The Annals of Applied Probability
2016, Vol. 26, No. 3, 1407–1442
DOI: 10.1214/15-AAP1120

PROPAGATION OF CHAOS FOR INTERACTING PARTICLES SUBJECT TO ENVIRONMENTAL NOISE

BY MICHELE COGHI AND FRANCO FLANDOLI

Scuola Normale Superiore and University of Pisa

A system of interacting particles described by stochastic differential equations is considered. As oppposed to the usual model, where the noise perturbations acting on different particles are independent, here the particles are subject to the same space-dependent noise, similar to the (noninteracting) particles of the theory of diffusion of passive scalars. We prove a result of propagation of chaos and show that the limit PDE is stochastic and of inviscid type, as opposed to the case when independent noises drive the different particles.

1. Introduction.

We prove a propagation of chaos result for the interacting particle system in $\mathbb{R}^d$ described by the equations

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} K\left(X_t^{i,N} - X_t^{j,N}\right)dt + \sum_{k=1}^{\infty}\sigma_k\left(X_t^{i,N}\right)\circ dB_t^k, \qquad i = 1, \dots, N, \tag{1}$$

where $K, \sigma_k : \mathbb{R}^d \to \mathbb{R}^d$, $k \in \mathbb{N}$, are uniformly Lipschitz continuous and $(B^k)_{k\in\mathbb{N}}$ are independent real-valued Brownian motions on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$; the additional assumption Hypothesis 1 will be imposed on σ_k's, in Section 2. In (1), we chose Stratonovich stochastic integration since the final result, in Stratonovich form and under Hypothesis 1, is more clear and elegant. However, at the price of additional terms, the results hold for the Itô case and under more general assumptions (e.g., time-dependent σ_k); see Section 2.3.

The classical propagation of chaos framework considered in the literature deals with the system

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} K\left(X_t^{i,N} - X_t^{j,N}\right)dt + dW_t^i, \qquad i = 1, \dots, N, \tag{2}$$

Received March 2014; revised March 2015.
MSC2010 subject classifications. 60K35, 82C22, 60K37, 60H15.
Key words and phrases. Interacting particle system, propagation of chaos, mean field limit, Kraichnan noise, Wasserstain metric.

where $(W^i)_{i\in\mathbb{N}}$ are independent $\mathbb{R}^d$-valued Brownian motions; see, for instance, [14]. Unlike this classical case, in (1) *the same* space-dependent delta-correlated-in-time noise $v(t,x)$, formally given by

$$v(t,x) = \sum_{k=1}^{\infty} \sigma_k(x) \frac{dB_t^k}{dt}$$

acts on each particle. This type of space correlated noise was introduced in physics to describe small scale motion in a turbulent fluid, as in the famous Kraichnan model of the sixties. The physical intuition in this case, for equation (1), is that the particles are embedded in a turbulent fluid with velocity $v(t,x)$. Each particle is subject to the transport effect of the fluid and to the motion caused by the interaction with the other particles. Among other examples, we may also think of the case of smoothed point vortices (think of relatively large scale vortex structures in ocean or atmosphere), subject to the transport effect of each other (the interaction) and of a background, small scale, turbulent perturbation. Instead of considering all fluid scales as a whole, described by classical equations of fluid dynamics, one could try, phenomenologically, to separate the large scale vortex structures from the small scale more irregular fluctuations and consider the small scales modeled independently a priori, and the vortices just influencing each other and influenced by the small scales without feedback on small scales. In such an example, to fit with the assumptions of model (1), we have to assume that the interaction between vortices is described by a smoothed Biot–Savart kernel since the singularity of the true Biot–Savart kernel introduces additional difficulties which cannot be handled with the techniques of this paper. On the other hand, the more classical model (2) is more suitable when each particle has its own internal origin of randomness (like certain living organisms) or the external sources of randomness can be considered to be totally uncorrelated at the scale of the particles, like for very light macroscopic particles interacting with the molecules of a gas.

If the covariance of the noise is suitably concentrated (see Hypothesis 1 in Section 2), the random field $v(t,x)$ is poorly space-correlated, except at very short distances, and thus particles which occupy sufficiently distant positions are subject to almost independent noise, a fact that makes the two systems (1) and (2) not so different when the collection of particles is sufficiently sparse.

However, in the limit when $N \to \infty$, the behavior is completely different. Let $(X^i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$ with law μ_0; assume that the families $(B_\cdot^k)_{k\in\mathbb{N}}$ $[(W^i)_{i\in\mathbb{N}}$ for equation (2)] and $(X^i)_{i\in\mathbb{N}}$ are independent and take $X_0^{i,N} = X^i$ as initial conditions for system (1). Denote by S_t^N the empirical measure defined as

$$S_t^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^{i,N}}. \tag{3}$$

The random probability measure S_0^N converges weakly to μ_0 in probability. In both cases of equations (1) and (2), one can prove [cf. [14] for case (2) and the present paper for case (1)] that S_t^N converges weakly, in probability, to a probability measure μ_t. However, in case (2), μ_t is deterministic, the weak convergence of S_t^N to μ_t is understood in probability with respect to both initial conditions and noise, and μ_t is a distributional solution of the nonlinear equation

$$\frac{\partial \mu_t}{\partial t} + \operatorname{div}(b_{\mu_t}\mu_t) = \frac{1}{2}\Delta\mu_t,$$

where, for a generic probability measure ν, the vector field $b_\nu : \mathbb{R}^d \to \mathbb{R}^d$ is defined as

$$b_\nu(x) = \int_{\mathbb{R}^d} K(x-y)\nu(dy).$$

On the contrary, in case (1), μ_t is a random probability measure and, under the particular assumptions of Section 2.1, it satisfies in the distributional sense the stochastic PDE

$$d\mu_t + \operatorname{div}(b_{\mu_t}\mu_t)\,dt + \sum_{k=1}^{\infty} \operatorname{div}(\sigma_k\mu_t) \circ dB_t^k = 0 \tag{4}$$

and the weak convergence of S_t^N to μ_t is understood in probability only with respect to the initial conditions. In Section 2.1, we give the Itô form of this stochastic partial differential equation and in Section 2.3 we show the modifications when we start from (1) in Itô form or when the assumptions on σ_k are more general than those of Section 2.1.

The main result of this paper is the following theorem, by which one can relate the convergence of the empirical measure of the system with the convergence of the empirical measure of the initial conditions.

THEOREM 1. *Let $T > 0$ and assume Hypothesis 1, given in Section 2, on the noise. There exists a constant $\widetilde{C}_T > 0$ such that*

$$\mathbb{E}[W_1(\mu, S_t^N)] \le \widetilde{C}_T \mathbb{E}[W_1(\mu_0, S_0^N)],$$

where W_1 is the Wasserstein distance (see Definition 9).

In Section 4 we give a more precise statement of Theorem 1, as well as a short discussion on recent results on quantitative estimates on the rate of convergence of S_0^N to μ_0 which can be applied in our model.

From Theorem 1 we deduce a *conditional* propagation of chaos result: Conditional to $(B^k)_{k\in\mathbb{N}}$, the particles tend to be independent as $N \to \infty$. One can find other works in literature dealing with conditional propagation of chaos, but referring to different objects and in different contexts. In [2] and [8], the authors treat propagation of chaos conditionally to produce measures on the Kac's sphere and in the latter are given quantitative estimates. In other works, the conditionality is given with respect to the σ-field of the permutable events; see, for example, [15] and [4].

The precise statement about conditional propagation of chaos in this work is given by the following theorem.

THEOREM 2. *Let $\mathcal{F}_t^B$ be the filtration associated to $(B^k)_{k\in\mathbb{N}}$. We suppose that the noise satisfies Hypothesis 1 in both equations (1) and (4). There exists a random measure-valued solution μ_t of equation (4) such that*

$$\lim_{N\to\infty} E\big[\big|\langle S_t^N, \phi\rangle - \langle \mu_t, \phi\rangle\big|\big] = 0$$

for all $\phi \in C_b(\mathbb{R}^d)$.

Moreover, given $r \in \mathbb{N}$ and $\phi_1, \dots, \phi_r \in C_b(\mathbb{R}^d)$, we have

$$\lim_{N\to\infty} E\big[\phi_1\big(X_t^{1,N}\big)\cdots\phi_r\big(X_t^{r,N}\big)|\mathcal{F}_t^B\big] = \prod_{i=1}^r \langle \mu_t, \phi_i\rangle$$

in $L^1(\Omega)$.

In particular, for every $r \in \mathbb{N}$ and $\phi \in C_b(\mathbb{R}^d)$, $\lim_{N\to\infty} E[\phi(X_t^{r,N})|\mathcal{F}_t^B] = \langle \mu_t, \phi\rangle$, namely the conditional law of $X_t^{r,N}$ given $\mathcal{F}_t^B$ converges weakly to μ_t. We can also prove the following.

THEOREM 3. *Given μ_t as in Theorem 2 and $r \in \mathbb{N}$, if X_t is the unique strong solution of the SDE*

$$dX_t = b_{\mu_t}(X_t)\,dt + \sum_{k=1}^\infty \sigma_k(X_t)\,dB_t^k, \qquad X_0 = X_0^r,$$

where the noise satisfies Hypotesis 1, then

$$\lim_{N\to\infty} E\big[\big|X_t^{r,N} - X_t\big|\big] = 0.$$

Moreover, μ_t is a version of the conditional law of X_t with respect to $\mathcal{F}_t^B$, namely

$$\langle \mu_t, \phi\rangle \in \mathbb{E}\big[\phi(X_t)|\mathcal{F}_t^B\big]$$

for every $\phi \in C_b^\infty(\mathbb{R}^d)$.

The result is similar to the case of a *deterministic* environment acting on the particles, which could be modeled by the equations

$$\frac{dX_t^{i,N}}{dt} = \frac{1}{N}\sum_{j=1}^N K\big(X_t^{i,N} - X_t^{j,N}\big) + v\big(t, X_t^{i,N}\big),$$

$$i = 1, \dots, N.$$

As shown by [5], this system satisfies a propagation of chaos property with the limit deterministic inviscid PDE

$$\frac{\partial \mu_t}{\partial t} + \operatorname{div}(b_{\mu_t}\mu_t)\,dt + \operatorname{div}\bigl(v(x)\mu_t\bigr) = 0.$$

Also some technical steps of our proof are strongly inspired by [5]. Moreover, with a different proof and partially a different purpose, some of the technical steps about existence and (especially) stability results for measure-valued stochastic equations have been proved before by [10, 11, 13].

We do not treat here a number of additional interesting questions that are postponed to future works, like: (i) the fact that μ_t should have a density with respect to Lebesgue measure if this is assumed for μ_0; (ii) the uniqueness of solutions to the SPDE (8) (which seems to be true in some class of integrable functions when μ_0 has an integrable density, but it is less clear in spaces of measure-valued solutions); (iii) possible generalizations to non-Lipschitz continuous interation kernel K. In particular, the problem of propagation of chaos for system (1) when $K(x) = \frac{x^\perp}{|x|^2}$, corresponding to point vortices in 2D inviscid fluids, has been posed by [7] and seems to be a challenging question.

In Section 2, we give some information about the settings in which we study the problem. Section 3 is devoted to the study of existence and uniqueness of equation (4) using its Itô version. Finally, in Section 4 we study the convergence and propagation of chaos results.

2. Precise setting of the problem.

2.1. *Assumptions on the noise.* We will now state the assumptions which we will consider on the noise. Recall that $\sigma_k : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field, for every $k \in \mathbb{N}$.

HYPOTHESIS 1. (i) $\sigma_k : \mathbb{R}^d \to \mathbb{R}^d$ *are measurable and satisfy* $\sum_{k=1}^{\infty} |\sigma_k(x)|^2 < +\infty$, *for every* $x \in \mathbb{R}^d$.

(ii) σ_k *is a* C^2 *divergence free vector fields, that is,*

$$\operatorname{div}\sigma_k = 0 \qquad \forall k \geq 1.$$

Define the matrix-valued function $Q : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ *as*

$$Q^{ij}(x,y) := \sum_{k=1}^{\infty} \sigma_k^i(x)\sigma_k^j(y). \tag{5}$$

(iii) *With a little abuse of notation, there exists a function* $Q : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ *such that*:

(a) $Q(x,y) = Q(x-y)$ [*space homogeneity of the random field* $\varphi(t,x) = \sum_{k=1}^{\infty} \sigma_k(x) B_t^k$];

(b) $Q(0) = \mathrm{Id}$;
(c) $Q(\cdot)$ *is of class* C^2 *with second derivatives uniformly bounded in the euclidean norm of* $\mathbb{R}^{d\times d}$, *that is*, $\sup_{x\in\mathbb{R}^d} |\partial^2_{x_i x_j} Q(x)| < +\infty$. *Here, we are using the Hilbert–Schmidt norm on the space of the matrices*.

One can find examples of this model in several references, for example, [3] and [9]. We recall here the most important properties of this type of noise and we give an explicit example.

REMARK 4. Under the previous assumptions, we have

$$\sum_{k=1}^{\infty} \left|\sigma_k(x) - \sigma_k(y)\right|^2 \le L_\sigma^2 |x-y|^2 \qquad \text{for all } x, y \in \mathbb{R}^d \tag{6}$$

for some constant $L_\sigma > 0$. Indeed,

$$\sum_{k=1}^{\infty} \left|\sigma_k(x) - \sigma_k(y)\right|^2 = 2\,\mathrm{Tr}\big(Q(0)\big) - 2\,\mathrm{Tr}\big(Q(x-y)\big).$$

The function $f(z) = \mathrm{Tr}(Q(z))$ has the property $f(-z) = f(z)$, hence from the identity $2f(z) = f(z) + f(-z)$ and Taylor development of both $f(z)$ and $f(-z)$ we get $2f(z) = 2f(0) + \langle D^2 f(0)z, z\rangle + o(|z|^2)$ which implies $\sum_{k=1}^{\infty} |\sigma_k(x) - \sigma_k(y)|^2 \le C_1|x-y|^2$ if $|x-y| \le 1$, for a suitable constant $C_1 > 0$. When $|z| > 1$ we have $f(z) \le C_2|z|^2$ for a suitable constant $C_2 > 0$, because $Q(\cdot)$ has bounded second derivative. Hence, $\sum_{k=1}^{\infty} |\sigma_k(x) - \sigma_k(y)|^2 \le C_2|x-y|^2$ when $|x-y| > 1$. This proves (6) with $L_\sigma^2 = \max(C_1, C_2)$.

It is also important to notice that the covariance function Q can be given first. Indeed Theorem 4.2.5 of [9] states that any matrix valued function $Q : (x, y) \to Q(x, y)$ satisfying (6) can be expressed in the form (5). A very common example of this kind of noise is the isotropic random field, which we present now.

