$$\left. ((X^1(s) - X^2(s))_i (X^1(s) - X^2(s))_j)(\sigma(X^1(s), P_s^1) - \sigma(X^2(s), P_s^2))\right] ds.$$

Of course, due to the Lipschitz property of the diffusion coefficient σ, this term gives rise to a new term linear in $\mathbf{E}[\sup_{t \in [t_0,t_1]} \|X^1(t) - X^2(t)\|]$, since we have, for a certain $K \geqslant 0$,

$$\sigma_\epsilon[t_0, t_1] \leqslant KL\left(\int_{t_0}^{t_1} \mathbf{E}\left[\sup_{s \in [t_0,t]} \big\| X^1(s) - X^2(s) \big\|\right] dt + (t_1 - t_0)\rho_{[t_0,t_1]}\big(P^1, P^2\big)\right).$$

In addition to the previous arguments, the treatment of the term $\Sigma_\epsilon[t_0, t_1]$ involves the Burkholder–Davis–Gundy inequalities and the global Lipschitz condition (LD):

$$\Sigma_\epsilon[t_0, t_1] \leqslant C_1 \mathbf{E}\left[\left(\int_{t_0}^{t_1} \big\| (\sigma(X^1(s), P_s^1) - \sigma(X^2(s), P_s^2))^T\right.\right.$$
$$\left.\left. (\sigma(X^1(s), P_s^1) - \sigma(X^2(s), P_s^2)) \big\| \, ds\right)^{1/2}\right]$$
$$\leqslant C_1 L \mathbf{E}\left[\left(\int_{t_0}^{t_1} \sup_{s \in [t_0,t_1]} \big\| X^1(s) - X^2(s) \big\|^2 dt + (t_1 - t_0)\rho_{[t_0,t_1]}(P^1, P^2)^2\right)^{1/2}\right]$$
$$\leqslant C_1 L \sqrt{(t_1 - t_0)}\left(\mathbf{E}\left[\sup_{t \in [t_0,t_1]} \big\| X^1(t) - X^2(t) \big\|\right] + \rho_{[t_0,t_1]}(P^1, P^2)\right)$$

for some constant C_1 not depending on t_0, t_1. To bound the term $\Lambda_{[t_0,t_1]}$, we make use of the properties of the process $\{\Lambda(t)\}_{t\in[0,T]}$, of the W_1 Wasserstein distance, as well as condition (L1) and the monotone convergence theorem:

$$\liminf_{\epsilon\downarrow 0} \Lambda_\epsilon[t_0, t_1] = \mathbf{E}\left[\sup_{t\in[t_0,t_1]} \int_{t_0}^t \int_{[0,1]} \int_0^\infty \left\| X^1(s) + \psi(X^1(s), P^1_s, h)\mathbb{1}_{u\leqslant\lambda(X^1(s),P^1_s)} - X^2(s) \right.\right.$$
$$\left.\left. -\psi(X^2(s), P^2_s, h)\mathbb{1}_{u\leqslant\lambda(X^2(s),P^2_s)} \right\| - \left\| X^1(s) - X^2(s) \right\| \, ds\,du\,\nu_1(dh) \right]$$

$$\leqslant \mathbf{E}\left[\int_{t_0}^{t_1} \int_{[0,\infty)\times[0,1]} \left\| \psi(X^1(s^-), P^1_{s^-}, h)1_{(0,\lambda(X^1(s^-),P^1_{s^-})}\right.\right.$$
$$\left.\left. -\psi(X^2(s^-), P^2_{s^-}, h)\mathbb{1}_{(0,\lambda(X^2(s^-),P^2_{s^-}))} \right\| \, ds\,du\,\nu(dh) \right]$$

$$\leqslant L\left(\int_{t_0}^{t_1} \mathbf{E}\left[\sup_{s\in[0,t]} \left\| X^1(s) - X^2(s) \right\| \right] dt + (t_1 - t_0)\rho_{[t_0,t_1]}(P^1, P^2) \right).$$

Therefore,

$$\mathbf{E}\left[\sup_{t\in[t_0,t_1]} \left\| X^1(t) - X^2(t) \right\| \right] \leqslant \mathbf{E}\left[\left\| X^1(t_0) - X^2(t_0) \right\| \right]$$
$$+ L\left((K+1)(t_1 - t_0) + C_1\sqrt{t_1 - t_0}\right)\rho_{[t_0,t_1]}(P^1, P^2)$$
$$+ C_1 L\sqrt{(t_1 - t_0)}\mathbf{E}\left[\sup_{t\in[t_0,t_1]} \left\| X^1(t) - X^2(t) \right\| \right]$$
$$+ L(1+K)(t_1 - t_0)\int_{t_0}^{t_1} \mathbf{E}\left[\sup_{s\in[0,t]} \left\| X^1(s) - X^2(s) \right\| \right] dt.$$

By hypothesis, $\mathbf{E}[\|X^1(0) - X^2(0)\|] = 0$. Choose $T_0 > 0$ such that $1 - C_1 L\sqrt{T_0} > 0$. Then

$$\mathbf{E}\left[\sup_{t\in[0,T_0\wedge T]} \|X^1(t) - X^2(t)\| \right] \leqslant \frac{L(1+K)T_0}{1 - C_1 L\sqrt{T_0}}\int_0^{T_0\wedge T} \mathbf{E}\left[\sup_{s\in[0,t]} \|X^1(s) - X^2(s)\| \right] dt$$
$$+ \frac{L\left((1+K)T_0 + C_1\sqrt{T_0}\right)}{1 - C_1 L\sqrt{T_0}}\rho_{T_0}(P^1, P^2). \tag{13}$$

Applying Gronwall's lemma to (13), there exists a $T_0 > 0$ sufficiently small such that

$$\rho_{T_0}(Q^1, Q^2) \leqslant \mathbf{E}\left[\sup_{t\in[0,T_0\wedge T]} \|X^1(t) - X^2(t)\| \right] < C_{T_0}\rho_{T_0}(P^1, P^2),$$

for some constant $C_{T_0} < 1$. Therefore, when $P^k \doteq Q^k$, this shows uniqueness of the McKean–Vlasov measure in $\mathcal{M}^1(\mathbf{D}([0, T_0], \mathbb{R}^d))$. However, since C_{T_0} depends only on the amplitude of the interval, the same procedure iterated over a finite number of intervals of the type $[T_0\wedge T, 2T_0\wedge T]$, $[2T_0\wedge T, 3T_0\wedge T]$, etc., yields uniqueness of the measure in $\mathcal{M}^1(\mathbf{D}([0, T], \mathbb{R}^d))$.

The proof of existence is obtained via a Picard iteration argument, starting from (11). Let $P^k \doteq Q^{k-1}$, then (11) gives a sequence of laws $\{Q^k\}_{k\in\mathbb{N}}$, that is a Cauchy sequence for the metric ρ_{T_0} on $\mathcal{M}^1(\mathbf{D}([0, T_0], \mathbb{R}^d))$. Consequently, it is a Cauchy sequence also for a weaker Wasserstein metric based on a complete Skorohod metric, that yields existence of a solution of (11) on $[0, T_0\wedge T]$. Again, iterating the procedure over a finite number of intervals gives the thesis.

4.3. Propagation of chaos

A propagation of chaos result for a particle system with superlinear drift term of type (U) and simultaneous jumps clearly follows along the lines of Section 3.2, with the techniques of Theorem 4.1. For this reason the following result is almost straightforward.

Theorem 4.2. *Grant Assumption 2 and Assumption 3. Let μ_0 be a probability measure on $\mathbb{R}^d$ such that $\int \|x\|^2 \mu_0(dx) < +\infty$. For $N \in \mathbb{N}$, let X^N be a solution of Equation (1) in $[0, T]$. Assume that $X^N(0) = (X_1^N(0), ..., X_N^N(0))$, $N \in \mathbb{N}$, form a sequence of square-integrable random vectors that is μ_0-chaotic in W_1. Let Q be the law of the solution of Equation (3) in $[0, T]$ with initial law $\mathbf{P} \circ X(0)^{-1} = \mu_0$. Then X^N is Q chaotic in W_1.*

We use again the trick of the sequence of *intermediate processes* $\{Y^N\}_{N \in \mathbb{N}}$, where each process $Y^N = \{Y^N(t)\}_{t \in [0,T]}$ is defined as the solution of the system (4). As before, the *collateral jumps* have been absorbed by a new drift term, that by the properties of the jump rate λ and its amplitude Θ is a globally Lipschitz drift term, that is added to b, giving rise to a new drift $\bar{F}$ that maintains condition (U) of F. For the proof of propagation of chaos, we apply again the procedure of Section 2.3, starting with the bound over the distance of the two particle systems (1) and (4) of the form

$$\mathbf{E}\left[\sup_{t \in [0,T]} \|X_i^N(t) - Y_i^N(t)\|\right] \leqslant \frac{C_T}{\sqrt{N}},$$

which is obtained by computations similar to those of Theorem 4.1. Then propagation of chaos follows by coupling the intermediate process (4) with N independent copies of the solution of (3).

5. Non-globally Lipschitz jump rate

Stochastic models in neuroscience often focus on the membrane potential of neurons and describe its spikes in terms of SDEs with jumps. The jump rates in those models are usually super-linear. It is therefore interesting to investigate systems where the jump coefficients are not required to be globally Lipschitz. We adapt the model presented in [10] to a d-dimensional framework and a slightly more general situation allowing for jumps with random amplitude. In order to get a tractable model with a super-linear jump rate, we are forced to make more restrictive assumptions on the other parts of the dynamics than in the previous sections. We consider a model where particles are subject to a linear attracting drift, we drop the diffusion part, and we assume the main jump to force the particles into a given compact set. Furthermore, the collateral jumps are of bounded random amplitude and do not depend on the positions of the affected jumping particle.

5.1. Particle system

We consider the Markov process X^N solution of the following SDE, similar to Equation (1),

$$dX_i^N(t) = -X_i^N(t)dt + \frac{1}{N}\sum_{j \neq i}\int_{[0,\infty]\times[0,1]^{\mathbb{N}}} V(h_j, h_i)\mathbb{1}_{[0,\lambda(X_j^N(t)))}(u)\mathcal{N}^j(dt, du, dh)$$

$$- \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} (X_i^N(t) - U(h_i))\mathbb{1}_{[0,\lambda(X_i^N(t)))}(u)\mathcal{N}^i(dt, du, dh) \tag{14}$$

for all $i = 1, ..., N$. As before, $(N^i)_{i \in \mathbb{N}}$ is an independent family of Poisson random measures N^i, each of them with characteristic measure $l \times l \times \nu$, l is the Lebesgue measure restricted to $[0, \infty)$, and ν is a symmetric probability measure on $[0, 1]^{\mathbb{N}}$ such that there exists $(\nu_N)_{N \in \mathbb{N}}$, a consistent family of symmetric probability measures, each of them defined respectively on $[0, 1]^N$ and coinciding with the projections of ν on N coordinates.

Assumption 4. The coefficients of the system (14) obey the following properties:

- the jump amplitudes, V and U, are two bounded functions from respectively $[0, 1]^2$ and $[0, 1]$ to $\mathbb{R}^d$ (since they represent two random variables with values in some bounded subsets of $\mathbb{R}^d$, with abuse of notation we will indicate as expectations their integrals w.r.t. the measure v);
- the jump rate of each particle is a non-negative C^1 function of its position, $\lambda : \mathbb{R}^d \to \mathbb{R}_+$, that is written as a sum of two functions:

$$\lambda(\cdot) \doteq b(\| \cdot \|) + h(\cdot).$$

  $- b$ is a C^1, positive, non-decreasing, unbounded function such that

$$b'(r) \leqslant \gamma b(r) + c \tag{15}$$

  for some $c > 0$ and $\gamma < \frac{1}{5\mathbf{E}[\|V\|]}$;
  $- h : \mathbb{R}^d \to \mathbb{R}$ is a C^1 bounded function, i.e. there exists $H > 0$ such that $\forall x \in \mathbb{R}^d$, $\|h(x)\| \leqslant H$.

