

RATE OF HOMOGENIZATION FOR FULLY-COUPLED MCKEAN-VLASOV SDES

Z.W. BEZEMEK, AND K. SPILIOPOULOS

ABSTRACT. We consider a fully-coupled slow-fast system of McKean-Vlasov SDEs with full dependence on the slow and fast component and on the law of the slow component and derive convergence rates to its homogenized limit. We do not make periodicity assumptions, but we impose conditions on the fast motion to guarantee ergodicity. In the course of the proof we obtain related ergodic theorems and we gain results on the regularity of Poisson type of equations and of the associated Cauchy-Problem on the Wasserstein space that are of independent interest.

1. INTRODUCTION

The goal of this paper is to study the behavior as $\epsilon \rightarrow 0$ of the system of slow-fast McKean-Vlasov SDEs

$$(1) \quad \begin{aligned} X_t^{\epsilon, \nu} &= \eta + \int_0^t \left[\frac{1}{\epsilon} b(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) + c(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right] ds + \int_0^t \sigma(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) dW_s \\ Y_t^{\epsilon, \nu} &= \zeta + \int_0^t \frac{1}{\epsilon} \left[\frac{1}{\epsilon} f(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) + g(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right] dt \\ &\quad + \frac{1}{\epsilon} \int_0^t \left[\tau_1(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) dW_s + \int_0^t \tau_2(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) dB_s \right]. \end{aligned}$$

Equation (1) is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ with $\{\mathcal{F}_t\}$ satisfying the usual conditions, where $b, c, f, g, : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$, $\sigma, \tau_1, \tau_2 : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times m}$, B_t, W_t are independent standard m -dimensional $\{\mathcal{F}_t\}$ -Brownian motions, $\eta \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$ with $\eta \sim \nu$, $\zeta \in L^p(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$ for all $p > 0$, and (η, ζ) is independent of (W, B) . Here and throughout $\mathcal{P}_2(\mathbb{R}^d)$ denotes the space of probability measures on \mathbb{R}^d with finite second moment, equipped with the 2-Wasserstein metric (see Appendix B).

Note the superscript $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ on $(X^{\epsilon, \nu}, Y^{\epsilon, \nu})$, which is parameterizing the distribution of $X_0^{\epsilon, \nu} = \eta$. This parameterization is important for the formulation of solutions to the Cauchy-Problem on Wasserstein Space (10), which is employed to prove our main result, Theorem 3.1. See Section 5 and in particular Remark 5.2 for further discussion of this choice of parameterization and its importance.

The theory of averaging for diffusion process with coefficients that do not depend on the law of the solution itself, i.e., that do not depend on $\mathcal{L}(X^{\epsilon, \nu})$, is a classical result by now and had been studied under different assumptions and settings, see for example [2, 5, 47, 48, 50, 53, 57] to name a few. Existing averaging results for slow-fast McKean-Vlasov SDEs can be found in [31, 39, 54, 55, 62]. In particular, only systems where L^2 convergence rates can be found, possible for instance when $b = 0$ and $\sigma(x, y, \mu) = \sigma(x, \mu)$, have been considered in the McKean-Vlasov setting, leaving the fully-coupled cases unsolved. Even for standard diffusions (where the coefficients do not depend on the law of the solution), the only existing result for rates of convergence in distribution can be found in Theorem 2.3 of [53]. We close this literature review mentioning the very recent preprints [32, 43, 51], dealing with various aspects of averaging for McKean-Vlasov SDE systems, that appeared on arXiv after this paper had appeared on arXiv and was submitted to the journal.

Let $a(x, y, \mu) = \frac{1}{2}[\tau_1 \tau_1^\top + \tau_2 \tau_2^\top](x, y, \mu)$. For $x \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)$, we define the differential operator $\mathcal{L}_{x, \mu}$ acting on $\phi \in C_b^2(\mathbb{R}^d)$ by

Date: August 24, 2022.

2010 *Mathematics Subject Classification.* 60F05, 60F17, 60G17, 60J60.

Key words and phrases. multiscale processes, empirical measure, McKean-Vlasov process, ergodic theorems, averaging, homogenization.

This work has been partially supported by the National Science Foundation (DMS 2107856) and Simons Foundation Award 672441.

$$(2) \quad \mathcal{L}_{x,\mu}\phi(y) = f(x, y, \mu) \cdot \nabla\phi(y) + a(x, y, \mu) : \nabla^2\phi(y).$$

This is the generator obtained from considering the $O(1/\epsilon^2)$ terms from the generator of $(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu})$ and “freezing” the terms associated to the slow process at fixed $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Under assumptions A1) and A2) below, by [47] Proposition 1 (see also [60]), there exists a $\pi(\cdot; x, \mu)$ which is the unique probability measure solving the distributional equation

$$(3) \quad \mathcal{L}_{x,\mu}^* \pi = 0.$$

Moreover, all moments of π are bounded uniformly in x and μ .

For π as in Equation (3), under the standard centering condition

$$(4) \quad \int_{\mathbb{R}^d} b(x, y, \mu) \pi(dy; x, \mu) = 0, \forall x \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

and other technical assumption to be stated later on (see Assumptions A1)-A3)), by Lemma A.1 we may also consider Φ the unique classical solution to:

$$(5) \quad \begin{aligned} \mathcal{L}_{x,\mu} \Phi_k(x, y, \mu) &= -b_k(x, y, \mu), k \in \{1, \dots, d\} \\ \int_{\mathbb{R}^d} \Phi(x, y, \mu) \pi(dy; x, \mu) &= 0. \end{aligned}$$

Define:

$$(6) \quad \begin{aligned} \gamma(x, y, \mu) &:= \gamma_1(x, y, \mu) + c(x, y, \mu) \\ \gamma_1(x, y, \mu) &:= \partial_x \Phi(x, y, \mu) b(x, y, \mu) + \partial_y \Phi(x, y, \mu) g(x, y, \mu) + \tau_1(x, y, \mu) \sigma^\top(x, y, \mu) : \partial_x \partial_y \Phi(x, y, \mu) \\ D(x, y, \mu) &:= D_1(x, y, \mu) + D_1^\top(x, y, \mu) + \frac{1}{2} \sigma(x, y, \mu) \sigma^\top(x, y, \mu) \\ D_1(x, y, \mu) &:= \frac{1}{2} [b(x, y, \mu) \otimes \Phi(x, y, \mu) + \partial_y \Phi(x, y, \mu) \tau_1(x, y, \mu) \sigma^\top(x, y, \mu)] \end{aligned}$$

and

$$(7) \quad \begin{aligned} \bar{\gamma}(x, \mu) &:= \left[\int_{\mathbb{R}^d} \gamma(x, y, \mu) \pi(dy; x, \mu) \right] \\ \bar{D}(x, \mu) &:= \left[\int_{\mathbb{R}^d} D(x, y, \mu) \pi(dy; x, \mu) \right]. \end{aligned}$$

Here

$$\tau_1 \sigma^\top : \partial_x \partial_y \Phi(x, y, \mu) := [\tau_1 \sigma^\top : \partial_x \partial_y \Phi_1(x, y, \mu), \dots, \tau_1 \sigma^\top : \partial_x \partial_y \Phi_d(x, y, \mu)]^\top.$$

In this paper, we will establish a rate of convergence of $\mathcal{L}(X_t^{\epsilon,\nu})$ to $\mathcal{L}(X_t^\nu)$ in terms of sufficiently smooth test functions on the space $\mathcal{P}_2(\mathbb{R}^d)$, where X_t^ν satisfies the averaged McKean-Vlasov SDE:

$$(8) \quad X_t^\nu = \eta^2 + \int_0^t \bar{\gamma}(X_s^\nu, \mathcal{L}(X_s^\nu)) ds + \int_0^t \sqrt{2} \bar{D}^{1/2}(X_s^\nu, \mathcal{L}(X_s^\nu)) dW_s^2.$$

Here the equation is posed on a possibly different filtered probability space satisfying the usual conditions and supporting a d -dimensional Brownian motion W^2 , η^2 is random variable on this new probability space independent from W^2 and equal in distribution to $X_0^{\epsilon,\nu} = \eta$ - that is $\eta^2 \sim \nu \in \mathcal{P}_2(\mathbb{R}^d)$ - and once again the superscript ν in the notation X^ν is parameterizing the distribution of the initial condition. In addition, $\bar{D}^{1/2}(x, \mu)$ is the unique positive semi-definite matrix such that $\bar{D}^{1/2}(x, \mu) \bar{D}^{1/2}(x, \mu) = \bar{D}(x, \mu)$. Note that while \bar{D} is symmetric, it is not clear a priori that it is positive semi-definite. Thus we make the following remark:

Remark 1.1. One can find that the diffusion coefficient \bar{D} can be written in the alternative form

$$(9) \quad \begin{aligned} \bar{D}(x, \mu) &= \frac{1}{2} \int_{\mathbb{R}^d} \left(\partial_y \Phi(x, y, \mu) \tau_2(x, y, \mu) \tau_2^\top(x, y, \mu) [\partial_y \Phi(x, y, \mu)]^\top \right. \\ &\quad \left. + [\sigma(x, y, \mu) + \partial_y \Phi(x, y, \mu) \tau_1(x, y, \mu)] [\sigma(x, y, \mu) + \partial_y \Phi(x, y, \mu) \tau_1(x, y, \mu)]^\top \right) \pi(dy; x, \mu), \end{aligned}$$

and hence is both symmetric and positive semi-definite. This reduces to showing

$$\begin{aligned} & \int_{\mathbb{R}^d} (b(x, y, \mu) \otimes \Phi(x, y, \mu) + \Phi(x, y, \mu) \otimes b(x, y, \mu)) \pi(dy; x, \mu) \\ &= 2 \int_{\mathbb{R}^d} \partial_y \Phi(x, y, \mu) a(x, y, \mu) [\partial_y \Phi(x, y, \mu)]^\top \pi(dy; x, \mu), \end{aligned}$$

which can be seen via an integration by parts argument as per [50] Remarks 11.4/11.5.

Further than simply establishing this new averaging principle, we also establish a rate of convergence in distribution. Our main result is Theorem 3.1, while Corollary 3.2 specializes the main result in the important case of convergence for linear test functionals of the law. We make our results concrete in Section 3.1 for a class of Aggregation-Diffusion equations where we also note that the effect of the multiple scales is to decrease the magnitude of the effective interaction potential in all directions. When using classical methods, such as the martingale problem, to show convergence in distribution of stochastic process, rates of convergence can only possibly be found after serious added effort. An added difficulty of the general setting studied in this paper is that Φ solving (5) depends on the measure parameter μ , and thus terms involving derivatives of Φ with respect to μ appear in the prelimit expression for functions of the slow process and its law; see Proposition 4.3 (and analogously Proposition 4.4). These terms are handled using a novel coupling argument and an extended Poisson equation (27) (a doubled corrector problem).

To our knowledge, the only result providing rates of convergence in distribution in the fully-coupled setting for standard diffusion processes (which do not depend on their law, i.e. functional derivatives with respect to μ and terms that we have to deal with in Propositions 4.3 and 4.4 do not appear there), are found in [53] as Theorem 2.3. The insight provided by the proof of that Theorem is to write the difference $\mathbb{E} \left[\phi(X_t^\epsilon) - \phi(X_t) \right]$ in terms of the solution to the Cauchy problem (backward Kolmogorov equation) associated to the limiting system (here X^ϵ and X are solutions of a standard SDEs, not McKean-Vlasov SDEs).

Though not stated explicitly in [53], the constant such that $\sup_{t \in [0, T]} \mathbb{E} \left[\phi(X_t^\epsilon) - \phi(X_t) \right] \leq C\epsilon$ (considering here Regime 4 and $\vartheta = 1$ in Theorem 2.3) can be seen to depend linearly on the norm of the test function ϕ in an appropriate function space. Viewing $\mathcal{L}(X_t^\epsilon)$ and $\mathcal{L}(X_t)$ as elements of the dual of this space, such estimates on the “operator norm” of these probability measures are key in establishing tightness for fluctuation processes which establish a functional CLT related to the propagation of chaos for standard McKean-Vlasov SDEs [19, 24, 30, 40, 58]. This is, in fact, a key source of inspiration for this work, as it provides a needed rate of convergence in distribution for an intermediary Slow-Fast McKean-Vlasov process to its averaged limit in the proof of tightness of the fluctuations process for which a large deviations principle for is established in [4].

As we will see, it is greatly beneficial in the McKean-Vlasov setting to consider, in lieu of the standard backward Kolmogorov equation associated to the averaged dynamics (8), the associated Cauchy-Problem on Wasserstein space:

$$\begin{aligned} (10) \quad \dot{U}(t, \mu) &= \int_{\mathbb{R}^d} \bar{\gamma}(z, \mu) \cdot \partial_\mu U(t, \mu)[z] + \bar{D}(z, \mu) : \partial_z \partial_\mu U(t, \mu)[z] \mu(dz), t \in (0, \tau], \mu \in \mathcal{P}_2 \\ U(0, \mu) &= G(\mu). \end{aligned}$$

The derivatives in the measure argument in the above equation are in the sense of Lions [7]. For the reader’s convenience, we have included in Appendix B a brief review on differentiation of functions on spaces of measures. For a more comprehensive exposition on this, we refer the interested reader to [8] Chapter 5.

Such equations were originally studied in [6]. They have been used to study propagation of chaos rates in [12], and a related but different PDE on Wasserstein space originally posed in [10] is used in the study Mean Field Games. We will use the recent result of [11] Theorem 2.15, which extends regularity of solutions to (10) beyond the two derivatives usually needed for these applications (stated here as Lemma 5.1).

To our knowledge, this is the first application of the Cauchy problem on Wasserstein space used to establish rates of convergence of the law of one McKean-Vlasov SDE to another. One benefit of our proof method via the use of the Cauchy Problem on Wasserstein space is that it allows for non-linear test functions on the space of measures, so that a Corollary of our proof method is somewhat of an extension of the current results on rates of averaging for Fully-Coupled standard diffusions - see Remark 3.3. For more discussion of the

reasoning behind and benefits of using solutions of Equation (10) over solutions to the standard backward Kolmogorov equation, see Section 5.

The rest of the paper is organized as follows. In Section 2 we go over notation and the assumptions that hold throughout the paper. Section 3 contains our main results together with illustrative examples. In Section 4 we present ergodic theorems relevant to the behavior of the system (1) as $\epsilon \rightarrow 0$. Section 5 discusses in detail the Cauchy Problem (10). The proof of Theorem 3.1 is in Section 6. Conclusions and a discussion on future work is the content of Section 7. The Appendix contains a number of technical results used in the paper. In particular, Appendix A contains regularity results on the Poisson equations studied in this paper. In Appendix B, we recall some notation and terminology associated to differentiation of functions on spaces of measures.

2. NOTATION AND ASSUMPTIONS

Let \mathbf{X} and \mathbf{Y} be a Polish spaces, and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mu)$ be a measure space. We will denote by $\mathcal{P}(\mathbf{X})$ the space of probability measures on \mathbf{X} with the topology of weak convergence, $\mathcal{P}_2(\mathbf{X}) \subset \mathcal{P}(\mathbf{X})$ the space of square integrable probability measures on \mathbf{X} with the 2-Wasserstein metric (see Definition B.1), $\mathcal{B}(\mathbf{X})$ the Borel σ -field of \mathbf{X} , $C(\mathbf{X}; \mathbf{Y})$ the space of continuous functions from \mathbf{X} to \mathbf{Y} , $C_b(\mathbf{X})$ the space of bounded, continuous functions from \mathbf{X} to \mathbb{R} with norm $\|\psi\|_\infty := \sup_{x \in \mathbf{X}} |\psi(x)|$, and $L^p(\tilde{\Omega}, \tilde{\mathcal{F}}, \mu; \mathbb{R}^d)$ the space of p -integrable functions on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mu)$ with values in \mathbb{R}^d (where if $\tilde{\Omega} = \mathbf{X}$ and no σ -algebra is provided we assume it is $\mathcal{B}(\mathbf{X})$). $C_b^k(\mathbb{R}^d)$ for $k \in \mathbb{N}$ will note the space of functions with k continuous and bounded derivatives on \mathbb{R}^d , with norm $\|\psi\|_{C_b^k(\mathbb{R}^d)} = \sum_{j=0}^k \|\nabla^j \psi\|_\infty$, and $C_b^{1,k}([0, T] \times \mathbb{R}^d)$ will denote continuous functions ψ on $[0, T] \times \mathbb{R}^d$ with a continuous, bounded time derivative on $(0, T)$, denoted $\dot{\psi}$, such that $\|\psi\|_{C_b^{1,k}([0, T] \times \mathbb{R}^d)} := \sup_{t \in [0, T], x \in \mathbb{R}^d} |\dot{\psi}(t, x)| + \sup_{t \in [0, T]} \|\psi(t, \cdot)\|_{C_b^k(\mathbb{R}^d)} < \infty$. $C_{b,L}^k(\mathbb{R}^d) \subset C_b^k(\mathbb{R}^d)$ is the space of functions in $C_b^k(\mathbb{R}^d)$ such that all k derivatives are Lipschitz continuous. For $\phi \in L^1(\mathbf{X}, \mu)$, $\mu \in \mathcal{P}(\mathbb{R}^d)$ we define $\langle \mu, \phi \rangle := \int_{\mathbf{X}} \phi(x) \mu(dx)$. For $a, b \in \mathbb{R}$, we will denote $a \wedge b = \min\{a, b\}$. C will be used for a constant which may change from line to line throughout, and when there are parameters a_1, \dots, a_n which C depends on in an important manner, will denote this dependence by $C(a_1, \dots, a_n)$. For all function spaces, the codomain is assumed to be \mathbb{R} unless otherwise denoted.

In the proof of the Theorem 3.1, we will be making use of regularity of solutions to a Cauchy Problem on Wasserstein space (see Equation (10)), for which we will need to establish existence, uniqueness, and regularity of some number of derivatives. In doing so, will be controlling many mixed derivatives of functions in the Lions sense and in the standard sense, it will be useful for us to borrow the multi-index notation proposed in [15] and employed in [11]. We will also need to ensure that the derivatives of Φ which appear in the definition the limiting coefficients in Equation (6) are well-defined and integrable against π . We thus extend the multi-index notation from the aforementioned papers to track specific collections of mixed partial derivatives, and to capture needed assumptions of local Hölder continuity and polynomial growth in y .

Definition 2.1. Let n, l, k be non-negative integers and $\beta = (\beta_1, \dots, \beta_n)$ be an n -dimensional vector of non-negative integers. We call any ordered tuple of the form (n, l, β) a **multi-index**. For a function $G : \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$, we will denote for a multi-index (n, l, β)

$$D^{(n, l, \beta)} G(x, \mu)[z_1, \dots, z_n] = \partial_{z_1}^{\beta_1} \dots \partial_{z_n}^{\beta_n} \partial_x^l \partial_\mu^n G(x, \mu)[z_1, \dots, z_n]$$

if this derivative is well defined. As noted in the Remark B.4, for such a derivative to be well defined we require for it to be jointly continuous in x, μ, z_1, \dots, z_n where the topology used in the measure component is that of $\mathcal{P}_2(\mathbb{R}^d)$.

Definition 2.2. For ζ a collection of multi-indices of the form $(n, l, \beta) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}^n$, we will call ζ a **complete** collection of multi-indices if for any $(n, l, \beta) \in \zeta$, $\{(k, j, \alpha(k)) : \alpha(k) = (\alpha_1, \dots, \alpha_k), \alpha_p \leq \beta(k)_p, \forall \beta(k) = (\beta(k)_1, \dots, \beta(k)_k) \in \binom{\beta}{k}, k \leq n, j \leq l\} \subset \zeta$. Here for a vector of positive integers $\beta = (\beta_1, \dots, \beta_n)$ and $k \in \mathbb{N}, k \leq n$, we are using the notation $\binom{\beta}{k}$ to represent the set of size $\binom{n}{k}$ containing all the k -dimensional vectors of positive integers which can be obtained from removing $n - k$ entries from β .

Remark 2.3. Definition 2.2 is essentially enforcing that if collection of multi-indices contains a multi-index representing some mixed derivative in (x, μ, z) as per Definition 2.1, then it also contains all lower-order mixed derivatives of the same type. For instance, let ζ be the collection of multi-indices containing

$(2, 0, (1, 1))$ (corresponding to $\partial_{z_1} \partial_{z_2} \partial_\mu^2 G(x, \mu)[z_1, z_2]$). Then, in order to be complete, ζ must also contain $(2, 0, (1, 0))$, $(2, 0, (0, 1))$, $(2, 0, 0)$, $(1, 0, 1)$, $(1, 0, 0)$, and $(0, 0, 0)$ (corresponding to $\partial_{z_1} \partial_\mu^2 G(x, \mu)[z_1, z_2]$, $\partial_{z_2} \partial_\mu^2 G(x, \mu)[z_1, z_2]$, $\partial_\mu^2 G(x, \mu)[z_1, z_2]$, $\partial_z \partial_\mu G(x, \mu)[z]$, $\partial_\mu G(x, \mu)[z]$, and $G(x, \mu)$ respectively). This is a technical requirement used in order to state the results in Appendix A in a way that allows the inductive arguments used therein to go through.

Using this multi-index notation, it will be useful for us to define some spaces regarding regularity of functions in regard to these mixed derivatives. We thus make the following modifications to Definition 2.13 in [11]:

Definition 2.4. For ζ a collection of multi-indices of the form $(n, l, \beta) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}^n$ and $k, j, d \in \mathbb{N}$ (or $k \in \mathbb{N} \times \mathbb{N}$ to denote matrix-valued functions), we define $\mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ to be the class of functions $G : \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ such that $D^{(n, l, \beta)} G(x, \mu)[z_1, \dots, z_n]$ exists and satisfies

$$(11) \quad \|G\|_{\mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)} := \sup_{(n, l, \beta) \in \zeta} \sup_{x \in \mathbb{R}^j, z_1, \dots, z_n \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |D^{(n, l, \beta)} G(x, \mu)[z_1, \dots, z_n]| \leq C.$$

We denote the class of functions $G \in \mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ such that:

$$(12) \quad |D^{(n, l, \beta)} G(x, \mu)[z_1, \dots, z_n] - D^{(n, l, \beta)} G(x', \mu')[z'_1, \dots, z'_n]| \leq C_L \left(|x - x'| + \sum_{i=1}^N |z_i - z'_i| + \mathbb{W}_2(\mu, \mu') \right)$$

for all $(n, l, \beta) \in \zeta$ and $x, x' \in \mathbb{R}^j, z_1, \dots, z_n, z'_1, \dots, z'_n \in \mathbb{R}^d, \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ by $\mathcal{M}_{b,L}^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$. We define $\mathcal{M}_b^\zeta(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ and $\mathcal{M}_{b,L}^\zeta(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ analogously, where instead here ζ a collection of multi-indices of the form $(n, \beta) \in \mathbb{N} \times \mathbb{N}^n$, and we take the $l = 0$ in the above multi-index notation for the derivatives.