EXAMPLE 5. Let $d \ge 2$ and $f \in L^1(\mathbb{R}_+)$ such that $\int_{\mathbb{R}^d} |y|^2 f(|y|)\,dy < +\infty$. Given $\pi(y)$ a $d \times d$ matrix defined as

$$\pi(y) = (1-p)\mathrm{Id}_d + |y|^{-2}(pd-1)\, y \otimes y \qquad \text{for } y \in \mathbb{R}^d,\ p \in [0,1],$$

we consider

$$Q(x) = \int_{\mathbb{R}^d} e^{iy\cdot x} \pi(y) f\big(|y|\big)\,dy, \qquad x \in \mathbb{R}^d.$$

It is easy to see that property (iii)(a) is satisfied. Property (iii)(c) is true after a renormalization in L^1 of f and (iii)(c) can be verified with a straightforward computation.

REMARK 6. A strong solution of system (1) is a continuous process $(X^{1,N}, \dots, X^{N,N})$, adapted to $(\mathcal{F}_t^B)_{t\geq 0}$, such that

$$P\left(\sum_{k=1}^{\infty}\int_0^T |\sigma_k(X_t^{i,N})|^2\,dt<\infty\right)=1$$

for every $i=1,\dots,N$ (so that the series of stochastic integrals converge in probability) and identity (1) holds in the integral sense. But $\sum_{k=1}^{\infty}|\sigma_k(X_t^{i,N})|^2 = \mathrm{Tr}(Q(0))=d$, hence the sum of stochastic integrals in equation (1) always converges, even in mean square.

2.2. *Itô formulation.* In the Introduction, for the benefit of interpretation, we have formulated the interacting particle system and the limit SPDE both in Stratonovich form. However, for the sake of rigor and mathematical simplicity, it is convenient to work in the corresponding Itô form. Under Hypothesis 1, the interacting particle system in Itô form is

$$dX_t^{i,N}=\frac{1}{N}\sum_{j=1}^{N}K(X_t^{i,N}-X_t^{j,N})\,dt+\sum_{k=1}^{\infty}\sigma_k(X_t^{i,N})\,dB_t^k, \qquad i=1,\dots,N \tag{7}$$

and the SPDE (4) in Itô form is

$$d\mu_t+\mathrm{div}(b_{\mu_t}\mu_t)\,dt+\sum_{k=1}^{\infty}\mathrm{div}(\sigma_k(x)\mu_t)\,dB_t^k=\frac{1}{2}\Delta\mu_t, \tag{8}$$

which will be interpreted in weak form in Definition 11 below. At the rigorous level, these are the equations to which the statements of the Introduction apply. Let us motivate the fact that (7) and (8) correspond to (1) and (4) under Hypothesis 1. This correspondence can be made rigorous but it requires [especially for (4)] proper definitions of solutions and a number of details. If we accept that (1) and (4) are given only for interpretation ad the rigorous setup is given by (7) and (8), an heuristic proof of their equivalence is sufficient. The correspondence between (1) and (7) is due to the fact that the Stratonovich integral $\int_0^t\sigma_k(X_s^{i,N})\circ dB_s^k$ is equal to

$$\int_0^t\sigma_k(X_s^{i,N})\,dB_s^k+\frac{1}{2}\int_0^t(D\sigma_k\cdot\sigma_k)(X_s^{i,N})\,ds$$

(see [9]) where $(D\sigma_k\cdot\sigma_k)_i(x)=\sum_{j=1}^d\sigma_k^j(x)\partial_j\sigma_k^i(x)$. This correction term vanishes thanks to the assumption

$$\mathrm{div}\,\sigma_k=0 \qquad \text{for each } k\in\mathbb{N}$$

[it is natural if we interpret $v(t,x)$ as the velocity field of an incompressible fluid] along with the assumptions on Q made above. Indeed,

$$0 = \left(\sum_{j=1}^{d} \partial_j\right) Q^{ij}(0) = \sum_{k=1}^{\infty} \sum_{j=1}^{d} \partial_j \big(\sigma_k^j(x)\sigma_k^i(x)\big)$$
$$= \sum_{k=1}^{\infty} \sum_{j=1}^{d} \sigma_k^j(x)\partial_j \sigma_k^i(x).$$

Therefore, the Stratonovich and Itô formulations coincide for the interacting particle system.

Let us discuss now the correspondence between (4) and (8). The Stratonovich integral $\int_0^t \operatorname{div}(\sigma_k(x)\mu_s) \circ dB_s^k$ is formally equal to (one should write all terms applied to test functions)

$$\int_0^t \operatorname{div}\big(\sigma_k(x)\mu_s\big)\, dB_s^k - \frac{1}{2}\int_0^t \operatorname{div}\big(\sigma_k(x) \operatorname{div}\big(\sigma_k(x)\mu_s\big)\big)\, ds$$

[the second term, with heuristic language, is initially given by $\frac{1}{2}\int_0^t \operatorname{div}(\sigma_k(x)\, d\langle \mu, B^k\rangle_s)$ where $\langle \mu, B^k\rangle_s$ is the mutual quadratic covariation; then we use again equation (4) to compute $d\langle \mu, B^k\rangle_s$ and get $d\langle \mu, B^k\rangle_s = \operatorname{div}(\sigma_k(x)\mu_s)\, ds$]. Now we see that

$$\begin{aligned} &\sum_{k=1}^{\infty} \operatorname{div}\big(\sigma_k(x) \operatorname{div}\big(\sigma_k(x)\mu_s\big)\big) \\ &\quad = \sum_{\alpha,\beta=1}^{d} \partial_\alpha \partial_\beta \big(Q^{\alpha\beta}(x,x)\mu_s\big) - \operatorname{div}\left(\left(\sum_{k=1}^{\infty} D\sigma_k \cdot \sigma_k\right)\mu_s\right), \end{aligned} \tag{9}$$

where $D\sigma_k \cdot \sigma_k$ is the vector field with components

$$(D\sigma_k \cdot \sigma_k)^\alpha = \sum_{\beta=1}^{d} \big(\partial_\beta \sigma_k^\alpha\big)\sigma_k^\beta.$$

Indeed,

$$\begin{aligned} \sum_{k=1}^{\infty} \operatorname{div}\big(\sigma_k(x) \operatorname{div}\big(\sigma_k(x)\mu_s\big)\big) &= \sum_{k=1}^{\infty} \sum_{\alpha,\beta=1}^{d} \partial_\alpha \big(\sigma_k^\alpha(x)\partial_\beta\big(\sigma_k^\beta(x)\mu_s\big)\big) \\ &= \sum_{k=1}^{\infty} \sum_{\alpha,\beta=1}^{d} \partial_\alpha \partial_\beta \big(\sigma_k^\alpha(x)\sigma_k^\beta(x)\mu_s\big) \\ &\quad - \sum_{k=1}^{\infty} \sum_{\alpha,\beta=1}^{d} \partial_\alpha \big(\big(\partial_\beta \sigma_k^\alpha\big)(x)\sigma_k^\beta(x)\mu_s\big) \end{aligned}$$

and $\sum_{k=1}^{\infty} \sigma_k^\alpha(x)\sigma_k^\beta(x) = Q^{\alpha\beta}(x,x)$. Moreover,

$$\begin{aligned}\sum_{k=1}^{\infty}\bigl(D\sigma_k(x)\cdot\sigma_k(x)\bigr)^\alpha &= \sum_{k=1}^{\infty}\sum_{\beta=1}^{d}\bigl(\partial_\beta\sigma_k^\alpha(x)\bigr)\sigma_k^\beta(x) \\ &= \sum_{\beta=1}^{d}\partial_\beta Q^{\alpha\beta}(x,x) - \sum_{k=1}^{\infty}\sigma_k^\alpha(x)\operatorname{div}\sigma_k(x).\end{aligned} \tag{10}$$

In view of the next section, we stress that until now we have not used Hypothesis 1. Under Hypothesis 1, we have $Q^{\alpha\beta}(x,x) = \delta_{\alpha\beta}$ and $\operatorname{div}\sigma_k = 0$, hence $\sum_{k=1}^{\infty}(D\sigma_k(x)\cdot\sigma_k(x))^\alpha = 0$ for all $\alpha = 1,\dots,d$, and finally

$$\sum_{k=1}^{\infty}\operatorname{div}\bigl(\sigma_k(x)\operatorname{div}\bigl(\sigma_k(x)\mu_s\bigr)\bigr) = \Delta\mu_s.$$

Therefore, the Itô formulation of equation (4) is (8).

2.3. *Extensions and variants*. As we remarked in the Introduction, we chose to work under Hypothesis 1 since it leads to particularly simple and elegant equations and relations between Itô and Stratonovich formulations. However, all the results hold in more general cases, some of which we discuss here.

Assume $u,\sigma_k : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$, $k\in\mathbb{N}$, are measurable vector fields such that, for some constants $C, L > 0$

$$\bigl|u(t,x)\bigr|^2 + \sum_{k=1}^{\infty}\bigl|\sigma_k(t,x)\bigr|^2 \le C\bigl(1+|x|^2\bigr),$$

$$\bigl|u(t,x)-u(t,y)\bigr|^2 + \sum_{k=1}^{\infty}\bigl|\sigma_k(t,x)-\sigma_k(t,y)\bigr|^2 \le L|x-y|^2$$

for all $x,y\in\mathbb{R}^d$ and all $t\in[0,T]$. Under these conditions, always with K Lipschitz continuous, consider the system of equations in Itô form

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} K\bigl(X_t^{i,N}-X_t^{j,N}\bigr)\,dt + u\bigl(t,X_t^{i,N}\bigr)\,dt + \sum_{k=1}^{\infty}\sigma_k\bigl(t,X_t^{i,N}\bigr)\,dB_t^k, \qquad i=1,\dots,N. \tag{11}$$

Set

$$\begin{aligned}Q_t^{\alpha\beta}(x,y) &:= \sum_{k=1}^{\infty}\sigma_k^\alpha(t,x)\sigma_k^\beta(t,y),\\ a^{\alpha\beta}(t,x) &:= Q_t^{\alpha\beta}(x,x).\end{aligned}$$

All results of the present paper hold true in this case with the corresponding SPDE given by

$$
\begin{aligned}
d\mu_t + \operatorname{div}\big((b_{\mu_t} + u)\mu_t\big)\,dt + \sum_{k=1}^{\infty} \operatorname{div}(\sigma_k \mu_t)\,dB_t^k \\
= \frac{1}{2}\sum_{\alpha,\beta=1}^{d} \partial_\alpha \partial_\beta\big(a^{\alpha\beta}(t,\cdot)\mu_t^N\big)\,dt
\end{aligned}
\tag{12}
$$

(to be interpreted in weak form similar to Definition 11 below). The connection between these two equations can be seen informally in a few lines by applying Itô formula to $\phi(X_t^{i,N})$, with $\phi \in C_c^\infty(\mathbb{R}^d)$; the result is that S_t^N satisfies

$$
\begin{aligned}
d\langle S_t^N, \phi\rangle = \langle S_t^N, \nabla\phi \cdot (b_{\mu_t} + u)\rangle\,dt + \sum_{k=1}^{\infty}\langle S_t^N, \nabla\phi\cdot\sigma_k(t,\cdot)\rangle\,dB_t^k \\
+ \left\langle S_t^N, \frac{1}{2}\sum_{\alpha,\beta=1}^{d} a^{\alpha\beta}(t,\cdot)\partial_\alpha\partial_\beta\phi\right\rangle dt,
\end{aligned}
$$

which is the weak formulation of the SPDE (12) above.

REMARK 7. Assuming a suitable differentiability of $\sigma_k(t,\cdot)$ in the t variable, we may rewrite the SPDE (12) in Stratonovich form. We keep this remark at heuristic level, to avoid unnecessary details. As in the previous section, the Stratonovich integral $\int_0^t \operatorname{div}(\sigma_k(s,x)\mu_s) \circ dB_s^k$ is equal to the Itô integral $\int_0^t \operatorname{div}(\sigma_k(s,x)\mu_s)\,dB_s^k$ plus the correction term

$$
\tfrac{1}{2}\big[\operatorname{div}\big(\sigma_k(\cdot,x)\mu_\cdot\big), B_\cdot^k\big]_t. \tag{13}
$$

Now, $\sigma_k(t,x)\mu_t$ formally satisfies the identity (by Itô's formula)

$$
d\big(\sigma_k(t,x)\mu_t\big) = \frac{\partial \sigma_k}{\partial t}(t,x)\mu_t\,dt + \sigma_k(t,x)\,d\mu_t
$$

hence only the term

$$
-\sigma_k(t,x)\sum_{k'=1}^{\infty}\operatorname{div}\big(\sigma_{k'}(t,x)\mu_t\big)\,dB_t^{k'}
$$

contributes to the quadratic covariation (13), which is thus equal (as in the previous section) to

$$
-\frac{1}{2}\int_0^t \operatorname{div}\big(\sigma_k(s,x)\operatorname{div}\big(\sigma_k(s,x)\mu_s\big)\big)\,ds.
$$

From identity (9), where now $Q^{\alpha\beta}(x,x)$ is replaced by $a^{\alpha\beta}(t,x)$, we get that μ_t satisfies (in weak form) the Stratonovich equation

$$
d\mu_t = -\operatorname{div}\big((b_{\mu_t}+u)\mu_t\big)\,dt - \sum_{k=1}^{\infty}\operatorname{div}\big(\sigma_k(t,\cdot)\mu_t\big)\circ dB_t^k + \mathcal{D}(t,\cdot)\mu_t\,dt, \tag{14}
$$

where the first-order differential operator $\mathcal{D}(t,x)$ is given by

$$\mathcal{D}f := \frac{1}{2}\operatorname{div}\left(\sum_{k=1}^{\infty} D\sigma_k \cdot \sigma_k f\right).$$

REMARK 8. The Stratonovich reformulation (14) reveals that the true nature of the SPDE (12) is not parabolic but of a first-order equation, informally speaking of hyperbolic type.