The assumption $\gamma < \frac{1}{5\mathbf{E}[\|V\|]}$ will allow to obtain a priori bounds on the moments of $\lambda(X(t))$ where $X(t)$ is the solution of the corresponding McKean–Vlasov equation, see (16) below; it is used in the proofs of Lemmas 5.4 and 5.5, which are postponed to the appendix.

Remark 5.1. The model and the form of the function b is suggested by [10]. It is interesting to notice that Assumption 4 allows to consider non-globally Lipschitz functions; in particular, this covers all the cases where $b(r)$ is of the form r^α, for $\alpha \geqslant 1$. We also remark that the condition on b here is a little stronger than in [10], due to the coupling method (vs. the martingale approach) in the proof, which in particular allows to identify the rate of convergence, which is of the order $O(\frac{1}{\sqrt{N}})$. This requires $\gamma < \frac{1}{K\mathbf{E}[\|V\|]}$ with $K = 5$ rather than $K = 3$, as in [10].

Remark 5.2. The neuronal model presented in [14, 15] is similar to this one, but it has a drift toward the barycenter of the system, instead of the origin. When dealing with initial conditions with bounded support, we could adapt our computations to that case, except for the fact that the function b has to be convex. In [15], the authors succeed in proving propagation of chaos with an explicit rate (namely, the expected $1/\sqrt{N}$) even for weaker conditions on the initial values, by defining an ad-hoc distance based on the rate function λ itself.

Existence and uniqueness of a non-explosive solution to the system with N fixed, both the initial particle system (14) and the intermediate process (19) below, rely on a truncation argument on the function λ, see Appendix B.

5.2. McKean–Vlasov equation

This section is devoted to analyzing the McKean–Vlasov equation whose law is the limit of the sequence of empirical measures corresponding to system (14), that is

$$dX(t) = \mathbf{E}[\lambda(X(t))]\mathbf{E}[V]dt - X(t)dt - \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \big(X(t) - U(h_1)\big)1_{[0,\lambda(X(t)))}(u)N(dt, du, dh),$$

(16)

with $\mathcal{N}$ a Poisson random measure with characteristic measure $l \times \nu \times l$. Since the model that we treat is basically an extension in d-dimension of the model presented in [10], techniques for proving existence and uniqueness of solutions for the nonlinear Markov process (16) are adaptations of the techniques presented in that paper. The procedure relies on a priori bounds on moments of the solution and of the expectation of $\lambda(X(t))$; we present the main steps here, while we gather the details in Appendix B.

Lemma 5.1. *Let* $f : \mathbb{R}_+ \to \mathbb{R}^d$ *be a locally bounded Borel function, then there exists a unique solution* $(Z_f(t))$ *to the SDE*

$$dZ_f(t) = -Z_f(t)dt + f(t)dt - \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \big(Z_f(t) - U(h_1)\big)\mathbb{1}_{[0,\lambda(Z_f(t)))}(u)\mathcal{N}(dt, du, dh) \quad (17)$$

with initial condition x *and coefficients satisfying Assumption 4. Moreover, for every pair of locally bounded Borel functions* f *and* g, *for every* $T > 0$ *there exists a constant* $C_T > 0$ *such that*

$$\mathbf{E}\left[\sup_{t\in[0,T]} \|Z_f(t) - Z_g(t)\|\right] \leqslant C_T \int_0^T \sup_{s\in[0,t]} \|f(s) - g(s)\|dt.$$

(18)

A priori bounds for any solution of (16) are necessary to perform the iteration that yields existence and uniqueness of the nonlinear process itself. The following lemma provides the required bounds.

Lemma 5.2. *Suppose Assumption 4 is satisfied. Let* X *be a solution of (16) with integrable initial condition* $X(0)$; *then we have that* $\sup_{t\geqslant 0} \mathbf{E}[\|X(t)\|] < \infty$. *Moreover, for* $p = 1, 2, 3, 4$, *if* $\mathbf{E}[\lambda^p(X(0))] < \infty$ *then* $\sup_{t\geqslant 0} \mathbf{E}[\lambda^p(X(t))] \leqslant C < \infty$, *where* C *only depends on* $\mathbf{E}[\lambda^p(X(0))]$ *and on the parameters of equations (16).*

The proof of this lemma is in Appendix B and it basically relies on the properties of the function b.

Theorem 5.3 (Solution of the McKean–Vlasov equation). *Under Assumption 4, for any initial condition* $X(0)$ *with bounded support and independent of* $\mathcal{N}$, *there exists a unique strong solution* $\{X(t)\}_{t\in[0,T]}$ *for (16).*

Proof. Fix a constant $C > 0$, and consider the following Picard iteration: $Z_0^C(t) \equiv X(0)$ and

$$\begin{cases} dZ_n^C(t) = -Z_n^C(t)dt + \big(\mathbf{E}\big[\lambda\big(Z_{n-1}^C(t)\big)\big] \wedge C\big)\mathbf{E}[V]dt \\ \qquad\qquad - \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \big(Z_n^C(t) - U(h_1)\big)\mathbb{1}_{[0,\lambda(Z_n^C(t)))}(u)\mathcal{N}(dt, du, dh), \\ Z_n^C(0) = X(0) \end{cases}$$

Since the jump term is forcing the process to return in a compact set, the following a priori bound is essentially obvious: for all $n \geqslant 1$ it holds

$$\|Z_n^C(t)\| \leqslant K + tC\mathbf{E}[\|V\|]$$

a.s. for a suitable $K > 0$ depending on the support of $X(0)$ and the range of $U(h)$. Indeed, when $\|Z^C(t)\|$ is large, the linear term $-Z_n^C(t)dt$ as well as the jumps can only decrease the norm. From Lemma 5.1 we know that there exists a constant C_T such that

$$\mathbf{E}\left[\sup_{t\in[0,T]} \|Z_{n+1}^C(t) - Z_n^C(t)\|\right] \leqslant C_T \mathbf{E}[\|V\|] \int_0^T \sup_{s\in[0,t]} \mathbf{E}\left[\lambda\big(Z_n^C(s)\big) - \lambda\big(Z_{n-1}^C(s)\big)\right] dt.$$

Thanks to the a.s. bounds on $\|Z_n^C(t)\|$, we can exploit the local Lipschitzianity of λ and get, for a certain constant $K_T > 0$,

$$\mathbf{E}\left[\sup_{t\in[0,T]} \|Z_{n+1}^C(t) - Z_n^C(t)\|\right] \leqslant C_T \mathbf{E}[\|V\|] K_T \int_0^T \mathbf{E}\left[\sup_{s\in[0,t]} \|Z_n^C(s) - Z_{n-1}^C(s)\|\right] dt$$

$$\leqslant ... \leqslant \frac{(K_T C_T \mathbf{E}[\|V\|]T)^n}{n!} \mathbf{E}\left[\sup_{s\in[0,T]} \|Z_1^C(s) - Z_0^C(s)\|\right].$$

Therefore the sequence $\{Z_n^C\}_{n\in\mathbb{N}}$ is a Cauchy sequence and its limit Z^C is a solution of the SDE

$$dZ^C(t) = -Z^C(t)dt + \big(\mathbf{E}\big[\lambda\big(Z^C(t)\big)\big] \wedge C\big)\mathbf{E}[V]dt$$
$$- \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \big(Z^C(t^-) - U(h_1)\big)\mathbb{1}_{[0,\lambda(Z^C(t^-)))}(u)N(dt, du, dh).$$

By Lemma 5.2, we can choose C such that $\mathbf{E}[\lambda(Z^C(t))] \leqslant C$ for all t, so that Z^C is indeed a solution of (16).

To prove uniqueness we can consider two solutions Z_1 and Z_2. Using the above a priori bound, (18) and Gronwall's lemma, their equality follows from standard arguments.

5.3. Propagation of chaos

As in the previous sections, we introduce an intermediate process $\{Y^N(t)\}_{t\in[0,T]}$ that is the solution of a system, similar to (14), that is

$$dY_i^N(t) = -Y_i^N(t)dt + \frac{1}{N}\sum_{j=1}^N \mathbf{E}[V]\lambda\left(Y_j^N(t)\right)dt \tag{19}$$
$$- \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \big(Y_i^N(t) - U(h_i)\big)\mathbb{1}_{[0,\lambda(Y_i^N(t)))}(u)N^i(dt, du, dh)$$

for all $i = 1, ..., N$. We indicate the empirical measure corresponding to the solution of (19) as μ_Y^N. In order to use a coupling procedure to prove propagation of chaos, we need to set some a priori bounds on the involved quantities. The proofs of the two following lemmas are in Appendix B, we state them here to highlight the quantities involved and the comparison with the a priori bounds for the nonlinear process (16).

Lemma 5.4. *For* $N>0$, *under Assumption 4, let* X^N *and* Y^N *be solutions, respectively, of* (14) *and* (19), *starting from initial conditions s.t.* $\mathbf{E}[\langle \mu_X^N(0), \lambda^4(\cdot)\rangle]<\infty$ *and* $\mathbf{E}[\langle \mu_Y^N(0), \lambda^4(\cdot)\rangle]<\infty$. *Then there exists a certain* $N_0>0$ *such that it holds*

$$\sup_{N\geqslant N_0}\sup_{t\geqslant 0}\mathbf{E}\Big[\langle \mu_X^N(t), \lambda^4(\cdot)\rangle\Big]<\infty \qquad \text{and} \qquad \sup_{N\geqslant N_0}\sup_{t\geqslant 0}\mathbf{E}\Big[\langle \mu_Y^N(t), \lambda^4(\cdot)\rangle\Big]<\infty.$$

Lemma 5.4 is crucial for proving that the number of jumps of the system in a compact time interval is proportional to N with probability increasing with N. This bound is stated in the following lemma.

Lemma 5.5 (Bound on the number of jumps)**.** *Assume that Assumption 4 is satisfied, that, for any* $N>0$, X^N *and* Y^N *are solutions, respectively, of* (14) *and* (19), *starting from initial conditions that are* μ_0*-chaotic. Here* μ_0 *is a probability measure on* $\mathbb{R}^d$ *s.t.* $\langle \mu_0, \lambda^3(\cdot)\rangle<\infty$. *Then, for any* $T>0$, *there exists a positive constant* H_T *and a natural number* $N_0>0$ *such that, for certain positive constants* K_T *and* $\widetilde{K}_T$

$$\mathbf{P}\left(\frac{C_N(T)}{N}\geqslant H_T\right)\leqslant \frac{K_T}{N} \quad \text{and} \quad \mathbf{P}\left(\int_0^T\langle \mu_Y^N(s), \lambda\rangle ds\geqslant H_T\right)\leqslant \frac{\widetilde{K}_T}{N},$$

for all $N>N_0$. *Here* $C_N(T)$ *is the number of jumps performed by system* (14) *up to time* T.