We will also make use of the class of functions $\mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ which contains $G : \mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ such that $G(\cdot, y, \cdot) \in \mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ for all $y \in \mathbb{R}^j$, for each multi-index $(n, l, \beta) \in \zeta$, there exists $m \in \mathbb{N}$ and $C > 0$ such that:

$$(13) \quad \sup_{x \in \mathbb{R}^j, z_1, \dots, z_n \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |D^{(n, l, \beta)} G(x, y, \mu)[z_1, \dots, z_n]| \leq C(1 + |y|^m),$$

and there exists $\theta \in (0, 1]$ such that for all $(n, l, \beta) \in \zeta$ and $y, y', x, x' \in \mathbb{R}^j, z_1, \dots, z_n, z'_1, \dots, z'_n \in \mathbb{R}^d, \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$:

$$(14) \quad |D^{(n, l, \beta)} G(x, y, \mu)[z_1, \dots, z_n] - D^{(n, l, \beta)} G(x', y', \mu')[z'_1, \dots, z'_n]| \leq C \left(1 \wedge |y - y'|^\theta + |x - x'| + \sum_{i=1}^N |z_i - z'_i| + \mathbb{W}_2(\mu, \mu') \right) \left(1 + |y|^m + |y'|^m \right).$$

We also define $\mathcal{M}_b^\zeta([0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ to be the class of functions $G : [0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ such that $G(\cdot, x, \mu)$ is continuously differentiable on $(0, T)$ for all $x \in \mathbb{R}^j, \mu \in \mathcal{P}_2(\mathbb{R}^d)$ with time derivative denoted by $\dot{G}(t, x, \mu)$, $G(t, \cdot, \cdot) \in \mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ for all $t \in [0, T]$, with (11) holding uniformly in t , and G, \dot{G} , and all derivatives involved in the definition of $\mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ are jointly continuous in time, measure, and space. We define for $G \in \mathcal{M}_b^\zeta([0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$

$$\|G\|_{\mathcal{M}_b^\zeta([0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)} := \sup_{t \in [0, T]} \|G(t, \cdot)\|_{\mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)} + \sup_{t \in [0, T], x \in \mathbb{R}^j, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\dot{G}(t, x, \mu)|.$$

We denote the class of functions $G \in \mathcal{M}_b^\zeta([0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ such that (12) holds uniformly in t by $\mathcal{M}_{b,L}^\zeta([0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$. Again, we define $\mathcal{M}_b^\zeta([0, T] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ and $\mathcal{M}_{b,L}^\zeta([0, T] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ analogously.

It will be useful to define the following complete collections of multi-indices in the sense of Definitions 2.1 and 2.2:

$$(15) \quad \begin{aligned} \hat{\zeta} &:= \{(0, j_1, 0), (1, j_2, j_3), (2, j_4, (j_5, j_6)), (3, 0, (j_7, 0, 0)) : j_1 \in \{0, 1, \dots, 4\}, j_2 + j_3 \leq 4, j_4 + j_5 + j_6 \leq 2, j_7 = 0, 1\} \\ \hat{\zeta}_1 &:= \{(0, j_1, 0), (1, j_2, j_3), (2, j_4, (j_5, j_6)), (3, j_7, (j_8, 0, 0)) \\ &\quad : j_1 \in \{0, 1, \dots, 5\}, j_3 \leq 4, j_2 + j_3 \leq 5, j_5 + j_6 \leq 2, j_4 + j_5 + j_6 \leq 3, j_7 + j_8 \leq 1\}. \end{aligned}$$

Next, we introduce the main assumptions to be used throughout this paper. For a detailed explanation of the use of these assumptions, and in particular of A5), which is stated in terms of the collections of multi-indices from Equation (15), see Remark 2.6.

- A1) There exist $\lambda_-, \lambda_+ > 0$ such that $0 < \lambda_- \leq \frac{z^\top [\tau_1 \tau_1^\top + \tau_2 \tau_2^\top](x, y, \mu) z}{|z|^2} \leq \lambda_+ < \infty$, $\forall x, y, z \in \mathbb{R}^d, z \neq 0, \mu \in \mathcal{P}_2(\mathbb{R}^d)$ and τ_1, τ_2 are bounded, have two uniformly bounded derivatives in y , and τ_1, τ_2 and both these derivatives are Hölder continuous in y uniformly in (x, μ) .
- A2) There exists constants $C, \beta > 0$ independent of x, y, μ such that:
- (16)
$$f(x, y, \mu) \cdot y \leq -\beta|y|^2 + C, \forall x, y \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

f grows at most linearly in $|y|$, f has two uniformly bounded derivatives in y , and f and both these derivatives are Hölder continuous in y uniformly in (x, μ) .

- A3) As is standard in averaging results, we assume also the centering condition. Namely, we assume that for π as in Equation (3), the condition (4) holds.

In terms of regularity of coefficients, we assume the following assumptions on the coefficients of (1):

- A4) For $F = b, c, \sigma, f, g, \tau_1$, or τ_2 , F is globally Lipschitz continuous in (x, y, μ) . That is, there exists $C > 0$ such that for all $x, x', y, y' \in \mathbb{R}^d$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$:
- (17)
$$|F(x, y, \mu) - F(x', y', \mu')| \leq C(|x - x'| + |y - y'| + \mathbb{W}_2(\mu, \mu')).$$

Moreover, g is uniformly bounded.

- A5) $g, c \in \mathcal{M}_p^{\dot{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, $\sigma, \tau_1 \in \mathcal{M}_p^{\dot{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times m})$, $f, b \in \mathcal{M}_p^{\dot{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, and $a \in \mathcal{M}_p^{\dot{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$.
- A6) For \bar{D} as in Equation (7), there exists $\bar{\lambda}_- > 0$ such that $0 < \bar{\lambda}_- \leq \frac{z^\top \bar{D}(x, \mu) z}{|z|^2}$, $\forall x, z \in \mathbb{R}^d, z \neq 0, \mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Remark 2.5. Note that under Assumption A4), for each choice of square integrable initial conditions (η, ζ) and each $\epsilon > 0$, there exists a unique solution $\{(X_t^{\epsilon, \nu}, Y_t^{\epsilon, \nu}), t \geq 0\}$ to the system (1) such that $\mathcal{L}(X_t^{\epsilon, \nu}) \in \mathcal{P}_2(\mathbb{R}^d)$ for each $t \geq 0$. See, e.g. [61] Theorem 2.1 and Section 6.1 in [54]. There are weaker assumptions under which existence and uniqueness for McKean-Vlasov SDEs have been established in the recent literature which may replace the global Lipschitz assumption (17) (see, e.g. [29, 35, 36, 44, 52]), but we chose the simplest of these in order to clearly illustrate our results.

Moreover, under Assumptions A1), A2), A3), A5), and A6), Proposition A.8 yields that $\bar{\gamma}, \bar{D}^{1/2}$ are globally Lipschitz in (x, μ) , so that the same result implies existence and uniqueness of solutions to the limiting averaged dynamics (8).

Remark 2.6. The high amount of regularity imposed on the coefficients and their derivatives in Assumption A5) is needed to establish the analogous regularity of the averaged coefficients appearing in the limiting Equation (8). This regularity of the averaged coefficients is needed for Lemma 5.1, which provides bounds on the derivatives of the solution to the Cauchy Problem on Wasserstein space (10) which appear in the proof of Theorem 3.1—that is, those which are contained in $\dot{\zeta}$ as defined in Theorem 3.1. Unpacking the multi-index notation, we need that U has bounded Lions derivatives up to order 3, that $\partial_\mu U(\mu)[z]$ has 3 bounded derivatives in z , that $\partial_\mu^2 U(\mu)[z_1, z_2]$ has bounded second order derivatives in z_1, z_2 , and that $\partial_\mu^3 U(\mu)[z_1, z_2, z_3]$ has bounded first order derivatives in z_1, z_2, z_3 (by symmetry we can just assume this in one of the auxiliary variables).

In Lemma 5.1, we use the results of [6] as extended in [11], where the proof method is via a “variational approach” that requires Lipschitz continuity in x, μ of many derivatives of the coefficients of the PDE (10) in order to establish existence and uniqueness of variational equations related to the derivatives of the associated process (8) in its initial conditions. The way the result is stated in [11], the requirement that $U \in \mathcal{M}_b^{\dot{\zeta}}(\mathcal{P}_2(\mathbb{R}^d))$ would be summarized as U having all derivatives in (μ, z) of order 4 bounded, and the sufficient requirement on the coefficients in $\bar{\gamma}, \bar{D}$ appearing in Equation (10) would be that $\bar{\gamma}, \bar{D}^{1/2}$ have all derivatives in (x, μ, z) of order 4 bounded. However, we don’t require U be 4 times differentiable in μ , so we can require slightly less regularity of $\bar{\gamma}, \bar{D}^{1/2}$, resulting in the collection of multi-indices $\dot{\zeta}$. The assumption that some $F \in \mathcal{M}_p^{\dot{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$, as in A5), is requiring regularity of F , $\partial_x^{j_1} F$ for $j_1 = 1, 2, 3, 4$, $\partial_z^{j_3} \partial_x^{j_2} \partial_\mu F$ for $j_2 + j_3 \leq 4$, $\partial_{z_1}^{j_5} \partial_{z_2}^{j_6} \partial_x^{j_4} \partial_\mu^2 F$ for $j_4 + j_5 + j_6 \leq 2$, $\partial_\mu^3 F$, and $\partial_{z_1} \partial_\mu^3 F$. Note the lack

of requirement of 4-times differentiability in μ . Also note that this assumption does not impose regularity in the y argument, other than the joint Hölder continuity imposed by Equation (14) in Definition 2.4. The collection of multi-indices $\hat{\zeta}_1$ is the result of adding one more derivative in x to the derivatives represented by $\hat{\zeta}$. This is needed to ensure the terms $\partial_x \Phi$ and $\partial_x \partial_y \Phi$ appearing in $\bar{\gamma}$ have the $\hat{\zeta}$ -regularity required for Lemma 5.1. See Proposition A.8 for how the regularity of the prelimit coefficients assumed in A5) implies the required regularity of the limiting coefficients.

It is likely that the regularity of the limiting, and hence prelimit, coefficients and of the initial condition G that is required in order to establish boundedness of these derivatives of solutions to (10) can be weakened to, e.g. Hölder continuity of some lesser number of derivatives, as in the case for the standard Cauchy problem [23, 25, 42]. Partial results in this direction via Malliavin calculus techniques [15] and an infinite-dimensional parametrix method [12, 13] already exist. Extending these results to the higher number of derivatives needed in Theorem 3.1 is an interesting avenue of future research, and is beyond the scope of this paper.

It is also worth noting that, with the exception of the boundedness of g, τ_1 , and τ_2 as well as the dissipativity assumption A2), which are used in Lemma 4.1, all of the other imposed regularity in the Assumptions A1)-A6) are only needed to provide (weak) existence and uniqueness of the prelimit and limiting system ((1) and (8) respectively), the aforementioned sufficient regularity for Lemma 5.1, and the needed regularity and existence/uniqueness of the auxiliary Poisson Equations used in Section 4. These are simply sufficient conditions, and are by no means necessary, and these properties can also be proved on a case-by-case basis. Observe that the only assumption imposed on the limiting coefficients here is A6), which by the representation provided in Equation (9) will hold in most situations.

3. MAIN RESULTS AND EXAMPLES

We are now ready to state our main results:

Theorem 3.1. *Assume A1)-A6). Define*

$$\hat{\zeta} := \{(0, 0), (1, j_1), (2, (j_2, j_3)), (3, (0, 0, j_4)) : j_1 \in \{0, 1, 2, 3\}, j_2 + j_3 \leq 2, j_4 = 0, 1\}.$$

Then for any $G \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and $T > 0$, there is $C(T)$ independent of G and ν such that for $\epsilon \in (0, 1]$:

$$\sup_{s \in [0, T]} \left| G(\mathcal{L}(X_s^{\epsilon, \nu})) - G(\mathcal{L}(X_s^\nu)) \right| \leq \epsilon C(T) \|G\|_{\mathcal{M}_b^{\hat{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}.$$

Proof. The proof is found in Section 6. □

Corollary 3.2. *In the setup of Theorem 3.1, for any $\phi \in C_{b,L}^4(\mathbb{R}^d)$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, we have there is $C(T)$ independent of ϕ and ν such that:*

$$\sup_{s \in [0, T]} \left| \mathbb{E}[\phi(X_s^{\epsilon, \nu})] - \mathbb{E}[\phi(X_s^\nu)] \right| \leq \epsilon C(T) |\phi|_{C_b^4(\mathbb{R}^d)}.$$

Proof. Considering G of the form $G_\phi(\mu) = \langle \mu, \phi \rangle$ for $\phi \in C_{b,L}^4(\mathbb{R}^d)$, $\partial_\mu G_\phi(\mu)[z] = \nabla \phi(z)$ (see, e.g. [8] Section 5.2.2 Example 1). Thus, $G_\phi \in \mathcal{M}_b^{\hat{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, and $\|G_\phi\|_{\mathcal{M}_b^{\hat{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})} \leq |\phi|_{C_b^4(\mathbb{R}^d)}$. The result then follows immediately via an application of Theorem 3.1. □

Remark 3.3. Theorem 3.1 holds in the situation where the coefficients in Equation (1) are independent of μ . In this setting, Assumption A5) is imposing that 4 derivatives of g, σ, τ_1, c in x and 5 derivatives of f, a, b in x grow at most polynomially in y and are jointly Lipschitz/Locally Hölder continuous in (x, y) in the sense of Equation (12). This thus extends the results of [53] Theorem 2.3 Regime 4 with $\vartheta = 1$ from test functions of the form G_ϕ as in Corollary 3.2 to the more general class of (possibly non-linear) test functions in $\mathcal{M}_{b,L}^{\hat{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$.

3.1. Examples: A Class of Aggregation-Diffusion Equations. A common form for interacting particle systems which are widely used in many settings such as in biology, ecology, social sciences, economics, molecular dynamics, and in study of spatially homogeneous granular media (see [1, 27, 38, 45] and the references therein) is:

$$(18) \quad dX_t^{i,N} = -\nabla V(X_t^{i,N})dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N})dt + \sigma dW_t^i, \quad X_0^{i,N} = \eta^i,$$

where $V : \mathbb{R}^d \rightarrow \mathbb{R}$ is a sufficiently smooth confining potential, $W : \mathbb{R}^d \rightarrow \mathbb{R}$ is a sufficiently smooth interaction potential, $\sigma > 0$, and W^i are iid d -dimensional Brownian motions. The class of systems (18) contains the system in the seminal paper [17], where many mathematical aspects of a model for cooperative behavior in a bi-stable confining potential with attraction to the mean are explored.

Under sufficient regularity on V and W and exchangeability assumptions on the initial conditions, as $N \rightarrow \infty$, the particles $X^{i,N}$ become IID and equal in distribution to \tilde{X} obeying the dynamics of the Aggregation-Diffusion Equation

$$(19) \quad d\tilde{X}_t = -\nabla V(\tilde{X}_t)dt - \bar{\mathbb{E}}[\nabla W(x - \tilde{X}_t)]|_{x=\tilde{X}_t} dt + \sigma dW_t, \quad \tilde{X}_0 = \eta,$$

where here we are denoting by \tilde{X}_t an independent copy of \tilde{X}_t on another probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$, and by $\bar{\mathbb{E}}$ the expectation on that space, and the distribution of η is determined by that of the η^i 's.

Inspired by the pervasiveness of such systems in the literature, we consider the fully-coupled slow-fast system of Aggregation-Diffusions Equations:

$$(20) \quad \begin{aligned} dX_t^{\epsilon,\nu} &= -[\nabla V_1(X_t^{\epsilon,\nu}) + \frac{1}{\epsilon} \nabla V_2(Y_t^{\epsilon,\nu})]dt - \bar{\mathbb{E}}[\nabla W_1(x - \bar{X}_t^{\epsilon,\nu})]|_{x=X_t^{\epsilon,\nu}} dt + \sigma dW_t \\ dY_t^{\epsilon,\nu} &= -\frac{1}{\epsilon} [\nabla V_3(X_t^{\epsilon,\nu}) + \frac{1}{\epsilon} \nabla V_4(Y_t^{\epsilon,\nu})]dt - \frac{1}{\epsilon} \bar{\mathbb{E}}[\nabla W_2(x - \bar{X}_t^{\epsilon,\nu})]|_{x=X_t^{\epsilon,\nu}} dt + \frac{1}{\epsilon} \tau_1 dW_t + \frac{1}{\epsilon} \tau_2 dB_t \\ (X_0^{\epsilon,\nu}, Y_0^{\epsilon,\nu}) &= (\eta, \zeta), \end{aligned}$$

where $V_k : \mathbb{R}^d \rightarrow \mathbb{R}, k = 1, \dots, 4$, $W_1, W_2 : \mathbb{R}^d \rightarrow \mathbb{R}$, $\bar{X}^{\epsilon,\nu}, \bar{\mathbb{E}}$ as in Equation (19), and the rest of the setup is as in Equation (1) with $d = m$.

This falls into the class of systems (1) with

$$\begin{aligned} f(x, y, \mu) &= -\nabla V_4(y) \\ b(x, y, \mu) &= -\nabla V_2(y) \\ c(x, y, \mu) &= -\nabla V_1(x) - \langle \mu, \nabla W_1(x - \cdot) \rangle \\ g(x, y, \mu) &= -\nabla V_3(x) - \langle \mu, \nabla W_2(x - \cdot) \rangle \\ \sigma(x, y, \mu) &\equiv \sigma I \\ \tau_1(x, y, \mu) &\equiv \tau_1 I \\ \tau_2(x, y, \mu) &\equiv \tau_2 I. \end{aligned}$$

Example 3.4. Consider the system (20).

Suppose $\alpha := \frac{1}{2}[\tau_1^2 + \tau_2^2] > 0$, there exists $C, \beta > 0$ such that $\nabla V_4(y) \cdot y + C \geq \beta|y|^2$, ∇V_4 grows at most linearly in y and has two locally Hölder continuous, bounded derivatives, ∇V_2 is Lipschitz continuous, and $V_1, V_3 \in C_{b,L}^5(\mathbb{R}^d)$, $W_1, W_2 \in C_{b,L}^6(\mathbb{R}^d)$.

Moreover, assume

$$\int_{\mathbb{R}^d} \nabla V_2(y) \exp(-V_4(y)/\alpha) dy = 0.$$

Then for any $G \in \mathcal{M}_{b,L}^\dagger(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and $T > 0$:

$$\sup_{s \in [0, T]} \left| G(\mathcal{L}(X_s^{\epsilon,\nu})) - G(\mathcal{L}(X_s^\nu)) \right| \leq \epsilon C(T) \|G\|_{\mathcal{M}_b^\dagger(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})},$$

where here X_t^ν satisfies:

$$(21) \quad dX_t^\nu = -[\alpha_1 \nabla V_3(X_t^\nu) + \nabla V_1(X_t^\nu)]dt - \bar{\mathbb{E}}[\alpha_1 \nabla W_2(x - \bar{X}_t^\nu) + \nabla W_1(x - \bar{X}_t^\nu)]|_{x=X_t^\nu} dt$$

$$\begin{aligned}
& + [\sigma^2 I + 2\alpha\alpha_2 + \sigma\tau_1[\alpha_1 + \alpha_1^\top]]^{1/2} dW_t^2 \\
X_0^\nu &= \eta^2 \sim \nu \\
\alpha_1 &= Z^{-1} \int_{\mathbb{R}^d} \nabla \Phi(y) \exp\left(\frac{-V_4(y)}{\alpha}\right) dy \\
\alpha_2 &= Z^{-1} \int_{\mathbb{R}^d} \nabla \Phi(y) (\nabla \Phi)^\top(y) \exp\left(\frac{-V_4(y)}{\alpha}\right) dy \\
Z &= \int_{\mathbb{R}^d} \exp\left(\frac{-V_4(y)}{\alpha}\right) dy.
\end{aligned}$$

Proof. In this setting the invariant measure π from Equation (3) admits a density of the form

$$\pi(y) = Z^{-1} \exp\left(\frac{-V_4(y)}{\alpha}\right).$$

Thus Assumptions A1)-A3) hold by supposition. In addition, noting that as per, e.g. [8] Section 5.2.2 Example 1, $\partial_\mu c(x, y, \mu)[z] = -\nabla^2 W_1(x - z)$ and similarly for g , we can see that the current assumptions also imply Assumptions A4) and A5). Here for the Lipschitz continuity of c and g in \mathbb{W}_2 , we use the boundedness of their Lions derivatives and Remark 5.27 in [8]. Note that Assumption A6) is only being used in Proposition A.8 in order to obtain that $\bar{D}^{1/2} \in \mathcal{M}_{b,L}^{\hat{S}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$ knowing $\bar{D} \in \mathcal{M}_{b,L}^{\hat{S}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, but here this fact is trivial since both are constant, and hence we need not worry about uniform positive-definiteness of \bar{D} . Thus, this is a direct application of Theorem 3.1. \square

An interesting subcase of the above example is when $V_2 = V_4 =: Q$ is 1-periodic in all directions, $V_1 = V_3 =: V, W_1 = W_2 =: W, \sigma = \tau_1 > 0, \tau_2 = 0$, so that $Y_t^{\epsilon, \nu}$ has the same dynamics as $X_t^{\epsilon, \nu}$ but order $1/\epsilon$ faster. In this situation, the dynamics of $Y^{\epsilon, \nu}$ are confined to the torus, so we need not worry about its integrability (in particular, we may drop Equation (16) from Assumption A2)). This corresponds to the standard Aggregation-Diffusion Equation (18), but where we replace the confining potential V with a rough potential $V^\epsilon(x) = V(x) + Q(x/\epsilon)$. Note that in this situation the centering condition A3) automatically holds.

In the case that we have a separable fluctuating part, that is $Q(y_1, \dots, y_d) = Q_1(y_1) + \dots + Q_d(y_d)$, the limiting equation (21) becomes more tractable, as everything is explicitly computable. It reads:

$$\begin{aligned}
(22) \quad dX_t^\nu &= -\Theta \nabla V(X_t^\nu) dt - \mathbb{E}[\Theta \nabla W(x - \bar{X}_t^\nu)]|_{x=X_t^\nu} dt + \sigma \Theta^{1/2} dW_t^2 \\
X_0^\nu &= \eta^2 \sim \nu \\
\Theta &:= \text{diag}\left[Z_1^{-1} \hat{Z}_1^{-1}, \dots, Z_d^{-1} \hat{Z}_d^{-1}\right] \\
Z_k &:= \int_0^1 \exp(-2Q_k(y)/\sigma^2) dy, \hat{Z}_k := \int_0^1 \exp(2Q_k(y)/\sigma^2) dy, k = 1, \dots, d.
\end{aligned}$$

Observe that $Z_k^{-1} \hat{Z}_k^{-1} \in (0, 1]$ for $k = 1, \dots, d$. Thus, the effect of averaging is not only that the effective diffusivity of the aggregation-diffusion equation is decreased and the magnitude of the effective confining potential is decreased in all directions (as is well understood- see e.g. [22, 63]), but also that the magnitude of the effective interaction potential is decreased in all directions. This is remarkable considering the fact that, considering the Aggregation-Diffusion Equation (18) as the limit of the particle system (19), the addition of multiscale structure through modifying V to V^ϵ is a priori only effecting the motion of each particle, not their means of interaction. Note that this is not a byproduct of the fact that we consider the limit as $\epsilon \downarrow 0$ after $N \rightarrow \infty$, as the limits have been shown to commute [3]. See Figure 1 for an example of this rescaling of both the confining and interaction potentials in the setting of the seminal paper [17], where the confining potential is Curie-Weiss and the interaction potential is quadratic. Note that in practice the interaction and diffusion potentials must be mollified so that they are bounded as $|x| \rightarrow \infty$ to fit into the regime of Example 3.4.

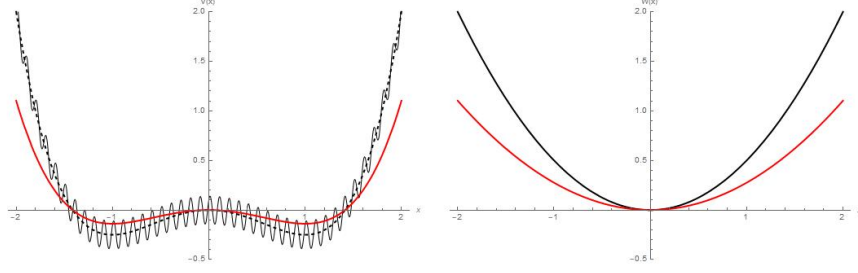


FIGURE 1. Left: A rough confining potential $V^\epsilon(x) = \frac{x^4}{4} - \frac{x^2}{2} + .1[\cos(2\pi x/\epsilon) + \sin(2\pi x/\epsilon)]$ with $\epsilon = .1$ (black) overlaid on its non-rough counterpart (dashed). The corresponding effective confining potential after averaging is given in red. Right: A prelimit interaction potential $W(x) = \frac{1}{2}x^2$ (black) and the corresponding effective interaction potential after averaging (red). The calculations are done with $\sigma = .5$.