If we start from the beginning with the Stratonovich equation,

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} K\left(X_t^{i,N} - X_t^{j,N}\right)dt + u\left(t, X_t^{i,N}\right)dt + \sum_{k=1}^{\infty}\sigma_k\left(t, X_t^{i,N}\right)\circ dB_t^k$$

in place of (11), we may rewrite it in the Itô form

$$\begin{aligned} dX_t^{i,N} &= \frac{1}{N}\sum_{j=1}^{N} K\left(X_t^{i,N} - X_t^{j,N}\right)dt + u\left(t, X_t^{i,N}\right)dt \\ &\quad + \sum_{k=1}^{\infty}\sigma_k\left(t, X_t^{i,N}\right)dB_t^k + \frac{1}{2}\sum_{k=1}^{\infty}(D\sigma_k \cdot \sigma_k)\left(t, X_t^{i,N}\right)dt, \end{aligned}$$

where $(D\sigma_k \cdot \sigma_k)^\alpha = \sum_{\beta=1}^{d} \partial_\beta \sigma_k^\alpha \sigma_k^\beta$. This is the case because the correction term of the α-component is

$$\frac{1}{2}\sum_{k=1}^{\infty} d\left[\sigma_k^\alpha\left(\cdot, X_\cdot^{i,N}\right), B_t^k\right]_t = \frac{1}{2}\sum_{k=1}^{\infty}\nabla\sigma_k^\alpha\left(t, X_t^{i,N}\right)\cdot\sigma_k\left(t, X_t^{i,N}\right)dt$$

since, under suitable differentiability assumptions on σ_k, we may apply Itô's formula to $\sigma_k^\alpha(t, X_t^{i,N})$ and see that for the quadratic covariation $[\sigma_k^\alpha(\cdot, X_\cdot^{i,N}), B_t^k]_t$ only the following term [part of $\nabla\sigma_k^\alpha(t, X_t^{i,N}) \cdot dX_t^{i,N}$] matters:

$$\nabla\sigma_k^\alpha\left(t, X_t^{i,N}\right)\cdot\sum_{k'=1}^{\infty}\sigma_{k'}\left(t, X_t^{i,N}\right)dB_t^{k'}.$$

Thus we see that under appropriate regularity and summability (in k) properties on σ_k, we may transform the Stratonovich equation into the Itô one (11) and apply the previous result. The additional drift

$$\frac{1}{2}\sum_{k=1}^{\infty}(D\sigma_k \cdot \sigma_k)(t,x) \tag{15}$$

appears in the Itô formulation.

Finally, we have seen that two annoying correction terms appear in the computations above, namely $\mathcal{D}(t,\cdot)\mu_t$ in the SPDE (14) and the additional drift (15).

Both are related to passages from Itô to Stratonovich forms. Both of them are equal to zero if we assume

$$\sum_{k=1}^{\infty} D\sigma_k \cdot \sigma_k = 0.$$

Similar to (10), this can be rewritten as

$$\sum_{\beta=1}^{d} \partial_\beta a^{\alpha\beta}(t,x) - \sum_{k=1}^{\infty} \sigma_k^\alpha \operatorname{div} \sigma_k = 0.$$

A sufficient condition thus is the pair of assumptions

$$\begin{aligned} &a^{\alpha\beta}(t,x) \qquad \text{independent of } x, \\ &\operatorname{div} \sigma_k = 0 \qquad \text{for every } k, \end{aligned}$$

which are part of Hypothesis 1.

2.4. *Some definitions*. Recall the definition of the empirical measure $S_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}$, which can be used, as we did in the Introduction, to rewrite the drift coefficient as $b_{S_t^N}(x) = K * S_t^N(x) = \frac{1}{N}\sum_{j=1}^N K(x - X_t^{j,N})$. We can thus write equation (7), for $i = 1, \dots, N$, as

$$dX_t^{i,N} = b_{S_t^N}(X_t^{i,N})\,dt + \sum_{k=1}^{\infty} \sigma_k(X_t^{i,N})\,dB_t^k.$$

If we take a test function $\phi \in C_b^2(\mathbb{R}^d)$ and we apply Itô's formula, from the assumptions on Q it follows, for $i = 1, \dots, N$,

$$\begin{aligned} d\phi(X_t^{i,N}) = {}& \left[\nabla\phi(X_t^{i,N}) \cdot b_{S_t^N}(X_t^{i,N}) + \frac{1}{2}\Delta\phi(X_t^{i,N})\right] dt \\ &+ \sum_{k=1}^{\infty} \nabla\phi(X_t^{i,N}) \cdot \sigma_k(X_t^{i,N})\,dB_t^k, \end{aligned}$$

which becomes, adding over N and dividing by N,

$$\langle S_t^N, \phi\rangle = \left[\langle S_t^N, \nabla\phi \cdot b_{S_t^N}\rangle + \frac{1}{2}\langle S_t^N, \Delta\phi\rangle\right] dt + \sum_{k=1}^{\infty} \langle S_t^N, \nabla\phi \cdot \sigma_k\rangle\,dB_t^k.$$

Hence, S_t^N is a measure-valued solution of equation (4), in the sense of Definition 11 below.

We define now the space over which we will study equation (4).

DEFINITION 9. $(\mathcal{P}_1(\mathbb{R}^d), W_1)$ is the space of probability measures μ_0 on $\mathbb{R}^d$ with finite first moment, that is,

$$\|\mu_0\| := \int_{\mathbb{R}^d} d\mu_0 = 1, \qquad M_1(\mu_0) := \int_{\mathbb{R}^d} |x|\, d\mu_0(x) < \infty$$

endowed with the 1-Wasserstein metric defined as

$$W_1(\nu_0, \mu_0) = \inf_{m \in \Gamma(\mu_0, \nu_0)} \int_{\mathbb{R}^{2d}} |x - y| m(dx, dy), \qquad \mu_0, \nu_0 \in \mathcal{P}_1(\mathbb{R}^d).$$

Here, $\Gamma(\mu_0, \nu_0)$ is the set of the finite measures on $\mathbb{R}^{2d}$ with first and second marginals equal respectively to μ_0 and ν_0, namely

$$\Gamma(\mu_0, \nu_0) = \{m \in \mathcal{P}_1(\mathbb{R}^{2d}) : m(A \times \mathbb{R}^d) = \mu_0(A), m(\mathbb{R}^d \times A) = \nu_0(A), \forall A \in \mathcal{B}(\mathbb{R}^d)\}.$$

$\mathcal{S}$ will be the space of the stochastic processes taking values on $(\mathcal{P}_1(\mathbb{R}^d), W_1)$,

$$\mu : [0, T] \times \Omega \to \mathcal{P}_1(\mathbb{R}^d)$$

such that $\mathbb{E}[\sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|\, d\mu_t(x)] < \infty$ and $\langle \mu_t, \phi \rangle$ is $\mathcal{F}_t$-adapted for every test function $\phi \in C_b^\infty(\mathbb{R}^d)$. We endow $\mathcal{S}$ with the following distance:

$$d_{\mathcal{S}}(\mu, \nu) := \mathbb{E}\Big[\sup_{t\in[0,T]} W_1(\mu_t, \nu_t)\Big],$$

where $\mu = (\mu_t)_{t\in[0,T]}, \nu = (\nu_t)_{t\in[0,T]} \in \mathcal{S}$.

REMARK 10. The metric space $(\mathcal{P}_1(\mathbb{R}^d), W_1)$ has been well studied in optimal transportation theory and extensive results on it can be found in the literature, (see, e.g., [1]). In particular, this space is complete and separable (Proposition 7.1.5 of [1]). Hence, follows from standard arguments that $(\mathcal{S}, d_{\mathcal{S}})$ is also a complete metric space.

HYPOTHESIS 2. *Concerning the initial condition $\mu_0 : \Omega \to \mathcal{P}_1(\mathbb{R}^d)$ of equation* (4) *we shall always assume that*:

(i) μ_0 *is $\mathcal{F}_0$-measurable*;
(ii) $\mathbb{E}[\int_{\mathbb{R}^d} |x|\, d\mu_0(x)] < \infty$.

For every μ_0 that satisfies the previous hypothesis, we call $\mathcal{S}_{\mu_0}$ the set of $\mu \in \mathcal{S}$ such that $\mu|_{t=0} = \mu_0$.

DEFINITION 11. A family $\{\mu_t(\omega); t \geq 0, \omega \in \Omega\}$ of random probability measures taking value in $\mathcal{P}_1(\mathbb{R}^d)$ is a measure-valued solution of equation (4) if:

(i) for all $\phi \in C_b(\mathbb{R}^d)$, $\langle \mu_t, \phi \rangle$ is an adapted process with a continuous version,

(ii) for all $\phi \in C_b^2(\mathbb{R}^d)$

$$\langle \mu_t, \phi \rangle = \langle \mu_0, \phi \rangle + \int_0^t \langle \mu_s, b_{\mu_s} \cdot \nabla \phi \rangle \, ds + \frac{1}{2} \int_0^t \langle \mu_s, \Delta \phi \rangle \, ds$$
$$+ \sum_{k=1}^{\infty} \int_0^t \langle \mu_s, \sigma_k \cdot \nabla \phi \rangle \, dB_s^k.$$

REMARK 12. Notice that the infinite sum in the previous equation converges under our assumptions. Indeed, if $\phi \in C_b^2(\mathbb{R}^d)$, it holds, by Itô isometry and Jensen inequality,

$$\mathbb{E}\left[\left|\sum_{k=1}^{\infty} \int_0^t \langle \mu_s, \sigma_k \cdot \nabla \phi \rangle \, dB_s^k\right|^2\right] = \mathbb{E}\left[\sum_{k=1}^{\infty} \int_0^t \langle \mu_s, \sigma_k \cdot \nabla \phi \rangle^2 \, ds\right]$$
$$\leq \mathbb{E}\left[\sum_{k=1}^{\infty} \int_0^t \langle \mu_s, |\sigma_k \cdot \nabla \phi|^2 \rangle \, ds\right].$$

Now, by the assumptions on σ_k, we have

$$\sum_{k=1}^{\infty} \left|\sigma_k(x) \cdot \nabla \phi(x)\right|^2 \leq \sum_{k=1}^{\infty} \left|\nabla \phi(x)\right|^2 \left|\sigma_k(x)\right|^2 = \left|\nabla \phi(x)\right|^2 \sum_{k=1}^{\infty} \left|\sigma_k(x)\right|^2$$
$$\leq C \left|\nabla \phi(x)\right|^2 < +\infty.$$

3. Well posedness of the stochastic PDE. In this chapter, we study the well posedness of equation (4), and thus we prove the following.

THEOREM 13. *Let $T \geq 0$ and $\mu_0 : \Omega \to \mathcal{P}_1(\mathbb{R}^d)$ be as in Hypothesis 2. There exists a unique solution $\mu = (\mu_t)_{t \in [0,T]}$ of equation (4) in the sense of Definition 11 starting from μ_0 and defined up to time T, that can be seen as the only fixed point of the operator (27) defined below.*

We have already seen that the empirical measure S_t^N defined in (3) satisfies in the distributional sense (4) for every test function ϕ, moreover it is a probability measure with finite first moment and the process

$$\langle S_t^N, \phi \rangle = \frac{1}{N} \sum_{i=1}^{N} \phi(X_t^{i,N})$$

is $\mathcal{F}_t$-adapted. This is true since the processes $X_t^{i,N}$ are solutions of the SDE (7), and hence are adapted and continuous. Hence, the empirical measure S_t^N satisfies (4) in the sense of Definition 11.

3.1. *Stochastic Liouville equation.* In order to investigate the solutions of equation (4), we first want to study what happens when the drift coefficient does not depend on the solution but it is instead a priori defined (but random). We hence consider the following stochastic differential equation:

$$
\begin{aligned}
dX_t &= b(t, X_t)\,dt + \sum_k^{\infty} \sigma_k(X_t)\,dB_t^k, \\
X_0 &= x \in \mathbb{R}^d,
\end{aligned}
\tag{16}
$$

where the σ_k's are defined as before. Here, $b = b(t, x, \omega)$ is an $\mathcal{F}_t$-adapted process, continuous in (t, x), which satisfies:

- b Lipschitz continuous in x uniformly in (t, ω), with Lipschitz constant L_b, not depending on ω and t, that is,

$$
\bigl|b(t, x, \omega) - b(t, y, \omega)\bigr| \le L_b|x - y| \qquad \forall x, y \in \mathbb{R}^d, \forall t \in \mathbb{R}, \mathbb{P}\text{-a.s.}
$$

- For every fixed w, b has linear growth in x uniformly t, that is,

$$
\bigl|b(t, x, \omega)\bigr| \le c_1|x| + c_2(\omega) \qquad \forall x \in \mathbb{R}^d, \forall t \in \mathbb{R}, \text{ for } \mathbb{P}\text{-a.e. } \omega,
$$

  where $c_1 \in \mathbb{R}$ and $c_2(\omega)$ is a random variable such that $\mathbb{E}[|c_2(\omega)|] < \infty$.

By classical results on SDEs (see, e.g., [9]), this equation admits a unique solution $X_t = X(t, x, \omega)$ which is continuous in time. Moreover, taking into account the following lemma, it follows from Kolmogorov continuity theorem that there exists a modification of $X(t, x)$ which is continuous in x. It is also jointly continuous in (t, x) by Kolmogorov theorem for processes taking values in Banach spaces, precisely in the space $C([0, T]; \mathbb{R}^d)$. This results on continuity of the stochastic flow of equation (16) can be found in the literature as in [9]. However, we want to stress in the following the dependence on the different parameters and outline more explicitly the constants.

We define now some constants depending on the coefficients b and σ_k of the problem, which we will use in the following results. For a fixed real number $p \ge 1$, we call C_p the constat which appears in the Burkholder–Davis–Gundy theorem. Moreover, for $t > 0$ and $p \ge 1$, we define

$$
C(p, t) := C_p T^{1/(2p)} L_\sigma + T^{1/p} L_b. \tag{17}
$$

Finally, for a fixed $T > 0$, let $n \in \mathbb{N}$ be the minimum such that $C(p, (T/n)) < 1$, so that we can define

$$
C_{p,T} := \bigl(1 - C\bigl(p, (T/n)\bigr)\bigr)^{-np}. \tag{18}
$$

From our choice of $n \in \mathbb{N}$, this last constant is well defined and depending only on T, p and the coefficients of problem (16).

LEMMA 14. *Let $p \geq 1$, $T \geq 0$ and let $X(t,x)$ be a solution of equation* (16) *up to time T. Then*

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|X(t,x)-X(t,x')|^p\big|\mathcal{F}_0\Big] \leq C_{p,T}|x-x'|^p, \tag{19}$$

where the constant $C_{p,T}$ is defined in (18).

PROOF. Let $n \in \mathbb{N}$ be the minimum such that $C(p,(T/n)) < 1$, where $C(\cdot,\cdot)$ is defined in (17). Now we divide the temporal interval $[0,T]$ in n subintervals. We set $X^{(0)}(t,x) = x$ and we call $X^{(m)}$, for $m = 1,\dots,n$, the solution to

$$dX_t = b(t,X_t)\,dt + \sum_k^{\infty}\sigma_k(X_t)\,dB_t^k,$$
$$X_{((m-1)/n)T} = X^{(m-1)}\left(\frac{m-1}{n}T, x\right)$$

on the interval $[\frac{m-1}{n}T, \frac{m}{n}T]$. We prove by induction that, for every $m = 1,\dots,n$,

$$\begin{aligned} &E\Big[\sup_{t\in[((m-1)/n)T,(m/n)T]}|X^{(m)}(t,x)-X^{(m)}(t,x')|^p\big|\mathcal{F}_0\Big]^{1/p} \\ &\quad \leq \frac{|x-x'|}{(1-C(p,(T/n)))^m}. \end{aligned} \tag{20}$$

It follows from the uniqueness of solution of the stochastic differential equations that the solution X_t of equation (16) coincides on each interval $[\frac{m-1}{n}T, \frac{m}{n}T]$ with the process $X_t^{(m)}$. The thesis follows noting that the worst constant in (20) appears when $m = n$ and it coincides with $C_{p,t}$.