The bounds on the number of *collateral* jumps and of the corresponding drift in a compact time interval play a role in the proof of propagation of chaos, since they let us exploit the local Lipschitzianity of the function λ when we start from initial conditions with bounded support. The proofs of these lemmas involve the form of the function λ; they are given in Appendix B. In the following, we state and prove the result on propagation of chaos; also in this case, the simultaneous jumps result in a rate of the order $1/\sqrt{N}$. As in the previous sections, we start with the comparison between the particle system X^N and the intermediate system Y^N.

Theorem 5.6. *Let Assumptions 4 and 2 be satisfied and let* X^N *and* Y^N *be the solution, respectively, of* (14) *and* (19) *with initial conditions* $X^N(0)= Y^N(0)$ *a.s. that are* μ_0*-chaotic, with* μ_0 *probability measure on* $\mathbb{R}^d$ *with compact support. We assume the two processes are driven by the same Poisson random measures, and start from the same permutation invariant initial condition with compact support. Then, for each fixed* $i \in \mathbb{N}$,

$$\lim_{N\to+\infty}\mathbf{E}\left[\sup_{t\in[0,T]}\|X_i^N(t) - Y_i^N(t)\|\right] = 0.$$

Proof. As in previous sections, by permutation invariance of the initial conditions and of the dynamics, we have

$$\mathbf{E}\left[\sup_{t\in[0,T]}\|X_i^N(t) - Y_i^N(t)\|\right] = \frac{1}{N}\sum_{i=1}^N\mathbf{E}\left[\sup_{t\in[0,T]}\|X_i^N(t) - Y_i^N(t)\|\right].$$

Let us start with

$$\mathbf{E}\left[\sup_{t\in[0,T]}\|X_i^N(t) - Y_i^N(t)\|\right]\leqslant\mathbf{E}\left[\int_0^T\|X_i^N(t) - Y_i^N(t)\|dt\right]+V_{X_i^N,Y_i^N}(T) + U_{X_i^N,Y_i^N}(T),$$

where, for simplicity, we have set:

$$V_{X_i^N, Y_i^N}(T) := \mathbf{E}\left[\sup_{t\in[0,T]}\left\|\frac{\mathbf{E}[V]}{N}\sum_{j=1}^{N}\int_0^t \lambda(X_j^N(s)) - \lambda(Y_j^N(s))ds\right.\right.$$

$$\left.\left.+\frac{1}{N}\sum_{j\neq i}\int_0^t\int_{[0,1]^{\mathbb{N}}}\int_0^\infty V(h_i, h_j)\mathbb{1}_{[0,\lambda(X_j^N(s))]}(u)\widetilde{N}^j(ds, du, dh)\right\|\right];$$

$$U_{X_i^N, Y_i^N}(T) := \mathbf{E}\left[\sup_{t\in[0,T]}\left\|-\int_0^t\int_{[0,1]^{\mathbb{N}}}\int_0^\infty (X_i^N(s) - U(h_i))\mathbb{1}_{[0,\lambda(X_i^N(s)))}(u)\right.\right.$$

$$\left.\left.-(Y_i^N(s)-U(h_i))\mathbb{1}_{[0,\lambda(Y_i^N(s)))}(u)N^i(ds, du, dh)\right\|\right].$$

With the notation of Lemma 5.5, we consider the positive constant H_T and the event

$$E_N := \left\{\frac{C_N(T)}{N}\leqslant H_T\right\} \cap \left\{\int_0^T\langle\mu_Y^N(s), \lambda\rangle ds\leqslant H_T\right\},$$

so that $\mathbf{P}(E_N^c) \to 0$ for $N \to \infty$. Obviously, under the event E_N, for all $i = 1, ..., N$, the quantities $\sup_{t\in[0,T]}\lambda(X_i^N(t))$ and $\sup_{t\in[0,T]}\lambda(Y_i^N(t))$ are uniformly bounded and we can exploit local Lipschitzianity of λ (we will indicate its Lipschitz constant as L_{H_T}). Thus, we bound the first terms in $V_{X_i^N, Y_i^N}(T)$ in the following way:

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left\|\frac{\mathbf{E}[V]}{N}\sum_{j=1}^{N}\int_0^t \lambda\left(X_j^N(s)\right) - \lambda\left(Y_j^N(s)\right)ds\right\|\right]$$

$$\leqslant\frac{\mathbf{E}[\|V\|]}{N}\sum_{j=1}^{N}\mathbf{E}\left[\left(\int_0^T L_{H_T}\|X_j^N(s) - Y_j^N(s)\|ds\right)\mathbb{1}_{E_N}\right]$$

$$+\mathbf{E}[\|V\|]\mathbf{E}\left[\left(\int_0^T\frac{1}{N}\sum_{j=1}^{N}\left|\lambda\left(X_j^N(s)\right)\right| + |\lambda\left(Y_j^N(s)\right)|ds\right)\mathbb{1}_{E_N^c}\right]$$

$$\leqslant L_{H_T}\mathbf{E}[\|V\|]\int_0^T\frac{1}{N}\sum_{j=1}^{N}\mathbf{E}\left[\sup_{s\in[0,t]}\|X_j^N(s) - Y_j^N(s)\|\right]dt$$

$$+\int_0^T\frac{\mathbf{E}[\|V\|]}{N}\sum_{j=1}^{N}\sqrt{\mathbf{P}\left(E_N^C\right)}\sqrt{\mathbf{E}[|\lambda(X_j^N(s))|^2]}ds$$

$$+\int_0^T\frac{\mathbf{E}[\|V\|]}{N}\sum_{j=1}^{N}\sqrt{\mathbf{P}\left(E_N^C\right)}\sqrt{\mathbf{E}[|\lambda(Y_j^N(s))|^2]}ds$$

$$\leqslant L_{H_T}\mathbf{E}[\|V\|]\int_0^T\frac{1}{N}\sum_{j=1}^{N}\mathbf{E}\left[\sup_{s\in[0,t]}\|X_j^N(s) - Y_j^N(s)\|\right]dt$$

$$+\int_0^T\mathbf{E}[\|V\|]\sqrt{\mathbf{P}(E_N^C)}\sqrt{\mathbf{E}[\langle\mu_X^N(s), |\lambda(\cdot)|^2\rangle]}ds$$

$$+\int_0^T\mathbf{E}[\|V\|]\sqrt{\mathbf{P}(E_N^C)}\sqrt{\mathbf{E}[\langle\mu_Y^N(s), |\lambda(\cdot)|^2\rangle]}ds.$$

By Lemma 5.4 there exists $N_0>0$ such that for all $N>N_0$ $\sup_{t\geqslant 0}\mathbf{E}[\langle\mu_X^N(s), |\lambda(\cdot)|^2\rangle]$ and $\sup_{t\geqslant 0}\mathbf{E}[\langle\mu_Y^N(s), |\lambda(\cdot)|^2\rangle]$ are bounded. By Lemma 5.5, there exists a constant $K_T\geqslant 0$ such that $\mathbf{P}(E_N^C)\leqslant\frac{K_T}{N}$. The second term in $V_{X_i^N, Y_i^N}(T)$ is bounded using Burkholder–Davis–Gundy inequality, the orthogonality of the martingales $\{\widetilde{N}^j\}_{j\in\mathbb{N}}$ and Lemma 5.4.

$$\mathbf{E}\left[\sup_{t\in[0,T]}\left\|\frac{1}{N}\sum_{j\neq1}^{N}\int_{0}^{t}\int_{0}^{t}\int_{[0,1]^{N}}\int_{0}^{\infty}V(h_{i},h_{j})\mathbb{1}_{\left(0,\lambda\left(X_{j}^{N}(s)\right)\right](u)}\widetilde{N}^{j}(ds,du,dh)\right\|\right]$$

$$\leqslant\frac{M}{N}\mathbf{E}\left[\left(\sum_{j\neq i}^{N}\int_{0}^{T}\mathbf{E}[\|V\|^{2}]\lambda\left(X_{j}^{N}(s)\right)ds\right)^{1/2}\right]\leqslant M\sqrt{\frac{\mathbf{E}[\|V\|^{2}]}{N}}\mathbf{E}\left[\left(\int_{0}^{T}\langle\mu_{X}^{N}(t),\lambda(\cdot)\rangle dt\right)^{1/2}\right].$$

Therefore we get that there exist two constants C_T and K_T such that, for all $N>N_0$,

$$V_{X_{i}^{N},Y_{i}^{N}}^{N}(T)\leqslant C_{T}\int_{0}^{T}\mathbf{E}\left[\sup_{s\in[0,t]}\|X^{N}(s)-Y^{N}(s)\|^{2}\right]dt+\frac{K_{T}}{\sqrt{N}}.$$

With a similar argument, we get a bound of the same type for $U_{X_{i}^{N},Y_{i}^{N}}(T)$:

$$\frac{1}{N}\sum_{i=1}^{N}U_{X_{i}^{N},Y_{i}^{N}}^{N}(T)\leqslant C_{T}\int_{0}^{T}\frac{1}{N}\sum_{i=1}^{N}\|X_{i}^{N}(t)-Y_{i}^{N}(t)\|dt+\mathbf{E}\left[\mathbb{1}_{E_{N}^{C}}\int_{0}^{T}\frac{1}{N}\sum_{i=1}^{N}\|X_{i}^{N}(t)\|\lambda(X_{i}^{N}(t))dt\right]$$

$$+\mathbf{E}\left[\mathbb{1}_{E_{N}^{C}}\int_{0}^{T}\frac{1}{N}\sum_{i=1}^{N}\|Y_{i}^{N}(t)\|\lambda(Y_{i}^{N}(t))dt\right]+\mathbf{E}[\|U\|]\mathbf{E}\left[\mathbb{1}_{E_{N}^{C}}\int_{0}^{T}\frac{1}{N}\sum_{i=1}^{N}\lambda(X_{i}^{N}(t))dt\right]$$

$$+\mathbf{E}[\|U\|]\mathbf{E}\left[\mathbb{1}_{E_{N}^{C}}\int_{0}^{T}\frac{1}{N}\sum_{i=1}^{N}\lambda(Y_{i}^{N}(t))dt\right].$$

As before, we wish to get a bound of the order $O(\frac{1}{\sqrt{N}})$ for the last terms. We do that by means of Cauchy–Schwartz inequality, Lemmas 5.4 and 5.5. We also exploit that, by definition of λ, it holds $\|x\|\leqslant B\lambda(x)+c$ for some positive constant B and a constant c. Take, for instance, the second term of the right-hand side:

$$\mathbf{E}\left[\mathbb{1}_{E_{N}^{C}}\int_{0}^{T}\frac{1}{N}\sum_{i=1}^{N}\left\|X_{i}^{N}(t)\right\|\lambda(X_{i}^{N}(t))dt\right]\leqslant\int_{0}^{T}\sqrt{\mathbf{P}(E_{N}^{C})}\sqrt{\mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\left\|X_{i}^{N}(s)\right\|\lambda(X_{i}^{N}(s))\right)^{2}\right]}ds$$

$$\leqslant T\sqrt{\mathbf{P}(E_{N}^{C})}\sqrt{\mathbf{E}\left[\sup_{t\in[0,T]}\langle\mu_{X}^{N}(t),\|\cdot\|^{2}\rangle\langle\mu_{X}^{N}(t),\lambda(\cdot)^{2}\rangle\right]}\leqslant T\sqrt{\mathbf{P}(E_{N}^{C})}\sqrt{\mathbf{E}\left[\sup_{t\in[0,T]}\langle\mu_{X}^{N}(t),\lambda(\cdot)^{4}\rangle\right]}.$$

The same holds for the remaining right-hand side terms. Thus, there exist two constants $\widetilde{C}_T$ and $\widetilde{K}_T$ and a $N_0>0$, such that for all $N>N_0$ it holds

$$\frac{1}{N}\sum_{i=1}^{N}U_{X_{i}^{N},Y_{i}^{N}}(T)\leqslant\widetilde{C}_{T}\int_{0}^{T}\mathbf{E}\left[\sup_{s\in[0,t]}\|X^{N}(s)-Y^{N}(s)\|\right]dt+\frac{\widetilde{K}_{T}}{\sqrt{N}}.$$

It follows that there exist two constants, depending only on T, which by abuse of notation we will indicate by C_T and K_T, and $N_0>0$ such that, for all $N>N_0$ it holds

$$\mathbf{E}\left[\sup_{t\in[0,T]}\|X^{N}(t)-Y^{N}(t)\|\right]\leqslant C_{T}\int_{0}^{T}\mathbf{E}\left[\sup_{s\in[0,t]}\|X^{N}(s)-Y^{N}(s)\|\right]dt+\frac{K_{T}}{\sqrt{N}}.$$

By applying Gronwall's lemma we get the thesis.