4. ERGODIC-TYPE THEOREMS FOR FULLY-COUPLED MCKEAN-VLASOV SDES

In this subsection, we use the method of auxiliary Poisson equations to derive rates of averaging in the form of Ergodic-Type Theorems for the Slow-Fast McKean-Vlasov System (1). This method is standard for averaging of diffusions. In particular, the analogous result to Proposition 4.2 is necessary to see what the form of the limiting equation should be, and an analogous result is commonplace in the literature on averaging fully-coupled standard diffusions - see e.g. [48] Theorem 4 and [53] Lemma 4.4. A major difference here is, since we consider the Cauchy-Problem on Wasserstein Space (10), the test-function ψ 's domain is not only time and space, but also a measure component. When in some sense rates of convergence of the slow-fast system to the limiting averaged system must be established (such as when fluctuations are being considered), an analogous result to Proposition 4.5 is often employed - see e.g. [57] Lemma 4.1, [46] Lemma B.5, and [53] Lemma 4.2. For an analogous result in the context of McKean-Vlasov SDEs, see the proof of Theorem 2.5 in [54], although there they do not consider a Fully-Coupled system, so there is no need for the inclusion of the test function ψ .

It is worth noting that the terms being controlled in Propositions 4.3 and 4.4 below are unique to slow-fast McKean-Vlasov SDEs (or their weakly interacting particle system counterparts), and thus do not appear in the one-particle setting. Thus the “doubled Poisson equation” construction (see Equations (27) and (28)) and the proof of Propositions 4.3 and 4.4 are novel to this paper and its inspiring work [4]. We begin with the following Lemma, which provides a necessary bound on the moments of the fast process needed for the Ergodic-Type Theorems:

Lemma 4.1. Assume A1), A2), and A4). Then any $p \in \mathbb{N}$:

$$\sup_{\epsilon \in (0,1]} \sup_{t \in [0,T]} \mathbb{E} \left[|Y_t^{\epsilon,\nu}|^{2p} \right] \leq C(p) + \mathbb{E}[|\zeta|^{2p}].$$

Proof. We first note that, by Itô's formula (suppressing the arguments of the coefficients for readability):

$$\begin{aligned} \mathbb{E} \left[|Y_t^{\epsilon,\nu}|^{2p} \right] &= \mathbb{E}[|\zeta|^{2p}] + 2p \int_0^t \mathbb{E} \left[\frac{1}{\epsilon^2} \left(f \cdot Y_s^{\epsilon,\nu} |Y_s^{\epsilon,\nu}|^2 + (p-1)[\tau_1^2 + \tau_2^2] : Y_s^{\epsilon,\nu} \otimes Y_s^{\epsilon,\nu} \right) |Y_s^{\epsilon,\nu}|^{2p-4} \right] ds \\ &\quad + \frac{p}{\epsilon^2} \int_0^t \mathbb{E} \left[\text{Tr}(\tau_1^2 + \tau_2^2) |Y_s^{\epsilon,\nu}|^{2p-2} \right] ds + \frac{2p}{\epsilon} \int_0^t \mathbb{E} \left[g \cdot Y_s^{\epsilon,\nu} |Y_s^{\epsilon,\nu}|^{2p-2} \right] ds \\ &\quad + \frac{2p}{\epsilon} \mathbb{E} \left[\int_0^t |Y_s^{\epsilon,\nu}|^{2p-2} (Y_s^{\epsilon,\nu})^\top \tau_1 dW_s + \int_0^t |Y_s^{\epsilon,\nu}|^{2p-2} (Y_s^{\epsilon,\nu})^\top \tau_2 dB_s \right] \\ &= \mathbb{E}[|\zeta|^{2p}] + 2p \int_0^t \mathbb{E} \left[\frac{1}{\epsilon^2} \left(f \cdot Y_s^{\epsilon,\nu} |Y_s^{\epsilon,\nu}|^2 + (p-1)[\tau_1^2 + \tau_2^2] : Y_s^{\epsilon,\nu} \otimes Y_s^{\epsilon,\nu} \right) |Y_s^{\epsilon,\nu}|^{2p-4} \right] ds \\ &\quad + \frac{p}{\epsilon^2} \int_0^t \mathbb{E} \left[\text{Tr}(\tau_1^2 + \tau_2^2) |Y_s^{\epsilon,\nu}|^{2p-2} \right] ds + \frac{2p}{\epsilon} \int_0^t \mathbb{E} \left[g \cdot Y_s^{\epsilon,\nu} |Y_s^{\epsilon,\nu}|^{2p-2} \right] ds, \end{aligned}$$

using that by the boundedness of τ_1, τ_2, g and linear growth of f that for each $\epsilon > 0$ and $m \in \mathbb{N}$, $\mathbb{E} \left[\int_0^T |Y_s^{\epsilon, \nu}|^m ds \right] \leq C(\epsilon, m, T)$. Continuing, we have by Equation (16) and boundedness of τ_1, τ_2, g :

$$\begin{aligned} \frac{d}{dt} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] &\leq -\frac{2p\beta}{\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] + \frac{C(p)}{\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p-2} \right] + \frac{C(p)}{\epsilon} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p-1} \right] \\ &\leq -\frac{2p\beta}{\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] + \frac{p\beta}{2\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] + \frac{C(p)}{\epsilon^2} + \frac{C(p)}{\epsilon} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p-1} \right]. \end{aligned}$$

Here we wrote for $\eta > 0$:

$$\frac{C(p)}{\epsilon^2} |Y_s^{\epsilon, \nu}|^{2p-2} = \frac{1}{\epsilon^2} \frac{C(p)}{\eta} \eta |Y_s^{\epsilon, \nu}|^{2p-2} \leq \frac{1}{\epsilon^2} \frac{1}{p} \left(\frac{C(p)}{\eta} \right)^p + \frac{1}{\epsilon^2} \frac{2p-2}{2p} \left(\eta |Y_s^{\epsilon, \nu}|^{2p-2} \right)^{2p/(2p-2)}$$

using Young's inequality, and let $\eta = \left(\frac{p^2 \beta}{2(p-1)} \right)^{(2p-2)/(2p)}$.

Similarly, writing

$$\frac{C(p)}{\epsilon} |Y_s^{\epsilon, \nu}|^{2p-1} = \frac{C(p)}{\epsilon^{1-\alpha} \eta} \frac{\eta |Y_s^{\epsilon, \nu}|^{2p-1}}{\epsilon^\alpha} \leq \frac{1}{2p} \left(\frac{C(p)}{\epsilon^{1-\alpha} \eta} \right)^{2p} + \frac{2p-1}{2p} \left(\frac{\eta |Y_s^{\epsilon, \nu}|^{2p-1}}{\epsilon^\alpha} \right)^{2p/(2p-1)}$$

and letting $\alpha = \frac{2p-1}{p}$ and $\eta = \left(\frac{p^2 \beta}{2p-1} \right)^{(2p-1)/(2p)}$, we get

$$\begin{aligned} \frac{d}{dt} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] &\leq -\frac{2p\beta}{\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] + \frac{p\beta}{2\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] + \frac{C(p)}{\epsilon^2} + \frac{p\beta}{2\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] + \frac{C(p)}{\epsilon^{2-2p}} \\ &= -\frac{p\beta}{\epsilon^2} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] + \frac{C(p)}{\epsilon^2} + \frac{C(p)}{\epsilon^{2-2p}}. \end{aligned}$$

Now, recalling that if $g'(s) \leq -\gamma g(s) + C, \forall s \in [0, t]$, then $g(t) \leq C \int_0^t e^{-\gamma(t-s)} ds + e^{-\gamma t} g(0)$, we have

$$\begin{aligned} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2p} \right] &\leq C(p) \left[\frac{1}{\epsilon^2} + \frac{1}{\epsilon^{2-2p}} \right] \exp(-\frac{p\beta}{\epsilon^2} t) \int_0^t \exp(\frac{p\beta}{\epsilon^2} s) ds + \mathbb{E} [|\zeta|^{2p}] \exp(-\frac{p\beta}{\epsilon^2} t) \\ &= C(p) \left[\frac{1}{\epsilon^2} + \frac{1}{\epsilon^{2-2p}} \right] \frac{\epsilon^2}{p\beta} \exp(-\frac{p\beta}{\epsilon^2} t) [\exp(\frac{p\beta}{\epsilon^2} t) - 1] + \mathbb{E} [|\zeta|^{2p}] \exp(-\frac{p\beta}{\epsilon^2} t) \\ &\leq C(p) [1 + \epsilon^{2p}] + \mathbb{E} [|\zeta|^{2p}] \\ &\leq C(p) + \mathbb{E} [|\zeta|^{2p}] \end{aligned}$$

since $p > 0$ and $\epsilon \in (0, 1]$. Since the bound is uniform in t , we are done. \square

We now provide the aforementioned Ergodic-Type Theorems and their proofs:

Proposition 4.2. Consider $\psi \in \mathcal{M}_b^{\tilde{\zeta}}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, where here

$$\tilde{\zeta} := \{(0, j_1, 0), (1, 0, j_2) : j_1 = 0, 1, 2, j_2 = 0, 1\}.$$

Assume A1) - A5). Then for any $t \in [0, T]$ and $\epsilon \in (0, 1]$:

$$\begin{aligned} &\left| \mathbb{E} \left[\int_0^t \frac{1}{\epsilon} b(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \cdot \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) ds - \int_0^t \gamma_1(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \cdot \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right. \right. \\ &\quad + 2D_1(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) ds \\ &\quad - \int_0^t \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \\ &\quad \left. \left. - \int_0^t \int_{\mathbb{R}^{2d}} \Phi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \right] \right| \\ &\leq C\epsilon(1+t) \|\psi\|_{\mathcal{M}_b^{\tilde{\zeta}}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)}. \end{aligned}$$

Proof. First we note that we may apply Itô's formula for measure-dependent functions (Proposition 5.102 in [8], see also [54] Section 4.2 for a similar computation) to $\Phi(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) \cdot \psi(t, X_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu}))$. We have more than enough differentiability of Φ in x, y , and μ to apply Itô's formula as per Proposition A.8 (see also Remark B.3). Thus, we have:

$$\begin{aligned} & \int_0^t \frac{1}{\epsilon} b(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \cdot \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) ds - \int_0^t \gamma_1(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \cdot \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \\ & + 2D_1(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) : \partial_x \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) ds \\ & - \int_0^t \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_\mu \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon,\nu})) \mathcal{L}(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu})(dx, dy) ds \\ & - \int_0^t \int_{\mathbb{R}^{2d}} \Phi^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_\mu \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon,\nu})) \mathcal{L}(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu})(dx, dy) ds \\ & = \sum_{k=1}^5 A_k^\epsilon(t), \end{aligned}$$

where

$$\begin{aligned} A_1^\epsilon(t) &= \epsilon \left[\Phi(X_0^{\epsilon,\nu}, Y_0^{\epsilon,\nu}, \mathcal{L}(X_0^{\epsilon,\nu})) \cdot \psi(0, X_0^{\epsilon,\nu}, \mathcal{L}(X_0^{\epsilon,\nu})) - \Phi(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) \cdot \psi(t, X_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) \right] \\ A_2^\epsilon(t) &= \epsilon \int_0^t \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \cdot \dot{\psi}(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \\ & + \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_x \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) c(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \\ & + \Phi^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_x \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) c(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \\ & + \frac{1}{2} \left[\psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \sigma \sigma^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) : \partial_x^2 \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \right. \\ & + 2\sigma \sigma^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) : [(\partial_x \Phi)^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_x \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu}))] \\ & \left. + \Phi^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \sigma \sigma^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) : \partial_x^2 \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \right] ds \\ A_3^\epsilon(t) &= \epsilon \int_0^t \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_\mu \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [x] c(x, y, \mathcal{L}(X_s^{\epsilon,\nu})) \\ & + \Phi^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_\mu \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [x] c(x, y, \mathcal{L}(X_s^{\epsilon,\nu})) \\ & + \frac{1}{2} \left[\psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon,\nu})) : \partial_z \partial_\mu \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [x] \right. \\ & \left. + \Phi^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon,\nu})) : \partial_z \partial_\mu \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [x] \right] \mathcal{L}(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu})(dx, dy) ds \\ A_4^\epsilon(t) &= \epsilon \int_0^t \left[\psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_x \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) + \Phi^\top(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_x \psi(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \right] \\ & \quad \sigma(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) dW_s \\ & + \int_0^t \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_y \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \tau_1(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) dW_s \\ & + \int_0^t \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_y \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \tau_2(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) dB_s. \end{aligned}$$

Via the same Proposition A.8 and Assumption A5), all the coefficients, as well as Φ and its derivatives, which appear in $A_k^\epsilon, k = 1, 2, 3, 4$ grow at most polynomially in y uniformly in their other arguments. Thus, by Lemma 4.1, the martingale terms collected in $A_4^\epsilon(t)$ vanish in expectation, and there exists $m \in \mathbb{N}$ such

that:

$$\begin{aligned}
& \left| \mathbb{E} \left[A_1^\epsilon(t) \right] \right| \leq \mathbb{E} \left[|A_1^\epsilon(t)| \right] \\
& \leq \epsilon \sup_{t \in [0, T]} \mathbb{E} \left[|\Phi(X_t^{\epsilon, \nu}, Y_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu}))| \right] \|\psi\|_\infty \\
& \leq \epsilon C \left(1 + \sup_{t \in [0, T]} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2m} \right] \right) \|\psi\|_\infty \\
& \leq \epsilon C \|\psi\|_\infty \\
& \left| \mathbb{E} \left[A_2^\epsilon(t) \right] \right| \leq \mathbb{E} \left[|A_2^\epsilon(t)| \right] \\
& \leq \epsilon C t \sup_{s \in [0, T]} \mathbb{E} \left[|\Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| + |\partial_x \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |c(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| \right. \\
& \quad + |\Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |c(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| \\
& \quad + |\sigma \sigma^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |\partial_x^2 \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| \\
& \quad + |\sigma \sigma^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |\partial_x \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| \\
& \quad \left. + |\Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |\sigma \sigma^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| \right] \sup_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \|\psi(\cdot, \cdot, \mu)\|_{C_b^{1,2}([0, T]; \mathbb{R}^d)} \\
& \leq \epsilon C t \left(1 + \sup_{t \in [0, T]} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2m} \right] \right) \sup_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \|\psi(\cdot, \cdot, \mu)\|_{C_b^{1,2}([0, T]; \mathbb{R}^d)} \\
& \leq \epsilon C t \sup_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \|\psi(\cdot, \cdot, \mu)\|_{C_b^{1,2}([0, T]; \mathbb{R}^d)} \\
& \left| \mathbb{E} \left[A_3^\epsilon(t) \right] \right| \leq \mathbb{E} \left[|A_3^\epsilon(t)| \right] \\
& \leq \epsilon C t \sup_{s \in [0, T]} \mathbb{E} \left[\int_{\mathbb{R}^{2d}} |\partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |c(x, y, \mathcal{L}(X_s^{\epsilon, \nu}))| \right. \\
& \quad + |\Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |c(x, y, \mathcal{L}(X_s^{\epsilon, \nu}))| \\
& \quad + |\sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu}))| |\partial_z \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |x| \\
& \quad \left. + |\Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |\sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu}))| |\mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy)| \right] (\|\psi\|_\infty + \|\partial_\mu \psi\|_\infty + \|\partial_z \partial_\mu \psi\|_\infty) \\
& \leq \epsilon C t \sup_{s \in [0, T]} \left\{ \mathbb{E} \left[\sup_{x \in \mathbb{R}^d} |\partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |x| + |\Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))|^2 \right. \right. \\
& \quad \left. \left. + \sup_{x \in \mathbb{R}^d} |\partial_z \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))| |x| \right]^2 \right\} \\
& \quad + \int_{\mathbb{R}^{2d}} |c(x, y, \mathcal{L}(X_s^{\epsilon, \nu}))|^2 + |\sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu}))|^2 |\mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy)| \Big\} (\|\psi\|_\infty + \|\partial_\mu \psi\|_\infty + \|\partial_z \partial_\mu \psi\|_\infty) \\
& \leq \epsilon C t \left(1 + \sup_{s \in [0, T]} \left\{ \mathbb{E} \left[|Y_s^{\epsilon, \nu}|^{2m} \right] + \int_{\mathbb{R}^{2d}} |y|^{2m} \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \right\} \right) (\|\psi\|_\infty + \|\partial_\mu \psi\|_\infty + \|\partial_z \partial_\mu \psi\|_\infty) \\
& = \epsilon C t \left(1 + \sup_{s \in [0, T]} \mathbb{E} \left[|Y_s^{\epsilon, \nu}|^{2m} \right] \right) (\|\psi\|_\infty + \|\partial_\mu \psi\|_\infty + \|\partial_z \partial_\mu \psi\|_\infty) \\
& \leq \epsilon C t (\|\psi\|_\infty + \|\partial_\mu \psi\|_\infty + \|\partial_z \partial_\mu \psi\|_\infty).
\end{aligned}$$

Combining the above bounds, we get the desired result. \square

Proposition 4.3. *In the setup of Proposition 4.2, for any $t \in [0, T]$ and $\epsilon \in (0, 1]$:*

$$\left| \mathbb{E} \left[\int_0^t \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \right] \right|$$

$$\leq C\epsilon(1+t)\|\psi\|_{\mathcal{M}_b^{\tilde{\zeta}}([0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d);\mathbb{R}^d)}.$$

Proof. To see this, we note for all $t \in [0, T]$:

$$\begin{aligned} & \mathbb{E} \left[\int_0^t \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_\mu \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon,\nu})) \mathcal{L}(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu})(dx, dy) ds \right] \\ &= \mathbb{E} \left[\int_0^t \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_\mu \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [\hat{X}_s^\epsilon] b(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon,\nu})) ds \right] \end{aligned}$$

where $(\hat{X}^\epsilon, \hat{Y}^\epsilon)$ is an independent copy of $(X^{\epsilon,\nu}, Y^{\epsilon,\nu})$. Thus it is enough to show that the expectation of $\int_0^t \psi^\top(s, X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \partial_\mu \Phi(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) [\hat{X}_s^\epsilon] b(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon,\nu})) ds$ is $O(\epsilon)$, where

$$\begin{aligned} dX_t^{\epsilon,\nu} &= \left[\frac{1}{\epsilon} b(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) + c(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) \right] dt + \sigma(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) dW_t \\ dY_t^{\epsilon,\nu} &= \frac{1}{\epsilon} \left[\frac{1}{\epsilon} f(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) + g(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) \right] dt \\ &\quad + \frac{1}{\epsilon} \left[\tau_1(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) dW_t + \tau_2(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu}, \mathcal{L}(X_t^{\epsilon,\nu})) dB_t \right] \\ d\hat{X}_t^\epsilon &= \left[\frac{1}{\epsilon} b(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \mathcal{L}(X_t^{\epsilon,\nu})) + c(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \mathcal{L}(X_t^{\epsilon,\nu})) \right] dt + \sigma(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \mathcal{L}(X_t^{\epsilon,\nu})) d\hat{W}_t \\ d\hat{Y}_t^\epsilon &= \frac{1}{\epsilon} \left[\frac{1}{\epsilon} f(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \mathcal{L}(X_t^{\epsilon,\nu})) + g(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \mathcal{L}(X_t^{\epsilon,\nu})) \right] dt \\ &\quad + \frac{1}{\epsilon} \left[\tau_1(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \mathcal{L}(X_t^{\epsilon,\nu})) d\hat{W}_t + \tau_2(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon, \mathcal{L}(X_t^{\epsilon,\nu})) d\hat{B}_t \right] \\ (X_0^{\epsilon,\nu}, Y_0^{\epsilon,\nu}, \hat{X}_0^\epsilon, \hat{Y}_0^\epsilon) &= (\eta, \zeta, \hat{\eta}, \hat{\zeta}) \end{aligned}$$

and W, B, \hat{W}, \hat{B} are independent m -dimensional standard Brownian motions, (η, ζ) and $(\hat{\eta}, \hat{\zeta})$ are IID with (η, ζ) from Equation (1), and $(\eta, \zeta, \hat{\eta}, \hat{\zeta})$ is independent of (W, B, \hat{W}, \hat{B}) .

Recall the operator $\mathcal{L}_{x,\mu}$ from Equation (2). For fixed $x \in \mathbb{R}, \mu \in \mathcal{P}(\mathbb{R})$, this is the generator of the frozen process

$$(23) \quad dY_t^{x,\mu} = f(x, Y_t^{x,\mu}, \mu) dt + \tau_1(x, Y_t^{x,\mu}, \mu) dW_t + \tau_2(x, Y_t^{x,\mu}, \mu) dB_t$$

for W_t, B_t independent m -dimensional standard Brownian motions. We now introduce a new generator $\mathcal{L}_{x,\bar{x},\mu}^2$ parameterized by $x, \bar{x} \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)$ which acts on $\psi \in C_b^2(\mathbb{R}^2)$ by

$$(24) \quad \begin{aligned} \mathcal{L}_{x,\bar{x},\mu}^2 \psi(y, \bar{y}) &= f(x, y, \mu) \cdot \partial_y \psi(y, \bar{y}) + f(\bar{x}, \bar{y}, \mu) \cdot \partial_{\bar{y}} \psi(y, \bar{y}) \\ &\quad + a(x, y, \mu) : \partial_y^2 \psi(y, \bar{y}) + a(\bar{x}, \bar{y}, \mu) : \partial_{\bar{y}}^2 \psi(y, \bar{y}). \end{aligned}$$

This is the generator associated to the 2-dimensional process solving 2 independent copies of Equation (23) where the same parameter μ enters both equations, but different x, \bar{x} enter each equation, i.e.

$$(25) \quad \begin{aligned} dY_t^{x,\mu} &= f(x, Y_t^{x,\mu}, \mu) dt + \tau_1(x, Y_t^{x,\mu}, \mu) dW_t + \tau_2(x, Y_t^{x,\mu}, \mu) dB_t \\ d\bar{Y}_t^{\bar{x},\mu} &= f(\bar{x}, \bar{Y}_t^{\bar{x},\mu}, \mu) dt + \tau_1(\bar{x}, \bar{Y}_t^{\bar{x},\mu}, \mu) d\bar{W}_t + \tau_2(\bar{x}, \bar{Y}_t^{\bar{x},\mu}, \mu) d\bar{B}_t. \end{aligned}$$

for $W_t, B_t, \bar{W}_t, \bar{B}_t$ independent m -dimensional standard Brownian motions.

One can see that the unique distributional solution of the adjoint equation

$$\begin{aligned} (\mathcal{L}_{x,\bar{x},\mu}^2)^* \bar{\pi}(\cdot; x, \bar{x}, \mu) &= 0 \\ \int_{\mathbb{R}^2} \bar{\pi}(dy, d\bar{y}; x, \bar{x}, \mu) &= 1, \forall x, \bar{x} \in \mathbb{R}, \mu \in \mathcal{P}(\mathbb{R}) \end{aligned}$$

is given by

$$(26) \quad \bar{\pi}(dy, d\bar{y}; x, \bar{x}, \mu) = \pi(dy; x, \mu) \otimes \pi(d\bar{y}; \bar{x}, \mu)$$

where π is as in Equation (3).