Step 1. Now we prove (20) for $m = 1$. By a triangular inequality, we get

$$\begin{aligned} &\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}|X^{(1)}(t,x)-X^{(1)}(t,x')|^p\big|\mathcal{F}_0\Big]^{1/p} \\ &\quad \leq |x-x'| \\ &\qquad + \mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\Big|\int_0^t b(s,X^{(1)}(s,x)) - b(s,X^{(1)}(s,x'))\,ds\Big|^p\Big|\mathcal{F}_0\Big]^{1/p} \\ &\qquad + \mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\Big|\int_0^t \sum_k \sigma_k(X^{(1)}(s,x)) - \sigma_k(X^{(1)}(s,x'))\,dB_s^k\Big|^p\Big|\mathcal{F}_0\Big]^{1/p}. \end{aligned}$$

In order to estimate this, we first notice that, by the Lipschitz continuity of b one can get

$$\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\Big|\int_0^t b\big(s, X^{(1)}(s,x)\big) - b\big(s, X^{(1)}(s,x')\big)\,ds\Big|^p\Big|\mathcal{F}_0\Big]^{1/p}$$
$$\le \big((T/n)\big)^{1/p} L_b \mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\big|X^{(1)}(t,x) - X^{(1)}\big(t,x'\big)\big|^p\big|\mathcal{F}_0\Big]^{1/p}.$$

Now, using the conditional Burkholder–Davis–Gundy inequality (Proposition 27), we obtain

$$\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\Big|\int_0^t \sum_k \sigma_k\big(X^{(1)}(s,x)\big) - \sigma_k\big(X^{(1)}(s,x')\big)\,dB_s^k\Big|^p\Big|\mathcal{F}_0\Big]^{1/p}$$
$$\le C_p\mathbb{E}\Big[\Big(\int_0^{(T/n)} \sum_k\big|\sigma_k\big(X^{(1)}(s,x)\big) - \sigma_k\big(X^{(1)}(s,x')\big)\big|^2\,ds\Big)^{p/2}\Big|\mathcal{F}_0\Big]^{1/p}$$
$$\le C_p\big((T/n)\big)^{1/(2p)} L_\sigma \mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\big|X^{(1)}(t,x) - X^{(1)}\big(t,x'\big)\big|^p\big|\mathcal{F}_0\Big]^{1/p}.$$

We have hence proved the base step of the induction.

Step 2. Now we suppose (20) true for m and we prove it for $m+1$. First, thanks to a triangular inequality we obtain

$$\mathbb{E}\Big[\sup_{t\in[(m/n)T,((m+1)/n)T]}\big|X^{(m+1)}(t,x) - X^{(m+1)}\big(t,x'\big)\big|^p\big|\mathcal{F}_0\Big]^{1/p}$$
$$\le \mathbb{E}\Big[\Big|X^{(m)}\Big(\frac{m}{n}T,x\Big) - X^{(m)}\Big(\frac{m}{n}T,x'\Big)\Big|^p\Big|\mathcal{F}_0\Big]^{1/p}$$
$$+ \mathbb{E}\Big[\sup_{t\in[(m/n)T,((m+1)/n)T]}\Big|\int_{(m/n)T}^t b\big(s, X^{(m+1)}(s,x)\big)$$
$$- b\big(s, X^{(m+1)}(s,x')\big)\,ds\Big|^p\Big|\mathcal{F}_0\Big]^{1/p}$$
$$+ \mathbb{E}\Big[\sup_{t\in[(m/n)T,((m+1)/n)T]}\Big|\int_{(m/n)T}^t \sum_k \sigma_k\big(X^{(m+1)}(s,x)\big)$$
$$- \sigma_k\big(X^{(m+1)}(s,x')\big)\,dB_s^k\Big|^p\Big|\mathcal{F}_0\Big]^{1/p.}$$

Now, as in step 1, we use the Lipschitz property of b and σ_k and Lemma 27 to get

$$\mathbb{E}\Big[\sup_{t\in[(m/n)T,((m+1)/n)T]}\big|X^{(m+1)}(t,x) - X^{(m+1)}\big(t,x'\big)\big|^p\big|\mathcal{F}_0\Big]^{1/p}$$
$$\le \frac{\mathbb{E}[|X^{(m)}((m/n)T,x) - X^{(m)}((m/n)T,x')|^p|\mathcal{F}_0]^{1/p}}{(1-C(p,(T/n)))^p}.$$

Now estimate the right-hand side using (20) for m, and we conclude this last step,

$$\mathbb{E}\Big[\sup_{t\in[(m/n)T,((m+1)/n)T]}\big|X^{(m+1)}(t,x)-X^{(m+1)}(t,x')\big|^p\big|\mathcal{F}_0\Big]^{1/p} \le \frac{|x-x'|}{(1-C(p,(T/n)))^m}\frac{1}{(1-C(p,(T/n)))}. \qquad \square$$

Using the continuous version in x of the solution of equation (16), we are going to define a solution for equation (4) in the case in which the drift coefficient is fixed. This is shown in the following proposition. The push forward described in the next statement has to be understood ω-wise: for a.e. ω and for each $t \in [0,T]$, we take the initial measure $\mu_0(\omega)=\mu_0(\omega,dx)$ and we consider its image measure (or push forward) under the continuous map $x \mapsto X(t,x,\omega)$, denoted by $\mu_t(\omega)$ or $\mu_t(\omega,dx)$.

PROPOSITION 15. *Given μ_0 which satisfies Hypotesis 2, the push forward of μ_0 with respect to the solution of* (16) *namely*

$$\mu_t(\omega)=X(t,\cdot,\omega)_{\#}\mu_0(\omega)$$

solves the following equation in the sense of Definition 11:

$$\begin{cases} d\mu_t = -\operatorname{div}(b\mu_t)\,dt - \sum_{k=0}^{\infty}\operatorname{div}(\sigma_k\mu_t)\,dB_t^k + \frac{1}{2}\Delta\mu_t, \\ \mu_{t|_{t=0}} = \mu_0. \end{cases}$$

PROOF. First, notice that $\mu \in \mathcal{S}$. By definition, for every $t \in [0,T]$ and $\mathbb{P}$-a.s., μ_t is a finite and positive measure. We show that the first moment of μ_t is finite,

$$\begin{aligned} \mathbb{E}\Big[\int_{\mathbb{R}^d}|x|\,d\mu_t(x)\Big] &= \mathbb{E}\Big[\int_{\mathbb{R}^d}|X_t|\,d\mu_0(x)\Big] \\ &\le \mathbb{E}\Big[\int_{\mathbb{R}^d}|x|\,d\mu_0(x)\Big] \qquad (21)\\ &\quad + \mathbb{E}\Big[\int_{\mathbb{R}^d}\int_0^t\big|b\big(X(s,x)\big)\big|\,ds\,d\mu_0(x)\Big] \qquad (22)\\ &\quad + \mathbb{E}\Big[\int_{\mathbb{R}^d}\Big|\int_0^t\sum_k^{\infty}\sigma_k\big(X(s,x)\big)\,dB_s^k\Big|\,d\mu_0(x)\Big]. \qquad (23) \end{aligned}$$

It follows from the choice of μ_0 that (21) is finite. We can bound (22) if we notice that the Lipschitz continuity assumption on b implies $|b(x)| \le 1+|x|$, which gives

$$\mathbb{E}\Big[\int_{\mathbb{R}^d}\int_0^t\big|b\big(X(s,x)\big)\big|\,ds\,d\mu_0(x)\Big] \le CT + C\int_0^t\mathbb{E}\Big[\int_{\mathbb{R}^d}|x|\,d\mu_s(x)\Big]\,ds. \qquad (24)$$

In order to bound (23), we use Propositions 28 and 27 and we do the following:

$$
\begin{aligned}
&\mathbb{E}\biggl[\mathbb{E}\biggl[\int_{\mathbb{R}^d}\biggl|\int_0^t\sum_{k=0}^{\infty}\sigma_k\bigl(X(s,x)\bigr)\,dB_s^k\biggr|\,d\mu_0(x)\Big|\mathcal{F}_0\biggr]\biggr]\\
&\quad=\mathbb{E}\biggl[\int_{\mathbb{R}^d}\mathbb{E}\biggl[\biggl|\int_0^t\sum_{k=0}^{\infty}\sigma_k\bigl(X(s,x)\bigr)\,dB_s^k\biggr|\Big|\mathcal{F}_0\biggr]\,d\mu_0(x)\biggr]\\
&\quad\le C\mathbb{E}\biggl[\int_{\mathbb{R}^d}\mathbb{E}\biggl[\int_0^t\sum_{k=0}^{\infty}\bigl|\sigma_k\bigl(X(s,x)\bigr)\bigr|^2\,ds\Big|\mathcal{F}_0\biggr]^{1/2}\,d\mu_0(x)\biggr]\\
&\quad\le C\sqrt{T}.
\end{aligned}
\tag{25}
$$

Here, we used $\sum_{k=0}^{\infty}|\sigma_k(X(s,x))|^2<+\infty$. Taking into account (24) and (25), we can apply the Gronwall lemma to deduce that the first moment of μ_t is finite for every t. Let us stress a detail. In order to apply Proposition 28 of the Appendix, we need to know that the random field (t here is fixed)

$$
f(x)=\int_0^t\sum_{k=0}^{\infty}\sigma_k\bigl(X(s,x)\bigr)\,dB_s^k
$$

is continuous, or it has a continuous modification. This is true because by the BDG inequality,

$$
\begin{aligned}
E\bigl[\bigl|f(x)-f(y)\bigr|^p\bigr]&=E\biggl[\biggl|\int_0^t\sum_{k=0}^{\infty}\bigl(\sigma_k\bigl(X(s,x)\bigr)-\sigma_k\bigl(X(s,y)\bigr)\bigr)\,dB_s^k\biggr|^p\biggr]\\
&\le C_pE\biggl[\biggl(\int_0^t\sum_{k=0}^{\infty}\bigl|\sigma_k\bigl(X(s,x)\bigr)-\sigma_k\bigl(X(s,y)\bigr)\bigr|^2\,ds\biggr)^{p/2}\biggr]\\
&\le C_pL_\sigma^pE\biggl[\biggl(\int_0^t\bigl|X(s,x)-X(s,y)\bigr|^2\,ds\biggr)^{p/2}\biggr]\\
&\le C_{p,T}C_pL_\sigma^pT|x-y|^p.
\end{aligned}
$$

This last inequality follows from Lemma 14. Thus, for $p>d$ we may apply Kolmogorov regularity theorem and deduce that f has a continuous version.

We show now that μ_t satisfies the conditions of Definition 11:

(i) to prove that $\langle\mu_t,\phi\rangle$ is continuous and adapted, it is sufficient to notice that

$$
\langle\mu_t,\phi\rangle=\int_{\mathbb{R}^d}\phi\bigl(X(t,x)\bigr)\mu_0(dx).
$$

(ii) Let $\phi\in C_b^2(\mathbb{R}^d)$, we apply Itô's formula

$$
d\phi(X_t)=\nabla\phi(X_t)\cdot dX_t+\frac{1}{2}\sum_{k}^{\infty}\sum_{i,j=1}^{d}\partial_{i,j}^2\phi(X_t)\sigma_k^i(X_t)\sigma_k^j(X_t)\,dt.
$$

Under the homogeneity assumption over σ_k, we obtain the following:

$$d\phi(X_t) = \left[\nabla\phi(X_t) \cdot b(X_t) + \frac{1}{2}\Delta\phi(X_t)\right] dt + \sum_k^\infty \nabla\phi(X_t)\sigma_k(X_t)\, dB_t^k.$$

Integrating now over μ_0, we get

$$d\langle\mu_t, \phi\rangle = \left[\langle\mu_t, \nabla\phi \cdot b\rangle + \frac{1}{2}\langle\mu_t, \Delta\phi\rangle\right] dt + \sum_k^\infty \int_{\mathbb{R}^d} \nabla\phi(X_t)\sigma_k(X_t)\, dB_t^k\, d\mu_0.$$

Using the stochastic Fubini's theorem, we interchange the stochastic integral and the integral in μ_0 and we obtain the desired equation. □

3.2. *The contraction mapping*. In this section, we will construct a solution of equation (4) by means of a fixed-point argument. Given $\mu_0 : \Omega \to \mathcal{P}_1(\mathbb{R}^d)$ as in Hypothesis 2, we define now an operator $\Phi_{\mu_0} : \mathcal{S} \to \mathcal{S}$. In Theorem 17, we prove that it is a contraction and we see that his unique fixed point is a solution to (4).

Let $\mu = (\mu_t)_{t\in[0,T]} \in \mathcal{S}$. We define the following as the convolution between μ_t and K:

$$b_\mu(t, x, \omega) := \int_{\mathbb{R}^d} K(x - y)\mu_t(\omega, dy).$$

Notice that $b_\mu(t, \cdot, \omega)$ is Lipschitz continuous with Lipschitz constant L_K, which is the Lipschitz constant of K and does not depend on t and ω. Moreover, since $|K(x)| \le L_K(K(0) + |x|)$,

$$\begin{aligned} |b_\mu(t, 0, \omega)| &\le \int_{\mathbb{R}^d} |K(-y)|\mu_t(\omega, dy) \le L_K \int_{\mathbb{R}^d} (K(0) + |y|)\mu_t(\omega, dy) \\ &\le L_K K(0) + L_K \int_{\mathbb{R}^d} |x|\mu_t(\omega, dx) \end{aligned}$$

and the random variable $\int_{\mathbb{R}^d} |x|\mu_t(\omega, dx)$ is integrable. Hence, b_μ satisfies the assumptions required in Section 3 to have strong existence and uniqueness of solutions. Let now X_t^μ be the solution to equation (16) with drift coefficient b_μ, namely

$$\begin{aligned} dX_t &= b_\mu(X_t)\, dt + \sum_k \sigma_k(X_t)\, dB_t^k, \\ X_0 &= x. \end{aligned} \tag{26}$$

Let $X^\mu(t, x, \omega)$ be a modification of X_t^μ continuous in x. We define, for every t,

$$(\Phi_{\mu_0}\mu)_t(\omega) := X^\mu(t, \cdot, \omega)_\# \mu_0(\omega), \qquad \omega\text{-a.s.} \tag{27}$$

REMARK 16. Notice that the range of Φ_{μ_0} is included in $\mathcal{S}_{\mu_0}$ and that $\Phi_{\mu_0}\mu$ is a solution of equation (4) in the sense of Definition 11, thanks to Proposition 15.

From Lemma 19 and Proposition 15, we deduce the following theorem, which is the main result of this section.