Theorem 5.7 (Propagation of Chaos for Y^N). *Grant Assumptions 4 and 2. Let μ be a probability measure on $\mathbb{R}^d$ with compact support. For $N\in\mathbb{N}$, let Y^N be a solution of*

Equation *(19)* in $[0, T]$. *Assume that* $Y^N(0) = (Y_1^N(0), ..., Y_N^N(0))$, $N \in \mathbb{N}$, *form a sequence of integrable random vectors that is µ-chaotic in* W_1. *Let Q be the law of the solution of* Equation *(16)* in $[0, T]$ *with initial law* $\mathbf{P} \circ X(0)^{-1} = \mu$. *Then* Y^N *is Q chaotic in* W_1.

The proof of this theorem is a combination of the computations done for proving Theorem 5.6 and the coupling techniques for propagation of chaos used in the previous sections. Again, this implies propagation of chaos for X^N.

Corollary 5.8 (Propagation of Chaos for X^N). *Grant Assumptions 4 and 2. Let µ be a probability measure on* $\mathbb{R}^d$ *with compact support. For* $N \in \mathbb{N}$, *let* X^N *be a solution of* Equation *(14)* in $[0, T]$. *Assume that* $X^N(0) = (X_1^N(0), ..., X_N^N(0))$, $N \in \mathbb{N}$, *form a sequence of integrable random vectors that is µ-chaotic in* W_1. *Let Q be the law of the solution of* Equation *(16)* in $[0, T]$ *with initial law* $\mathbf{P} \circ X(0)^{-1} = \mu$. *Then* X^N *is Q chaotic in* W_1.

References

[1] Kac, M. (1956). Foundations of kinetic theory. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3, pp. 171–197, University of California Press.

[2] McKean, H. P. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* 56(6):1907–1911.

[3] Dawson, D. A. (1983). Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Stat. Phys.* 31(1):29–85.

[4] Graham, C. (1992). McKean–Vlasov Itô–Skorohod equations, and nonlinear diffusions with discrete jump sets. *Stoch. Process. Appl.* 40(1):69–82.

[5] Graham, C. (1992). Nonlinear diffusion with jumps. *Ann. Inst. Henri Poincaré Probab. Stat.* 28(3):393–402.

[6] Sznitman, A.-S. (1991). Topics in propagation of chaos. In: *Ecole D'Eté De Probabilités De Saint-Flour XIX—1989*. Berlin, Heidelberg: Springer, pp. 165–251.

[7] Baladron, J., Fasoli, D., Faugeras, O., Touboul, J. (2012). Mean-field description and propagation of chaos in networks of Hodgkin–Huxley and FitzHugh–Nagumo neurons. *J. Math. Neurosci.* 2(1):10.

[8] Bertini, L., Giacomin, G., Pakdaman, K. (2010). Dynamical aspects of mean field plane rotators and the Kuramoto model. *J. Stat. Phys.* 138(1–3):270–290.

[9] Borile, C., Dai Pra, P., Fischer, M., Formentin, M., Maritan, A. (2014). Time to absorption for a heterogeneous neutral competition model. *J. Stat. Phys.* 156(1):119–130.

[10] Robert, P., Touboul, J. (2016). On the dynamics of random neuronal networks. *J. Stat. Phys.* 165(3):545–584.

[11] Dai Pra, P., Runggaldier, W. J., Sartori, E., Tolotti, M. (2009). Large portfolio losses: a dynamic contagion model. *Ann. Appl. Probab.* 19(1):347–394.

[12] Garnier, J., Papanicolaou, G., Yang, T.-W. (2013). Large deviations for a mean field model of systemic risk. *SIAM J. Finan. Math.* 4(1):151–184.

[13] Giesecke, K., Spiliopoulos, K., Sowers, R. B., Sirignano, J. A. (2015). Large portfolio asymptotics for loss from default. *Math. Finance* 25(1):77–114.

[14] De Masi, A., Galves, A., Löcherbach, E., Presutti, E. (2015). Hydrodynamic limit for inter-acting neurons. *J. Stat. Phys.* 158(4):866–902.

[15] Fournier, N., Löcherbach, E. (2016). On a toy model of interacting neurons. *Ann. Inst. Henri Poincaré Probab. Stat.* 52(4):1844–1876.

[16] Scheutzow, M. (1987). Uniqueness and non-uniqueness of solutions of Vlasov–McKean equations. *J. Aust. Math. Soc.* 43(2):246–256.

[17] Budhiraja, A., Kira, E., Saha, S. (2017). Central limit results for jump-diffusions with mean field interaction and a common factor. *Stoch. Anal. Appl.* 35(5):767–802.

[18] Fournier, N., Guillin, A. (2015). On the rate of convergence in wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* 162(3–4):707–738.

[19] Billingsley, P. (1995). *Probability and Measure*. New York: John Wiley & Sons.

[20] Villani, C. (2009). *Optimal Transport, Old and New*. Berlin, Heidelberg: Springer-Verlag.

[21] Boissard, E., Le Gouic, T. (2014). On the mean speed of convergence of empirical and occupa-tion measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* 50(2):539–563.

[22] Ikeda, N., Nagasawa, M., Watanabe, S. (1966). A construction of branching Markov proc-esses. *Proc. Jpn. Acad.* 42(4):380–384.

[23] Ikeda, N., Nagasawa, M., Watanabe, S. (1966). A construction of Markov processes by piecing out. *Proc. Jpn. Acad.* 42(4):370–375.

[24] Löcherbach, E. (2002). Likelihood ratio processes for Markovian particle systems with kill-ing and jumps. *Stat. Inference Stoch. Process.* 5(2):153–177.

[25] McKean, H. P. (1969). *Stochastic Integrals*. New York: Academic Press.

[26] Ikeda, N., Watanabe, S. (2014). *Stochastic Differential Equations and Diffusion Processes*, Vol. 24. Amsterdam: Elsevier.

Appendix A

Appendix for Section 4

We gather in this subsection some lemmas useful to the analysis of the nonlinear stochastic differen-tial equation in the case of the non-globally Lipschitz drift condition stated in (U). These lemmas rely on the procedure of constructing iteratively the solution by pasting together the diffusive paths between two successive jump times and exploiting the finite intensity of the jump process. The same approach has been used to define branching Markov processes [22–24]. Here, we consider a proced-ure of this type where the process has an additional dependence on an external flow of measures that parametrizes the solution and will play a crucial role in the proof of Theorem 4.1.

Lemma A.1. *Consider the SDE parametrized by a measure* $\alpha \in \mathcal{M}^1(\mathbf{D}([0, T], \mathbb{R}^d))$

$$dX(t) = F(X(t), \alpha_t)dt + \sigma(X(t), \alpha_t)dB_t$$
$$+ \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \psi(X(t^-), \alpha_{t^-}, h_1)\mathbb{1}_{(0,\lambda(X(t^-),\alpha_{t^-}))}(u)\mathcal{N}(dt, du, dh). \tag{20}$$

If the coefficients satisfy Assumption 3, then for every $\alpha \in \mathbf{M}^1(\mathbf{D}([0, T], \mathbb{R}^d))$ *and every square-integrable initial condition, there exists a unique strong solution to* Equation (20).

Proof. First let X^1 and X^2 be two integrable stochastic processes on $[0, T]$ with values in $\mathbb{R}^d$. We define the map that associates the law of X^k to the law of the solution of

$$dY^k(t) = F(Y^k(t), \alpha_t)dt + \sigma(Y^k(t), \alpha_t)dB_t$$
$$+ \int_{[0,\infty)\times[0,1]^{\mathbb{N}}} \psi(X^k(t^-), \alpha_{t^-}, h_1)\mathbb{1}_{(0,\lambda(X^k(t^-),\alpha_{t^-}))}(u)\mathcal{N}(dt, du, dh), \tag{21}$$

which is well defined by Lemma A.2. With the same computation of the proof of Theorem 4.1, we get that, for a small enough $T_0 > 0$, there exists a constant $C_{T_0} < 1$ such that

$$\mathbf{E}\left[\sup_{t\in[0,T_0]}\|Y^1(t) - Y^2(t)\|\right] \leqslant C_{T_0}\mathbf{E}\left[\sup_{t\in[0,T_0]}\|X^1(t) - X^2(t)\|\right].$$

This shows *pathwise uniqueness* for solution of (20). By means of (21), we define a Picard iteration argument that gives a sequence of laws $\{Q^n\}_{n\in\mathbb{N}}$ on $\mathbf{D}([0,T],\mathbb{R}^d)$. Again, there exists a $T_0 > 0$ small enough such that $\{Q^n\}_{n\in\mathbb{N}}$ is a Cauchy sequence for ρ_{T_0} and hence for a weaker but complete Wasserstein metric on $\mathcal{M}^1(\mathbf{D}([0,T_0],\mathbb{R}^d))$. Iterating the procedure over a finite number of time intervals, to cover $[0,T]$, yields the thesis.

The integrability property is proved as in the proof of Lemma A.2.

Remark A.1. Notice that, in the proof of Lemma A.1, we need to define the map by means of (21) and not simply by substituting X^k in the whole right-hand side of (20). In fact, we need to control the jumps by means of a known process, but at the same time, we need to have the same variable as argument of the drift coefficient to exploit the convexity of the potential function U.

Lemma A.2. *Consider the SDE parametrized by two measures α and $\beta \in \mathcal{M}^1(\mathbf{D}([0,T],\mathbb{R}^d))$*

$$dX(t) = F(X(t),\alpha_t)dt + \sigma(X(t^-),\alpha_t)dB_t$$
$$+ \int_{[0,\infty)\times[0,1]^{\mathbb{N}}}\psi(Y(t^-),\alpha_{t^-},h_1)1_{(0,\lambda(Y(t^-),\alpha_{t^-})]}(u)N(dt,du,dh), \tag{22}$$

where $\{Y(t)\}_{t\in[0,T]}$ is a progressively measurable process with law β. If the coefficients satisfy Assumption 3, then for every α and $\beta \in \mathcal{M}^1(\mathbf{D}([0,T],\mathbb{R}^d))$, every square-integrable initial condition, there exists a unique strong solution to Equation (22).

Moreover, let $\mu \doteq \text{Law}((X(t))_{t\in[0,T]})$ be the law of the solution of (22) starting from the square-integrable initial condition $X(0)$ μ_0-distributed, then $\mu \in \mathcal{M}^1(\mathbf{D}([0,T],\mathbb{R}^d))$.