We now consider $\chi(x, \bar{x}, y, \bar{y}, \mu) : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ solving

$$(27) \quad \mathcal{L}_{x, \bar{x}, \mu}^2 \chi_l(x, \bar{x}, y, \bar{y}, \mu) = -[\partial_\mu \Phi(\bar{x}, \bar{y}, \mu)[x]b(x, y, \mu)]_l, l = 1, \dots, d$$

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \chi(x, \bar{x}, y, \bar{y}, \mu) \pi(dy; x, \mu) \pi(d\bar{y}, \bar{x}, \mu) = 0.$$

Note that by the centering condition, Equation (4), the right hand side of Equation (27) integrates against $\bar{\pi}$ from Equation (26) to 0 for all x, \bar{x}, μ . By Proposition A.8 and Lemma A.1/Remark A.7 we have existence and uniqueness for the solution χ to Equation (27), and moreover χ enjoys more than enough regularity in x, \bar{x}, y, \bar{y} , and μ to apply Itô's formula for measure dependent functions to $\chi(\hat{X}_t^\epsilon, X_t^{\epsilon, \nu}, \hat{Y}_t^\epsilon, Y_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) \cdot \psi(t, X_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu}))$. Thus:

$$\int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] b(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon, \nu})) ds = \sum_{k=1}^6 B^\epsilon_k(t)$$

where

$$\begin{aligned} B_1^\epsilon(t) &= \epsilon^2 \left[\chi(\hat{X}_0^\epsilon, X_0^{\epsilon, \nu}, \hat{Y}_0^\epsilon, Y_0^{\epsilon, \nu}, \mathcal{L}(X_0^{\epsilon, \nu})) \cdot \psi(0, X_0^{\epsilon, \nu}, \mathcal{L}(X_0^{\epsilon, \nu})) - \chi(\hat{X}_t^\epsilon, X_t^{\epsilon, \nu}, \hat{Y}_t^\epsilon, Y_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) \cdot \psi(t, X_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) \right] \\ B_2^\epsilon(t) &= \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_x \chi \hat{b} + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_{\bar{x}} \chi \tilde{b} + \chi^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \tilde{b} \\ &\quad + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_y \chi \hat{g} + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_{\bar{y}} \chi \tilde{g} + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \hat{\tau}_1 \hat{\sigma}^\top : \partial_x \partial_y \chi \\ &\quad + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \tilde{\tau}_1 \tilde{\sigma}^\top : \partial_{\bar{x}} \partial_{\bar{y}} \chi + \tilde{\tau}_1 \tilde{\sigma}^\top : [(\partial_x \psi)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_{\bar{y}} \chi] \\ &\quad + \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \chi(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\ &\quad + \int_{\mathbb{R}^{2d}} \chi^\top(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \\ B_3^\epsilon(t) &= \epsilon^2 \int_0^t \chi \cdot \dot{\psi}(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_x \chi \hat{c} + \frac{1}{2} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \hat{\sigma} \hat{\sigma}^\top : \partial_x^2 \chi \\ &\quad + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_{\bar{x}} \chi \tilde{c} + \chi^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \tilde{c} \\ &\quad + \frac{1}{2} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \tilde{\sigma} \tilde{\sigma}^\top : \partial_{\bar{x}}^2 \chi + \tilde{\sigma} \tilde{\sigma}^\top : [(\partial_{\bar{x}} \chi)^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))] \\ &\quad + \frac{1}{2} \chi^\top \tilde{\sigma} \tilde{\sigma}^\top : \partial_x^2 \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) ds \\ B_4^\epsilon(t) &= \epsilon^2 \int_0^t \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \chi(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] c(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\ &\quad + \int_{\mathbb{R}^{2d}} \chi^\top(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] c(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\ &\quad + \frac{1}{2} \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \chi(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\ &\quad + \frac{1}{2} \int_{\mathbb{R}^{2d}} \chi^\top(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \\ B_5^\epsilon(t) &= \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_y \chi \hat{\tau}_1 d\hat{W}_s + \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_{\bar{y}} \chi \tilde{\tau}_1 dW_s + \\ &\quad + \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_y \chi \hat{\tau}_2 d\hat{B}_s + \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_{\bar{y}} \chi \tilde{\tau}_2 dB_s \\ B_6^\epsilon(t) &= \epsilon^2 \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_x \chi \hat{\sigma} d\hat{W}_s + \epsilon^2 \int_0^t \left[\psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_{\bar{x}} \chi + \chi^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right] \tilde{\sigma} dW_s. \end{aligned}$$

where \hat{c} denotes $(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon, \nu}))$ as an argument, \tilde{c} denotes $(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))$ as an argument, and similarly for the other coefficients. The argument of χ where suppressed is $(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))$.

Then, by computations along the same lines as Proposition 4.2, using the polynomial growth of χ and its derivatives in y, \bar{y} uniformly in x, \bar{x} , and μ from Proposition A.8, we get there is $m \in \mathbb{N}$ such that:

$$\begin{aligned} & \left| \mathbb{E} \left[\int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] b(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon, \nu})) ds \right] \right| \\ & \leq C[\epsilon^2 + (\epsilon + \epsilon^2)t] (1 + \sup_{t \in [0, T]} \mathbb{E} [|Y_t^{\epsilon, \nu}|^{2m}]) \|\psi\|_{\mathcal{M}_b^{\tilde{\zeta}}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)} \\ & \leq C\epsilon(1+t) \|\psi\|_{\mathcal{M}_b^{\tilde{\zeta}}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)} \end{aligned}$$

by Lemma 4.1 and the fact that $\epsilon \in (0, 1]$. □

Proposition 4.4. *Assume A1) - A5). Consider $\psi \in \mathcal{M}_b^{\tilde{\zeta}}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, where*

$$\tilde{\zeta}_2 := \{(1, j, 0), (1, 0, j), (2, 0, (0, k)) : j = 0, 1, 2, k = 0, 1\}$$

such that $\partial_\mu \dot{\psi}$ exists, is jointly continuous in all arguments, and $\sup_{t \in [0, T], x, z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_\mu \dot{\psi}(t, x, \mu)[z]| < \infty$. Then we have for any $t \in [0, T]$ and $\epsilon \in (0, 1]$:

$$\begin{aligned} & \left| \mathbb{E} \left[\int_0^t \int_{\mathbb{R}^{2d}} \Phi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \right] \right| \\ & \leq C\epsilon(1+t) \left[\sup_{t \in [0, T]} \|\psi(t, \cdot, \cdot)\|_{\mathcal{M}_b^{\tilde{\zeta}_2}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)} + \sup_{t \in [0, T], x, z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_\mu \dot{\psi}(t, x, \mu)[z]| \right]. \end{aligned}$$

Proof. As in Proposition 4.3, it suffices to show the bound for

$$\left| \mathbb{E} \left[\int_0^t \Phi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] b(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon, \nu})) ds \right] \right|,$$

where $(\hat{X}^\epsilon, \hat{Y}^\epsilon)$ is an independent copy of $(X^{\epsilon, \nu}, Y^{\epsilon, \nu})$. We consider the new doubled Poisson equation

$$(28) \quad \begin{aligned} \mathcal{L}_{x, \bar{x}, \mu}^2 \tilde{\chi}_{k, l}(x, \bar{x}, y, \bar{y}, \mu) &= -b_l(x, y, \mu) \Phi_k(\bar{x}, \bar{y}, \mu), l, k = 1, \dots, d \\ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \tilde{\chi}(x, \bar{x}, y, \bar{y}, \mu) \pi(dy; x, \mu) \pi(d\bar{y}, \bar{x}, \mu) &= 0z \end{aligned}$$

where $\mathcal{L}_{x, \bar{x}, \mu}^2$ is as in Equation (24). We note once again that by assumption (4), the inhomogeneity is centered with respect to the invariant measure associated to $\mathcal{L}_{x, \bar{x}, \mu}^2$ given in Equation (26). Applying Itô's formula for measure dependent functions to

$$\tilde{\chi}(\hat{X}_t^\epsilon, X_t^{\epsilon, \nu}, \hat{Y}_t^\epsilon, Y_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) : \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon]$$

(using here the required differentiability of $\tilde{\chi}$ granted by Proposition A.8),

$$\int_0^t \Phi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] b(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon, \nu})) ds = \sum_{j=1}^6 C_j^\epsilon(t)$$

where, for $k = 1, \dots, d$, denoting by $\tilde{\chi}_k$ the vector comprising the k 'th row of $\tilde{\chi}$:

$$\begin{aligned} C_1^\epsilon(t) &= \epsilon^2 \left[\tilde{\chi}(\hat{X}_0^\epsilon, X_0^{\epsilon, \nu}, \hat{Y}_0^\epsilon, Y_0^{\epsilon, \nu}, \mathcal{L}(X_0^{\epsilon, \nu})) : \partial_\mu \psi(0, X_0^{\epsilon, \nu}, \mathcal{L}(X_0^{\epsilon, \nu})) [\hat{X}_0^\epsilon] \right. \\ & \quad \left. - \tilde{\chi}(\hat{X}_t^\epsilon, X_t^{\epsilon, \nu}, \hat{Y}_t^\epsilon, Y_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) : \partial_\mu \psi(t, X_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) [\hat{X}_t^\epsilon] \right] \\ C_2^\epsilon(t) &= \epsilon \sum_{k=1}^d \left\{ \int_0^t (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_x \tilde{\chi}_k \hat{b} + (\tilde{\chi}_k)^\top \partial_z \partial_\mu \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \hat{b} \right. \\ & \quad + (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{x}} \tilde{\chi}_k \tilde{b} + (\tilde{\chi}_k)^\top (\partial_\mu \partial_x \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \tilde{b} \\ & \quad + (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{y}} \tilde{\chi}_k \hat{g} + (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{y}} \tilde{\chi}_k \tilde{g} \\ & \quad + (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \hat{\tau}_1 \hat{\sigma}^\top : \partial_x \partial_y \tilde{\chi}_k + \hat{\tau}_1 \hat{\sigma}^\top : [(\partial_z \partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_y \tilde{\chi}_k] \\ & \quad \left. + (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \tilde{\tau}_1 \tilde{\sigma}^\top : \partial_{\bar{x}} \partial_{\bar{y}} \tilde{\chi}_k + \tilde{\tau}_1 \tilde{\sigma}^\top : [\partial_\mu \partial_x \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{y}} \tilde{\chi}_k] \right\} \end{aligned}$$

$$\begin{aligned}
& + \int_{\mathbb{R}^{2d}} (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_\mu \tilde{\chi}_k(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
& + \int_{\mathbb{R}^{2d}} (\tilde{\chi}_k)^\top(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu^2 \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon, x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \Big\} \\
C_3^\epsilon(t) &= \epsilon^2 \sum_{k=1}^d \left\{ \int_0^t \tilde{\chi}_k \cdot \partial_\mu \dot{\psi}_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] + (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_x \tilde{\chi}_k \hat{c} \right. \\
& + (\tilde{\chi}_k)^\top \partial_z \partial_\mu \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \hat{c} + \frac{1}{2} (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \hat{\sigma} \hat{\sigma}^\top : \partial_x^2 \tilde{\chi}_k \\
& + \hat{\sigma} \hat{\sigma}^\top : [(\partial_z \partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_x \tilde{\chi}_k] + \frac{1}{2} (\tilde{\chi}_k)^\top \hat{\sigma} \hat{\sigma}^\top : \partial_z^2 \partial_\mu \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \\
& + (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{x}} \tilde{\chi}_k \tilde{c} + (\tilde{\chi}_k)^\top (\partial_\mu \partial_x \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \tilde{c} \\
& + \frac{1}{2} (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \tilde{\sigma} \tilde{\sigma}^\top : \partial_{\bar{x}} \partial_x \tilde{\chi}_k + \tilde{\sigma} \tilde{\sigma}^\top : [\partial_\mu \partial_x \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{x}} \tilde{\chi}_k] \\
& \left. + \frac{1}{2} (\tilde{\chi}_k)^\top \tilde{\sigma} \tilde{\sigma}^\top : \partial_x^2 \partial_\mu \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] ds \right\} \\
C_4^\epsilon(t) &= \epsilon^2 \sum_{k=1}^d \left\{ \int_0^t \int_{\mathbb{R}^{2d}} (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_\mu \tilde{\chi}_k(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] \right. \\
& \quad \quad \quad c(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
& + \int_{\mathbb{R}^{2d}} (\tilde{\chi}_k)^\top(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu^2 \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon, x] c(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
& + \frac{1}{2} \int_{\mathbb{R}^{2d}} (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \\
& \quad \quad \quad : \partial_z \partial_\mu \tilde{\chi}_k(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
& + \frac{1}{2} \int_{\mathbb{R}^{2d}} (\tilde{\chi}_k)^\top(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \\
& \quad \quad \quad : \partial_{z_2} \partial_\mu^2 \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon, x] \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds \Big\} \\
C_5^\epsilon(t) &= \epsilon \sum_{k=1}^d \left\{ \int_0^t (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_y \tilde{\chi}_k \hat{\tau}_1 d\hat{W}_s + \epsilon \int_0^t (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{y}} \tilde{\chi}_k \tilde{\tau}_1 dW_s \right. \\
& + \epsilon \int_0^t (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_y \tilde{\chi}_k \hat{\tau}_2 d\hat{B}_s \\
& \left. + \epsilon \int_0^t (\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{y}} \tilde{\chi}_k \tilde{\tau}_2 dB_s \right\} \\
C_6^\epsilon(t) &= \epsilon^2 \sum_{k=1}^d \left\{ \int_0^t \left[(\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_x \tilde{\chi}_k + (\tilde{\chi}_k)^\top \partial_z \partial_\mu \psi_k(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \right] \hat{\sigma} d\hat{W}_s \right. \\
& \left. + \epsilon^2 \int_0^t \left[(\partial_\mu \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \partial_{\bar{x}} \tilde{\chi}_k + (\tilde{\chi}_k)^\top (\partial_\mu \partial_x \psi_k)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] \right] \tilde{\sigma} dW_s \right\},
\end{aligned}$$

where \hat{c} denotes $(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s))$ as an argument, \tilde{c} denotes $(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))$ as an argument, and the other coefficients. The argument of χ where suppressed is $(\hat{X}_s^\epsilon, X_s^{\epsilon, \nu}, \hat{Y}_s^\epsilon, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))$.

Then, by computations along the same lines as Proposition 4.2, using the polynomial growth of $\tilde{\chi}$ and its derivatives in y, \bar{y} uniformly in x, \bar{x} , and μ from Proposition A.8, we get there is $m \in \mathbb{N}$ such that:

$$\begin{aligned}
& \left| \mathbb{E} \left[\int_0^t \Phi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [\hat{X}_s^\epsilon] b(\hat{X}_s^\epsilon, \hat{Y}_s^\epsilon, \mathcal{L}(X_s^{\epsilon, \nu})) ds \right] \right| \\
& \leq C[\epsilon^2 + (\epsilon + \epsilon^2)t] (1 + \sup_{t \in [0, T]} \mathbb{E} [|Y_t^{\epsilon, \nu}|^{2m}]) [\sup_{t \in [0, T]} \|\psi(t, \cdot, \cdot)\|_{\mathcal{M}_b^{\xi_2}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)} +
\end{aligned}$$

$$\begin{aligned}
& + \sup_{t \in [0, T], x, z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_\mu \dot{\psi}(t, x, \mu)[z]| \\
& \leq C\epsilon(1+t) \left[\sup_{t \in [0, T]} \|\psi(t, \cdot, \cdot)\|_{\mathcal{M}_b^{\tilde{\zeta}_2}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)} + \sup_{t \in [0, T], x, z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_\mu \dot{\psi}(t, x, \mu)[z]| \right]
\end{aligned}$$

by Lemma 4.1 and the fact that $\epsilon \in (0, 1]$. \square

Proposition 4.5. *Assume A1) - A4), and let $F : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ be any function such that there exists Ξ the unique classical solution to Equation (29) with $\Xi \in \mathcal{M}_b^{\tilde{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $\partial_y \Xi \in \mathcal{M}_b^{\zeta_x}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{k \times d})$ where $\tilde{\zeta}$ is as in Proposition 4.2 and*

$$\zeta_x := \{(0, j, 0) : j \in \{0, 1\}\}$$

(in particular this holds with $F = \gamma, D$ under the additional Assumption A5) via Proposition A.8). Then for $\bar{F}(x, \mu) := \int_{\mathbb{R}^d} F(x, y, \mu) \pi(dy; x, \mu)$, with π as in Equation (3), any $\psi \in \mathcal{M}_b^{\tilde{\zeta}}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ and $t \in [0, T]$, and $\epsilon \in (0, 1]$:

$$\left| \mathbb{E} \left[\int_0^t \left(F(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) - \bar{F}(X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right) \cdot \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) dt \right] \right| \leq C\epsilon(1+t) \|\psi\|_{\mathcal{M}_b^{\tilde{\zeta}}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)}.$$

Proof. By assumption, we can consider $\Xi : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ the unique classical solution to

$$\begin{aligned}
(29) \quad \mathcal{L}_{x, \mu} \Xi_l(x, y, \mu) &= -[F_l(x, y, \mu) - \int_{\mathbb{R}^d} F_l(x, y, \mu) \pi(dy; x, \mu)], l = 1, \dots, k \\
\int_{\mathbb{R}} \Xi(x, y, \mu) \pi(dy; x, \mu) &= 0.
\end{aligned}$$

Applying Itô's formula for measure-dependent functions to $\Xi(X_t^{\epsilon, \nu}, Y_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) \cdot \psi(t, X_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu}))$, we get:

$$\int_0^t \left(F(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) - \bar{F}(X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right) \cdot \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) dt = \sum_{j=1}^6 D_j^\epsilon(t)$$

where

$$D_1^\epsilon(t) = \epsilon^2 \left[\Xi(X_0^{\epsilon, \nu}, Y_0^{\epsilon, \nu}, \mathcal{L}(X_0^{\epsilon, \nu})) \cdot \psi(0, X_0^{\epsilon, \nu}, \mathcal{L}(X_0^{\epsilon, \nu})) - \Xi(X_t^{\epsilon, \nu}, Y_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) \cdot \psi(t, X_t^{\epsilon, \nu}, \mathcal{L}(X_t^{\epsilon, \nu})) \right]$$

$$\begin{aligned}
D_2^\epsilon(t) &= \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_x \Xi b + \Xi^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) b + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_y \Xi g \\
&+ \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \tau_1 \sigma^\top : \partial_x \partial_y \Xi + \tau_1 \sigma^\top : [(\partial_x \psi)^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_y \Xi] \\
&+ \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \Xi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
&+ \int_{\mathbb{R}^{2d}} \Xi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds
\end{aligned}$$

$$\begin{aligned}
D_3^\epsilon(t) &= \epsilon^2 \int_0^t \Xi \cdot \dot{\psi}(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) + \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_x \Xi c + \Xi^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) c \\
&+ \frac{1}{2} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \sigma \sigma^\top : \partial_x^2 \Xi + \sigma \sigma^\top : [(\partial_x \Xi)^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))] \\
&+ \frac{1}{2} \Xi^\top \sigma \sigma^\top : \partial_x^2 \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) ds
\end{aligned}$$

$$\begin{aligned}
D_4^\epsilon(t) &= \epsilon^2 \int_0^t \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \Xi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] c(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
&+ \int_{\mathbb{R}^{2d}} \Xi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] c(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
&+ \frac{1}{2} \int_{\mathbb{R}^{2d}} \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \Xi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \\
&+ \frac{1}{2} \int_{\mathbb{R}^{2d}} \Xi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \sigma \sigma^\top(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) [x] \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) ds
\end{aligned}$$

$$D_5^\epsilon(t) = \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_y \Xi \tau_1 dW_s + \epsilon \int_0^t \psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_y \Xi \tau_2 dB_s$$

$$D_6^\epsilon(t) = \epsilon^2 \int_0^t \left[\psi^\top(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_x \Xi + \Xi^\top \partial_x \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right] \sigma dW_s.$$

Here argument for the coefficients and Ξ and its derivatives where it is suppressed is $(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))$. By computations along the same lines as Proposition 4.2, we get there is $m \in \mathbb{N}$ such that:

$$\begin{aligned} & \left| \mathbb{E} \left[\int_0^t \left(F(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) - \bar{F}(X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right) \psi(s, X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) dt \right] \right| \\ & \leq C[\epsilon + (\epsilon + \epsilon^2)t] \left(1 + \sup_{t \in [0, T]} \mathbb{E} \left[|Y_t^{\epsilon, \nu}|^{2m} \right] \right) \|\psi\|_{\mathcal{M}_b^\xi([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)} \\ & \leq C\epsilon(1+t) \|\psi\|_{\mathcal{M}_b^\xi([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)} \end{aligned}$$

by Lemma 4.1 and the fact that $\epsilon \in (0, 1]$. \square

5. ON THE CAUCHY PROBLEM (10)

With the results of Section 4, we are almost ready to use the Cauchy Problem on Wasserstein Space in order to prove our main result, Theorem 3.1. Before doing so, we need a final Lemma, which provides the needed existence, uniqueness, and regularity of solutions to Equation (10), and is a refinement of [11] Theorem 2.15:

Lemma 5.1. Let $\tau \in [0, T]$. Assume A1)-A6) and let $G \in \mathcal{M}_{b,L}^\xi(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, where ξ is as in Theorem 3.1. Then $U(t, \mu) = G(\mathcal{L}(X_t^\mu))$ is the unique solution to the PDE (10), and $U \in \mathcal{M}_{b,L}^\xi([0, \tau] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, with

$$(30) \quad \|U\|_{\mathcal{M}_b^\xi([0, \tau] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})} \leq C(T) \|G\|_{\mathcal{M}_b^\xi(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}.$$

Here $C(T)$ is independent of G , and depends only on T and $\|\bar{\gamma}\|_{\mathcal{M}_b^\xi(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)}, \|\bar{D}^{1/2}\|_{\mathcal{M}_b^\xi(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})}$ and their Lipschitz constants, where $\hat{\xi}$ is as in Equation (15).

Moreover, $\partial_\mu \dot{U}(s, \mu)[z], \partial_z \partial_\mu \dot{U}(s, \mu)[z], \partial_\mu^2 \dot{U}(s, \mu)[z_1, z_2]$ all exist and are jointly continuous in t, μ, z_1, z_2 , and

$$(31) \quad \begin{aligned} \sup_{t \in [0, \tau], z_1, z_2 \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} \max\{|\partial_\mu \dot{U}(t, \mu)[z_1]|, |\partial_z \partial_\mu \dot{U}(t, \mu)[z_1]|, |\partial_\mu^2 \dot{U}(t, \mu)[z_1, z_2]|\} & \leq C \sup_{t \in [0, \tau]} \|U(t, \cdot)\|_{\mathcal{M}_b^\xi(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})} \\ & \leq C(T) \|G\|_{\mathcal{M}_b^\xi(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}. \end{aligned}$$

Recall here that the superscript $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ is denoting that X_t^μ from Equation (8) is initialized at a random variable ξ independent from W^2 with $\mathcal{L}(\xi) = \mu$. Thus, varying μ in $U(t, \mu)$ is varying the initial distribution of the McKean-Vlasov SDE (8).

Proof. First we note that indeed $\bar{\gamma} \in \mathcal{M}_{b,L}^\xi(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d), \bar{D}^{1/2} \in \mathcal{M}_{b,L}^\xi(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$ by Proposition A.8.

The representation for the solution on terms of G , uniqueness, and time differentiability is the subject of Theorem 7.2 in [6]. Note that by the flow property discussed in Remark 5.2, one can adapt the terminal condition formulation of the Cauchy Problem found in that paper to our form for the initial condition (see [15] Equation 1.2 and Theorem 5.8). [11] Theorem 2.15 outlines how to extend that theorem to higher derivatives, but there they do not track entirely what derivatives of the coefficients and initial condition are needed to control each specific derivative of U , and just write things in terms of the order of the derivatives of U and the coefficients. We could directly apply this result, but that would be requiring $\bar{\gamma}, \bar{D}^{1/2}$ have all mixed derivatives in (x, μ, z) of order 4 bounded and Lipschitz, and would imply that U have all derivatives of order 4 in (μ, z) which are bounded and Lipschitz. We don't require order 4 differentiability of U in μ , and hence make a slight refinement.

If one repeats the computations of [11] in the full setting (using the proof of [6] Lemma 6.2 for guidance), we see $U \in \mathcal{M}_{b,L}^\xi([0, \tau] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ corresponds to the coefficients satisfying $\bar{\gamma} \in \mathcal{M}_{b,L}^\xi(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d), \bar{D}^{1/2} \in$

$\mathcal{M}_{b,L}^{\tilde{\xi}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$. One should note that a small typo is made when going from (6.23) to (6.24) in [6], and in (6.25) the terms multiplied by $\partial_x \partial_\mu \tilde{X}$ should have $\tilde{\xi}$ replaced by z , those multiplied by $\partial_x \tilde{X}$ should have $\tilde{\xi}$ replaced by z , and those multiplied by $\partial_x \tilde{X}$ should have $\tilde{\xi}$ replaced by z . It is also evident that the dependence is linear in terms of derivatives of the initial condition, even though neither of these results are framed as such. This establishes (30).