THEOREM 17. *Given $T > 0$, the operator Φ_{μ_0} has a unique fixed point $\mu = \{\mu_t\}_{t\in[0,T]}$ in $\mathcal{S}_{\mu_0}$. This fixed point is a solution of equation* (4).

PROOF. From Lemma 19, we have

$$d_{\mathcal{S}}(\Phi_{\mu_0}\mu, \Phi_{\mu_0}\nu) \le \gamma_T d_{\mathcal{S}}(\mu, \nu) \qquad \forall \mu, \nu \in \mathcal{S},$$

where γ_T is defined in (28) as $\gamma_T := L_K T C_{1,T}$. Hence, there exists a time t^* up to which the operator Φ_{μ_0} is a contraction, thus it has a unique fixed point $\mu = (\mu_t)_{t\in[0,t^*]}$. It follows from Proposition 15 that μ is a solution, in the sense of Definition 11, to equation (4) on the interval $[0, t^*]$, starting from μ_0. We can repeat this method on the interval $[t^*, 2t^*]$ with initial condition μ_{t^*}, and iterate it up to any finite time T because t^* depends only on the Lipschitz constants of the coefficients, and not on the initial condition. In this way, we have shown that we can construct a solution μ on the interval $[0, T]$ which is a fixed point for the operator $\Phi_{\mu_{nt^*}}$ on the interval $[nt^*, (n + 1)t^*]$, for every $n \in \mathbb{N}$ such that $nt^* < T$. Moreover, we can prove that any two fixed points μ, ν of the map Φ_{μ_0} on the interval $[0, T]$ coincide. Indeed, if $t_0 \in [0, T]$ is the largest time such that $\mu = \nu$, one proves $t_0 = T$ by contradiction, by applying the contraction argument on $[t_0, t_0 + \delta]$ for a suitable $\delta > 0$, if $t_0 < T$. $\square$

LEMMA 18. *Set $T > 0$. Let $\mu = \{\mu_t\}_{t\ge0}, \nu = \{\nu_t\}_{t\ge0} \in \mathcal{S}$ and let X^μ, X^ν be the solutions of equation* (26) *with drift coefficients b_μ and b_ν, respectively. The following holds true*:

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X^\mu(t, x) - X^\nu(t, x)| \big|\mathcal{F}_0\Big] \le \gamma_T \mathbb{E}\Big[\sup_{t\in[0,T]} W_1(\mu_t, \nu_t) \big|\mathcal{F}_0\Big],$$

where

$$\gamma_T := L_K T C_{1,T}. \tag{28}$$

The constant $C_{1,T}$ is defined in (18).

PROOF. Given $T > 0$, we call n the smallest positive integer such that $C(1, (T/n)) < 1$ [see (17)]. We split the interval $[0, T]$ in n subintervals, namely $[\frac{m-1}{n}T, \frac{m}{n}T]$, for $m \le n$. We will give the proof by induction over m.

First, we prove our claim on the interval $[0, (T/n)]$. We start our estimation by giving bounds for the drift and the noise of equation (26). It holds, $\mathbb{P}$-a.s.,

$$\begin{aligned}
&\int_0^t |b_\mu(s, X^\mu(s, x)) - b_\nu(s, X^\nu(s, x))| \, ds \\
&\qquad \le \int_0^t |b_\mu(s, X^\mu(s, x)) - b_\mu(s, X^\nu(s, x))| \, ds
\end{aligned} \tag{29}$$

$$
\begin{aligned}
&+ \int_0^t \left| b_\mu\big(s, X^\nu(s,x)\big) - b_\nu\big(s, X^\nu(s,x)\big)\right| ds \\
&\le L_K \int_0^t \left| X^\mu(s,x) - X^\nu(s,x)\right| ds + L_K \int_0^t W_1(\mu_s, \nu_s)\, ds.
\end{aligned}
$$

Here, we used that, for every $t \in [0, (T/n)]$, $x \in \mathbb{R}^d$ and $\mathbb{P}$-a.s.,

$$
\left| b_\mu(t,x) - b_\nu(t,x)\right| \le L_K W_1(\mu_t, \nu_t). \tag{30}
$$

To prove this, we apply first the definition of b_μ:

$$
\begin{aligned}
&\left| b_\mu(t,x) - b_\nu(t,x)\right| \\
&\quad = \left| \int_{\mathbb{R}^d} K(x-y)\, d\mu_t(y) - \int_{\mathbb{R}^d} K\big(x-y'\big)\, d\nu_t\big(y'\big)\right|.
\end{aligned}
$$

Given $\omega \in \Omega$ a.s. and $t \in [0, (T/n)]$ for every $m \in \Gamma(\mu_t(\omega), \nu_t(\omega))$ so we can rewrite the right-hand side as follows and then apply the Lipshitz continuity of K to obtain, for $\mathbb{P}$-a.e. ω,

$$
\begin{aligned}
&\left| b_\mu(s,x) - b_\nu(s,x)\right| \\
&\quad = \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} K(x-y)\, dm\big(y, y'\big) - \int_{\mathbb{R}^d \times \mathbb{R}^d} K\big(x-y'\big)\, dm\big(y, y'\big)\right| \\
&\quad \le \int_{\mathbb{R}^d \times \mathbb{R}^d} \left| K(x-y) - K\big(x-y'\big)\right| dm\big(y, y'\big) \\
&\quad \le L_K \int_{\mathbb{R}^d \times \mathbb{R}^d} \left| y - y'\right| dm\big(y, y'\big).
\end{aligned}
$$

Now (30) follows since m is arbitrary.

Using the conditional Burkholder–Davis–Gundy inequality (see Proposition 27) and the Lipschitz continuity of the noise, we can estimate the following:

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{t \in [0,(T/n)]} \Big| \int_0^t \sum_k \sigma_k\big(X^\mu(s,x)\big) - \sigma_k\big(X^\nu(s,x)\big)\, dB_s^k \Big| \Big| \mathcal{F}_0 \Big] \\
&\quad \le C_1 \mathbb{E}\Big[\Big(\int_0^{(T/n)} \sum_k \big(\sigma_k\big(X^\mu(s,x)\big) - \sigma_k\big(X^\nu(s,x)\big)\big)^2\, ds \Big)^{1/2} \Big| \mathcal{F}_0 \Big] \\
&\quad \le C_1 L_\sigma \mathbb{E}\Big[\Big(\int_0^{(T/n)} \left| X^\mu(t,x) - X^\nu(t,x)\right|^2 dt \Big)^{1/2} \Big| \mathcal{F}_0 \Big] \qquad (31) \\
&\quad \le C_1 (T/n)^{1/2} L_\sigma \mathbb{E}\Big[\Big(\sup_{t \in [0,(T/n)]} \left| X^\mu(t,x) - X^\nu(t,x)\right|^2 \Big)^{1/2} \big| \mathcal{F}_0 \Big] \\
&\quad \le C_1 (T/n)^{1/2} L_\sigma \mathbb{E}\Big[\sup_{t \in [0,(T/n)]} \left| X^\mu(t,x) - X^\nu(t,x)\right| \big| \mathcal{F}_0 \Big].
\end{aligned}
$$

We now use (29) and (31) to estimate the following:

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\big|X^{\mu}(t,x)-X^{\nu}(t,x)\big|\big|\mathcal{F}_0\Big]\\
&\quad\le \mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\int_0^t\big|b_\mu\big(s,X^{\mu}(t,x)\big)-b_\nu\big(s,X^{\nu}(t,x)\big)\big|\,ds\Big|\mathcal{F}_0\Big]\\
&\qquad+\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\Big|\int_0^t\sum_k\big(\sigma_k\big(X^{\mu}(s,x)\big)-\sigma_k\big(X^{\nu}(s,x)\big)\big)\,dB_s^k\Big|\Big|\mathcal{F}_0\Big]\\
&\quad\le\big(L_K(T/n)+C_1L_\sigma(T/n)^{1/2}\big)\mathbb{E}\Big[\sup_{t\in[0,T]}\big|X^{\mu}(s,x)-X^{\nu}(s,x)\big|\big|\mathcal{F}_0\Big]\\
&\qquad+L_K(T/n)\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}W_1(\mu_t,\nu_t)\big|\mathcal{F}_0\Big].
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}\big|X^{\mu}(t,x)-X^{\nu}(t,x)\big|\big|\mathcal{F}_0\Big]\\
&\qquad\le\frac{1}{1-C(1,(T/n))}L_K(T/n)\mathbb{E}\Big[\sup_{t\in[0,(T/n)]}W_1(\mu_t,\nu_t)\big|\mathcal{F}_0\Big],
\end{aligned}$$

where $C(1,(T/n))$ is defined in (18).

We now prove the inductive step. Suppose that for some $m-1\le n$, it holds

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[((m-2)/n)T,((m-1)/n)T]}\big|X^{\mu}(t,x)-X^{\nu}(t,x)\big|\big|\mathcal{F}_0\Big]\\
&\qquad\le\left(L_K(T/n)\sum_{i=1}^{m-1}\left(\frac{1}{1-C(1,(T/n))}\right)^i\right)\mathbb{E}\Big[\sup_{t\in[0,T]}W_1(\mu_t,\nu_t)\big|\mathcal{F}_0\Big],
\end{aligned}\tag{32}$$

we will prove the same for m. In the same way as in the first step, one can deduce

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[((m-1)/n)T,(m/n)T]}\big|X^{\mu}(t,x)-X^{\nu}(t,x)\big|\big|\mathcal{F}_0\Big]\\
&\qquad\le\mathbb{E}\big[\big|X^{\mu}\big((m-1)T/n,x\big)-X^{\nu}\big((m-1)T/n,x\big)\big|\big|\mathcal{F}_0\big] \qquad (33)\\
&\qquad\quad+\big(L_K(T/n)+C_1L_\sigma(T/n)^{1/2}\big)\\
&\qquad\quad\times\mathbb{E}\Big[\sup_{t\in[((m-1)/n)T,(m/n)T]}\big|X^{\mu}(s,x)-X^{\nu}(s,x)\big|\big|\mathcal{F}_0\Big] \qquad (34)\\
&\qquad\quad+L_K(T/n)\mathbb{E}\Big[\sup_{t\in[((m-1)/n)T,(m/n)T]}W_1(\mu_t,\nu_t)\big|\mathcal{F}_0\Big]. \qquad (35)
\end{aligned}$$

Now we use the inductive hypothesis (32) to estimate (33). We put (34) on the left-hand side and we note that the supremum in (33) is less than the supremum over the whole interval

$$
\begin{aligned}
C^{-1}_{1,(T/n)}&\mathbb{E}\Big[\sup_{t\in[((m-1)/n)T,(m/n)T]}\big|X^{\mu}(t,x)-X^{\nu}(t,x)\big|\big|\mathcal{F}_0\Big]\\
&\le\left(L_K T\sum_{i=1}^{(m-1)}\left(\frac{1}{1-C(1,(T/n))}\right)^i\right)\mathbb{E}\Big[\sup_{t\in[0,T]}W_1(\mu_t,\nu_t)\big|\mathcal{F}_0\Big]\\
&\quad+L_K(T/n)\mathbb{E}\Big[\sup_{t\in[0,T]}W_1(\mu_t,\nu_t)\big|\mathcal{F}_0\Big]\\
&=L_K(T/n)\left(\sum_{i=1}^{(m-1)}\left(\frac{1}{1-C(1,(T/n))}\right)^i+1\right)\mathbb{E}\Big[\sup_{t\in[0,T]}W_1(\mu_t,\nu_t)\big|\mathcal{F}_0\Big].
\end{aligned}
$$

So, (32) is proved for m.

Finally, to obtain the constant of Lemma 18 notice that $\frac{1}{1-C(1,(T/n))}>1$, hence $(\frac{1}{1-C(1,(T/n))})^i\le(\frac{1}{1-C(1,(T/n))})^n$ when $i\le n$. Thus, the constant in (32), in the case $m=n$, can be further estimate by

$$
\begin{aligned}
\left(L_K(T/n)\sum_{i=1}^{m}\left(\frac{1}{1-C(1,(T/n))}\right)^i\right)&\le\left(L_K(T/n)\sum_{i=1}^{n}\left(\frac{1}{1-C(1,(T/n))}\right)^i\right)\\
&\le L_K(T/n)n\left(\frac{1}{1-C(1,(T/n))}\right)^n.
\end{aligned}
$$

This last term is exactly γ_T because of the definition of $C_{1,T}$ [see (18)]. $\square$

LEMMA 19. *For every $T>0$, we have*

$$
d_{\mathcal{S}}(\Phi_{\mu_0}\mu,\Phi_{\mu_0}\nu)\le\gamma_T d_{\mathcal{S}}(\mu,\nu)\qquad\forall\mu,\nu\in\mathcal{S},
$$

where γ_T is defined in (28).

PROOF. Let $\omega\in\Omega$ and $t\in[0,T]$ be fixed. The measure $m=(X^{\mu}(t,\cdot,\omega),X^{\nu}(t,\cdot,\omega))_{\#}\mu_0$ belongs to $\Gamma((\Phi_{\mu_0}\mu)_t(\omega),(\Phi_{\mu_0}\nu)_t(\omega))$. Indeed, for every $A\in\mathbb{R}^{2d}$, it holds $m(B)=\mu_0\{x\in\mathbb{R}^d:X^{\mu}(t,x,\omega),X^{\nu}(t,x,\omega)\in B\}$, which implies, for every $A\in\mathcal{B}(\mathbb{R}^d)$,

$$
\begin{aligned}
m\big(A\times\mathbb{R}^d\big)&=\mu_0\big\{x\in\mathbb{R}^d:X^{\mu}(t,x,\omega)\in A\big\}\\
&=X^{\mu}(t,\cdot,\omega)_{\#}\mu_0(A)=(\Phi_{\mu_0}\mu)_t(\omega)(A).
\end{aligned}
$$

In the same way, $m(\mathbb{R}^d\times A)=(\Phi_{\mu_0}\nu)_t(\omega)(A)$. Thus, from the definition of the Wasserstein metric W_1, it is easy to see that

$$
d_{\mathcal{S}}(\Phi_{\mu_0}\mu,\Phi_{\mu_0}\nu)\le\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{\mathbb{R}^d}\big|X^{\mu}(t,x)-X^{\nu}(t,x)\big|\,d\mu_0\Big].
$$

From the $\mathcal{F}_0$-measurability of the initial condition μ_0 and applying Proposition 28, we have the following:

$$\mathbb{E}\Big[\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{\mathbb{R}^d}|X^\mu(t,x)-X^\nu(t,x)|\,d\mu_0\Big|\mathcal{F}_0\Big]\Big]$$
$$=\mathbb{E}\Big[\int_{\mathbb{R}^d}\mathbb{E}\Big[\sup_{t\in[0,T]}|X^\mu(t,x)-X^\nu(t,x)|\Big|\mathcal{F}_0\Big]\,d\mu_0\Big].$$

Now we complete the proof applying Lemma 18 as follows:

$$d_{\mathcal{S}}(\Phi_{\mu_0}\mu,\Phi_{\mu_0}\nu)\le\mathbb{E}\Big[\int_{\mathbb{R}^d}\mathbb{E}\Big[\sup_{t\in[0,T]}|X^\mu(t,x)-X^\nu(t,x)||\mathcal{F}_0\Big]\,d\mu_0\Big]$$
$$\le\gamma_T d_{\mathcal{S}}(\mu,\nu).\qquad\square$$

4. Convergence and propagation of chaos. In this section, we will show that the distance between two solutions of (4) can be estimated by the distance between the respective initial conditions. Since we have shown in Section 2 that the empirical measure solves (4) with the appropriate initial condition, we will be able to deduce from 20 some results of propagation of chaos.