Proof. Let B be an (F_t)-Brownian motion, p be a (F_t)-stationary Poisson point process with characteristic measure $l \times \nu$ and $D_p \doteq \{t \in [0,T] : p(t) \in (0,\infty)\times[0,1]^{\mathbb{N}}\}$ and ξ be a F_0-measurable square-integrable r.v. Let $D \doteq \{s \in D_p \text{ s.t. } p(s) \in \bar{U}_s = (0,\lambda(Y(s^-),\alpha_{s^-})]\times[0,1]\times[0,1]\times\cdots\}$. Let us call $\sigma_1 < \sigma_2 < \dots$ the elements of D. Each σ_n is an F_t-stopping time and $\lim_{n\to\infty}\sigma_n = \infty$ a.s. Indeed, for every $T > 0$ and for a fixed $n \in \mathbb{N}^*$,

$$\mathbf{P}(\sigma_n \leqslant T) = \mathbf{P}\left(\int_0^T\int_{[0,\infty)\times[0,1]^{\mathbb{N}}}\mathbb{1}_{(0,\lambda(Y(t^-),\alpha_{t^-})]}(u)N(du,dh,dt) \geqslant n\right) \leqslant \frac{\mathbf{E}\left[\int_0^T\lambda(Y(t),\alpha_t)dt\right]}{n} \leqslant \frac{C_T}{n},$$

for a certain constant C_T and by Borel–Cantelli, we get the claim. Then we start by showing existence and uniqueness of solutions to (22) on $[0,\sigma_1]$. Consider the equation

$$Z(t) = \xi + \int_0^t F(Z(s),\alpha_s)ds + \int_0^t \sigma(Z(s^-),\alpha_s)dB_s. \tag{23}$$

Existence and uniqueness of a strong solution for (23) are ensured by the classical Hasminskii's test for non-explosion (see e.g. [25] with the Lyapunov function $V(z) = \|z\|^2$). The test's conditions are guaranteed by the inequality

$$\sup_{\alpha\in\mathcal{M}^1(\mathbb{R}^d)} z\cdot F(z,\alpha) + \text{tr}\left(\sigma(z,\alpha)\sigma^T(z,\alpha)\right) \leqslant C\left(1 + \|z\|^2\right), \tag{24}$$

for some $C > 0$, for all $z \in \mathbb{R}^d$. Indeed, fix an $\alpha \in \mathcal{M}^1(\mathbb{R}^d)$. Then, under (U) from Assumption 3, we have

$$z\cdot F(z,\alpha) = -(z-\_0)\cdot\left(\nabla U(z) - \nabla U(\_0)\right) + z\cdot\nabla U(\_0) + z\cdot b(z,\alpha) \leqslant C\left(\|z\|^2 + 1\right),$$

due to the convexity of U and the linear growth of b. A similar bound is obtained for the second

summand in the l.h.s of (24), which has uniform quadratic growth in the z variable. Then, for every integrable initial condition, there exists a unique strong solution to (23). Let π_1 be the projection defined as

$$\pi_1 : \quad [0,1]^{\mathbb{N}} \times [0,\infty) \quad \mapsto \quad [0,1]$$
$$(h, u) \quad \rightarrow \quad h_1.$$

Set

$$X_1(t) := \begin{cases} Z^1(t) & t \in [0, \sigma_1), \\ Z^1(\sigma_1^-) + \psi(Y(\sigma_1^-), \alpha(\sigma_1^-), \pi_1 \circ p(\sigma_1)) & t = \sigma_1, \end{cases} \tag{25}$$

where $\{Z^1(t)\}_{t \geqslant 0}$ is the solution of (23) with initial condition $Z^1(0) = \xi$ a.s. We see that $X^1(t)$ is solution of (22) for $t \in [0, \sigma_1]$. We iterate the procedure by setting $\bar{\bar{\xi}} \doteq X_1(\sigma_1)$, $\bar{B} \doteq (B(t + \sigma_1) - B(\sigma_1))_{t \geqslant 0}$ and $\bar{p} \doteq (p(t + \sigma_1))_{t \geqslant 0}$. We define $\bar{X}_1(t)$ for $t \in [0, \bar{\sigma}_1]$ as we did for $X_1(t)$ in (25), where $\bar{\sigma}_1$ is the smallest time such that $\bar{p}_s$ belongs to $\bar{U}_{\sigma_1 + s}$ and coincides with $\sigma_2 - \sigma_1$. We define

$$X_2(t) := \begin{cases} X_1(t) & t \in [0, \sigma_1], \\ \bar{X}_1(t - \sigma_1) & t \in [\sigma_1, \sigma_2]. \end{cases}$$

Clearly X_2 is solution of (22) for $t \in [0, \sigma_2]$. Since $\lim_{n \to \infty} \sigma_n = \infty$ a.s., we can iterate this procedure to cover the entire time interval $[0, T]$.

To prove that the law μ of a solution of (22) belongs to $\mathcal{M}^1(\mathbf{D}([0, T], \mathbb{R}^d))$, we will show that there exists a filtered probability space $(\Omega, \mathbf{P}, (\mathcal{F}_t), \mathcal{F})$, with a $\mathcal{F}_t$-Brownian motion B, an adapted F_t Poisson random measure N with characteristic measure $l \times l \times \nu$ and a F_0-measurable initial condition $X(0)$ μ_0-distributed such that $\mathbf{E}[\sup_{t \in [0,T]} \|X(t)\|] < \infty$. We consider the process $X(t)$, for all $t \geqslant 0$, solution of (22) and for all $K > 0$ we define the stopping time $\tau_K := \inf\{t \geqslant 0 : \|X(t)\| \geqslant K\}$. We apply Ito's rule to the smooth approximation f^ϵ of $\|\cdot\|$ and taking the limit for $\epsilon \downarrow 0$, to exploit the properties of the potential function U. For the details of the approach, see the proof of Theorem 4.1. By monotone convergence of $f^\epsilon(x)$ to $\|x\|$, we get, for all $t \in [0, T]$,

$$\mathbf{E}\big[\mathbb{1}(t \leqslant \tau_K)\|X(t)\|\big] \leqslant \liminf_{\epsilon \downarrow 0} \mathbf{E}\big[\mathbb{1}(t \leqslant \tau_K) f^\epsilon(X(t))\big] \leqslant \liminf_{\epsilon \downarrow 0} \mathbf{E}\big[f^\epsilon(X(t \wedge \tau_K))\big],$$

and by the optional stopping theorem, we have that there exist three positive constants D_1, D_2 and D_3 such that

$$\mathbf{E}\big[\mathbb{1}(t \leqslant \tau_K)\|X(t)\|\big] \leqslant \mathbf{E}\big[\|X(0)\|\big] + D_1 T + D_2 T \mathbf{E}\Big[\sup_{t \in [0,T]} \|Y(t)\|\Big] + D_1 \int_0^t \mathbf{E}\big[\mathbb{1}(s \leqslant \tau_K)\|X(s)\|\big] ds.$$

By Gronwall's lemma, we get the desired bound on $\mathbf{E}[\mathbb{1}(t \leqslant \tau_K)\|X(t)\|]$, but since it does not depend on K, we take the limit $K \nearrow \infty$ and get the thesis.

Appendix B

Appendix for Section 5

We collect here auxiliary lemmas and proofs for Section 5. First, Lemma B.1 concerns existence and uniqueness of solutions for the particle system (14) under Assumption 4. Notice that the same result holds also for the intermediate particle system (19). Then, thanks to two technical lemmas, we give the proof of Lemma 5.1, crucial for the existence and uniqueness of solution of the nonlinear process (16). Finally, we give the proofs of Lemma 5.4 and Lemma 5.5 that we use in the propagation of chaos section. Notice that, the key ingredient here is represented by the fact that all the main jumps of the processes are such that they make the process go back inside a compact set (the support of U). To exploit that, we need to apply Ito's rule for a process with

jumps (notice that here we do not have a diffusion term). Since all the functions of interest ($\|\cdot\|$ and $\lambda(\cdot)$) have singularities in the origin, we use a smooth approximation of the norm $\|\cdot\|$. As in the proof of Theorem 4.1, for all $\epsilon > 0$, we define

$$f^\epsilon(x) \doteq \sqrt{\|x\|^2 + \epsilon}.$$

Lemma B.1. *Under Assumption 4, for every integrable initial condition $X^N(0) \in \mathbb{R}^{d \times N}$, the SDE (14) admits a unique solution.*

Proof. The main issue is represented by the fact that the function λ is unbounded and not globally Lipschitz continuous, indeed when λ is bounded existence and uniqueness of solutions for (14) are consequences of standard results, see [26]. Therefore, let us consider the truncate function $\lambda^K \doteq \lambda \wedge K$, for $K \in \mathbb{N}$, and the solution $X^{N,K}(t)$ of (14) with the function λ^K instead of λ. This solution exists and it is unique for all $t \in [0, T]$, moreover, by pathwise uniqueness, it holds $X^{N,K}(t) = X^{N,K+1}(t)$ for all $t \in \tau^K$, where $\tau^K \doteq \inf\{t \ / \|X^{N,K}(t)\| \geqslant K\}$. Therefore $\tau^K \leqslant \tau^{K+1}$ a.s. and there exists a pathwise unique solution $X(t)$ to (14), defined for all $t \in [0, \tau)$, where $\tau \doteq \sup_{K \in \mathbb{N}} \tau^K$. We are left to prove that $\mathbf{P}(\tau > T) = 1$.

Let us fix $i \in \{1, ..., N\}$ and $\epsilon > 0$. By computing $f^\epsilon(X_i^N(t))$ by means of Ito's formula, we get

$$f^\epsilon(X_i^N(t)) \leqslant f^\epsilon(X_i^N(0)) + \frac{1}{N} \sum_{j=1}^N \int_0^t \int_{[0,1]^N} \int_0^\infty f^\epsilon(V(h_j, h_i)) \mathbb{1}_{(0, \lambda(X_j^N(s)))}(u) N^j(ds, du, dh)$$
$$+ \int_0^t \int_{[0,1]^N} \int_0^\infty \left(f^\epsilon(U(h_i)) - f^\epsilon(X_i^N(s))\right) \mathbb{1}_{(0, \lambda(X_i^N(s)))}(u) N^i(ds, du, dh).$$

Therefore, summing on all $i = 1, ..., N$ and taking expectation, by the application of Fatou's lemma we get:

$$\mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \|X_i^N(t)\|\right] \leqslant \liminf_{\epsilon \downarrow 0} \left(\mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N f^\epsilon(X_i^N(0))\right]\right.$$
$$+ \int_0^t (\mathbf{E}[f^\epsilon(V)] + \mathbf{E}[f^\epsilon(U)]) \mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \lambda(X_i^N(s))\right] - \mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N f^\epsilon(X_i^N(s)) \lambda(X_i^N(s))\right] ds\right)$$

Then, by monotone convergence, we have

$$\mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \|X_i^N(t)\|\right] \leqslant \mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \|X_i^N(0)\|\right]$$
$$+ \int_0^t (\mathbf{E}[\|V\|] + \mathbf{E}[\|U\|]) \mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \lambda(X_i^N(s))\right] - \mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \|X_i^N(s)\| \lambda(X_i^N(s))\right] ds.$$

Since b is increasing and h is bounded, there exists a positive constant C, depending on $\mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \|X_i^N(0)\|\right]$, such that

$$\sup_{t \geqslant 0} \mathbf{E}\left[\frac{1}{N} \sum_{i=1}^N \|X_i^N(t)\|\right] \leqslant C,$$

implying $\mathbf{P}(\tau > T) = 1$. ∎

The proof of existence and uniqueness of solutions of (16) for compact support initial condition relies on a straightforward adaptation of the arguments of [10] to our framework, therefore we write the proof of Lemma 5.1 only for completeness.