Finally, using that U solves (10), we can use a computation similar to Example 5 in Section 5.2.2 in [8] to see:

$$\begin{aligned} \partial_\mu \dot{U}(t, \mu)[z] &= (\partial_x \bar{\gamma})^\top(z, \mu) \partial_\mu U(t, \mu)[z] + (\partial_z \partial_\mu U)^\top(t, \mu)[z] \bar{\gamma}(z, \mu) + \partial_z \partial_\mu U(\mu)[z] : \partial_x \bar{D}(z, \mu) \\ &\quad + \bar{D}(z, \mu) : \partial_z^2 \partial_\mu U(\mu)[z] + \int_{\mathbb{R}^d} (\partial_\mu \bar{\gamma})^\top(y, \mu)[z] \partial_\mu U(t, \mu)[y] + (\partial_\mu^2 U(t, \mu))^\top[y, z] \bar{\gamma}(y, \mu) \\ &\quad + \partial_z \partial_\mu U(\mu)[y] : \partial_\mu \bar{D}(y, \mu)[z] + \bar{D}(y, \mu) : \partial_{z_1} \partial_\mu^2 U(\mu)[y, z] \mu(dy) \end{aligned}$$

and can arrive at similar representations for $\partial_z \partial_\mu \dot{U}(t, \mu)[z]$ and $\partial_\mu^2 \dot{U}(t, \mu)[z_1, z_2]$, which yield the bounds (31). \square

Before providing the proof of Theorem 3.1, let us make a few observations about the role of the Equation (10) and compare the proof method of this paper to that of [53] Theorem 2.3, which uses the standard Cauchy Problem associated to SDEs (the backward Kolmogorov equation) to prove the analogous fully-coupled averaging result in the setting without measure dependence of the coefficients.

Therein, under sufficient regularity of the coefficients in the limiting equation and of the test function ϕ , standard PDE results can be used to show that the solution of the backward Kolmogorov equation is in $C_b^{1,4}([0, T] \times \mathbb{R}^d)$, and that the rate of convergence is $\left| \mathbb{E} \left[\phi(X_t^\epsilon) - \phi(X_t) \right] \right| \leq C \epsilon \sup_{t \in [0, T]} \|u(t, \cdot)\|_{C_b^4(\mathbb{R}^d)} + \|\dot{u}(t, \cdot)\|_{C_b^2(\mathbb{R}^d)}$, where C does not depend on ϕ or the initial condition of X^ϵ (here X^ϵ and X are solutions of a standard SDEs, not McKean-Vlasov SDEs). Though this is not explicitly stated, due to the representation of solutions of the Cauchy problem as the initial condition integrated against the fundamental solution of the Cauchy problem, one can find that in fact $\sup_{t \in [0, T]} \|u(t, \cdot)\|_{C_b^4(\mathbb{R}^d)} + \|\dot{u}(t, \cdot)\|_{C_b^2(\mathbb{R}^d)} \leq \|\phi\|_{C_b^4(\mathbb{R}^d)}$ (for standard PDE methods see e.g. [42] Chapter 4 Sections 5 and 14 and [25] Chapter 9 Theorem 3, and for a probabilistic method see [9] Theorem 1.7.5). This is effectively our Corollary 3.2 in the case of standard Fully-Coupled Slow-Fast SDEs.

In order to establish Corollary 3.2, one might be tempted to treat the coefficients of X_t^ν from Equation (8) as time dependent coefficients for a standard SDE, i.e. to let $\tilde{\gamma}(t, x) = \bar{\gamma}(x, \mathcal{L}(X_t^{\delta_x}))$ and $\tilde{D}(t, x) = \bar{D}(x, \mathcal{L}(X_t^{\delta_x}))$, and consider $\tilde{u} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ solving

$$\begin{aligned} \dot{\tilde{u}}(t, x) &= \tilde{L}\tilde{u}(t, x) := \tilde{\gamma}(t, x) \cdot \partial_x \tilde{u}(t, x) + \tilde{D}(t, x) : \partial_x^2 \tilde{u}(t, x), t \in (0, T], x \in \mathbb{R}^d, \\ \tilde{u}(0, x) &= \phi(x) \end{aligned}$$

and use the method of [53] to establish a rate of convergence of X^{ϵ, δ_x} to X^{δ_x} . However, there are a few disadvantages to this approach compared when considering the convergence of McKean-Vlasov SDEs.

The first is that the PDE results regarding differentiability of the fundamental solution of the Cauchy problem depends on having regularity (differentiability and Hölder continuity in x and uniform continuity in t) of the coefficients $\tilde{\gamma}(t, x), \tilde{D}(t, x)$. Since these coefficients are dependent implicitly on $\mathcal{L}(X_t^{\delta_x})$, this would require understanding regularity properties of the Law of $X_t^{\delta_x}$, and understanding how these regularity properties transfer through the second argument of $\tilde{\gamma}$ and \tilde{D} , which a priori may have any form of dependence on their measure argument as long as it is smooth enough. Recently, such regularity was studied in [15, 34] and [12, 13] through the use of Malliavin Calculus and a parametrix method respectively. However, the first two results don't extend completely to cover the derivatives that we need since they are formed specifically for functions of the form $G_\phi(\mu) = \langle \mu, \phi \rangle$, and thus some extra work would be needed to use these for the required bounds on mixed derivatives in time and space of \tilde{u} . In addition, in the latter two, the authors are only interested in proving enough regularity in order to establish strong solutions of the Cauchy problem on Wasserstein Space (see Equation (10) below) and to apply it to establish rates for the Propagation of Chaos for systems without multiscale structure (in other words they only need up to 2 derivatives in space, where

we need 4). We should also mention [59], where such regularity is studied via an alternative method using the Linear Functional Derivative, and the associated McKean-Vlasov SDE has additive noise and is posed on the torus.

The other issues are related to the following remark regarding the “flow” property of solutions to McKean-Vlasov SDEs with respect to their initial conditions:

Remark 5.2. It is important that we parameterize the initial condition in Equations (1) and (8) by $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ rather than $x \in \mathbb{R}^d$, since due to the dependence of the coefficients on their law, McKean-Vlasov SDEs do not enjoy the same flow property on the reals as standard SDEs. In other words, while for $\tilde{X}^{\epsilon, x}$ corresponding to Equation (1) with all coefficients independent of μ (that is, when Equation (1) is a standard SDE and not a McKean-Vlasov SDE) and deterministic initial condition $x \in \mathbb{R}^d$, $\tilde{X}_t^{\epsilon, y} \Big|_{y=\tilde{X}_s^{\epsilon, x}} = \tilde{X}_{s+t}^{\epsilon, x}$, in general

$X_t^{\epsilon, \delta_y} \Big|_{y=X_s^{\epsilon, \delta_x}} \neq X_{s+t}^{\epsilon, \delta_x}$, where here we mean equality in distribution. We do, however, have via strong

existence and uniqueness (see Remark 2.5) that $X_t^{\epsilon, \bar{\nu}} \Big|_{\bar{\nu}=\mathcal{L}(X_s^{\epsilon, \nu})} = X_{s+t}^{\epsilon, \nu}$ for all $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. In other words,

the flow property for solutions of the McKean-Vlasov SDE holds viewed as a flow of measures, not as a flow on \mathbb{R}^d .

To illustrate why one cannot hope to have a flow property on \mathbb{R}^d , we note that for a random variable $\eta \sim \nu$, $X_t^{\epsilon, \delta_x} \Big|_{x=\eta} \neq X_t^{\epsilon, \nu}$. This becomes evident by setting all coefficients equal to 0 but c , and letting $c(x, y, \mu) = c(\mu) = \langle \mu, \phi \rangle$ for some deterministic, non-constant $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Then we no longer have ϵ dependence, and

$$\begin{aligned} X_t^{\delta_x} &= x + \int_0^t \langle \mathcal{L}(X_s^{\delta_x}), \phi \rangle ds \\ &= x + \int_0^t \langle \delta_{X_s^{\delta_x}}, \phi \rangle ds \\ &= x + \int_0^t \phi(X_s^{\delta_x}) ds \end{aligned}$$

is deterministic, with

$$X_t^{\delta_x} \Big|_{x=\eta} = \eta + \int_0^t \phi(X_s^{\delta_x}) \Big|_{x=\eta} ds,$$

while

$$\begin{aligned} X_t^\nu &= \eta + \int_0^t \langle \mathcal{L}(X_s^\nu), \phi \rangle ds \\ &= \eta + \int_0^t \mathbb{E}[\phi(X_s^\nu)] ds, \end{aligned}$$

so

$$X_t^{\delta_x} \Big|_{x=\eta} = X_t^\nu \Leftrightarrow \int_0^t \phi(X_s^{\delta_x}) \Big|_{x=\eta} ds = \int_0^t \mathbb{E}[\phi(X_s^\nu)] ds,$$

the left of which is a random variable and the right of which is deterministic. Thus, this is requiring that η be deterministic to hold in general. Thus, via the aforementioned flow property on measures, $X_{s+t}^{\epsilon, \delta_x} = X_t^{\epsilon, \nu} \Big|_{\nu=\mathcal{L}(X_s^{\epsilon, \delta_x})} \neq X_t^{\epsilon, \delta_y} \Big|_{y=X_s^{\epsilon, \delta_x}}$. See [61] p.3-4 for a good further exposition on this, and how it relates to the non-linearity of the generator of McKean-Vlasov SDEs.

This leads us to the next major disadvantage of using the standard Cauchy Problem over the Cauchy Problem on Wasserstein space for studying convergence of McKean-Vlasov SDEs. Where in the standard

averaging case, we have for $\xi \sim \nu$,

$$\begin{aligned} \left| \mathbb{E} \left[\phi(\tilde{X}_t^{\epsilon, \xi}) - \phi(\tilde{X}_t^\xi) \right] \right| &= \left| \int_{\mathbb{R}^d} \mathbb{E} \left[\phi(\tilde{X}_t^{\epsilon, x}) - \phi(\tilde{X}_t^x) \right] \nu(dx) \right| \\ &\leq \int_{\mathbb{R}^d} \left| \mathbb{E} \left[\phi(\tilde{X}_t^{\epsilon, x}) - \phi(\tilde{X}_t^x) \right] \right| \nu(dx) \\ &\leq \int_{\mathbb{R}^d} C\epsilon |\phi|_{C_b^4(\mathbb{R}^d)} \nu(dx) \\ &= C\epsilon |\phi|_{C_b^4(\mathbb{R}^d)}, \end{aligned}$$

so the result established via the standard Cauchy Problem for deterministic initial conditions $x \in \mathbb{R}^d$ also holds in the case of random initial conditions. Meanwhile, for the McKean-Vlasov setting, due to the non-linearity of the Fokker-Plank equations associated to $X^{\epsilon, \nu}$ and X^ν (see p.4 in [61]):

$$\mathbb{E} \left[\phi(X_t^{\epsilon, \nu}) - \phi(X_t^\nu) \right] \neq \int_{\mathbb{R}^d} \mathbb{E} \left[\phi(X_t^{\epsilon, \delta_x}) - \phi(X_t^{\delta_x}) \right] \nu(dx)$$

in general, so the result does not have the same upshot, and a proof of convergence of $\mathcal{L}(X_t^{\epsilon, \delta_x})$ to $\mathcal{L}(X_t^{\delta_x})$ for each $x \in \mathbb{R}^d$, which is what would be obtained via the standard Cauchy problem, does not immediately lift to the case of random initial conditions. Lastly, even assuming sufficient regularity on the coefficients to apply Itô's formula to $\tilde{u}(t, X_t^{\epsilon, \delta_x})$, we would end up needing to control terms of the form $[\tilde{\gamma}(X_t^{\epsilon, \delta_x}, \mathcal{L}(X_t^{\delta_y}))]|_{y=X_t^{\epsilon, \delta_x}} - \gamma(X_t^{\epsilon, \delta_x}, Y_t^{\epsilon, \delta_x}, \mathcal{L}(X_t^{\epsilon, \delta_x})) \cdot \partial_x \tilde{u}(t, X_t^{\epsilon, \delta_x})$ (see the last equation on p.1230 of [53]) rather than the term $[\tilde{\gamma}(X_t^{\epsilon, \delta_x}, \mathcal{L}(X_t^{\delta_x})) - \gamma(X_t^{\epsilon, \delta_x}, Y_t^{\epsilon, \delta_x}, \mathcal{L}(X_t^{\epsilon, \delta_x}))] \cdot \partial_\mu U(t, \mathcal{L}(X_t^{\epsilon, \delta_x}))[X_t^{\epsilon, \delta_x}]$. The latter is what appears when using the Cauchy-Problem on Wasserstein space (see the term R_4 in the proof of Theorem 3.1 below), and is such that we can get vanishing bounds in expectation on its time integral via auxiliary Poisson equation methods (see Proposition 4.5). The former would require some kind of artificial Lipschitz estimates for $\tilde{\gamma}(X_t^{\epsilon, \delta_x}, \mathcal{L}(X_t^{\delta_x})) - \tilde{\gamma}(X_t^{\epsilon, \delta_x}, \mathcal{L}(X_t^{\delta_x}))$ (note the difference in the measure argument). This is due to an effective decoupling of the evolution of the measure which appears in the coefficients $\tilde{\gamma}, \tilde{D}$ from the process whose generator is given by the right hand side of the equation which \tilde{u} satisfies. In fact, this is the very decoupling used in [6] Equation (3.2) in order to define the real-valued component of the evolution of the Cauchy Problem on Wasserstein space in the full setting.

These issues reflect that if we were going to attempt to use \tilde{u} as above, we would be attempting to re-frame the McKean-Vlasov SDE (8) in a way so that we are considering it as a flow on \mathbb{R}^d , rather than $\mathcal{P}_2(\mathbb{R}^d)$. Using instead the Cauchy Problem on Wasserstein space, we are “lifting” our view of the process to a deterministic evolution of a measure on $\mathcal{P}_2(\mathbb{R}^d)$, that is, its Law. This is reflected by the fact that in Equation (10), we are using the infinitesimal generator of the strongly continuous semi-group $\mathcal{P}_t, t \geq 0$ on the space of uniformly bounded, uniformly continuous functions $G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ such that $[\mathcal{P}_t G](\mu) = G(\mathcal{L}(X_t^\mu))$, rather than the infinitesimal generator associated to the nonlinear semi-group $T_t, t \geq 0$ acting on functions $\phi \in C_b^2(\mathbb{R}^d)$ by $[T_t \phi](x) = \mathbb{E}[\phi(X_t^{\delta_x})]$, as is the case for \tilde{u} (see Section 5.7.4 in [8]). This viewpoint is also what allows us to consider possibly nonlinear test functions G , where if we were to use \tilde{u} , we would be restricting ourselves to linear interactions with the measure.

Other than the modification of using the Cauchy Problem on Wasserstein space, the main idea of the proof of Theorem 3.1 is similar to that of [53] Theorem 2.3. That is, we can re-express the distance between any sufficiently regular test function G evaluated at the law of $X_\tau^{\epsilon, \nu}$ and the law of X_τ^ν for a given time τ as the distance between the time-reversed solution $\tilde{U}(t, \mu; \tau) := U(\tau - t, \mu)$ to Equation (10) evaluated at $(t, \mu) = (\tau, \mathcal{L}(X_\tau^{\epsilon, \nu}))$ and $(0, \nu)$ respectively. Then we can apply Itô's formula and use the Propositions of Section 4 to conclude this distance is $O(\epsilon)$, and to express it linearly in terms of Lions derivatives of U , and hence in terms of the test function G . Intuitively what is going on with this approach to rates of convergence in distribution is, via the representation for $U(t, \mu)$ provided in Lemma 5.1, for any fixed $\tau \in [0, T]$, $t \mapsto \tilde{U}(t, \mathcal{L}(X_t^\nu); \tau)$ is constant and equal to $U(\tau - t, \mathcal{L}(X_t^\nu)) = G(\mathcal{L}(X_{\tau-t}^{\mathcal{L}(X_t^\nu)})) = G(\mathcal{L}(X_\tau^\nu)) = U(\tau, \nu) = \tilde{U}(0, \nu; \tau)$ for any $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $t \in [0, \tau]$. Thus, the change of $t \mapsto \tilde{U}(t, \mathcal{L}(X_t^{\epsilon, \nu}); \tau)$ from its initial value $\tilde{U}(0, \nu; \tau)$ as t varies from 0 to τ is in a sense measuring the distance of $\mathcal{L}(X_\tau^{\epsilon, \nu})$ from $\mathcal{L}(X_\tau^\nu)$.

6. PROOF OF THEOREM 3.1

Proof. We use the Cauchy problem on Wasserstein Space defined in Equation (10).

Fix $G \in \mathcal{M}_{b,L}^{\dot{}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, $\tau \in (0, T]$ and consider the solution $U : [0, \tau] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ to (10), with this choice of G . Then, by Lemma 5.1, we have that U is given uniquely by $U(t, \mu) = G(\mathcal{L}(X_t^\mu))$, where we recall here the superscript μ is denoting that X_t^μ is initialized with a random variable ξ independent from W^2 such that $\mathcal{L}(\xi) = \mu$, and that $U \in \mathcal{M}_{b,L}^{\dot{}}([0, \tau] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$. Then, letting $\tilde{U}(t, \mu) = U(\tau - t, \mu)$ for $t \in [0, \tau]$, we have $\tilde{U}(\tau, \mu) = U(0, \mu) = G(\mu)$. Thus, for any initial distribution $\nu \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\begin{aligned} \left| G(\mathcal{L}(X_\tau^{\epsilon, \nu})) - G(\mathcal{L}(X_\tau^\nu)) \right| &= \left| G(\mathcal{L}(X_\tau^{\epsilon, \nu})) - U(\tau, \nu) \right| \\ &= \left| \tilde{U}(\tau, \mathcal{L}(X_\tau^{\epsilon, \nu})) - \tilde{U}(0, \nu) \right|. \end{aligned}$$

Note that \tilde{U} is parameterized by the end time of the interval, τ , but we suppress this in the notation for presentation purposes.

Now we can apply the Chain Rule for Measure Dependent Functions from Proposition 5.102 in [8] to express the above in terms of Lions derivatives of \tilde{U} in a way that does not depend on X^ν or its Law. We get, using $\dot{\tilde{U}}(s, \nu) = -\dot{U}(\tau - s, \nu)$ and that U satisfies (10):

$$\begin{aligned} \left| G(\mathcal{L}(X_\tau^{\epsilon, \nu})) - G(\mathcal{L}(X_\tau^\nu)) \right| &= \left| \int_0^\tau \dot{\tilde{U}}(s, \mathcal{L}(X_s^{\epsilon, \nu})) \right. \\ &\quad + \mathbb{E} \left[\left[\frac{1}{\epsilon} b(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) + c(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right] \cdot \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right. \\ &\quad \left. + \frac{1}{2} \sigma \sigma^\top (X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right] ds \Big| \\ &= \left| \int_0^\tau -\mathbb{E} \left[\bar{\gamma}(X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \cdot \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right. \right. \\ &\quad \left. + \bar{D}(X_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right] \\ &\quad + \mathbb{E} \left[\left[\frac{1}{\epsilon} b(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) + c(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \right] \cdot \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right. \\ &\quad \left. + \frac{1}{2} \sigma \sigma^\top (X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right] ds \Big| \\ &= \left| R_1(\tau) + R_2(\tau) + R_3(\tau) + R_4(\tau) + R_5(\tau) \right| \end{aligned}$$

where

$$\begin{aligned} R_1(\tau) &= \int_0^\tau \mathbb{E} \left[\frac{1}{\epsilon} b(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \cdot \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] - \gamma_1(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \cdot \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right. \\ &\quad \left. - 2D_1(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) : \partial_z \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right. \\ &\quad \left. - \int_{\mathbb{R}^{2d}} (\partial_\mu \tilde{U})^\top(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))[x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \right. \\ &\quad \left. - \int_{\mathbb{R}^{2d}} \Phi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu^2 \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}, x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \right] ds \\ R_2(\tau) &= \int_0^\tau \mathbb{E} \left[\int_{\mathbb{R}^{2d}} (\partial_\mu \tilde{U})^\top(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \partial_\mu \Phi(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu}))[x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \right] ds \\ R_3(\tau) &= \int_0^\tau \mathbb{E} \left[\int_{\mathbb{R}^{2d}} \Phi^\top(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \partial_\mu^2 \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}, x] b(x, y, \mathcal{L}(X_s^{\epsilon, \nu})) \mathcal{L}(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu})(dx, dy) \right] ds \\ R_4(\tau) &= \int_0^\tau \mathbb{E} \left[\gamma(X_s^{\epsilon, \nu}, Y_s^{\epsilon, \nu}, \mathcal{L}(X_s^{\epsilon, \nu})) \cdot \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon, \nu}))[X_s^{\epsilon, \nu}] \right] \end{aligned}$$

$$\begin{aligned}
& - \int_{\mathbb{R}^d} \gamma(X_s^{\epsilon,\nu}, y, \mathcal{L}(X_s^{\epsilon,\nu})) \pi(dy; X_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) \cdot \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon,\nu}))[X_s^{\epsilon,\nu}] ds \\
R_5(\tau) &= \int_0^\tau \mathbb{E} \left[D(X_s^{\epsilon,\nu}, Y_s^{\epsilon,\nu}, \mathcal{L}(X_s^{\epsilon,\nu})) : \partial_z \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon,\nu}))[X_s^{\epsilon,\nu}] \right. \\
& \quad \left. - \int_{\mathbb{R}^d} D(X_s^{\epsilon,\nu}, y, \mathcal{L}(X_s^{\epsilon,\nu})) \pi(dy; X_s^{\epsilon,\nu}, \mathcal{L}(X_s)) : \partial_z \partial_\mu \tilde{U}(s, \mathcal{L}(X_s^{\epsilon,\nu}))[X_s^{\epsilon,\nu}] \right] ds.
\end{aligned}$$

Here the martingale terms from Itô's formula vanish in expectation thanks to Lemma 4.1, and we use that $\partial_z \partial_\mu \tilde{U}(s, \mu)[z]$ is a symmetric matrix for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $s \in [0, \tau]$, and $z \in \mathbb{R}^d$ (see [8] Corollary 5.89) to write:

$$\begin{aligned}
\frac{1}{2} \sigma \sigma^\top(x, y, \mu) : \partial_z \partial_\mu \tilde{U}(s, \mu)[z] &= D(x, y, \mu) : \partial_z \partial_\mu \tilde{U}(s, \mu)[z] - [D_1(x, y, \mu) + D_1^\top(x, y, \mu)] : \partial_z \partial_\mu \tilde{U}(s, \mu)[z] \\
&= D(x, y, \mu) : \partial_z \partial_\mu \tilde{U}(s, \mu)[z] - 2D_1(x, y, \mu) : \partial_z \partial_\mu \tilde{U}(s, \mu)[z].
\end{aligned}$$

Applying Proposition 4.2 with $\psi(s, x, \mu) = \partial_\mu \tilde{U}(s, \mu)[x]$, we get, letting $\tilde{\zeta}$ be as in the statement of that Proposition:

$$\begin{aligned}
|R_1(\tau)| &\leq C\epsilon[1 + \tau] \left\| (t, x, \mu) \mapsto \partial_\mu \tilde{U}(t, \mu)[x] \right\|_{\mathcal{M}_b^{\tilde{\zeta}}([0, \tau] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)} \\
&\leq C\epsilon[1 + \tau] \left[\sup_{t \in [0, \tau]} \|U(t, \cdot)\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})} + \sup_{t \in [0, \tau], z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_\mu \dot{U}(t, \mu)[z]| \right] \\
&\leq \epsilon C(T) \|G\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}
\end{aligned}$$

by Lemma 5.1. Applying Proposition 4.3 with $\psi(s, x, \mu) = \partial_\mu \tilde{U}(s, \mu)[x]$, we get in the same way:

$$|R_2(\tau)| \leq \epsilon C(T) \|G\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}.$$

Applying Proposition 4.4 with $\psi(s, x, \mu) = \partial_\mu \tilde{U}(s, \mu)[x]$, letting $\tilde{\zeta}_2$ be as in the statement of that Proposition, we get:

$$\begin{aligned}
|R_3(\tau)| &\leq C\epsilon[1 + \tau] \left[\sup_{t \in [0, \tau]} \left\| (x, \mu) \mapsto \partial_\mu \tilde{U}(t, \mu)[x] \right\|_{\mathcal{M}_b^{\tilde{\zeta}_2}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)} + \sup_{t \in [0, \tau], z_1, z_2 \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_\mu^2 \dot{U}(t, \mu)[z_1, z_2]| \right] \\
&\leq C\epsilon[1 + \tau] \left[\sup_{t \in [0, \tau]} \|U(t, \cdot)\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})} + \sup_{t \in [0, \tau], z_1, z_2 \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_\mu^2 \dot{U}(t, \mu)[z_1, z_2]| \right] \\
&\leq \epsilon C(T) \|G\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}
\end{aligned}$$

by Lemma 5.1.