Last we will give a review on recent quantitative results that can be applied together with Theorem 20 to obtain a more explicit rate of convergence to approximate the solution of SPDE (4) with the solution of SDE (1).

THEOREM 20. *Given $T>0$, let $\mu_0,\nu_0:\Omega\to\mathcal{P}_1(\mathbb{R}^d)$ be as in Hypothesis 2, and let $\mu\in\mathcal{S}_{\mu_0}$, $\nu\in\mathcal{S}_{\nu_0}$ be the respective solutions of equation (4) given by the contraction method described before, there exists a constant $\widetilde{C}_T>0$, such that*

$$d_{\mathcal{S}}(\mu,\nu)\le\widetilde{C}_T\mathbb{E}[W_1(\mu_0,\nu_0)].$$

PROOF. Given $T>0$, we define

$$\widetilde{C}_T:=\Big(\frac{1}{(1-\gamma_{(T/n)})(1-C(1,(T/n)))}\Big)^n,$$

where $n\in\mathbb{N}$ is the smallest integer such that $\gamma_{(T/n)}=L_kTC_{1,T}<1$; see (18) for the definition of $C_{1,T}$, and $C(1,(T/n))<1$, defined in (17). We will give the proof in the case when T is small enough such that $n=1$ and we refer to the inductive procedure used in Lemma 14 for the general case. Notice that under this assumption

$$\widetilde{C}_T:=\frac{C_{1,T}}{1-\gamma_T},$$

where $C_{1,T}$ is defined in (18).

Notice that, since $\|\mu_0\|=\|\nu_0\|=1$, the Lipschitz constants of b_μ and b_ν are the same, L_K. Moreover, recalling the definition of the operator Φ_{μ_0} (resp., Φ_{ν_0}), it holds that its fixed point μ (resp., ν) can be written as $\mu_t = X^\mu(t,\cdot)_\#\mu_0$ [resp., $\nu_t = X^\nu(t,\cdot)_\#\nu_0$] where $X^\mu(t,x,\omega)$ [resp., $X^\nu(t,x,\omega)$] is a continuous version of the solution of equation (16) with drift coefficient b_μ (resp., b_ν). Let now ω be fixed. Notice that the infimum in the definition of the Wasserstein metric is indeed a minimum (see [1], Chapter 6), that is, there exists a measure $m(\omega) \in \Gamma(\mu_0(\omega), \nu_0(\omega))$ such that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-x'| m(\omega, dx, dx') = W_1(\nu_0(\omega), \mu_0(\omega)). \tag{36}$$

Moreover, the function $\omega \mapsto m(\omega)$ is $\mathcal{F}_0$-measurable. Indeed, for every couple of measures $(\mu,\nu) \in \mathcal{P}_1 \times \mathcal{P}_1$ we can construct a measurable map $(\mu,\nu) \mapsto m \in \Gamma_0(\mu,\nu)$ using Proposition 29 in the Appendix, and then we can see that the function $\omega \mapsto (\mu_0(\omega), \nu_0(\omega)) \mapsto m(\omega)$ is $\mathcal{F}_0$-measurable since it is a composition of measurable functions. If we define $m_t(\omega) = (X^\mu(t,\cdot,\omega), X^\nu(t,\cdot,\omega))_\# m(\omega)$, we get $m_t \in \Gamma((\Phi_{\mu_0}\mu)_t, (\Phi_{\nu_0}\nu)_t)$.

As a particular case of Lemma 14, we have that

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \big|X^\mu(t,x) - X^\mu(t,x')\big| \big|\mathcal{F}_0\Big] \le C_{1,T}|x-x'|, \tag{37}$$

where $x, x' \in \mathbb{R}^d$ are two initial condition for equation (26).

In the following estimates, we use the definition of the Wasserstein metric, the definition of m_t, Proposition 28, inequality (37) and identity (36),

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,T]} W_1\big((\Phi_{\mu_0}\mu)_t, (\Phi_{\nu_0}\mu)_t\big)\Big] \\
&\quad\le \mathbb{E}\Big[\sup_{t\in[0,T]} \int_{\mathbb{R}^{2d}} |x-x'|\, dm_t(x,x')\Big] \\
&\quad= \mathbb{E}\Big[\mathbb{E}\Big[\sup_{t\in[0,T]} \int_{\mathbb{R}^{2d}} \big|X^\mu(t,x) - X^\mu(t,x')\big|\, dm(x,x') \Big|\mathcal{F}_0\Big]\Big] \\
&\quad\le \mathbb{E}\Big[\int_{\mathbb{R}^{2d}} \mathbb{E}\Big[\sup_{t\in[0,T]} \big|X^\mu(t,x) - X^\mu(t,x')\big| \big|\mathcal{F}_0\Big]\, dm(x,x')\Big] \\
&\quad\le \mathbb{E}\Big[\int_{\mathbb{R}^{2d}} C_{1,T}|x-x'|\, dm(x,x')\Big] \\
&\quad= C_{1,T}\mathbb{E}\big[W_1(\mu_0,\nu_0)\big].
\end{aligned} \tag{38}$$

Using now the definition of the operators Φ_{μ_0}, Φ_{ν_0} and a triangular inequality, we obtain

$$\begin{aligned}
d_{\mathcal{S}}(\mu,\nu) &= d_{\mathcal{S}}(\Phi_{\mu_0}\mu, \Phi_{\nu_0}\nu) \\
&\le d_{\mathcal{S}}(\Phi_{\mu_0}\mu, \Phi_{\nu_0}\mu) + d_{\mathcal{S}}(\Phi_{\nu_0}\mu, \Phi_{\nu_0}\nu) \\
&\le C_{1,T}\mathbb{E}\big[W_1(\mu_0,\nu_0)\big] + \gamma_T d_{\mathcal{S}}(\mu,\nu).
\end{aligned} \tag{39}$$

In the last inequality, we have used (38) and Lemma 19. Inequality (39) leads to

$$d_{\mathcal{S}}(\mu,\nu) \le \frac{C_{1,T}}{1-\gamma_T}\mathbb{E}\big[W_1(\mu_0,\nu_0)\big]. \qquad \square$$

Reading the proof of this theorem, one may wonder if it is really necessary to add the complication of splitting the time interval in subintervals. Indeed a more simple calculation can lead to a global estimate, although it can only be obtained if the initial conditions belong W_2, which is a stronger assumption. Nevertheless, we will give now the proof in that case so that the reader can compare the two different approaches. Moreover, if one is interested in the W_2 norm, one can apply this method to other results within this paper. We are indebted to an anonymous referee for suggesting us this idea.

At the end of this subsection, we will stress what is the difficulty encountered using W_1 which prevents us to obtain a straightforward global estimation in time.

THEOREM 21. *Under the same assumptions of Theorem 20, suppose that the random measures μ_0,ν_0 take values in $\mathcal{P}_2(\mathbb{R}^d)$, namely they have finite second moments. Then it holds, for all $t \le T$,*

$$\mathbb{E}\big[W_2^2(\mu_t,\nu_t)\big] \le 4e^{4t(2tL_k^2+C_2L_\sigma^2)}\mathbb{E}\big[W_2^2(\mu_0,\nu_0)\big],$$

where L_K and L_σ are the Lipschitz constants of the coefficients of the system and C_2 is the constant appearing in Burkholder–Davis–Gundy inequality with exponent 2.

PROOF. Proceeding as in the proof of Theorem 20, we can find a random measure $m \in \Gamma_0(\mu_0,\nu_0)$, such that $W_2^2(\mu_0,\nu_0) = \int_{\mathbb{R}^d\times\mathbb{R}^d}|x-x'|\,dm(x,x')$. Moreover, it holds

$$\begin{aligned}\mathbb{E}\big[W_2^2(\mu_t,\nu_t)\big] &\le \mathbb{E}\bigg[\int \big|X^\mu(t,x)-X^\nu(t,x')\big|^2\,dm\bigg]\\ &= \mathbb{E}\bigg[\int \mathbb{E}\big[\big|X^\mu(t,x)-X^\nu(t,x')\big|^2\big|\mathcal{F}_0\big]\,dm(x,x')\bigg].\end{aligned}$$

Hence, we proceed estimating the conditional expectation in the last term using that $X^\mu(t,x)$ and $X^\nu(t,x')$ solve (26) and a parallelogram inequality,

$$\begin{aligned}&\mathbb{E}\big[\big|X^\mu(t,x)-X^\nu(t,x')\big|^2|\mathcal{F}_0\big]\\ &\qquad\le 2|x-x'|^2 \qquad (40)\\ &\qquad\quad + 2\mathbb{E}\bigg[\bigg(\int_0^t \big|b_{\mu_s}(X^\mu(s,x))-b_{\nu_s}(X^\nu(s,x'))\big|\,ds\bigg)^2\bigg|\mathcal{F}_0\bigg] \qquad (41)\\ &\qquad\quad + 2\mathbb{E}\bigg[\bigg(\int_0^t \sum_k\big|\sigma_k(X^\mu(s,x))-\sigma_k(X^\nu(s,x'))\big|\,dB_s^k\bigg)^2\bigg|\mathcal{F}_0\bigg]. \qquad (42)\end{aligned}$$

Using a Burkholder–Davis–Gundy inequality and the Lipschitz continuity of σ_k, we can estimate (42) as follows:

$$
\begin{aligned}
&2\mathbb{E}\bigg[\bigg(\int_0^t \sum_k \big|\sigma_k\big(X^\mu(s,x)\big) - \sigma_k\big(X^\nu(s,x')\big)\big|\, dB_s^k\bigg)^2\bigg|\mathcal{F}_0\bigg] \\
&\qquad \le 2C_2 L_\sigma^2 \mathbb{E}\bigg[\int_0^t \big|X^\mu(t,x) - X^\nu(t,x')\big|^2\, ds\bigg|\mathcal{F}_0\bigg].
\end{aligned} \tag{43}
$$

To estimate (41), we first apply the Jensen inequality, then we need to split the drift using a triangular inequality and then use the Lipschitz continuity of K,

$$
\begin{aligned}
&2\mathbb{E}\bigg[\bigg(\int_0^t \big|b_{\mu_s}\big(X^\mu(s,x)\big) - b_{\nu_s}\big(X^\nu(s,x')\big)\big|\, ds\bigg)^2\bigg|\mathcal{F}_0\bigg] \\
&\quad \le 2t\mathbb{E}\bigg[\int_0^t \big|b_{\mu_s}\big(X^\mu(s,x)\big) - b_{\nu_s}\big(X^\nu(s,x')\big)\big|^2\, ds\bigg|\mathcal{F}_0\bigg] \\
&\quad \le 4t\mathbb{E}\bigg[\int_0^t ds \int \big|K\big(X^\mu(s,x) - y\big) - K\big(X^\nu(s,x') - y\big)\big|^2\, d\mu_s(y) \\
&\qquad + \bigg|\int \big(K\big(X^\nu(s,x') - y\big) \\
&\qquad - K\big(X^\nu(s,x') - y'\big)\big)\, d\big(\mu_s(y) - \nu_s(y')\big)\bigg|^2\bigg|\mathcal{F}_0\bigg] \\
&\quad \le 4tL_k^2 \int_0^t \mathbb{E}\big[\big|X^\mu(s,x) - X^\nu(s,x')\big|^2 |\mathcal{F}_0\big]\, ds \qquad (44) \\
&\qquad + 4tL_k^2 \int_0^t \mathbb{E}\big[W_2^2(\mu_s,\nu_s)|\mathcal{F}_0\big]\, ds. \qquad (45)
\end{aligned}
$$

We used here a property of the Wassertein metric which we already used and proved in the proof of Lemma 18 [see (29)] for W_1, but which can be straightforwardly readapted to W_2.

We now put together (41), (44), (45) and (43) to obtain

$$
\begin{aligned}
&\mathbb{E}\big[W_2^2(\mu_t,\nu_t)\big] \\
&\quad \le \mathbb{E}\bigg[\int \big|X^\mu(t,x) - X^\nu(t,x')\big|^2\, dm(x,x')\bigg] \\
&\quad \le 2\mathbb{E}\big[W_2^2(\mu_0,\nu_0)\big] \\
&\qquad + 4tL_k^2 \int_0^t \mathbb{E}\bigg[W_2^2(\mu_s,\nu_s) + \int \big|X^\mu(s,x) - X^\nu(s,x')\big|^2\, dm(x,x')\bigg]\, ds \\
&\qquad + 2C_2L_\sigma^2 \int_0^t \mathbb{E}\bigg[\int \big|X^\mu(s,x) - X^\nu(s,x')\big|^2\, dm(x,x')\bigg]\, ds.
\end{aligned}
$$

Adding at the end the positive term $2C_2L_\sigma^2\int_0^t\mathbb{E}[W_2^2(\mu_s,\nu_s)]\,ds$, we can apply the Gronwall inequality and obtain

$$\begin{aligned}\mathbb{E}\Big[W_2^2(\mu_t,\nu_t)+\int\big|X^\mu(t,x)-X^\nu(t,x')\big|^2\,dm(x,x')\Big]&\\ \le 4e^{4t(2tL_k^2+C_2L_\sigma^2)}\mathbb{E}\big[W_2^2(\mu_0,\nu_0)\big].&\qquad\square\end{aligned}$$

REMARK 22. Reading the proof of the previous theorem, one can be led to think that it is possible to do the same calculations using the norm W_1, which is true up to some point. In particular, following the idea of the proof of Theorem 21 one can reach the inequality

$$\begin{aligned}\mathbb{E}\big[W_1(\mu_t,\nu_t)\big]&\le\mathbb{E}\Big[\int|X_t^\mu-X_t^\nu|\,dm\Big]\\&\le\mathbb{E}\big[W_1(\mu_0,\nu_0)\big]+L_k\int_0^t\mathbb{E}\Big[W_1(\mu_s,\nu_s)+\int|X_s^\mu-X_s^\nu|\,dm\Big]ds\\&\quad+C_1L_\sigma\mathbb{E}\Big[\int\Big(\int_0^t|X_s^\mu-X_s^\nu|^2\,ds\Big)^{1/2}dm\Big].\end{aligned}$$

The difficult term is the last one, indeed we do not see a way to get rid of the powers or to switch them with the integrals. What we indeed do in most of the proofs in this paper is to take the supremum in time inside the integrals to obtain

$$\mathbb{E}\Big[\int\Big(\int_0^t|X_s^\mu-X_s^\nu|^2\,ds\Big)^{1/2}dm\Big]\le tC_1L_\sigma\mathbb{E}\Big[\int\sup_{s\in[0,t]}|X_s^\mu-X_s^\nu|\,dm\Big].$$

At this point, it is no longer possible to apply the Gronwall lemma, but this last term can be subtracted in both sides of the estimations to get something of the form $(1-tC_1L_\sigma)\mathbb{E}[\int\sup_{s\in[0,t]}|X_s^\mu-X_s^\nu|\,dm]\le\cdots$, from which the need to do the estimations in small intervals first.