Proof of Lemma 5.1. We want to get an almost sure bound for $\|Z_f(t)\|$, in order to use locally Lipschitzianity of λ in the following computations. Intuitively, the jumps have an increasing role

only if we are inside the support of the random variable U, otherwise they force the norm to decrease. Therefore, a.s., we can bound the process $\|Z_f(t)\|$ with the deterministic expression

$$K_0 + \int_0^t \|f(s)\| ds,$$

where $K_0 := \max\{\|x\|, \sup_{h \in [0,1]} \|U(h)\|\}$. This almost sure bound for $\|Z_f(t)\|$ and the continuity of the coefficients ensure the existence and uniqueness of a non-explosive solution Z_f on $[0, T]$. Let Z_f and Z_g two solutions of (17) corresponding to two different locally bounded Borellian functions f and g. The almost sure bounds on $\|Z_f(t)\|$ and $\|Z_g(t)\|$ let us define two positive constant $b_{f,g}(T)$ and $L_{f,g}(T)$, such that we have

$$\mathbf{E}\left[\sup_{t \in [0,T]} \|Z_f(t) - Z_g(t)\|\right] \leqslant \int_0^T \mathbf{E}\left[\sup_{s \in [0,t]} \|Z_f(s) - Z_g(s)\|\right] ds + \int_0^T \sup_{s \in [0,t]} \|f(s) - g(s)\| dt$$

$$+ \mathbf{E}\left[\int_0^T \int_{[0,1] \times [0,\infty)} \|\left(Z_f(s^-) - U(h)\right) \mathbb{1}_{[0,\lambda(Z_f(s^-)))}(u) - \left(Z_g(s^-) - U(h)\right) \mathbb{1}_{[0,\lambda(Z_g(s^-)))}(u)\| ds du \nu_1(dh)\right]$$

$$\leqslant \int_0^T \mathbf{E}\left[\sup_{s \in [0,t]} \|Z_f(s) - Z_g(s)\|\right] ds + \int_0^T \sup_{s \in [0,t]} \|f(s) - g(s)\| dt$$

$$+ \left(b_{f,g}(T) + H\right) \int_0^T \mathbf{E}\left[\sup_{s \in [0,t]} \|Z_f(s) - Z_g(s)\|\right] ds$$

$$+ L_{f,g}(T) \left(K_0 + \int_0^t \|f(s)\| ds\right) \int_0^T \mathbf{E}\left[\sup_{s \in [0,t]} \|Z_f(s) - Z_g(s)\|\right] ds.$$

We apply now Gronwall lemma and we obtain (18).

The proof of Lemma 5.2 requires two technical lemmas adapted to our case from [10].

Lemma B.2. *Let $x(t)$ be a non-negative C^1 function on $\mathbb{R}_+$. If the following inequality holds for any $0 \leqslant s \leqslant t$:*

$$x(t) \leqslant x(s) - \bar{K} \int_s^t x^k(u) du + \int_s^t P_\delta(x(u)) du$$

where $k, \bar{K} > 0$ and $P_\delta(\cdot)$ is a polynomial of degree $\delta < k$, then

$$\sup_{t \geqslant 0} x(t) \leqslant C_0 < \infty.$$

Proof. Consider that for $x \to \infty$, then

$$-\bar{K} x^k + P_\delta(x) \to -\infty.$$

Therefore it exists a value $\bar{C}_0$ such that, as soon as the trajectory exceeds $\bar{C}_0 \geqslant 0$ its derivative becomes strictly negative and the trajectory is forced toward zero. Thus, defining

$$C_0 := \max\{\bar{C}_0, x(0)\},$$

we get the desired bound.

Lemma B.3. *If the function b satisfies the assumption (15), then for any $\epsilon > 0$ and $p \in [1, 4 + 2\epsilon]$, there exists a constant $\gamma_1 < (4 + 2\epsilon)\gamma$, $c_1 > 0$ and a value $\eta > 0$, such that, for all $a \in \mathbb{R}^d$ with $\|a\| \leqslant \eta$ and for all $x \in \mathbb{R}^d$, it holds*

$$|b^p(\|x + a\|) - b^p(\|a\|)| \leqslant \|a\| \left(\gamma_1 b^p(\|x\|) + c_1\right).$$

Proof. The proof of this lemma comes directly from Lemma 8 in the appendix of [10].

Notice that the constant $\gamma_1 < (4 + 2\epsilon)\gamma$, together with the condition of Lemma B.2 on the positivity of the coefficient $\bar{K}$, cause the condition on γ w.r.t $\mathbf{E}[\|V\|]$ in Assumption 4. This condition plays a crucial role in all the proofs of the boundedness for the moments of $\lambda(X(t))$ and of $\lambda(X_i^N(t))$ for all i. Now that we have stated these two results, we are ready to prove Lemma 5.2, that provides a priori uniform bounds on the first moment of the solution to (16) and on the moments of $\lambda(X(t))$.

Proof of Lemma 5.2. Fix $\epsilon > 0$, by means of Ito's rule, we have

$$\mathbf{E}[f^\epsilon(X(t))] \leqslant \mathbf{E}[f^\epsilon(X(0))] - \int_0^t \mathbf{E}\big[f^\epsilon(\|X(s)\|)\big]ds + \int_0^t \mathbf{E}\big[(\mathbf{E}[\|V\|] + \mathbf{E}[\|U\|] - f^\epsilon(X(s)))h(X(s))\big]ds$$
$$+ \int_0^t \mathbf{E}\Big[b\big(\|X(s)\|\big)\big(\mathbf{E}[\|V\|] + \mathbf{E}[\|U\|] - f^\epsilon(X(s))\big)\Big]ds.$$

From the monoticity assumption on b, we know that there exist $\Lambda > 0$ and $\beta \geqslant 0$ such that $b(r)(\mathbf{E}[\|V\|] + \mathbf{E}[\|U\|] - r) \leqslant -\Lambda r + \beta$. Therefore, by Fatou's lemma and the monotone convergence theorem,

$$\mathbf{E}\big[\|X(t)\|\big] \leqslant \mathbf{E}\big[\|X(0)\|\big] + \int_0^t [H(\mathbf{E}[\|V\|] + \mathbf{E}[\|U\|]) + \beta]ds - \Lambda \int_0^t \mathbf{E}\big[\|X(s)\|\big]ds,$$

that gives the boundedness of $\sup_{t \geqslant 0} \mathbf{E}[\|X(t)\|]$.

Let $p = 1$, clearly, to get a bound for $\mathbf{E}[\lambda(X(t))]$, it is sufficient to bound $\mathbf{E}[b(\|X(t)\|)]$. Thus, again, we use Ito's rule to compute $b(f^\epsilon(X(t)))$ for $\epsilon > 0$.

$$\mathbf{E}[b(f^\epsilon(X(t)))] \leqslant \mathbf{E}[b(f^\epsilon(X(0)))] - \int_0^t \mathbf{E}\left[b'\big(f^\epsilon(X(s))\big)\frac{\|X(s)\|^2}{\sqrt{\|X(s)\|^2 + \epsilon}}\right]ds$$

$$+ \int_0^t \mathbf{E}\left[b'\big(f^\epsilon(X(s))\big)\mathbf{E}\big[b\big(\|X(s)\|\big)\big]\frac{X(s) \cdot \mathbf{E}[V]}{\sqrt{\|X(s)\|^2 + \epsilon}}\right]ds + H \int_0^t \mathbf{E}\left[b'\big(f^\epsilon(X(s))\big)\frac{X(s) \cdot \mathbf{E}[V]}{\sqrt{\|X(s)\|^2 + \epsilon}}\right]ds$$

$$+ \int_0^t \mathbf{E}\big[b\big(\|X(s)\|\big)\big]\mathbf{E}[b(f^\epsilon(U))]ds + \int_0^t \mathbf{E}[h(X(s))]\mathbf{E}[b(f^\epsilon(U))]ds$$

$$- \int_0^t \mathbf{E}\Big[b\big(f^\epsilon(X(s))\big)b\big(\|X(s)\|\big)\Big]ds - \int_0^t \mathbf{E}\big[b\big(f^\epsilon(X(s))\big)\big]\mathbf{E}[h(X(s))]ds.$$

Again we use Fatou's lemma and monotone convergence theorem (indeed $b(f^\epsilon(\cdot))$ converges monotonically to $b(\|\cdot\|)$, thanks to the increasing property of b). Since b' is positive, we disregard the second term, we use properties of b' to bound the remaining terms and we get

$$\mathbf{E}\big[b\big(\|X(t)\|\big)\big] \leqslant \mathbf{E}\big[b\big(\|X(0)\|\big)\big] + (Hc\mathbf{E}[\|V\|] + H\mathbf{E}[b(\|U\|)])t + (\gamma\mathbf{E}[\|V\|] - 1)\int_0^t \mathbf{E}\big[b\big(\|X(s)\|\big)\big]^2 ds$$
$$(c\mathbf{E}[\|V\|] + H\gamma\mathbf{E}[\|V\|] + \mathbf{E}[b(\|U\|)] + H)\int_0^t \mathbf{E}\big[b\big(\|X(s)\|\big)\big]ds.$$

With Lemma B.2 we conclude the boundedness for $\mathbf{E}[b(\|X(t)\|)]$. The same argument is used to get a uniform bound for $\mathbf{E}[b^p(\|X(t)\|)]$ when $p = 2, 3, 4$.

While the uniform bounds for $\mathbf{E}[\|X(t)\|]$ and $\mathbf{E}[b(\|X(t)\|)]$ are needed for the well-posedness of the nonlinear process itself, higher moments of λ are needed only for the proof of propagation of chaos. The same a priori bounds for the moments of λ appear also in the case of the particle system. Their proof is similar to the nonlinear case, relies on Lemmas B.2 and B.3, together with an argument based on orthogonal martingales.