Applying Proposition 4.5 with $\psi(s, x, \mu) = \partial_\mu \tilde{U}(s, \mu)[x]$, $F = \gamma$, and $k = d$, we get in the same way as for $R_1(\tau), R_2(\tau)$ that:

$$|R_4(\tau)| \leq \epsilon C(T) \|G\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}.$$

Lastly, Proposition 4.5 with $\psi(s, x, \mu) = \partial_z \partial_\mu \tilde{U}(s, \mu)[x]$, $F = D$, and $k = d \times d$, we get:

$$\begin{aligned}
|R_5(\tau)| &\leq C\epsilon[1 + \tau] \left\| (t, x, \mu) \mapsto \partial_z \partial_\mu \tilde{U}(t, \mu)[x] \right\|_{\mathcal{M}_b^{\tilde{\zeta}}([0, \tau] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})} \\
&\leq C\epsilon[1 + \tau] \left[\sup_{t \in [0, \tau]} \|U(t, \cdot)\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})} + \sup_{t \in [0, \tau], z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)} |\partial_z \partial_\mu \dot{U}(t, \mu)[z]| \right] \\
&\leq \epsilon C(T) \|G\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})}
\end{aligned}$$

by Lemma 5.1. Then, noting that these bounds are all uniform in $\tau \in [0, T]$, we have

$$\sup_{s \in [0, T]} \left| G(\mathcal{L}(X_s^{\epsilon,\nu})) - G(\mathcal{L}(X_s^\nu)) \right| \leq \epsilon C(T) \|G\|_{\mathcal{M}_b^{\tilde{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})},$$

as desired. \square

7. CONCLUSIONS AND FUTURE WORK

In this paper we have derived an averaging principle for fully coupled McKean-Vlasov SDEs, along with an associated rate of weak convergence.

In this paper, due to limitations in the literature regarding regularity of solutions to the Cauchy Problem on Wasserstein space, we made strong assumptions on the regularity of the coefficients of Equation (1). It is known that under weaker assumptions, similar rates of convergence can be derived in case of fully coupled standard SDEs (without coefficients which have explicit dependence on the law of the process). See Remarks 2.6 and 3.3 for further discussion of this. An interesting avenue of future research would be to see if the method proposed in this paper can be extended to weaker assumptions via improving the existing regularity results for PDEs of the type (10).

Another interesting extension would be to establish an averaging principle for fully-coupled SDEs in the setting where the coefficients the fast and slow process in Equation (1) depend on the law of the fast process, $\mathcal{L}(Y_t^{\epsilon,\nu})$. In this setting, the solution of the Poisson Equation Φ whose derivatives appear in the coefficients of limiting equation (8) will have to solve a PDE on Wasserstein Space, since the generator obtained from considering the $O(1/\epsilon^2)$ terms from the generator of $(X_t^{\epsilon,\nu}, Y_t^{\epsilon,\nu})$ and “freezing” the terms associated to the slow process will be that of a McKean-Vlasov SDE. See [20, 43, 62] for related results in this direction.

APPENDIX A. REGULARITY OF THE POISSON EQUATIONS

Throughout this subsection we assume A1) and A2), and prove the needed regularity of the Poisson Equations (5), (27), (28), and (29) and the averaged coefficients from (7) in order for the results in Sections 4, 5, and 6 to go through. The final result containing this needed regularity is Proposition A.8.

The proofs here are analogous to those found in [53]. Thus, for brevity, we choose only to sketch the proofs and focus on the necessary additional steps which arise from the fact that we take derivatives in $\mathcal{P}_2(\mathbb{R}^d)$ in addition to standard spatial derivatives.

Lemma A.1. Consider $B : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ such that

$$\int_{\mathbb{R}^d} B(x, y, \mu) \pi(dy; x, \mu) = 0, \forall x \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

B grows at most polynomially in y uniformly in x, μ as $|y| \rightarrow \infty$, and B is locally Hölder continuous in y uniformly in (x, μ) . Then there exists a unique solution $u : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ to

$$\mathcal{L}_{x,\mu} u_l(x, y, \mu) = B_l(x, y, \mu), l \in \{1, \dots, k\}$$

$\int_{\mathbb{R}^d} u(x, y, \mu) \pi(dy; x, \mu) = 0$, and $u, \partial_y u, \partial_y^2 u$ are locally Hölder continuous in y uniformly in x, μ and have at most polynomial growth as $|y| \rightarrow \infty$.

Proof. This follows as in the proof of Theorem 2.1 i) in [53] after noting that the estimates on the transition density associated to $\mathcal{L}_{x,\mu}$ collected in Lemma 3 extend to our situation where the f has linear growth in y and $\mathcal{L}_{x,\mu}^*$ is dissipative via the transformation argument provided in [23] Chapter 1 Section 5. \square

Lemma A.2. Consider $h : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$. Suppose h grows at most polynomially in y and is locally Hölder continuous uniformly in x, μ . Then

$$\begin{aligned} \int_{\mathbb{R}^d} h(x, y, \mu_1) \pi(dy; x, \mu_1) - \int_{\mathbb{R}^d} h(x, y, \mu_2) \pi(dy; x, \mu_2) &= \\ (32) \quad &= \int_{\mathbb{R}^d} h(x, y, \mu_1) - h(x, y, \mu_2) - [\mathcal{L}_{x,\mu_1} - \mathcal{L}_{x,\mu_2}]v(x, y, \mu_2) \pi(dy; x, \mu_1) \end{aligned}$$

for all $x \in \mathbb{R}^d, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, and

$$\begin{aligned} \int_{\mathbb{R}^d} h(x_1, y, \mu) \pi(dy; x_1, \mu) - \int_{\mathbb{R}^d} h(x_2, y, \mu) \pi(dy; x_2, \mu) &= \\ (33) \quad &= \int_{\mathbb{R}^d} h(x_1, y, \mu) - h(x_2, y, \mu) - [\mathcal{L}_{x_1,\mu} - \mathcal{L}_{x_2,\mu}]v(x_2, y, \mu) \pi(dy; x_1, \mu) \end{aligned}$$

for all $x_1, x_2 \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)$.

In addition, consider v solving

$$(34) \quad \mathcal{L}_{x,\mu} v(x, y, \mu) = h(x, y, \mu) - \int_{\mathbb{R}^d} h(x, \bar{y}, \mu) \pi(d\bar{y}; x, \mu),$$

and $\mathcal{L}_{x,\mu}^{(k,j,\alpha(\mathbf{p}_k))}[z_{\mathbf{p}_k}]$ is the differential operator acting on $\phi \in C_b^2(\mathbb{R}^d)$ by

$$\mathcal{L}_{x,\mu}^{(k,j,\alpha(\mathbf{p}_k))}[z_{\mathbf{p}_k}]\phi(y) = D^{(k,j,\alpha(\mathbf{p}_k))} f(x, y, \mu)[z_{\mathbf{p}_k}] \cdot \nabla \phi(y) + D^{(k,j,\alpha(\mathbf{p}_k))} a(x, y, \mu)[z_{\mathbf{p}_k}] : \nabla^2 \phi(y),$$

where the inner products are taken in each d -dimensional component of the derivative matrices of f and each $d \times d$ -dimensional component of the derivative matrices of a .

Suppose that for some complete collection of multi-indices ζ that $h \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, $f \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, $a \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, and that $\partial_y v \in \mathcal{M}_p^{\zeta'}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, $\partial_y^2 v \in \mathcal{M}_p^{\zeta'}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, where ζ' is obtained from removing any multi-indices which contain the maximal first and second values from ζ . Then for any multi-index $(n, l, \beta) \in \zeta$:

$$(35) \quad D^{(n,l,\beta)} \int_{\mathbb{R}^d} h(x, y, \mu) \pi(dy; x, \mu)[z_1, \dots, z_n] = \int_{\mathbb{R}^d} \left(D^{(n,l,\beta)} h(x, y, \mu)[z_1, \dots, z_n] - \sum_{k=0}^n \sum_{j=0}^l \sum_{\mathbf{p}_k} C_{(\mathbf{p}_k, j, n, l)} \mathcal{L}_{x,\mu}^{(k,j,\alpha(\mathbf{p}_k))}[z_{\mathbf{p}_k}] D^{(n-k, l-j, \alpha(\mathbf{p}'_{n-k}))} v(x, y, \mu)[z_{\mathbf{p}'_{n-k}}] \right) \pi(dy; x, \mu)$$

where here $\mathbf{p}_k \in \binom{\{1, \dots, n\}}{k}$ with $\mathbf{p}'_{n-k} = \{1, \dots, n\} \setminus \mathbf{p}_k$, for $\mathbf{p}_k = \{p_1, \dots, p_k\}$, the argument $[z_{\mathbf{p}_k}]$ denotes $[z_{p_1}, \dots, z_{p_k}]$, and $\alpha(\mathbf{p}_k) \in \mathbb{N}^k$ is determined by $\beta = (\beta_1, \dots, \beta_n)$ by $\alpha(\mathbf{p}_k) = (\alpha_1, \dots, \alpha_k)$, $\alpha_j = \beta_{p_j}$, $j \in \{1, \dots, k\}$, and similarly for $\alpha(\mathbf{p}'_{n-k})$. Also here $C_{(\mathbf{p}_0, 0, n, l)} = 0$, and $C_{(\mathbf{p}_k, j, n, l)} > 0$, $C_{(\mathbf{p}_k, j, n, l)} \in \mathbb{N}$ for $(k, j) \in \mathbb{N}^2$, $(k, j) \neq (0, 0)$. See Remark A.3 for an iterative way to define the constants $C_{(p, j, n, \ell)}$.

Proof. Note that by the current assumptions, the integrand on the right hand side of (35) grows at most polynomially in $|y|$ uniformly in x, μ , and hence is integrable against π . This kind of “transfer formula” on the regularity of derivatives in the coefficients of averaged functions against an invariant measure is the subject of Lemmas 3.2 and 3.7 in [53] and [56] respectively.

The result follows almost directly from the aforementioned Lemmas, taking a bit of care to account for the fact that we are dealing with Lions derivatives, and that our assumptions are a bit different than those found in those papers. Namely, in [53], boundedness of f is assumed. Tracking where this assumption is being used in Lemma 3.2 i), we see they come up only when employing the Equation (28) from [48]. In the proof of that result, however, the boundedness assumption on the coefficients is only needed in order to obtain the regularity for the transition density in Proposition 2 of [48]. Upon inspecting the results in the PDE literature that they are using, one can see that in fact the same regularity holds in our regime, where f itself has linear growth but $\partial_y f$ is bounded (under the additional assumption of local Hölder continuity of derivatives of the coefficients in y and dissipativity). In particular, the parametrix method used in Chapter 9 of [25] to prove the bounds collected as Proposition 2 in [48] can be extended to the case of dissipative parabolic PDEs with growth in their coefficients. See [23] Chapter 1 Section 5 (this result is also partially stated in terms of the transition density for SDEs as Theorem 4.1 in [49]).

Similarly, in Lemma 3.2 ii), the boundedness of f is only being used to appeal to the expression given for \mathcal{K}_2 in Lemma 4.1 in [55], wherein this boundedness assumption again is only used for the same estimates on the transition density (collected as Lemma 3.3 there), so again by the extension provided by [23] Chapter 1 Section 5, the result applies in our setting as well. Note that the expression for \mathcal{K}_2 in that paper is coming from the same computation as [48] Equations (25) and (26), except without dividing by h .

Thus, by this discussion, Equation (35) holds in the case that $n = 0$ (i.e. only derivatives in x) and (33), follow directly from Lemma 3.2 in [53].

Once we know that the result holds for $n = 0, l = 1$, so

$$(36) \quad D^{(0,1,0)} \left[\int_{\mathbb{R}^d} h(x, y, \mu) \pi(dy; x, \mu) \right] = \int_{\mathbb{R}^d} D^{(0,1,0)} h(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,1,0)} v(x, y, \mu) \pi(dy; x, \mu),$$

we can reapply the result for one derivative to the functions $p_k(x, y, \mu) = \left[D^{(0,1,0)} h(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,1,0)} v(x, y, \mu) \right]_k$, $k = 1, \dots, d$. The result applies since by assumption, each p_k is jointly continuous and grows at most polynomially in y uniformly in x and μ .

Then

$$(37) \quad D^{(0,1,0)} \left[\int_{\mathbb{R}^d} p_k(x, y, \mu) \pi(dy; x, \mu) \right] = \int_{\mathbb{R}^d} D^{(0,1,0)} p_k(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,1,0)} w_k(x, y, \mu) \pi(dy; x, \mu)$$

where

$$\mathcal{L}_{x,\mu} w_k(x, y, \mu) = p_k(x, y, \mu) - \int_{\mathbb{R}^d} p_k(x, y, \mu) \pi(dy; x, \mu), k \in \{1, \dots, d\}.$$

In other words,

$$\begin{aligned} \mathcal{L}_{x,\mu} w_k(x, y, \mu) &= \partial_{x_k} h(x, y, \mu) - \partial_{x_k} f(x, y, \mu) \cdot \partial_y v(x, y, \mu) - \partial_{x_k} a(x, y, \mu) : \partial_y^2 v(x, y, \mu) \\ &\quad - \int_{\mathbb{R}^d} \partial_{x_k} h(x, y, \mu) - \partial_{x_k} f(x, y, \mu) \cdot \partial_y v(x, y, \mu) - \partial_{x_k} a(x, y, \mu) : \partial_y^2 v(x, y, \mu) \pi(dy; x, \mu) \\ &= \partial_{x_k} h(x, y, \mu) - \partial_{x_k} f(x, y, \mu) \cdot \partial_y v(x, y, \mu) - \partial_{x_k} a(x, y, \mu) : \partial_y^2 v(x, y, \mu) \\ &\quad - \partial_{x_k} \left[\int_{\mathbb{R}^d} h(x, y, \mu) \pi(dy; x, \mu) \right], \end{aligned}$$

where in the second equality we used the expression (36).

But, differentiating the expression which v satisfies (34) in x_k , we have

$$\begin{aligned} \mathcal{L}_{x,\mu} \partial_{x_k} v &= \partial_{x_k} h(x, y, \mu) - \partial_{x_k} \left[\int_{\mathbb{R}^d} h(x, y, \mu) \pi(dy; x, \mu) \right] - \partial_{x_k} f(x, y, \mu) \cdot \partial_y v(x, y, \mu) \\ &\quad - \partial_{x_k} a(x, y, \mu) : \partial_y^2 v(x, y, \mu), \end{aligned}$$

and by the uniqueness granted by Lemma A.1, we have in fact that $w_k = \partial_{x_k} v$. By rewriting the expression (37) in terms of h and v , it reads:

$$\begin{aligned} D^{(0,2,0)} \left[\int_{\mathbb{R}^d} h(x, y, \mu) \pi(dy; x, \mu) \right] \\ = \int_{\mathbb{R}^d} D^{(0,2,0)} h(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,2,0)} v(x, y, \mu) - 2\mathcal{L}_{x,\mu}^{(0,1,0)} D^{(0,1,0)} v(x, y, \mu) \pi(dy; x, \mu). \end{aligned}$$

In addition, since we take no derivatives in the expression (32), the proof follows in the exact same way as when the difference is taken in the parameters which live in Euclidean space. In fact, it also holds that if $\hat{h} : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ satisfies the same polynomial growth and Hölder continuity assumptions in y , then

$$(38) \quad \begin{aligned} &\int_{\mathbb{R}^d} \hat{h}(x, y, z_1, \mu_1) \pi(dy; x, \mu_1) - \int_{\mathbb{R}^d} \hat{h}(x, y, z_2, \mu_2) \pi(dy; x, \mu_2) \\ &= \int_{\mathbb{R}^d} \hat{h}(x, y, z_1, \mu_1) - \hat{h}(x, y, z_2, \mu_2) - [\mathcal{L}_{x,\mu_1} - \mathcal{L}_{x,\mu_2}] \hat{v}(x, y, z_2, \mu_2) \pi(dy; x, \mu_1) \end{aligned}$$

where

$$\mathcal{L}_{x,\mu} \hat{v}(x, y, z, \mu) = \hat{h}(x, y, z, \mu) - \int_{\mathbb{R}^d} \hat{h}(x, y, z, \mu) \pi(dy; x, \mu).$$

As we will see, this is useful for dealing with the Lions derivatives when proving Equation (35).

In order to arrive at the full expression in Equation (35), we first examine the case of $n = 1, l, \beta = 0$. Then the result will follow in the iterative manner which we just outlined above.

Let $\tilde{h}, \tilde{f}, \tilde{a}, \tilde{\pi}$ be the lifted functions on some $L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$ as per Definition B.1. For $\phi : L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d) \rightarrow \mathbb{R}^k$ Gateaux differentiable, we denote by $D[\phi(\tilde{X}); \tilde{Y}] = \lim_{t \downarrow 0} [\phi(\tilde{X} + t\tilde{Y}) - \phi(\tilde{X})]/t$ the Gateaux derivative of ϕ at \tilde{X} in the direction \tilde{Y} .

Fixing x, \tilde{X}, \tilde{Y} , we can define $H(t, y) = \tilde{h}(x, y, \tilde{X} + t\tilde{Y}), F(t, y) = \tilde{f}(x, y, \tilde{X} + t\tilde{Y}), A(t, y) = \tilde{a}(x, y, \tilde{X} + t\tilde{Y}), \Pi(dy; t) = \tilde{\pi}(x, y, \tilde{X} + t\tilde{Y})$. Then, since x and μ are only parameters, we have for all $t \in \mathbb{R}$ $\Pi(\cdot; t)$ is the unique invariant measure associated to the linear operator with first order coefficients given by $F(t, y)$ and second order coefficients given by $A(t, y)$, and $\int_{\mathbb{R}^d} H(t, y) \Pi(dy; t) = 0$. By definition of the Lions derivative,

we also have the derivative of the coefficients with respect to t have at most polynomial growth in y uniformly in t by assumption. Then by Lemma 3.2 1) in [53] with $\eta = 1$, we have

$$\partial_t \left[\int_{\mathbb{R}^d} H(t, y) \Pi(dy; t) \right] = \int_{\mathbb{R}^d} \partial_t H(t, y) - (\partial_t F(t, y)) \cdot V(t, y) - (\partial_t A(t, y)) : V(t, y) \Pi(dy; t)$$

where

$$F(t, y) \cdot \partial_y V(t, y) + A(t, y) : \partial_y^2 V(t, y) = H(t, y) - \int_{\mathbb{R}^d} H(t, \bar{y}) \Pi(\bar{y}; t).$$

Thus, evaluating at $t = 0$, we have:

$$\begin{aligned} D \left[\int_{\mathbb{R}^d} \tilde{h}(x, y, \tilde{X}) \tilde{\pi}(y; x, \tilde{X}) dy; \tilde{Y} \right] &= \int_{\mathbb{R}^d} D[\tilde{h}(x, y, \tilde{X}); \tilde{Y}] - D[\tilde{f}(x, y, \tilde{X}); \tilde{Y}] \cdot \partial_y \tilde{v}(x, y, \tilde{X}) \\ &\quad - D[\tilde{a}(x, y, \tilde{X}); \tilde{Y}] : \partial_y^2 \tilde{v}(x, y, \tilde{X}) \pi(dy; x, \tilde{X}) \end{aligned}$$

for all $x, \in \mathbb{R}^d, \tilde{X}, \tilde{Y} \in L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$. Here

$$\tilde{\mathcal{L}}_{x, \tilde{X}} \tilde{v}(x, y, \tilde{X}) := \tilde{f}(x, y, \tilde{X}) \cdot \partial_y \tilde{v}(x, y, \tilde{X}) + \tilde{a}(x, y, \tilde{X}) : \partial_y^2 \tilde{v}(x, y, \tilde{X}) = \tilde{h}(x, y, \tilde{X}) - \int_{\mathbb{R}^d} \tilde{h}(x, y, \tilde{X}) \pi(dy; x, \tilde{X}).$$

Now we note that by definition of the liftings and the Lions derivative:

$$\begin{aligned} D \left[\int_{\mathbb{R}^d} \tilde{h}(x, y, \tilde{X}) \tilde{\pi}(y; x, \tilde{X}) dy; \tilde{Y} \right] &= \tilde{\mathbb{E}} \left[\int_{\mathbb{R}^d} D\tilde{h}(x, y, \tilde{X}) - D\tilde{f}(x, y, \tilde{X}) \cdot \partial_y v(x, y, \mu) \right. \\ &\quad \left. - D\tilde{a}(x, y, \tilde{X}) : \partial_y^2 v(x, y, \mu) \pi(dy; x, \mu) \cdot \tilde{Y} \right] \\ &= \tilde{\mathbb{E}} \left[\int_{\mathbb{R}^d} \partial_\mu h(x, y, \mu)[\tilde{X}] - \partial_\mu f(x, y, \mu)[\tilde{X}] \cdot \partial_y v(x, y, \mu) \right. \\ &\quad \left. - \partial_\mu a(x, y, \mu)[\tilde{X}] : \partial_y^2 v(x, y, \mu) \pi(dy; x, \mu) \cdot \tilde{Y} \right], \end{aligned}$$

where $D\phi(\tilde{X})$ is the Fréchet derivative of $\phi : L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d) \rightarrow \mathbb{R}^k$ at \tilde{X} , and we denote the expectation with respect to $\tilde{\mathbb{P}}$ by $\tilde{\mathbb{E}}$.