4.1. *Propagation of chaos.* Let $(\Omega,\mathcal{F},\mathcal{F}_t,\mathbb{P})$ be a filtered probability space, and $(X_0^i)_{i\in\mathbb{N}}$ be a sequence of symmetric $\mathbb{R}^d$-valued random variable on this space that are measurable with respect to $\mathcal{F}_0$. We consider a collection B_t^k, $k\ge 1$, of independent Brownian motions on this space, independent from the X_0^i, and we call $(\mathcal{F}_t^B)_{t\ge0}$ the filtration generated by $(B_t^k)_{k\ge1}$. For every $N\in\mathbb{N}$, $X^N=(X_t^{1,N},\dots,X_t^{N,N})_{t\ge0}$ is the solution of equation (1) with initial condition $(X_0^1,\dots,X_0^N)$. We will further suppose that the empirical measure $S_0^N:=\frac{1}{N}\sum_{i=0}^N\delta_{X_0^i}$ converges to a random probability measure μ_0, in the metric $\mathbb{E}[W_1(\cdot,\cdot)]$. Under these settings, we will now prove Theorems 24 (which is slightly more general then Theorem 2) and 3, but first we need the following lemma.

LEMMA 23. *Let $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$ be a permutation. Then*

$$\mathbb{E}[f(X_t^{1,N}, \dots, X_t^{N,N})|\mathcal{F}_t^B] = \mathbb{E}[f(X_t^{\sigma(i),N}, \dots, X_t^{\sigma(N),N})|\mathcal{F}_t^B], \tag{46}$$

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

PROOF. Let $X^{\sigma,N} := (X_t^{\sigma(1),N}, \dots, X_t^{\sigma(N),N})_{t\geq 0}$. Since X^N is a strong solution of equation (1) with initial condition $(X_1, \dots, X_N)$ it is easy to see that $X^{\sigma,N}$ is a strong solution of equation (1) with initial condition $(X_{\sigma(1)}, \dots, X_{\sigma(N)})$. Since the coefficients b and σ_k have the necessary Lipschitz properties (see [9]), we have strong uniqueness at fixed initial data $x \in \mathbb{R}^d$. Thus, we can apply Proposition 1.4 of [12] (notice that X^N and $X^{\sigma,N}$ have the same initial law) and we obtain uniqueness in law. More precisely we have

$$(X_t^N, (B_t^k)_{k\in\mathbb{N}})_{\#}\mathbb{P} = (X_t^{\sigma,N}, (B_t^k)_{k\in\mathbb{N}})_{\#}\mathbb{P} \qquad \forall t \geq 0.$$

This implies, for every $A \in \mathcal{F}_t^B$ such that $A = \{(B_t^k)_{k\geq 1} \in \widetilde{A}\}$ with $\widetilde{A} \in \mathcal{B}((\mathbb{R}^d)^\infty)$ and for every $\phi \in C_b((\mathbb{R}^d)^N)$,

$$\begin{aligned}\mathbb{E}[\mathbb{1}_A f(X_t^N)] &= \mathbb{E}[\mathbb{1}_{\{(B_t^k)_{k\geq 1}\in\widetilde{A}\}} f(X_t^N)] = \mathbb{E}[\mathbb{1}_{\{(B_t^k)_{k\geq 1}\in\widetilde{A}\}} f(X_t^{N,\sigma})] \\ &= \mathbb{E}[\mathbb{1}_A f(X_t^{N,\sigma})].\end{aligned}$$

Since the integrals of $f(X_t^N)$ and $f(X_t^{N,\sigma})$ coincide on every element of a basis of $\mathcal{F}_t^B$, their conditional expectation coincide also; hence, (46) follows. $\square$

Using the previous result, we can now prove Theorem 2 which we restate here for simplicity.

THEOREM 24. *There exists a random measure-valued solution μ_t of equation (8) such that*

$$\lim_{N\to\infty} E[|\langle S_t^N, \phi\rangle - \langle \mu_t, \phi\rangle|] = 0$$

for all $\phi \in C_b(\mathbb{R}^d)$.

Moreover, given $r \in \mathbb{N}$ and $\phi_1, \dots, \phi_r \in C_b(\mathbb{R}^d)$, we have

$$\lim_{N\to\infty} E[\phi_1(X_t^{1,N}) \cdots \phi_r(X_t^{r,N})|\mathcal{F}_t^B] = \mathbb{E}\left[\prod_{i=1}^r \langle \mu_t, \phi_i\rangle \middle| \mathcal{F}_t^B\right]$$

in $L^1(\Omega)$.

PROOF. Since the convergence in the Wasserstein metric W_1 implies the weak convergence, the first statement follows from Theorem 20.

Without loss of generality, we prove the second statement in the case $r = 2$. Let $\phi_1, \phi_2 \le M$. By a triangular inequality, we obtain

$$
\begin{aligned}
&|\mathbb{E}[\phi_1(X_t^{1,N})\phi_2(X_t^{2,N})|\mathcal{F}_t^B] - \mathbb{E}[\langle \mu_t, \phi_1\rangle\langle \mu_t, \phi_2\rangle|\mathcal{F}_t^B]| \\
&\quad \le |\mathbb{E}[\phi_1(X_t^{1,N})\phi_2(X_t^{2,N})|\mathcal{F}_t^B] - \mathbb{E}[\langle S_t^N, \phi_1\rangle\langle S_t^N, \phi_2\rangle|\mathcal{F}_t^B]| \qquad (47)\\
&\qquad + |\mathbb{E}[\langle S_t^N, \phi_1\rangle\langle S_t^N, \phi_2\rangle|\mathcal{F}_t^B] - \mathbb{E}[\langle \mu_t, \phi_1\rangle\langle \mu_t, \phi_2\rangle|\mathcal{F}_t^B]|. \qquad (48)
\end{aligned}
$$

Using Lemma 23, we can estimate (47) as follows:

$$
\begin{aligned}
&|\mathbb{E}[\phi_1(X_t^{1,N})\phi_2(X_t^{2,N})|\mathcal{F}_t^B] - \mathbb{E}[\langle S_t^N, \phi_1\rangle\langle S_t^N, \phi_2\rangle|\mathcal{F}_t^B]| \\
&\quad = \Bigg| \frac{1}{N^2 - N} \sum_{i,j=1, i\neq j}^{N} \mathbb{E}[\phi_1(X_t^{i,N})\phi_2(X_t^{j,N})|\mathcal{F}_t^B] \\
&\qquad - \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}[\phi_1(X_t^{i,N})\phi_2(X_t^{j,N})|\mathcal{F}_t^B] \Bigg| \\
&\quad \le \left| \left(\frac{1}{N^2 - N} - \frac{1}{N^2} \right)(N^2 - N)M^2 \right| + \left| \frac{1}{N} M^2 \right| \\
&\quad = 2\frac{M^2}{N} \to 0 \qquad \text{as } N \to \infty.
\end{aligned}
$$

The convergence to zero of (47) follows from the first statement of this theorem. Indeed,

$$
\begin{aligned}
&\mathbb{E}[|\mathbb{E}[\langle S_t^N, \phi_1\rangle\langle S_t^N, \phi_2\rangle|\mathcal{F}_t^B] - \mathbb{E}[\langle \mu_t, \phi_1\rangle\langle \mu_t, \phi_2\rangle|\mathcal{F}_t^B]|] \\
&\quad \le \mathbb{E}[|\langle S_t^N, \phi_1\rangle - \langle \mu_t, \phi_1\rangle||\langle S_t^N, \phi_2\rangle|] + \mathbb{E}[|\langle S_t^N, \phi_2\rangle - \langle \mu_t, \phi_2\rangle||\langle \mu_t, \phi_1\rangle|] \\
&\quad \le M\mathbb{E}[|\langle S_t^N, \phi_1\rangle - \langle \mu_t, \phi_1\rangle|] + M\mathbb{E}[|\langle S_t^N, \phi_2\rangle - \langle \mu_t, \phi_2\rangle|] \\
&\quad = 2M\mathbb{E}[|\langle S_t^N, \phi_1\rangle - \langle \mu_t, \phi_1\rangle|] \to 0 \qquad \text{as } N \to \infty. \qquad \square
\end{aligned}
$$

PROOF OF THEOREM 3. First, notice that $X^{r,N}$ is the strong solution of equation (26) with drift coefficient b_ν, where $\nu = S^N = \{S_t^N\}_{t\in[0,T]}$, and initial condition X_0^r. We can thus write $X^{r,N} = X^\nu(t, X_0^r(\omega), \omega)$.

If we apply Lemma 18, we obtain

$$E[|X_t^{r,N} - X_t|] \le \gamma_T d_{\mathcal{S}}(\mu, S^N).$$

This last quantity goes to 0 as $N \to \infty$ thanks to Theorem 20. $\square$

4.2. *Quantitative estimates*. As already mentioned, there are several recent results in literature that deal with the rate of convergence of an empirical measure. In this section, we want to give some examples of how these results can be applied in our model using Theorem 20. Under the assumption in the beginning of the section, we further define G_0^N the law of the initial condition $(X_0^1, \dots, X_0^N)$ and we denote by $G_{0,2}^N$ its first two marginals. Given a $p > 0$, we suppose that G_0^N and μ_0 have finite first p moments $M_p(G_0^N)$ and $M_p(\mu_0)$.

Using Theorem 2.4 of [8] on the initial conditions and our estimates of Theorem 20, we can compare the rate of convergence of the empirical measure of the solution to the rate of convergence of just two initial particles.

COROLLARY 25. *For every exponent $\gamma < (d+1+\frac{d}{p})^{-1}$, there exists a finite positive constant Γ depending only on p and d such that, for every $N \geq 1$,*

$$\mathbb{E}\big[W_1\big(S_t^N, \mu\big)\big] \leq \widetilde{C}\Gamma\big(M_p\big(G_0^N\big) + M_p(\mu_0)\big)^{1/p}\bigg(W_1\big(G_{0,2}^N, \mu_0\big) + \frac{1}{N}\bigg)^{\gamma}.$$

When the initial condition consists of a sequence of i.i.d. μ_0-distributed random variables $(X_0^i)_{i\in\mathbb{N}}$, a quantitative estimate can be derived from [6]. Under this stronger assumptions one can obtain a slightly stronger result, however in this case we must suppose that the measures which we are working on have finite p moments with p strictly greater than one.

COROLLARY 26. *Let $p > 1$. There exists a constant Γ depending on p and d such that, for all $N \geq 1$,*

$$\begin{aligned}&\mathbb{E}\big[W_1\big(S_t^N, \mu_t\big)\big]\\ &\quad\leq \widetilde{C}\Gamma M_p(\mu_0)^{1/p}\\ &\qquad\times\begin{cases} N^{-1/2}\log(1+N) + N^{-(p-1)/p}, & \text{if } d = 2 \text{ and } p \neq 2,\\ N^{-1/d} + N^{-(p-1)/p}, & \text{if } d > 2 \text{ and } p \neq \dfrac{d}{(d-1)}.\end{cases}\end{aligned}$$

APPENDIX

PROPOSITION 27. *Given $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, let M_t be a continuous martingale with respect to $\mathcal{F}_t$. If we define $M_t^* = \sup_{0\leq s\leq t}|M_s|$, it holds*

$$\mathbb{E}\big[|M_t^*|^p|\mathcal{F}_0\big] \leq C_p\mathbb{E}\big[[M]_t^{p/2}|\mathcal{F}_0\big],$$

for some constant $C_p > 0$.

PROOF. We fix an $A \in \mathcal{F}_0$ and we prove the following:

$$\mathbb{E}\big[\mathbb{1}_A\big|M_t^*\big|^p\big] \leq C_p\mathbb{E}\big[\mathbb{1}_A[M]_t^{p/2}\big].$$

First, we note that $N_t := M_t \mathbb{1}_A$ is a continuous $\mathcal{F}_t$-martingale, indeed $A \in \mathcal{F}_0 \subset \mathcal{F}_s$ implies

$$\mathbb{E}[\mathbb{1}_A M_t | \mathcal{F}_s] = \mathbb{1}_A \mathbb{E}[M_t | \mathcal{F}_s] = \mathbb{1}_A M_s.$$

We can thus apply the Burkholder–Davis–Gundy inequality to N_t and we obtain

$$\mathbb{E}[|N_t^*|^p] \le C_p \mathbb{E}[[N]_t^{p/2}].$$

Notice that $\mathbb{1}_A$ commute with $\sup_{t\in[0,T]}$. The thesis follows from the equality

$$[\mathbb{1}_A M]_t = \mathbb{1}_A [M]_t. \tag{49}$$

□

Throughout the paper, we repeatedly used an identity of the form

$$E\left[\int_{\mathbb{R}^d} f(x)\, d\mu_0(x) \middle| \mathcal{F}_0\right] = \int_{\mathbb{R}^d} E[f(x)|\mathcal{F}_0]\, d\mu_0(x). \tag{50}$$

This identity may look at first sight completely general but it requires appropriate assumptions of continuity in x and integrability. Just in order that all objects are well defined, we need:

(i) $f : \Omega \to C(\mathbb{R}^d)$ measurable,
(ii) $E[\int_{\mathbb{R}^d} |f(x)|\, d\mu_0(x)] < \infty$,
(iii) $E[\sup_{x\in K} |f(x)|] < \infty$ for every compact set $K \subset \mathbb{R}^d$.