Proof of Lemma 5.4. We only prove it for μ_X^N, then for μ_Y^N the steps are basically the same. Of course it is sufficient to prove the boundedness of $\sup_{N \geqslant N_0} \sup_{t \geqslant 0} \mathbf{E}[\langle \mu_X^N(t), b^4(f^\delta(\cdot))\rangle]$. Let us define for $K > 0$ the stopping time $\tau_K := \inf\{t \geqslant 0 : \langle \mu_X^N(t), b^5(f^\delta(\cdot))\rangle \geqslant K\}$. Obviously the random variables $\langle \mu_X^N(t \wedge \tau_K), b^p(f^\delta(\cdot))\rangle$ for $1 \leqslant p \leqslant 5$ and $\langle \mu_X^N(t \wedge \tau_K), f^\delta(\cdot)\rangle$ are integrable. Recall that, for all $\epsilon > 0$, the process $\{M_\epsilon^N(t)\}_{t \geqslant [0,T]}$, where, for $t \in [0, T]$ we have

$$M_\epsilon^N(t) \doteq \langle \mu_X^N(t), f^\epsilon(\cdot) \rangle - \langle \mu_X^N(0), f^\epsilon(\cdot) \rangle + \frac{1}{N} \sum_{i=1}^N \left(\int_0^t \frac{\|X_i^N(s)\|^2}{\sqrt{\|X_i^N(s)\|^2 + \epsilon}} ds \right)$$

$$- \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} \int_0^t \int_{[0,1]^{\mathbb{N}}} \lambda\left(X_j^N(s)\right) \left(f^\epsilon\left(X_i^N(s) + \frac{V(h_i, h_j)}{N}\right) - f^\epsilon\left(X_i^N(s)\right) \right) \nu(dh) ds$$

$$- \frac{1}{N} \sum_{i=1}^N \int_0^t \int_{[0,1]^{\mathbb{N}}} \lambda\left(X_i^N(s)\right) \left(f^\epsilon(U(h_i)) - f^\epsilon\left(X_i^N(s)\right) \right) \nu(dh) ds,$$

is a martingale. Then, from the optional stopping theorem, we have

$$\mathbf{E}\left[\langle \mu_X^N(t \wedge \tau_K), f^\epsilon(\cdot) \rangle \right] \leqslant \mathbf{E}\left[\langle \mu_X^N(0), f^\epsilon(\cdot) \rangle \right] - \mathbf{E}\left[\int_0^{t \wedge \tau_K} \langle \mu_X^N(s), f^\epsilon(\cdot) \rangle ds \right]$$

$$+ N \mathbf{E}\left[\int_0^{t \wedge \tau_K} \langle \mu_X^N(s), \lambda(\cdot) \rangle \langle \mu_X^N(s), \int_{[0,1]^2} f^\epsilon\left(\cdot + \frac{V(h_1, h_2)}{N} \right) - f^\epsilon(\cdot) \nu_2(dh) \rangle ds \right]$$

$$- \mathbf{E}\left[\int_0^{t \wedge \tau_K} \langle \mu_X^N(s), \lambda(\cdot) \int_{[0,1]} f^\epsilon\left(\cdot + \frac{V(h_1, h_1)}{N} \right) - f^\epsilon(\cdot) \nu_1(dh) \rangle ds \right]$$

$$+ \mathbf{E}\left[\int_0^{t \wedge \tau_K} \mathbf{E}[f^\epsilon(U)] \langle \mu_X^N(s), \lambda(\cdot) \rangle - \langle \mu_X^N(s), \lambda(\cdot) f^\epsilon(\cdot) \rangle ds \right].$$

Again, we use the monotone convergence of $f^\epsilon(x)$ to $\|x\|$, to get

$$\mathbf{E}\left[1(t \leqslant \tau_K) \langle \mu_X^N(t), \|\cdot\| \rangle \right] \leqslant \liminf_{\epsilon \downarrow 0} \mathbf{E}\left[1(t \leqslant \tau_K) \langle \mu_X^N(t), f^\epsilon(\cdot) \rangle \right] \leqslant \liminf_{\epsilon \downarrow 0} \mathbf{E}\left[\langle \mu_X^N(t \wedge \tau_K), f^\epsilon(\cdot) \rangle \right].$$

By arguments close to the one in the proof of Lemma 5.2, there exists $\Lambda > 0$ and $\beta > 0$, such that we get the following inequality

$$\mathbf{E}\left[1(t \leqslant \tau_K) \langle \mu_X^N(t), \|\cdot\| \rangle \right] \leqslant \mathbf{E}\left[\langle \mu_X^N(0), \|\cdot\| \rangle \right]$$

$$+ \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), \left(\mathbf{E}[\|V\|] + \frac{\mathbf{E}[\|V\|]}{N} + \mathbf{E}[\|U\|] - \|\cdot\| \right) \lambda(\cdot) \rangle \right] ds$$

$$\leqslant \mathbf{E}\left[\langle \mu_X^N(0), \|\cdot\| \rangle \right] + \left[H\left(\mathbf{E}[\|V\|] + \frac{\mathbf{E}[\|V\|]}{N} + \mathbf{E}[\|U\|] \right) + \beta \right] t$$

$$- \Lambda \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), \|\cdot\| \rangle \right] ds.$$

This, together with Lemma B.2, gives the boundedness of $\sup_{t \geqslant 0} \mathbf{E}[1(t \leqslant \tau_K) \langle \mu_X^N(t), \|\cdot\| \rangle]$. Since this bound does not depend on K, letting K go to infinity gives the bound on $\sup_{t \geqslant 0} \mathbf{E}[\langle \mu_X^N(t), \|\cdot\| \rangle]$. Now we apply the same argument to the martingale $\{M_{b^4}^N(t)\}_{t \geqslant [0,T]}$. By deleting some of the negative terms, applying Lemma B.3 and repeating the previous steps, we obtain the following bound

$$\mathbf{E}\left[1(\tau_K \leqslant t) \langle \mu_X^N(t), b^4(\|\cdot\|) \rangle \right] \leqslant \mathbf{E}\left[\langle \mu_X^N(0), b^4(\|\cdot\|) \rangle \right]$$

$$+ \gamma_1 \mathbf{E}[\|V\|] \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b^4(\|\cdot\|) \rangle \langle \mu_X^N(s), b(\|\cdot\|) \rangle \right] ds$$

$$+ H \gamma_1 \mathbf{E}[\|V\|] \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b^4(\|\cdot\|) \rangle \right] ds$$

$$+ c_1 \mathbf{E}[\|V\|] \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b(\|\cdot\|) \rangle \right] ds + c_1 H \mathbf{E}[\|V\|] \int_0^t \mathbf{E}[1(s \leqslant \tau_K)] ds$$

$$+ \gamma_1 \frac{\mathbf{E}[\|V\|]}{N} \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b^5(\|\cdot\|) \rangle \right] ds + c_1 \frac{\mathbf{E}[\|V\|]}{N} \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b(\|\cdot\|) \rangle \right] ds$$

$$+ H \gamma_1 \frac{\mathbf{E}[\|V\|]}{N} \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b^4(\|\cdot\|) \rangle \right] ds + c_1 H \frac{\mathbf{E}[\|V\|]}{N} \int_0^t \mathbf{E}[1(s \leqslant \tau_K)] ds$$

$$+ \mathbf{E}\left[b^4(\|U\|) \right] \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b(\|\cdot\|) \rangle \right] ds + \mathbf{E}\left[b^4(\|U\|) \right] H t$$

$$- \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b^5(\|\cdot\|) \rangle \right] ds + H \int_0^t \mathbf{E}\left[1(s \leqslant \tau_K) \langle \mu_X^N(s), b^4(\|\cdot\|) \rangle \right] ds.$$

By Hölder and Jensen inequalities, we get the following expression

$$\mathbf{E}\big[\langle \mathbb{1}(\tau_K \leqslant t)\mu_X^N(t), b^4(\|\cdot\|)\rangle\big] \leqslant \mathbf{E}\big[\langle \mu_X^N(0), b^4(\|\cdot\|)\rangle\big]$$

$$+\gamma_1\mathbf{E}[\|V\|]\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]^{5/4}ds + H\gamma_1\mathbf{E}[\|V\|]\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]ds$$

$$+c_1\mathbf{E}[\|V\|]\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]^{1/4}ds + c_1 H\mathbf{E}[\|V\|]t$$

$$+c_1\frac{\mathbf{E}[\|V\|]}{N}\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]^{1/4}ds + H\gamma_1\frac{\mathbf{E}[\|V\|]}{N}\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]ds$$

$$+\Big(c_1 H\frac{\mathbf{E}[\|V\|]}{N} + \mathbf{E}\big[b^4(\|U\|)\big]\Big)t + \mathbf{E}\big[b^4(\|U\|)\big]\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle^{1/4}\big]ds$$

$$+H\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]ds + \Big(\gamma_1\frac{\mathbf{E}[\|V\|]}{N} - 1\Big)\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]^{5/4}ds,$$

where we have exploited the fact that $(\gamma_1\frac{\mathbf{E}[\|V\|]}{N} - 1)<0$, for N large enough, and that $\langle \mu_X^N(s), b^5\rangle \geqslant \langle \mu_X^N(s), b^4\rangle^{5/4}$. Reordering, we get

$$\mathbf{E}\big[\mathbb{1}(\tau_K\leqslant t)\langle \mu_X^N(t), b^4(\|\cdot\|)\rangle\big] \leqslant \langle \mathbf{E}\big[\mu_X^N(0), b^4(\|\cdot\|)\rangle\big]$$

$$+\Big(c_1\mathbf{E}[\|V\|] + c_1\frac{\mathbf{E}[\|V\|]}{N} + \mathbf{E}\big[b^4(\|U\|)\big]\Big)\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]^{1/4}ds$$

$$+\Big(H\gamma_1\mathbf{E}[\|V\|] + H\gamma_1\frac{\mathbf{E}[\|V\|]}{N} + H\Big)\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]ds$$

$$+\Big(\gamma_1\mathbf{E}[\|V\|] + \gamma_1\frac{\mathbf{E}[\|V\|]}{N} - 1\Big)\int_0^t \mathbf{E}\big[\mathbb{1}(s\leqslant\tau_K)\langle \mu_X^N(s), b^4(\|\cdot\|)\rangle\big]^{5/4}ds.$$

Since, by hypothesis, there exists N_0 such that, for all $N \geqslant N_0$ it holds

$$\Big(\gamma_1\mathbf{E}[\|V\|] + \gamma_1\frac{\mathbf{E}[\|V\|]}{N} - 1\Big)<0,$$

we use Proposition B.2 and this gives a bound on $\mathbf{E}[\mathbb{1}(t\leqslant\tau_K)\langle \mu_X^N(t), b^4(\|\cdot\|)\rangle]$ independent of N and K; therefore letting K go to infinity proves the thesis.

As mentioned before, Lemma 5.4 plays a crucial role in the proof of Lemma 5.5, where we bound the number of jumps of a single particle for the particle system (14) and the contribution of the collateral drift term for the particle system (19).

Proof of Lemma 5.5. We develop the computations for the proof just in the case of (14), since for the system (19) they are almost the same. Let us start by describing the quantity $C_N(T)$, that is

$$C_N(T) = \sum_{i=1}^N \int_0^T \int_{[0,1]^N} \int_0^\infty \mathbb{1}_{[0,\lambda(X_i^N(s))(u)}N^i(ds,du,dh).$$

We can rewrite this quantity as the sum of orthogonal martingales, that we will indicate as $M^N(t)$, plus a term depending on the empirical measure, as follows:

$$\frac{C_N(T)}{N} = \frac{1}{N}\sum_{i=1}^N \int_0^T \int_{[0,1]^N} \int_0^\infty \mathbb{1}_{[0,\lambda(X_i^N)(s)(u)}\widetilde{N}^i(ds,du,dh) + \int_0^T \langle \mu_X^N(s),\lambda(\cdot)\rangle ds \doteq M^N(T) + \int_0^T \langle \mu_X^N(s),\lambda(\cdot)\rangle ds.$$

Let us consider a positive constant $H_T > 0$, then

$$\mathbf{P}\left(\frac{C_N(T)}{N} \geqslant H_T\right) \leqslant \mathbf{P}\left(\sup_{t \leqslant T} M^N(t) \geqslant \frac{H_T}{2}\right) + \mathbf{P}\left(\int_0^T \langle \mu_X^N(s), \lambda \rangle ds \geqslant \frac{H_T}{2}\right).$$

Of course, since $\{M^N(t)\}_{t \in [0,T]}$ is a martingale, by Chebychev and Doob inequalities we have that

$$\mathbf{P}\left(\sup_{t \in [0,T]} M^N(t) \geqslant \frac{H_T}{2}\right) \leqslant \frac{4}{H_T^2} \mathbf{E}\left[\sup_{t \in [0,T]} \left(M^N(t)\right)^2\right] \leqslant \frac{4}{H_T^2} \mathbf{E}\left[\langle M^N(T)\rangle\right].$$

By the orthogonality of the martingales, however, we have that

$$\left[\langle M^N(T)\rangle\right] \leqslant \frac{\mathbf{E}\left[\int_0^T \langle \mu_X^N(s), \lambda \rangle ds\right]}{N}.$$