Now it is useful to note that by assumption, the integrand on the right hand side of the above display is locally Hölder continuous and has at most polynomial growth in y uniformly in x, μ, z . So we have, letting $\hat{h}(x, y, z, \mu) = \partial_\mu h(x, y, \mu)[z] - \partial_\mu f(x, y, \mu)[z] \cdot \partial_y v(x, y, \mu) - \partial_\mu a(x, y, \mu)[z] : \partial_y^2 v(x, y, \mu)$, the expression (38) holds with this choice of \hat{h} . By Lemma A.1, $\partial_y \hat{v}$ and $\partial_y^2 \hat{v}$ grow at most polynomially in y uniformly in x, μ, z (it makes no difference to add another space parameter). Thus, by the assumptions on the coefficients:

$$\begin{aligned} &\left| \int_{\mathbb{R}^d} \hat{h}(x, y, z_1, \mu_1) \pi(dy; x, \mu_1) - \int_{\mathbb{R}^d} \hat{h}(x, y, z_2, \mu_2) \pi(dy; x, \mu_2) \right| \\ &\leq \int_{\mathbb{R}^d} |\hat{h}(x, y, z_1, \mu_1) - \hat{h}(x, y, z_2, \mu_2)| + |\mathcal{L}_{x, \mu_1} - \mathcal{L}_{x, \mu_2}] \hat{v}(x, y, z_2, \mu_2)| \pi(dy; x, \mu_1) \\ &\leq C \int_{\mathbb{R}^d} [\mathbb{W}_2(\mu_1, \mu_2) + |z_1 - z_2|] (1 + |y|^p) \pi(dy; x, \mu_1), \text{ for some } p \in \mathbb{N} \\ &\leq C[\mathbb{W}_2(\mu_1, \mu_2) + |z_1 - z_2|]. \end{aligned}$$

Then, taking $\{\tilde{X}^N\}_{N \in \mathbb{N}} \subset L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$ such that $\tilde{X}^N \rightarrow \tilde{X}$ and $\tilde{X}^N \sim \mu^N, \tilde{X} \sim \mu$, we have

$$\begin{aligned} &\lim_{N \rightarrow \infty} \sup_{\tilde{Y} : \|\tilde{Y}\|_{L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)} \neq 0} \left(\tilde{\mathbb{E}} \left[\int_{\mathbb{R}^d} \partial_\mu h(x, y, \mu^N)[\tilde{X}^N] - \partial_\mu f(x, y, \mu^N)[\tilde{X}^N] \cdot \partial_y v(x, y, \mu^N) \right. \right. \\ &\quad \left. \left. - \partial_\mu a(x, y, \mu^N)[\tilde{X}^N] : \partial_y^2 v(x, y, \mu^N) \pi(dy; x, \mu^N) \cdot \tilde{Y} \right] \right. \\ &\quad \left. - \tilde{\mathbb{E}} \left[\int_{\mathbb{R}^d} \partial_\mu h(x, y, \mu)[\tilde{X}] - \partial_\mu f(x, y, \mu)[\tilde{X}] \cdot \partial_y v(x, y, \mu) \right. \right. \end{aligned}$$

$$\begin{aligned}
& -\partial_\mu a(x, y, \mu)[\tilde{X}] : \partial_y^2 v(x, y, \mu) \pi(dy; x, \mu) \cdot \tilde{Y} \Big] / \|\tilde{Y}\|_{L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)} \\
& \leq \lim_{N \rightarrow \infty} \tilde{\mathbb{E}} \left[\left| \int_{\mathbb{R}^d} \left\{ \partial_\mu h(x, y, \mu^N)[\tilde{X}^N] - \partial_\mu f(x, y, \mu^N)[\tilde{X}^N] \cdot \partial_y v(x, y, \mu^N) - \right. \right. \right. \\
& \quad \left. \left. \left. - \partial_\mu a(x, y, \mu^N)[\tilde{X}^N] \partial_y^2 v(x, y, \mu^N) \right\} \pi(y; x, \mu^N) \right. \right. \\
& \quad \left. \left. - \left\{ \partial_\mu h(x, y, \mu)[\tilde{X}] - \partial_\mu f(x, y, \mu)[\tilde{X}] \cdot \partial_y v(x, y, \mu) - \partial_\mu a(x, y, \mu)[\tilde{X}] : \partial_y^2 v(x, y, \mu) \right\} \pi(y; x, \mu) dy \right|^2 \right]^{1/2} \\
& \text{by Hölder's inequality} \\
& \leq C \lim_{N \rightarrow \infty} \left\{ \tilde{\mathbb{E}} \left[|\tilde{X}^N - \tilde{X}|^2 \right] + \mathbb{W}_2(\mu^N, \mu) \right\} \\
& = 0.
\end{aligned}$$

Then in fact, by, e.g. Proposition 3.2.15 in [21], $\int_{\mathbb{R}^d} h(x, y, \tilde{X}) \pi(y; x, \tilde{X}) dy$ is Fréchet differentiable, and

$$\begin{aligned}
\tilde{\mathbb{E}} \left[D \int_{\mathbb{R}^d} \tilde{h}(x, y, \tilde{X}) \tilde{\pi}(y; x, \tilde{X}) dy \cdot \tilde{Y} \right] &= \tilde{\mathbb{E}} \left[\int_{\mathbb{R}^d} \partial_\mu h(x, y, \mu)[\tilde{X}] - \partial_\mu f(x, y, \mu)[\tilde{X}] \cdot \partial_y v(x, y, \mu) \right. \\
&\quad \left. - \partial_\mu a(x, y, \mu)[\tilde{X}] : \partial_y^2 v(x, y, \mu) \pi(dy; x, \mu) \cdot \tilde{Y} \right].
\end{aligned}$$

Since this holds for all $\tilde{Y} \in L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$, we have

$$\begin{aligned}
D \int_{\mathbb{R}^d} \tilde{h}(x, y, \tilde{X}) \tilde{\pi}(y; x, \tilde{X}) dy &= \int_{\mathbb{R}^d} \partial_\mu h(x, y, \mu)[\tilde{X}] - \partial_\mu f(x, y, \mu)[\tilde{X}] \cdot \partial_y v(x, y, \mu) \\
&\quad - \partial_\mu a(x, y, \mu)[\tilde{X}] : \partial_y^2 v_{yy}(x, y, \mu) \pi(dy; x, \mu)
\end{aligned}$$

$\tilde{\mathbb{P}}$ -almost surely, for all $x \in \mathbb{R}^d$, $\tilde{X} \in L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$, $\mu = \mathcal{L}(\tilde{X})$.

Then, by definition,

$$D^{(1,0,0)} \int_{\mathbb{R}^d} h(x, y, \mu) \pi(y; x, \mu) dy[z] = \int_{\mathbb{R}^d} D^{(1,0,0)} h(x, y, \mu)[z] - \mathcal{L}_{x,\mu}^{(1,0,0)}[z] v(x, y, \mu) \pi(dy; x, \mu),$$

for all $x \in \mathbb{R}^d$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, and μ -almost every $z \in \mathbb{R}^d$.

Then, using the same iterative argument as outlined above in the case of the x derivatives both in l and n along with this same argument to handle the Fréchet derivatives, we arrive at the expression for the mixed derivatives given in (35), with $\beta = 0$. Finally, differentiating the expression in z_1, \dots, z_n according to β and using uniform integrability to pass the derivatives into the integral against π , we obtain the full expression. \square

Remark A.3. The non-negative integers $C_{(\mathbf{p}_k, j, n, l)}$ in the statement of Lemma A.2 can be iteratively computed according to the following rules:

$$C_{(\mathbf{p}_0, 0, 0, 0)} = 0.$$

and to go up in l (taking an x derivative), we have for any $l, n \in \mathbb{N}$, $\mathbb{N} \ni j \leq l+1$, $\mathbb{N} \ni k \leq n$, and $\mathbf{p}_k \in \binom{\{1, \dots, n\}}{k}$:

$$C_{(\mathbf{p}_k, j, n, l+1)} = \begin{cases} C_{(\mathbf{p}_k, l, n, l)}, & \text{if } j = l+1 \\ C_{(\mathbf{p}_k, 0, n, l)}, & \text{if } j = 0 \\ C_{(\mathbf{p}_k, 1, n, l)} + 1, & \text{if } j = 1 \\ C_{(\mathbf{p}_k, j-1, n, l)} + C_{(\mathbf{p}_k, j, n, l)}, & \text{otherwise} \end{cases}.$$

To go up in n (taking a measure derivative), we have for any $l, n \in \mathbb{N}$, $\mathbb{N} \ni j \leq l$, $\mathbb{N} \ni k \leq n+1$, and $\mathbf{p}_k \in \binom{\{1, \dots, n+1\}}{k}$

$$C_{(\mathbf{p}_k, j, n+1, l)} = \begin{cases} C_{(\{1, \dots, n\}, j, n, l)} & \text{if } k = n+1 \\ C_{(\mathbf{p}_0, j, n, l)} & \text{if } k = 0 \\ \mathbb{1}_{\mathbf{p}_1 = \{n+1\}} + C_{(\mathbf{p}_1, j, n, l)} \mathbb{1}_{\mathbf{p}_1 \neq \{n+1\}} & \text{if } k = 1 \\ C_{(\mathbf{p}_k \setminus \{n+1\}, j, n, l)} \mathbb{1}_{\{n+1\} \in \mathbf{p}_k} + C_{(\mathbf{p}_k, j, n, l)} \mathbb{1}_{\{n+1\} \notin \mathbf{p}_k}, & \text{otherwise} \end{cases}.$$

This can be seen from tracking the constants in the iterative argument outlined in the proof.

In dimension $d = 1$, one can actually strengthen Lemma A.2 using the available exact formula for the invariant distribution. In fact in $d = 1$ we have the Lemma A.4.

Lemma A.4. Consider the case where $h, f, a : \mathbb{R} \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \rightarrow \mathbb{R}$, i.e., $d = 1$. Then, the results of Lemma A.2 hold for h that is jointly continuous in (x, y, μ) and grows at most polynomially in y uniformly in x, μ , and dropping the local Hölder continuity assumption from the Assumptions A1) and A2) and the definition of \mathcal{M}_p^ζ (that is, Equation (14)).

Proof. The essential idea of this argument is that appealing to Equation (28) from [48] and the expression given for \mathcal{K}_2 in Lemma 4.1 in [55] is unnecessary, and we can instead use a direct argument using the explicit form of π and v that we have in the 1D situation (see [26] p. 105). Thus the use of local Hölder continuity in order to establish bounds on the fundamental solution of the Cauchy problem associated to $\mathcal{L}_{x, \mu}$ is unnecessary.

We have (with some abuse of notation also denoting the density of π by π):

$$(39) \quad \pi(y; x, \mu) = \frac{Z(x, \mu)}{a(x, y, \mu)} \exp\left(\int_0^y \frac{f(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y}\right)$$

$$(40) \quad v(x, y, \mu) = \int_{-\infty}^y \frac{1}{a(x, \bar{y}, \mu)\pi(\bar{y}; x, \mu)} \left[\int_{-\infty}^{\bar{y}} h(x, \tilde{y}, \mu)\pi(\tilde{y}; x, \mu) d\tilde{y} - \bar{h}(x, \mu) \int_{-\infty}^{\bar{y}} \pi(\tilde{y}; x, \mu) d\tilde{y} \right] d\bar{y}$$

where $Z^{-1}(x, \mu) := \int_{\mathbb{R}} \frac{1}{a(x, y, \mu)} \exp\left(\int_0^y \frac{f(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y}\right) dy$ is the normalizing constant, and we use $\bar{h}(x, \mu) := \int_{\mathbb{R}} h(x, \bar{y}, \mu)\pi(\bar{y}; x, \mu) d\bar{y}$. In the proof we will at times use the subscript notation e.g. f_x rather than $\partial_x f$ to denote partial derivatives for presentation purposes.

First we see that Equation (35) holds with $n = 0$ and $l = 1$.

Using the dominated convergence theorem to pass derivatives inside the integrals where necessary, we have

$$\begin{aligned} \partial_x \int_{\mathbb{R}} h(x, y, \mu)\pi(y; x, \mu) dy &= \int_{\mathbb{R}} h_x(x, y, \mu)\pi(y; x, \mu) dy + \int_{\mathbb{R}} h(x, y, \mu)\pi_x(y; x, \mu) dy \\ \pi_x(y; x, \mu) &= \pi(y; x, \mu) \left[\int_0^y \frac{1}{a(x, \bar{y}, \mu)} [f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu)a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)] d\bar{y} + \frac{Z_x(x, \mu)}{Z(x, \mu)} \right. \\ &\quad \left. - \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \right] \\ v_y(x, y, \mu) &= \frac{1}{a(x, y, \mu)\pi(y; x, \mu)} \left[\int_{-\infty}^y h(x, \bar{y}, \mu)\pi(\bar{y}; x, \mu) d\bar{y} - \bar{h}(x, \mu) \int_{-\infty}^y \pi(\bar{y}; x, \mu) d\bar{y} \right] \\ v_{yy}(x, y, \mu) &= -\frac{f(x, y, \mu)}{a^2(x, y, \mu)\pi(y; x, \mu)} \left[\int_{-\infty}^y h(x, \bar{y}, \mu)\pi(\bar{y}; x, \mu) d\bar{y} - \bar{h}(x, \mu) \int_{-\infty}^y \pi(\bar{y}; x, \mu) d\bar{y} \right] \\ &\quad + \frac{h(x, y, \mu) - \bar{h}(x, \mu)}{a(x, y, \mu)}. \end{aligned}$$

We need to establish that

$$\int_{\mathbb{R}} h(x, y, \mu)\pi_x(y; x, \mu) dy = - \int_{\mathbb{R}} \mathcal{L}_{x, \mu}^{(0, 1, 0)} v(x, y, \mu)\pi(y; x, \mu) dy$$

so

$$\begin{aligned} & \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \int_{\mathbb{R}} h(x, y, \mu) \left[\int_0^y \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y} - \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \right] \pi(y; x, \mu) dy \\ &= \int_{\mathbb{R}} \frac{f_x(x, y, \mu) - f(x, y, \mu) a_x(x, y, \mu) / a(x, y, \mu)}{a(x, y, \mu)} \left[\bar{h}(x, \mu) \int_{-\infty}^y \pi(\bar{y}; x, \mu) d\bar{y} - \int_{-\infty}^y h(x, \bar{y}, \mu) \pi(\bar{y}; x, \mu) d\bar{y} \right] \\ &+ a_x(x, y, \mu) \frac{\bar{h}(x, \mu) - h(x, y, \mu)}{a(x, y, \mu)} \pi(y; x, \mu) dy. \end{aligned}$$

Noting the last terms are already the same, we can just establish

$$\begin{aligned} & \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \int_{\mathbb{R}} h(x, y, \mu) \int_0^y \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y} \pi(y; x, \mu) dy \\ &= \int_{\mathbb{R}} \frac{f_x(x, y, \mu) - f(x, y, \mu) a_x(x, y, \mu) / a(x, y, \mu)}{a(x, y, \mu)} \left[\bar{h}(x, \mu) \int_{-\infty}^y \pi(\bar{y}; x, \mu) d\bar{y} - \int_{-\infty}^y h(x, \bar{y}, \mu) \pi(\bar{y}; x, \mu) d\bar{y} \right] \\ &+ \bar{h}(x, \mu) \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \pi(y; x, \mu) dy. \end{aligned}$$

We work from the first term and get to the second. Changing the order of integration, we have

$$\begin{aligned} & \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \int_{\mathbb{R}} h(x, y, \mu) \int_0^y \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y} \pi(y; x, \mu) dy \\ &= \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \int_0^\infty \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{\bar{y}}^\infty h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y} \\ &- \int_{-\infty}^0 \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y}. \end{aligned}$$

Then using

$$\int_{\bar{y}}^\infty h(x, y, \mu) \pi(y; x, \mu) dy = \bar{h}(x, \mu) - \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy$$

we can continue

$$\begin{aligned} &= \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \int_0^\infty \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} [\bar{h}(x, \mu) - \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy] d\bar{y} \\ &- \int_{-\infty}^0 \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y} \\ &= \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \bar{h}(x, \mu) \int_0^\infty \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y} \\ &- \int_{\mathbb{R}} \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y}. \end{aligned}$$

Then using that π integrates to 1, we have

$$\begin{aligned} &= \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \bar{h}(x, \mu) \int_0^\infty \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{\mathbb{R}} \pi(y; x, \mu) dy d\bar{y} \\ &- \int_{\mathbb{R}} \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y} \\ &= \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \\ &+ \bar{h}(x, \mu) \int_0^\infty \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \left[\int_{-\infty}^{\bar{y}} \pi(y; x, \mu) dy + \int_{\bar{y}}^\infty \pi(y; x, \mu) dy \right] d\bar{y} \\ &- \int_{\mathbb{R}} \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu) / a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y} \end{aligned}$$

$$\begin{aligned}
&= \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \bar{h}(x, \mu) \int_0^\infty \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} \pi(y; x, \mu) dy d\bar{y} \\
&+ \bar{h}(x, \mu) \int_{-\infty}^0 \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} \pi(y; x, \mu) dy d\bar{y} \\
&+ \bar{h}(x, \mu) \int_0^\infty \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{\bar{y}}^\infty \pi(y; x, \mu) dy d\bar{y} \\
&- \bar{h}(x, \mu) \int_{-\infty}^0 \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} \pi(y; x, \mu) dy d\bar{y} \\
&- \int_{\mathbb{R}} \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y} \\
&= \frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \bar{h}(x, \mu) \int_{\mathbb{R}} \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} \pi(y; x, \mu) dy d\bar{y} \\
&+ \bar{h}(x, \mu) \int_{\mathbb{R}} \int_0^y \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y} \pi(y; x, \mu) dy \\
&- \int_{\mathbb{R}} \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy d\bar{y},
\end{aligned}$$

where in the last step we again changed the order of integration.

Now using

$$\pi_x(y; x, \mu) = \pi(y; x, \mu) \left[\int_0^y \frac{1}{a(x, \bar{y}, \mu)} [f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)] d\bar{y} + \frac{Z_x(x, \mu)}{Z(x, \mu)} - \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \right],$$

we have

$$\begin{aligned}
&\frac{Z_x(x, \mu)}{Z(x, \mu)} \bar{h}(x, \mu) + \bar{h}(x, \mu) \int_{\mathbb{R}} \int_0^y \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y} \pi(y; x, \mu) dy \\
&= \bar{h}(x, \mu) \int_{\mathbb{R}} \left[\int_0^y \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} d\bar{y} + \frac{Z_x(x, \mu)}{Z(x, \mu)} \right] \pi(y; x, \mu) dy \\
&= \bar{h}(x, \mu) \int_{\mathbb{R}} \pi(y; x, \mu) dy + \bar{h}(x, \mu) \int_{\mathbb{R}} \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \pi(y; x, \mu) dy \\
&= \bar{h}(x, \mu) \partial_x \int_{\mathbb{R}} \pi(y; x, \mu) dy + \bar{h}(x, \mu) \int_{\mathbb{R}} \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \pi(y; x, \mu) dy \\
&= \bar{h}(x, \mu) \int_{\mathbb{R}} \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \pi(y; x, \mu) dy,
\end{aligned}$$

where in the first and last step we again used π integrates to 1 for all x, μ .

Thus we can replace the first and third term in our chain of equalities to get:

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{f_x(x, \bar{y}, \mu) - f(x, \bar{y}, \mu) a_x(x, \bar{y}, \mu)/a(x, \bar{y}, \mu)}{a(x, \bar{y}, \mu)} \left[\bar{h}(x, \mu) \int_{-\infty}^{\bar{y}} \pi(y; x, \mu) dy - \int_{-\infty}^{\bar{y}} h(x, y, \mu) \pi(y; x, \mu) dy \right] d\bar{y} \\
&+ \bar{h}(x, \mu) \int_{\mathbb{R}} \frac{a_x(x, y, \mu)}{a(x, y, \mu)} \pi(y; x, \mu) dy
\end{aligned}$$

as desired.

For the case that $n = 1, l = 0$, we use the same argument, but using Gateaux derivatives of the lifted functions $\tilde{h}, \tilde{a}, \tilde{f}, \tilde{v}, \tilde{\pi}$ in the direction of some $\tilde{Y} \in L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R})$, and a similar argument to that in Lemma A.2 to establish Fréchet differentiability of the lifted function $\int_{\mathbb{R}} \tilde{h}(x, y, \tilde{X}) \tilde{\pi}(y; x, \tilde{X})$ and hence Lions differentiability of $\int_{\mathbb{R}} h(x, y, \mu) \pi(y; x, \mu) dy$.

Once the equality (35) is established in these two base cases, the same iterative argument as outlined in Lemma A.2 applies to arrive at the full expression.

Now we show the equality (32). We have

$$\begin{aligned}
& \int_{\mathbb{R}} h(x, y, \mu_1) - h(x, y, \mu_2) - [\mathcal{L}_{x, \mu_1} - \mathcal{L}_{x, \mu_2}]v(x, y, \mu_2)\pi(dy; x, \mu_1) \\
&= \bar{h}(x, \mu_1) - \int_{\mathbb{R}} h(x, y, \mu_2) + \mathcal{L}_{x, \mu_1}v(x, y, \mu_2)\pi(dy; x, \mu_1) + \int_{\mathbb{R}} h(x, y, \mu_2)\pi(dy; x, \mu_1) - \bar{h}(x, \mu_2) \\
&= \bar{h}(x, \mu_1) - \bar{h}(x, \mu_2) - \int_{\mathbb{R}} \mathcal{L}_{x, \mu_1}v(x, y, \mu_2)\pi(dy; x, \mu_1),
\end{aligned}$$

so to show

$$\begin{aligned}
& \int_{\mathbb{R}} h(x, y, \mu_1)\pi(dy; x, \mu_1) - \int_{\mathbb{R}} h(x, y, \mu_2)\pi(dy; x, \mu_2) = \\
&= \int_{\mathbb{R}} h(x, y, \mu_1) - h(x, y, \mu_2) - [\mathcal{L}_{x, \mu_1} - \mathcal{L}_{x, \mu_2}]v(x, y, \mu_2)\pi(dy; x, \mu_1)
\end{aligned}$$

it is sufficient to show

$$\int_{\mathbb{R}} \mathcal{L}_{x, \mu_1}v(x, y, \mu_2)\pi(dy; x, \mu_1) = 0.$$

Using the previous calculation for v_y and v_{yy} , we have

$$\begin{aligned}
& \int_{\mathbb{R}} \mathcal{L}_{x, \mu_1}v(x, y, \mu_2)\pi(dy; x, \mu_1) = \int_{\mathbb{R}} \frac{f(x, y, \mu_1) - f(x, y, \mu_2)a(x, y, \mu_1)/a(x, y, \mu_2)}{a(x, y, \mu_2)} \frac{\pi(y; x, \mu_1)}{\pi(y; x, \mu_2)} \times \\
& \quad \times \left[\int_{-\infty}^y h(x, \bar{y}, \mu_2)\pi(\bar{y}; x, \mu_2)d\bar{y} - \bar{h}(x, \mu_2) \int_{-\infty}^y \pi(\bar{y}; x, \mu_2)d\bar{y} \right] dy \\
& \quad + \int_{\mathbb{R}} a(x, y, \mu_1) \frac{h(x, y, \mu_2) - \bar{h}(x, \mu_2)}{a(x, y, \mu_2)} \pi(y; x, \mu_1) dy \\
&= \int_{\mathbb{R}} g(x, y, \mu_1, \mu_2) \exp\left(\int_0^y g(x, \bar{y}, \mu_1, \mu_2)d\bar{y}\right) \left[\int_{-\infty}^y h(x, \bar{y}, \mu_2)\pi(\bar{y}; x, \mu_2)d\bar{y} - \bar{h}(x, \mu_2) \int_{-\infty}^y \pi(\bar{y}; x, \mu_2)d\bar{y} \right] dy \\
& \quad + \int_{\mathbb{R}} a(x, y, \mu_1) \frac{h(x, y, \mu_2) - \bar{h}(x, \mu_2)}{a(x, y, \mu_2)} \pi(y; x, \mu_1) dy,
\end{aligned}$$

where $g(x, y, \mu_1, \mu_2) = \frac{f(x, y, \mu_1)}{a(x, y, \mu_1)} - \frac{f(x, y, \mu_2)}{a(x, y, \mu_2)}$, and here we used the explicit form of $\pi(y; x, \mu)$. Then by integration by parts, using that

$$\lim_{y \rightarrow \pm\infty} \int_{-\infty}^y h(x, \bar{y}, \mu_2)\pi(\bar{y}; x, \mu_2)d\bar{y} - \bar{h}(x, \mu_2) \int_{-\infty}^y \pi(\bar{y}; x, \mu_2)d\bar{y} = 0,$$

we get

$$\begin{aligned}
& \int_{\mathbb{R}} \mathcal{L}_{x, \mu_1}v(x, y, \mu_2)\pi(dy; x, \mu_1) = \\
&= - \int_{\mathbb{R}} \exp\left(\int_0^y \frac{f(x, \bar{y}, \mu_1)}{a(x, \bar{y}, \mu_1)} - \frac{f(x, \bar{y}, \mu_2)}{a(x, \bar{y}, \mu_2)} d\bar{y}\right) [h(x, y, \mu_2) - \bar{h}(x, \mu_2)]\pi(y; x, \mu_2) dy \\
& \quad + \int_{\mathbb{R}} a(x, y, \mu_1) \frac{h(x, y, \mu_2) - \bar{h}(x, \mu_2)}{a(x, y, \mu_2)} \pi(y; x, \mu_1) dy \\
&= - \int_{\mathbb{R}} \frac{a(x, y, \mu_1)}{a(x, y, \mu_2)} \pi(y; x, \mu_1) [h(x, y, \mu_2) - \bar{h}(x, \mu_2)] dy + \int_{\mathbb{R}} a(x, y, \mu_1) \frac{h(x, y, \mu_2) - \bar{h}(x, \mu_2)}{a(x, y, \mu_2)} \pi(y; x, \mu_1) dy \\
&= 0
\end{aligned}$$

as desired. The proof of the equality (33) follows in the exact same way. \square

Lemma A.5. Consider $B : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ such that

$$\int_{\mathbb{R}^d} B(x, y, \mu)\pi(dy; x, \mu) = 0, \forall x \in \mathbb{R}^d, \mu \in \mathbb{P}_2(\mathbb{R}^d).$$

Suppose that for some complete collection of multi-indices ζ that $B \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, $f \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, $a \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$. Then for the unique classical solution $u : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ to

$$\mathcal{L}_{x,\mu} u(x, y, \mu) = B(x, y, \mu)$$

such that $\int_{\mathbb{R}^d} u(x, y, \mu) \pi(dy; x, \mu) = 0$, and u has at most polynomial growth as $|y| \rightarrow \infty$ (which exists by Lemma A.1): $u, \partial_{y_i} u, \partial_{y_i} \partial_{y_j} u \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ for all $i, j = 1, \dots, d$.