Indeed, under (i)–(ii), the integral $\int_{\mathbb{R}^d} f(x)\, d\mu_0(x)$ is first well defined and finite a.s. (f has to be continuous in x since μ_0 is a general probability measure), and also $L^1(\Omega)$, so the conditional expectation $E[\int_{\mathbb{R}^d} f(x)\, d\mu_0(x)|\mathcal{F}_0]$ is well defined. As to the right-hand side of (50), on any compact set $K \subset \mathbb{R}^d$, from (i) and (iii), we have $\omega \mapsto f(\omega, \cdot)$ of class $L^1(\Omega; C(K))$ [the space $C(K)$ of continuous functions on K endowed with the uniform topology], hence by the definition of conditional expectation of random variables with values in Banach spaces, $E[f|_K|\mathcal{F}_0]$ is again a well-defined element of $L^1(\Omega; C(K))$; and, as shown below in the proof of next proposition, taking as compact sets the sequence of closed balls $B(0,n)$ one gets a definition of $E[f(x)|\mathcal{F}_0]$ as a measurable function from Ω to $C(\mathbb{R}^d)$; notice in particular that continuity in x of $E[f(x)|\mathcal{F}_0]$ is essential to define $\int_{\mathbb{R}^d} E[f(x)|\mathcal{F}_0]\, d\mu_0(x)$ because μ_0 is a general probability measure. Finally, the finiteness of $\int_{\mathbb{R}^d} E[f(x)|\mathcal{F}_0]\, d\mu_0(x)$ is ultimately a consequence of (ii) again, as proved in the next proposition.

PROPOSITION 28. *Under assumptions* (i), (ii) *and* (iii), *identity* (50) *holds true almost surely.*

PROOF. As already noticed, given $n \in \mathbb{N}$, $E[f|_{B(0,n)}|\mathcal{F}_0]$ is a well-defined element of $L^1(\Omega; C(B(0,n)))$. Moreover, if g is in the equivalence class of

$E[f|_{B(0,n)}|\mathcal{F}_0]$, then at any $x \in B(0,n)$ we have that $g(x)$ is in the equivalence class of $E[f(x)|\mathcal{F}_0]$ [understood as the conditional expectation of the r.v. $\omega \mapsto f(\omega,x)$, x given]. Indeed, for every $A \in \mathcal{F}_0$,

$$\mathbb{E}\big[g(x)\mathbb{1}_A\big] = \mathbb{E}[g\mathbb{1}_A](x) = \mathbb{E}[f\mathbb{1}_A](x) = \mathbb{E}\big[f(x)\mathbb{1}_A\big].$$

We can choose a sequence $f^{(m)} = \sum_{i=1}^m f_i \mathbb{1}_{A_i}$ such that $f_i \in C(B(0,n))$, $A_i \in \mathcal{F}$ and $f^{(m)} \to f$ in $L^1(\Omega, C(B(0,n)))$, as $m \to \infty$. Moreover one can choose, up to subsequences, $f^{(m)}$ such that the convergence is almost sure and $\|f^{(m)}\|_\infty \le \|f|_{B(0,n)}\|_\infty$, a.s. It is easy to see that $\mathbb{E}[f^{(m)}|\mathcal{F}_0] = \sum_i \mathbb{E}[f_i|\mathcal{F}_0]\mathbb{1}_{A_i}$. From this it follows that

$$\mathbb{E}\bigg[\int_{B(0,n)} f^{(m)}\,d\mu_0|\mathcal{F}_0\bigg] = \int_{B(0,n)} \mathbb{E}\big[f^{(m)}\big|\mathcal{F}_0\big]\,d\mu_0, \qquad \mathbb{P}\text{-a.s.}$$

Notice that, for every fixed ω, it holds $f^{(m)}(\omega) \to f(\omega)$ uniformly in x on the compact $B(0,n)$, and hence, by the dominated convergence theorem

$$\int_{B(0,n)} f^{(m)}(\omega)(x)\mu_0(\omega,dx) \to \int_{B(0,n)} f(\omega)(x)\mu_0(\omega,dx).$$

Thus, $\int_{B(0,n)} f^{(n)}\,d\mu_0 \to \int_{B(0,n)} f\,d\mu_0$ in L^1 from which follows that, up to a subsequence, $\mathbb{E}[\int_K f^{(n)}\,d\mu_0|\mathcal{F}_0] \to \mathbb{E}[\int_{B(0,n)} f\,d\mu_0|\mathcal{F}_0]$, $\mathbb{P}$-a.s. On the other hand, we can first apply conditional dominated convergence and then the traditional version of it to obtain $\int_{B(0,n)} \mathbb{E}[f^{(n)}|\mathcal{F}_0]\,d\mu_0 \to \int_{B(0,n)} \mathbb{E}[f|\mathcal{F}_0]\,d\mu_0$.

We have proven (50) on a closed ball of $\mathbb{R}^d$, we want to extend it on the whole space. Given $n \in \mathbb{N}$, we call f_n the restriction of f on $B(0,n)$. It holds, as already noted, $f_n \in L^1(\Omega, C(B(0,n)))$ for every $n \in \mathbb{N}$.

We construct now the sequence $\{g_n\}_{n\in\mathbb{N}}$ such that $g_n : \Omega \to C(B(0,n))$ and

$$g_n \in L^1\big(\Omega; C\big(B(0,n)\big)\big) \qquad \text{for every } n \in \mathbb{N},$$
$$g_n \in E[f_n|\mathcal{F}_0] \qquad \text{for every } n \in \mathbb{N}.$$

We will show that there exists a function $g : \Omega \to C(\mathbb{R}^d)$, such that for every $x \in \mathbb{R}^d$, $g(x) \in E[f(x)|\mathcal{F}_0]$ and $g|_{\Omega\times B(0,n)} = g_n$. Moreover, if $g, g' : \Omega \to C(\mathbb{R}^d)$ have the same properties, then $g = g'$ a.s.

First, let us prove that $g_{n+1}|_{\Omega\times B(0,n)}$, as a function from Ω to $C(B(0,n))$, is equal to g_n on a set Ω_n of measure one. The function g_{n+1} is characterized by two properties: it is $\mathcal{F}_0$-measurable, and $E[g_{n+1}\mathbb{1}_A] = E[f_{n+1}\mathbb{1}_A]$ for every $A \in \mathcal{F}_0$. Here, $E[g_{n+1}\mathbb{1}_A]$ and $E[f_{n+1}\mathbb{1}_A]$ are elements of $C(B(0,n+1))$. Similarly, g_n is $\mathcal{F}_0$-measurable, and $E[g_n\mathbb{1}_A] = E[f_n\mathbb{1}_A]$ for every $A \in \mathcal{F}_0$. Obviously, $g_{n+1}|_{\Omega\times B(0,n)}$ is $\mathcal{F}_0$-measurable. Moreover,

$$E[g_{n+1}|_{\Omega\times B(0,n)}\mathbb{1}_A] = E[g_{n+1}\mathbb{1}_A]|_{B(0,n)}.$$

To show this, notice that the function

$$G_n(x) := \mathbb{E}\big[g_{n+1}(x)|_{\Omega\times B(0,n)}\mathbb{1}_A\big]$$

is well defined by Fubini theorem as a function from $B(0,n)$ to $\mathbb{R}^d$. In the same way, one can define $G(x) := \mathbb{E}[g_{n+1}(x)\mathbb{1}_A]$ as a function on $B(0,n+1)$. Now $G_n(x) = G(x)$ for every $x \in B(0,n)$, hence $G_n = G|_{B(0,n)}$. Now,

$$\begin{aligned} E[g_{n+1}\mathbb{1}_A]|_{B(0,n)} &= E[f_{n+1}\mathbb{1}_A]|_{B(0,n)} = E[f_{n+1}|_{\Omega\times B(0,n)}\mathbb{1}_A] \\ &= E[f_n\mathbb{1}_A] = E[g_n\mathbb{1}_A] \end{aligned}$$

and thus $g_{n+1}|_{\Omega\times B(0,n)}$ is almost surely equal to g_n.

On the set $\bigcap_n \Omega_n$, we have $g_m|_{\Omega\times B(0,k)} = g_k$ for every $m \geq k \geq 0$. Let $g : \Omega \times \mathbb{R}^d \to \mathbb{R}$ be defined on $\bigcap_n \Omega_n$ as $g(x,\omega) = g_m(x,\omega)$ where m is the smallest integer such that $x \in B(0,m)$ (and arbitrarily on the complementary of $\bigcap_n \Omega_n$). For every $\omega \in \bigcap_n \Omega_n,$ the function $x \mapsto g(x,\omega)$ is continuous on each $B(0,m)$ (easy to check by the previous properties). Hence, $g : \Omega \to C(\mathbb{R}^d)$.

Now, if $g' : \Omega \to C(\mathbb{R}^d)$ is such that, for every $n \in \mathbb{N}$, it holds $g'|_{\Omega\times B(0,n)} \in \mathbb{E}[f_n|\mathcal{F}_0]$, then there exists a set $\Omega_n \subset \Omega$, such that $\mathbb{P}(\Omega_n) = 1$ and $g_n = g'_n$ on Ω_n. Then for every $\omega \in \bigcap_n \Omega_n$, and for every $x \in B(0,n)$, $g(\omega,x) = g_n(\omega,x) = g'_n(\omega,x) = g'(\omega,x)$; hence, $g = g'$ a.e. Finally, if $x \in B(0,n)$, and $A \in \mathcal{F}_0$,

$$\begin{aligned} \mathbb{E}\big[g(x)\mathbb{1}_A\big] &= \mathbb{E}\big[g_n(x)\mathbb{1}_A\big] = \mathbb{E}[g_n\mathbb{1}_A](x) = \mathbb{E}[f_n\mathbb{1}_A](x) = \mathbb{E}\big[f_n(x)\mathbb{1}_A\big] \\ &= \mathbb{E}\big[f(x)\mathbb{1}_A\big]. \end{aligned}$$

Hence, $g(x) \in \mathbb{E}[f(x)|\mathcal{F}_0]$. To conclude, we notice that applying Lebesgue dominate convergence theorem to the sequence f_n, the random variables $\int_{B(0,n)} f_n\,d\mu_0$ converges a.s. to the random variable $\int_{\mathbb{R}^d} f\,d\mu_0$, as $n \to \infty$. Thus, by the conditional version of dominated convergence theorem,

$$\mathbb{E}\biggl[\int_{R^d} f\,d\mu_0\Big|\mathcal{F}_0\biggr] = \lim_{n\to\infty} \mathbb{E}\biggl[\int_{B(0,n)} f_n\,d\mu_0\Big|\mathcal{F}_0\biggr]. \tag{51}$$

By the definition of g, we have that, as $n \to \infty$, the positive part g_n^+ increases to g^+ a.s., and the negative g_n^- increases to g^-. Thus, by monotone convergence theorem, it holds a.s.

$$\begin{aligned} \int_{\mathbb{R}^d} g\,d\mu_0 &= \int_{\mathbb{R}^d} g^+\,d\mu_0 - \int_{\mathbb{R}^d} g^-\,d\mu_0 \\ &= \lim_{n\to\infty}\int_{B(0,n)} g_n^+\,d\mu_0 - \lim_{n\to\infty}\int_{B(0,n)} g_n^-\,d\mu_0. \end{aligned} \tag{52}$$

The thesis follows from the equalities (51) and (52). Notice that this also implies that $\int_{\mathbb{R}^d} \mathbb{E}[f(x)|\mathcal{F}_0]\,d\mu_0(x)$ is finite, because it is equal to a finite quantity. □

PROPOSITION 29. *Let $(\mu,\nu) \in \mathcal{P}_1(\mathbb{R}^d)$. If we define the set*

$$\begin{aligned} &\Gamma_0(\mu,\nu) \\ &\quad := \biggl\{\bar{m} \in \Gamma(\mu,\nu) \Big| \int_{\mathbb{R}^{2d}} |x-y|\,d\bar{m}(x,y) = \inf_{m\in\Gamma(\mu,\nu)} \int_{\mathbb{R}^{2d}} |x-y|\,dm(x,y)\biggr\} \end{aligned}$$

then there exists a measurable function $f : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \to \mathcal{P}_1(\mathbb{R}^{2d})$ *such that* $f(\mu, \nu) \in \Gamma_0(\mu, \nu)$.

PROOF. The set $\{(\mu, \nu, m) | m \in \Gamma_0(\mu, \nu)\}$ is closed in $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^{2d})$ endowed with the weak topology (see, e.g., [1], Proposition 7.1.3), thus the proposition follows from Von Neumann theorem on measurable selections. $\square$

Acknowledgments. The authors wish to thank the anonymous referees for the careful revision which helped to clarify and improve considerably the initial version of this paper.

REFERENCES

[1] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. Birkhäuser, Basel. MR2401600

[2] CARLEN, E. A., CARVALHO, M. C., LE ROUX, J., LOSS, M. and VILLANI, C. (2010). Entropy and chaos in the Kac model. *Kinet*. *Relat*. *Models* **3** 85–122. MR2580955

[3] DELARUE, F., FLANDOLI, F. and VINCENZI, D. (2014). Noise prevents collapse of Vlasov–Poisson point charges. *Comm*. *Pure Appl*. *Math*. **67** 1700–1736. MR3251910

[4] DERMOUNE, A. (2003). Propagation and conditional propagation of chaos for pressureless gas equations. *Probab. Theory Related Fields* **126** 459–476. MR2001194

[5] DOBRUŠIN, R. L. (1979). Vlasov equations. *Funktsional*. *Anal*. *i Prilozhen*. **13** 48–58, 96. MR0541637

[6] FOURNIER, N. and GUILLIN, A. (2013). On the rate of convergence in wasserstein distance of the empirical measure. Available at arXiv:1312.2128v1.

[7] FOURNIER, N., HAURAY, M. and MISCHLER, S. (2014). Propagation of chaos for the 2D viscous vortex model. *J*. *Eur. Math*. *Soc*. (*JEMS*) **16** 1423–1466. MR3254330

[8] HAURAY, M. and MISCHLER, S. (2014). On Kac's chaos and related problems. *J*. *Funct*. *Anal*. **266** 6055–6157. MR3188710

[9] KUNITA, H. (1990). *Stochastic Flows and Stochastic Differential Equations*. *Cambridge Studies in Advanced Mathematics* **24**. Cambridge Univ. Press, Cambridge. MR1070361

[10] PILIPENKO, A. Y. (2005). Measure-valued diffusions and continuous systems of interacting particles in a random environment. *Ukraïn*. *Mat*. *Zh*. **57** 1289–1301. MR2216047

[11] PILIPENKO, A. Y. (2006). A functional central limit theorem for flows generated by stochastic equations with interaction. *Nelīnīĭnī Koliv*. **9** 85–97. MR2369777

[12] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* **293**. Springer, Berlin. MR1725357

[13] SKOROKHOD, A. V. (1997). Measure-valued diffusion. *Ukraïn*. *Mat*. *Zh*. **49** 458–464. MR1472230

[14] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math*. **1464** 165–251. Springer, Berlin. MR1108185

[15] ZHENG, W. A. (1995). Conditional propagation of chaos and a class of quasilinear PDE's. *Ann*. *Probab*. **23** 1389–1413. MR1349177

SCUOLA NORMALE SUPERIORE
PIAZZA DEI CAVALIERI 7
56126 PISA
ITALY
E-MAIL: michele.coghi@sns.it

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI PISA
PIAZZA BRUNO PONTECORVO 5
56127 PISA
ITALY
E-MAIL: flandoli@dma.unipi.it