This, together with Lemma 5.4, implies that, for all N sufficiently large, there exists a constant C_T such that

$$\mathbf{P}\left(\sup_{t \leqslant T} M^N(t) \geqslant \frac{H_T}{2}\right) \leqslant \frac{C_T}{N}.$$

Therefore, we want to get a bound for the probability $\mathbf{P}(\int_0^T \langle \mu_X^N(s), \lambda \rangle ds \geqslant H_T)$. Let $\delta > 0$ be fixed, the first step consists in proving that there exists $C_T > 0$ such that exists $C_T > 0$ such that

$$\mathbf{E}\left[\sup_{t \in [0,T]} M_{b,\delta}^N(t)^2\right] \leqslant \mathbf{E}\left[\langle M_{b,\delta}^N(T)\rangle\right] \leqslant \frac{C_T}{N},$$

where $\{M_{b,\delta}^N(t)\}_{t \in [0,T]}$ is the martingale arising from the compensated Poisson measure in the computation of $\langle \mu_X^N(t), b(f^\delta(\cdot))\rangle$ with Ito rule, that is

$$M_{b,\delta}^N(t) \doteq \frac{1}{N} \sum_{i=1}^N \int_0^t \int_{[0,1]^N} \int_0^\infty \mathbb{1}_{(0,\lambda(X_i^N(s))]}[b(f^\delta(U(h_i))) - b(f^\delta(X_i^N(s)))$$
$$+ \sum_{j \neq i} b(f^\delta(X_j^N(s) + \frac{V(h_i, h_j)}{N})) - b(f^\delta(X_j^N(s)))]\widetilde{N}^i(ds, du, dh)$$

and $\langle M_{b,\delta}^N(t)\rangle$ is its quadratic variation. We use the fact that $\{\widetilde{N}^i\}_{i=1,2\ldots}$ is a family of orthogonal martingales, therefore

$$\langle M_{b,\delta}^N(t)\rangle = \frac{1}{N^2} \sum_{i=1}^N \int_0^t \int_{[0,1]^N} \lambda(X_i^N(s))[b(f^\delta(U(h_i))) - b(f^\delta(X_i^N(s)))$$
$$+ \sum_{j \neq i} b(f^\delta(X_j^N(s) + \frac{V(h_i,h_j)}{N})) - b(f^\delta(X_j^N(s)))]^2 \nu(dh)ds.$$

Let us write $\langle M_{b,\delta}^N(t)\rangle \doteq \frac{1}{N^2}\sum_{i=1}^N M_{b,\delta,i}^N(t)$, we fix i and we compute $M_{b,\delta,i}(t)$ as follows.

$$M_{b,\delta,i}(t) \leqslant 2\int_0^t\int_{[0,1]^N} b\Big(f^\delta\big(X_i^N(s)\big)\Big)b^2\big(f^\delta(U)\big)+Hb^2\big(f^\delta(U)\big)+b^3\Big(f^\delta\big(X_i^N(s)\big)\Big)+Hb^2\Big(f^\delta\big(X_i^N(s)\big)\Big)\nu(dh)ds$$

$$+\int_0^t\int_{[0,1]^N} b\Big(f^\delta\big(X_i^N(s)\big)\Big)(N-1)\sum_{j\neq i}\left(\frac{f^\delta(V)}{N}\Big(\gamma_1 b\big(f^\delta\big(X_j^N(s)\big)\big)+c_1\Big)\right)^2\nu(dh)ds$$

$$+\int_0^t\int_{[0,1]^N} H(N-1)\sum_{j\neq i}\left(\frac{f^\delta(V)}{N}\Big(\gamma_1 b\big(f^\delta\big(X_j^N(s)\big)\big)+c_1\Big)\right)^2\nu(dh)ds$$

$$+2\int_0^t\int_{[0,1]^N} b\Big(f^\delta\big(X_i^N(s)\big)\Big)\Big(b\big(f^\delta(U)\big)-b\big(f^\delta\big(X_i^N(s)\big)\big)\Big)\sum_{j\neq i}\left(\frac{f^\delta(V)}{N}\Big(\gamma_1 b\big(f^\delta\big(X_j^N(s)\big)\big)+c_1\Big)\right)\nu(dh)ds$$

$$+2H\int_0^t\int_{[0,1]^N}\Big(b\big(f^\delta(U)\big)-b\big(f^\delta\big(X_i^N(s)\big)\big)\Big)\sum_{j\neq i}\left(\frac{f^\delta(V)}{N}\Big(\gamma_1 b\big(f^\delta\big(X_j^N(s)\big)\big)+c_1\Big)\right)\nu(dh)ds$$

$$\leqslant \left(2H\mathbf{E}\big[b^2\big(f^\delta(U)\big)\big]+Hc_1^2\mathbf{E}\big[f^\delta(V)^2\big]\frac{N-1}{N}+2c_1\mathbf{E}\big[b\big(f^\delta(U)\big)\big]\mathbf{E}\big[f^\delta(U)\big]H\right)t$$

$$+\left(2\mathbf{E}\big[b^2\big(f^\delta(U)\big)\big]+c_1^2\mathbf{E}\big[f^\delta(V)^2\big]\frac{N-1}{N}+2c_1\mathbf{E}\big[b\big(f^\delta(U)\big)\big]\mathbf{E}\big[f^\delta(V)\big]+2c_1\mathbf{E}\big[f^\delta(V)\big]H\right)\int_0^t b\Big(f^\delta\big(X_i^N(s)\big)\Big)ds$$

$$+\big(2H+2c_1\mathbf{E}\big[f^\delta(V)\big]\big)\int_0^t b^2\Big(f^\delta\big(X_i^N(s)\big)\Big)ds+\int_0^t b^3\Big(f^\delta\big(X_i^N(s)\big)\Big)ds$$

$$+\Big(2\gamma_1\mathbf{E}\big[b\big(f^\delta(U)\big)\big]\mathbf{E}\big[f^\delta(V)\big]+2\gamma_1\mathbf{E}\big[f^\delta(V)\big]H\Big)\int_0^t b\Big(f^\delta\big(X_i^N(s)\big)\Big)\langle\mu_X^N(s),b\big(f^\delta(\cdot)\big)\rangle ds$$

$$+\gamma_1^2\mathbf{E}\big[f^\delta(V)^2\big]\frac{N-1}{N}\int_0^t b\Big(f^\delta\big(X_i^N(s)\big)\Big)\langle\mu_X^N(s),b^2\big(f^\delta(\cdot)\big)\rangle ds$$

$$+H\gamma_1\mathbf{E}\big[f^\delta(V)^2\big]\frac{N-1}{N}\int_0^t\langle\mu_X^N(s),b^2\big(f^\delta(\cdot)\big)\rangle ds+2\gamma_1\mathbf{E}\big[b\big(f^\delta(U)\big)\big]\mathbf{E}\big[f^\delta(V)\big]\int_0^t\langle\mu_X^N(s),b\big(f^\delta(\cdot)\big)\rangle ds$$

$$+2\gamma_1\mathbf{E}\big[f^\delta(V)\big]\int_0^t b^2\Big(f^\delta\big(X_i^N(s)\big)\Big)\langle\mu_X^N(s),b\big(f^\delta(\cdot)\big)\rangle ds$$

Summing over all $i=1,...,N$ and dividing by N^2, we can find four positive constants K_1, K_2, K_3 and K_4 such that $\langle M_b^N(t)\rangle$ is bounded by the expression

$$\frac{K_1}{N}t+\frac{K_2}{N}\int_0^t\langle\mu_X^N(s),b^3\big(f^\delta(\cdot)\big)\rangle^{1/3}ds+\frac{K_3}{N}\int_0^t\langle\mu_X^N(s),b^3\big(f^\delta(\cdot)\big)\rangle^{2/3}ds+\frac{K_4}{N}\int_0^t\langle\mu_X^N(s),b^3\big(f^\delta(\cdot)\big)\rangle ds$$

Using the result of Lemma 5.4, we know that there exists a certain N_0, such that the expectation of all the terms involved is bounded uniformly in $N>N_0$. Therefore, for such N we have

$$\mathbf{E}\left[\sup_{t\in[0,T]}M_{b,\delta}^N(t)\right]\leqslant\frac{C_T}{N}.$$

By Chebychev and Doob inequalities this leads to

$$\mathbf{P}\left(\sup_{t\in[0,T]}M_{b,\delta}^N(t)\geqslant 1\right)\leqslant\mathbf{E}\left[\sup_{t\in[0,T]}\Big(M_{b,\delta}^N(t)\Big)^2\right]\leqslant\mathbf{E}\Big[\langle M_{b,\delta}^N(T)\rangle\Big]\leqslant\frac{C_T}{N}.$$

Now, we compute $\langle\mu_X^N(t),b(f^\delta(\cdot))\rangle$ with Ito's rule that gives the following bound:

$$\langle\mu_X^N(t),b\big(f^\delta(\cdot)\big)\rangle\leqslant\langle\mu_X^N(0),b\big(f^\delta(\cdot)\big)\rangle+M_{b,\delta}^N(t)+\left(\mathbf{E}\big[f^\delta(V)\big]\gamma_1\Big(1+\frac{1}{N}\Big)-1\right)\int_0^t\langle\mu_X^N(t),b^2\big(f^\delta(\cdot)\big)\rangle ds$$

$$+\left(H\mathbf{E}\big[f^\delta(V)\big]\gamma_1\Big(1+\frac{1}{N}\Big)+H+\mathbf{E}\big[f^\delta(V)\big]c_1\Big(1+\frac{1}{N}\Big)+\mathbf{E}\big[b^2\big(f^\delta(U)\big)\big]\right)\int_0^t\langle\mu_X^N(t),b^2\big(f^\delta(\cdot)\big)\rangle^{1/2}ds$$

$$+H\Big(\mathbf{E}\big[f^\delta(V)\big]c_1+\mathbf{E}\big[b^2\big(f^\delta(U)\big)\big]\Big)t.$$

Since, by hypothesis, $b(f^\delta(\cdot))$ is integrable with respect to the law of $X(0)$, from the law of large numbers we know that

$$\mathbf{P}\left(\left\langle \mu_X^N(0), b\left(f^\delta(\cdot)\right)\right\rangle \geqslant 1 + \mathbf{E}\left[b\left(f^\delta(X(0))\right)\right]\right) \leqslant \frac{\operatorname{Var}\left(b\left(f^\delta(X(0))\right)\right)}{N}.$$

Let us consider the event

$$\left\{\left\langle \mu_X^N(0), b\left(f^\delta(\cdot)\right)\right\rangle < 1 + \mathbf{E}\left[b\left(f^\delta(X(0))\right)\right]\right\} \cup \left\{\sup_{t\in[0,T]} M_{b,\delta}^N(t) < 1\right\},$$

that has a probability greater than $1-2\frac{C}{N}$. Under this event, we apply Lemma B.2 to get a bound for $\langle\mu_X^N(T), b(f^\delta(\cdot))\rangle$.

Since, for all $\delta>0$, $\lambda(\cdot)\leqslant b(f^\delta(\cdot)) + H$ a.s., this is equivalent to a bound for $\sup_{t\in[0,T]}\langle\mu_X^N(t), \lambda(\cdot)\rangle$, that leads to the existence of a positive constant K_T such that

$$\mathbf{P}\left(\int_0^T \langle\mu_X^N(s), \lambda\rangle ds \geqslant H_T\right) \leqslant \frac{K_T}{N},$$

and therefore to the desired bound for $\mathbf{P}(\frac{C_N(T)}{N} \geqslant H_T)$. □