Proof. The proof essentially uses the same tools and a similar method to Lemma A.2 here and Theorem 2.1 Step 3 in [53], so we will only check this in the case for $(n, l, \beta) = (0, 1, 0)$ and then comment on how the rest of the terms follow. Importantly, Lemma A.2 only assumes existence and polynomial growth of derivatives up to one order less than the derivative we obtain from Equation (35).

The result for $(n, l, \beta) = (0, 0, 0)$ is just another way of writing Lemma A.1, once we establish continuity of $u, \partial_y u, \partial_y^2 u$ in x, μ . For this, the proof is similar to Step 4 in the proof of Theorem 2.1 in [53].

We first note that

$$\begin{aligned} \mathcal{L}_{x,\mu_1}[u(x, y, \mu_1) - u(x, y, \mu_2)] &= B(x, y, \mu_1) - \mathcal{L}_{x,\mu_1} u(x, y, \mu_2) \\ &= B(x, y, \mu_1) - B(x, y, \mu_2) - [\mathcal{L}_{x,\mu_1} - \mathcal{L}_{x,\mu_2}]u(x, y, \mu_2). \end{aligned}$$

By the equality (32) from Lemma A.2, we have

$$\begin{aligned} \int_{\mathbb{R}^d} B(x, y, \mu_1) - B(x, y, \mu_2) - [\mathcal{L}_{x,\mu_1} - \mathcal{L}_{x,\mu_2}]u(x, y, \mu_2) \pi(dy, x, \mu_1) &= \\ = \int_{\mathbb{R}^d} B(x, y, \mu_1) \pi(dy; x, \mu_1) - \int_{\mathbb{R}^d} B(x, y, \mu_2) \pi(dy; x, \mu_2) &= 0, \end{aligned}$$

so in fact the inhomogeneity in the above Poisson equation is centered. Now, via the assumptions on B, f, a and the fact that the bound on the growth of u in Lemma A.1 depends linearly on the local Hölder semi-norm of the Poisson Equation's inhomogeneity as per the Proof of Theorem 2.1 i) in [53], we have:

$$\begin{aligned} |u(x, y, \mu_1) - u(x, y, \mu_2)| &\leq C \mathbb{W}_2(\mu_1, \mu_2)(1 + |y|)^p \\ |\partial_y u(x, y, \mu_1) - \partial_y u(x, y, \mu_2)| &\leq C \mathbb{W}_2(\mu_1, \mu_2)(1 + |y|)^{p'} \\ |\partial_y^2 u(x, y, \mu_1) - \partial_y^2 u(x, y, \mu_2)| &\leq C \mathbb{W}_2(\mu_1, \mu_2)(1 + |y|)^{p''}, \end{aligned}$$

for some $C > 0$ and $p, p', p'' \in \mathbb{N}$.

The proof with μ_1, μ_2 replaced by x_1, x_2 follows in the same way, and thus the desired continuity in x, \mathbb{W}_2 is established.

To obtain continuity of and a rate of polynomial growth for $\partial_x u$, we differentiate the equation that u satisfies to get

$$\begin{aligned} \mathcal{L}_{x,\mu} D^{(0,1,0)} u(x, y, \mu) &= D^{(0,1,0)} B(x, y, \mu) - D^{(0,1,0)} f(x, y, \mu) \cdot \partial_y u(x, y, \mu) - D^{(0,1,0)} a(x, y, \mu) : \partial_y^2 u(x, y, \mu) \\ &= D^{(0,1,0)} B(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,1,0)} u(x, y, \mu) \end{aligned}$$

in the notation of Lemma A.2, where a priori the derivative of u in x is in the weak sense.

But by the centering condition on B , we have by letting $B = h$ in Lemma A.2, that $u = v$ in the statement of that same lemma. Thus, we have

$$\int_{\mathbb{R}^d} D^{(0,1,0)} B(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,1,0)} u(x, y, \mu) \pi(dy; x, \mu) = D^{(0,1,0)} \left[\int_{\mathbb{R}^d} B(x, y, \mu) \pi(dy; x, \mu) \right] = 0,$$

and the inhomogeneity of the elliptic PDE that $D^{(0,1,0)} u$ solves in fact obeys the centering condition, and since we already know that u is locally Hölder continuous with polynomial growth in y , Lemma A.1 applies. This establishes that $D^{(0,1,0)} u$ grows at most polynomially and is locally Hölder continuous in y uniformly in x, μ . Then continuity in x, μ follows as above, but with $D^{(0,1,0)} B(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,1,0)} u(x, y, \mu)$ in the place of B , and $D^{(0,1,0)} u$ in the place of u . The same process applies to $D^{(1,0,0)} u$. To establish its continuity in z , we first recall that for all $x, y, z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)$

$$\mathcal{L}_{x,\mu} D^{(1,0,0)} u(x, y, \mu)[z] = D^{(1,0,0)} B(x, y, \mu)[z] - \mathcal{L}_{x,\mu}^{(1,0,0)} [z] u(x, y, \mu)$$

so

$$\begin{aligned} \mathcal{L}_{x,\mu} \left[D^{(1,0,0)} u(x, y, \mu)[z_1] - D^{(1,0,0)} u(x, y, \mu)[z_2] \right] &= D^{(1,0,0)} B(x, y, \mu)[z_1] - \mathcal{L}_{x,\mu}^{(1,0,0)}[z_1] u(x, y, \mu) \\ &\quad - \left[D^{(1,0,0)} B(x, y, \mu)[z_2] - \mathcal{L}_{x,\mu}^{(1,0,0)}[z_2] u(x, y, \mu) \right]. \end{aligned}$$

Then by the equality (35) from Lemma A.2, we have for all $x, z \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\int_{\mathbb{R}^d} D^{(1,0,0)} B(x, y, \mu)[z] - \mathcal{L}_{x,\mu}^{(1,0,0)}[z] u(x, y, \mu) \pi(dy; x, \mu) = D^{(0,1,0)} \int_{\mathbb{R}^d} B(x, y, \mu) \pi(dy; x, \mu)[z] = 0,$$

so the inhomogeneity in the Poisson equation above is centered. Thus, using the same argument as for the other continuity proofs and the assumed continuity of $D^{(1,0,0)} B, D^{(1,0,0)} f, D^{(1,0,0)} a$ in z with the fact that $\partial_y u, \partial_y^2 u$ grow at most polynomially in y , we get there is $p \in \mathbb{N}$ and $C > 0$ such that

$$\left| D^{(1,0,0)} u(x, y, \mu)[z_1] - D^{(1,0,0)} u(x, y, \mu)[z_2] \right| \leq C |z_1 - z_2| (1 + |y|)^p,$$

and similarly for $D^{(1,0,0)} \partial_y u$ and $D^{(1,0,0)} \partial_y^2 u$.

All of the remaining bounds work in the same way, with the inhomogeneity of the elliptic PDE of the desired derivative of u solves being the integrand of the expression for the corresponding derivative of $\bar{B}(x, y, \mu)$ from Equation (35) in Lemma A.2. Put explicitly:

(41)

$$\begin{aligned} &\mathcal{L}_{x,\mu} D^{(n,l,\beta)} u(x, y, \mu)[z_1, \dots, z_n] \\ &= D^{(n,l,\beta)} B(x, y, \mu)[z_1, \dots, z_n] - \sum_{k=0}^n \sum_{j=0}^l \sum_{\mathbf{p}_k} C_{(\mathbf{p}_k, j, n, l)} \mathcal{L}_{x,\mu}^{(k,j,\alpha(\mathbf{p}_k))} [z_{\mathbf{p}_k}] D^{(n-k, l-j, \alpha(\mathbf{p}'_{n-k}))} u(x, y, \mu)[z_{\mathbf{p}'_{n-k}}]. \end{aligned}$$

The inhomogeneity is always a jointly continuous (in the sense of Equation (14)) function which grows at most polynomially in y uniformly in x and μ , and only depends on lower order derivatives of u . Thus, it is clear the result follows by proceeding inductively on n, l . \square

Lemma A.6. Suppose that for some complete collection of multi-indices ζ that $h \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, $f \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, $a \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$. Then $\bar{h}(x, \mu) := \int_{\mathbb{R}^d} h(x, y, \mu) \pi(dy; x, \mu) \in \mathcal{M}_{b,L}^\zeta(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$.

Proof. This follows via Lemmas A.2 and A.1 in a similar way to Lemma A.5.

Boundedness and continuity of the derivatives when the coefficients are in $\mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$, $k \in \{1, d, d \times d\}$ follows in the same way as Lipschitz continuity. Thus, we only show the latter. For the Lipschitz property when $(n, l, \beta) = (0, 0, 0)$, we have via Equations (32) and (33), using the polynomial growth in y of $v, \partial_y v$, and $\partial_y^2 v$ from Equation (34) gained by Lemma A.1 with $B = h - \bar{h}$ and the assumed Lipschitz continuity of f, a, h :

$$\begin{aligned} |\bar{h}(x, \mu_1) - \bar{h}(x, \mu_2)| &\leq \int_{\mathbb{R}^d} C \mathbb{W}_2(\mu_1, \mu_2) (1 + |y|)^p \pi(dy; x, \mu_1) \text{ for some } p \in \mathbb{N} \\ &\leq C \mathbb{W}_2(\mu_1, \mu_2) \end{aligned}$$

and

$$\begin{aligned} |\bar{h}(x_1, \mu) - \bar{h}(x_2, \mu)| &\leq \int_{\mathbb{R}^d} C |x_1 - x_2| (1 + |y|)^p \pi(dy; x_1, \mu) \text{ for some } p \in \mathbb{N} \\ &\leq C |x_1 - x_2|, \end{aligned}$$

for all $x_1, x_2, x \in \mathbb{R}^d, \mu_1, \mu_2, \mu \in \mathcal{P}_2(\mathbb{R}^d)$, so the result follows. Now, as with the previous two results, the rest follows via an induction argument on n, l . For the case $(n, l, \beta) = (0, 1, 0)$, via Equation A.2, we have

$$|D^{(0,1,0)} \bar{h}(x, \mu)| = \left| \int_{\mathbb{R}^d} D^{(0,1,0)} h(x, y, \mu) - \mathcal{L}_{x,\mu}^{(0,1,0)} v(x, y, \mu) \pi(dy; x, \mu) dy \right|$$

$$\begin{aligned} &\leq \int_{\mathbb{R}^d} C(1+|y|)^p \pi(dy; x, \mu) \text{ for some } p \in \mathbb{N} \\ &\leq C \end{aligned}$$

from the fact that $\partial_x h, \partial_x f, \partial_x a, \partial_y v, \partial_y^2 v$ grow at most polynomially in y uniformly in x, μ by assumption and Lemma A.1.

In addition, from Equation (41) with $B = h - \bar{h}$, we know

$$\begin{aligned} \mathcal{L}_{x,\mu} D^{(0,1,0)} v &= D^{(0,1,0)} h - D^{(0,1,0)} \bar{h} - \mathcal{L}_{x,\mu}^{(0,1,0)} v(x, y, \mu) \\ &= D^{(0,1,0)} h - \mathcal{L}_{x,\mu}^{(0,1,0)} v(x, y, \mu) - \int_{\mathbb{R}^d} D^{(0,1,0)} h - \mathcal{L}_{x,\mu}^{(0,1,0)} v(x, y, \mu) \pi(dy; x, \mu) \end{aligned}$$

the right side of which we now know grows at most polynomially in y uniformly in x, μ . Then

$$\begin{aligned} D^{(0,1,0)} \bar{h}(x, \mu_1) - D^{(0,1,0)} \bar{h}(x, \mu_2) &= \int_{\mathbb{R}^d} D^{(0,1,0)} h(x, y, \mu_1) - \mathcal{L}_{x,\mu_1}^{(0,1,0)} v(x, y, \mu_1) \pi(y; x, \mu_1) dy \\ &\quad - \int_{\mathbb{R}^d} D^{(0,1,0)} h(x, y, \mu_2) - \mathcal{L}_{x,\mu_2}^{(0,1,0)} v(x, y, \mu_2) \pi(y; x, \mu_2) dy \end{aligned}$$

so using Equation (33) with $h(x, y, \mu) = D^{(0,1,0)} h(x, y, \mu) - D^{(0,1,0)} \mathcal{L}_{x,\mu} v(x, y, \mu)$, we get

$$\begin{aligned} D^{(0,1,0)} \bar{h}(x, \mu_1) - D^{(0,1,0)} \bar{h}(x, \mu_2) &= \int_{\mathbb{R}^d} D^{(0,1,0)} h(x, y, \mu_1) - \mathcal{L}_{x,\mu_1}^{(0,1,0)} v(x, y, \mu_1) - [D^{(0,1,0)} h(x, y, \mu_2) - \mathcal{L}_{x,\mu_2}^{(0,1,0)} v(x, y, \mu_2)] \\ &\quad - [\mathcal{L}_{x,\mu_1} - \mathcal{L}_{x,\mu_2}] D^{(0,1,0)} v(x, y, \mu_2) \pi(dy; x, \mu_1), \end{aligned}$$

so

$$\begin{aligned} |D^{(0,1,0)} \bar{h}(x, \mu_1) - D^{(0,1,0)} \bar{h}(x, \mu_2)| &\leq \int_{\mathbb{R}^d} C(1+|y|)^p \mathbb{W}_2(\mu_1, \mu_2) \pi(dy; x, \mu_1) \text{ for some } p \in \mathbb{N} \\ &\leq C \mathbb{W}_2(\mu_1, \mu_2), \end{aligned}$$

where here we used from Lemma A.5 that $D^{(0,1,0)} \partial_y v$ and $D^{(0,1,0)} \partial_y^2 v$ grow at most polynomially in y uniformly in x, μ . In the same way, we can get

$$|D^{(0,1,0)} \bar{h}(x_1, \mu) - D^{(0,1,0)} \bar{h}(x_2, \mu)| \leq C|x_1 - x_2|.$$

The proof for $(k, l, \mu) = (1, 0, 0)$ being bounded Lipschitz in x, μ follows in essentially the same way. To see that $D^{(1,0,0)} h(x, y, \mu)[z]$ is Lipschitz in z , we use the representation

$$D^{(1,0,0)} h(x, y, \mu)[z] = \int_{\mathbb{R}^d} \partial_\mu h(x, y, \mu)[z] - \mathcal{L}_{x,\mu}^{(1,0,0)}[z] v(x, y, \mu) \pi(dy; x, \mu),$$

so

$$\begin{aligned} D^{(1,0,0)} h(x, y, \mu)[z_1] - D^{(1,0,0)} h(x, y, \mu)[z_2] &= \int_{\mathbb{R}^d} \partial_\mu h(x, y, \mu)[z_1] - \mathcal{L}_{x,\mu}^{(1,0,0)}[z_1] v(x, y, \mu) \\ &\quad - [\partial_\mu h(x, y, \mu)[z_2] - \mathcal{L}_{x,\mu}^{(1,0,0)}[z_2] v(x, y, \mu)] \pi(dy; x, \mu), \end{aligned}$$

and by the Lipschitz properties of $D^{(1,0,0)} h, D^{(1,0,0)} f, D^{(1,0,0)} a$ in z , we see

$$\begin{aligned} |D^{(1,0,0)} h(x, y, \mu)[z_1] - D^{(1,0,0)} h(x, y, \mu)[z_2]| &\leq \int_{\mathbb{R}^d} C(1+|y|)^p |z_1 - z_2| \pi(dy; x, \mu_1) \text{ for some } p \in \mathbb{N} \\ &\leq C|z_1 - z_2|. \end{aligned}$$

Once again, the result for higher derivatives follows from iterating on the above method. \square

Remark A.7. Although Lemmas A.1, A.2.A.5, and A.6 were stated for simplicity in terms of $f, a, \mathcal{L}_{x,\mu}$, and π , the only assumptions needed other than those posed in the statement of each Lemma are those on f and a from Assumptions A1) and A2). Thus, if we take care to change the domains of the functions in the statements of these Lemmas, we can also apply them to gain regularity of the “doubled” Poisson Equations (27) and (28). Put explicitly:

Consider some $\tilde{f} : \mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^l) \rightarrow \mathbb{R}^j$ such that there exists constants $C', \beta' > 0$ independent of $x, y \in \mathbb{R}^j$ and $\mu \in \mathcal{P}_2(\mathbb{R}^l)$ such that

$$(42) \quad \tilde{f}(x, y, \mu) \cdot y \leq -\beta'|y|^2 + C', \forall x, y \in \mathbb{R}^j, \mu \in \mathcal{P}_2(\mathbb{R}^l),$$

\tilde{f} grows at most linearly in $|y|$, \tilde{f} has two uniformly bounded derivatives in y , and \tilde{f} and both these derivatives are Hölder continuous in y uniformly in (x, μ) and $\tilde{a} : \mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}(\mathbb{R}^l) \rightarrow \mathbb{R}^{j \times j}$ such that there exists $\lambda'_-, \lambda'_+ > 0$ such that $0 < \lambda'_- \leq 2 \frac{z^\top \tilde{a}(x, y, \mu) z}{|z|^2} \leq \lambda'_+ < \infty, \forall x, y, z \in \mathbb{R}^j, z \neq 0, \mu \in \mathcal{P}_2(\mathbb{R}^l)$ and \tilde{a} is bounded, has two uniformly bounded derivatives in y , and \tilde{a} and both these derivatives are Hölder continuous in y uniformly in (x, μ) .

We can then conclude that Lemmas A.1, A.2.A.5, and A.6 hold replacing $\mathcal{L}_{x, \mu}$ by $\tilde{\mathcal{L}}_{x, \mu}$ which acts on $\phi \in C_b^2(\mathbb{R}^j)$ by

$$(43) \quad \tilde{\mathcal{L}}_{x, \mu} \phi(y) = \tilde{f}(x, y, \mu) \cdot \nabla \phi(y) + \tilde{a}(x, y, \mu) : \nabla^2 \phi(y),$$

π by $\tilde{\pi}$ the unique (by [47] Proposition 1) probability measure satisfying $\tilde{\mathcal{L}}_{x, \mu}^* \tilde{\pi} = 0$, f by \tilde{f} , a by \tilde{a} , \mathbb{R}^d by \mathbb{R}^j , $\mathbb{R}^{d \times d}$ by $\mathbb{R}^{j \times j}$, and $\mathcal{P}_2(\mathbb{R}^d)$ by $\mathcal{P}_2(\mathbb{R}^l)$ in their statements.

In particular, considering $\tilde{f} : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{2d}$ given by $\tilde{f}(x, y, \mu) = [f(x_1, y_1, \mu), f(x_2, y_2, \mu)]^\top$ and $\tilde{a} : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{2d \times 2d}$ given by $\tilde{a}(x, y, \mu) = \text{diag}[a(x_1, y_1, \mu), a(x_2, y_2, \mu)]$ where $x \in \mathbb{R}^{2d} = (x_1, x_2), x_1, x_2 \in \mathbb{R}^d$ and $y \in \mathbb{R}^{2d} = (y_1, y_2), y_1, y_2 \in \mathbb{R}^d$, we have $\tilde{\mathcal{L}}_{x, \mu}$ in the above discussion is equal to $\mathcal{L}_{x_1, x_2, \mu}^2$ from Equation (24), and $\tilde{\pi}(dy; x, \mu)$ is equal to $\tilde{\pi}(dy_1, dy_2; x_1, x_2, \mu)$ from Equation (26).

Moreover, under Assumption A2)

$$\begin{aligned} \tilde{f}(x, y, \mu) \cdot y &= f(x_1, y_1, \mu) \cdot y_1 + f(x_2, y_2, \mu) \cdot y_2 \\ &\leq -\beta[|y_1|^2 + |y_2|^2] + 2C \\ &= -\beta|y|^2 + 2C, \forall x, y \in \mathbb{R}^{2d}, \mu \in \mathcal{P}_2(\mathbb{R}^d) \end{aligned}$$

and under assumption A1), writing $z \in \mathbb{R}^{2d} \setminus \{0\}$ as $(z_1, z_2) \in \mathbb{R}^d$ and taking any $x, y \in \mathbb{R}^{2d}$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ as before, we know

$$0 < \lambda_- |z_i|^2 \leq 2z_i^\top a(x_i, y_i, \mu) z_i \leq \lambda_+ |z_i|^2, i = 1, 2$$

and

$$z^\top \tilde{a}(x, y, \mu) z = z_1^\top a(x_1, y_1, \mu) z_1 + z_2^\top a(x_2, y_2, \mu) z_2$$

so

$$0 < \lambda_- |z|^2 = \lambda_- [|z_1|^2 + |z_2|^2] \leq 2z^\top \tilde{a}(x, y, \mu) z \leq \lambda_+ [|z_1|^2 + |z_2|^2] = \lambda_+ |z|^2.$$

Lastly, the desired regularity and growth properties of \tilde{a} and \tilde{f} are clearly inherited from those of a and f . So indeed we can apply Lemmas A.1 and A.5 to gain regularity of the Poisson Equations (27) and (28) by keeping these minor changes to the domains of the functions in the statements of the Lemmas in mind.

Proposition A.8. *Under Assumptions A1)-A5),*

$$\begin{aligned} \Phi, \partial_{y_i} \Phi, \partial_{y_i} \partial_{y_j} \Phi &\in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d) \\ \tilde{\chi}, \partial_{y_i} \tilde{\chi}, \partial_{\bar{y}_i} \tilde{\chi}, \partial_{y_i} \partial_{y_j} \tilde{\chi}, \partial_{\bar{y}_i} \partial_{y_j} \tilde{\chi}, \partial_{\bar{y}_i} \partial_{\bar{y}_j} \tilde{\chi} &\in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d}) \\ \chi, \partial_{y_i} \chi, \partial_{\bar{y}_i} \chi, \partial_{y_i} \partial_{y_j} \chi, \partial_{\bar{y}_i} \partial_{y_j} \chi, \partial_{\bar{y}_i} \partial_{\bar{y}_j} \chi &\in \mathcal{M}_p^{\hat{\zeta}_2}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d) \end{aligned}$$

and

$$\Xi, \partial_{y_i} \Xi, \partial_{y_i} \partial_{y_j} \Xi \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$$

for $F = \gamma$,

$$\Xi, \partial_{y_i} \Xi, \partial_{y_i} \partial_{y_j} \Xi \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$$

for $F = D$, for all $i, j = 1, \dots, d$, where $\Phi, \chi, \tilde{\chi}$, and Ξ are the the unique classical solutions to the PDEs (5), (27), (28), and (29) respectively, and F is denoting the function which enters the right-hand side of

Equation (29). Moreover, $\bar{\gamma} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $\bar{D} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, where both these coefficients are defined in Equation (6). Here $\hat{\zeta}, \hat{\zeta}_1$ are as in Equation (15), and

$$\hat{\zeta}_2 \ni \{(0, j_1, 0), (1, j_2, j_3), (2, j_4, (j_5, 0)) : j_1 \in \{0, 1, \dots, 4\}, j_2 + j_3 \leq 2, j_4 + j_5 \leq 1\}.$$

Additionally assuming A6), we get further that $\bar{D}^{1/2} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$.

Proof. The fact that $\Phi, \partial_{y_i} \Phi, \partial_{y_i} \partial_{y_j} \Phi \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ is immediate from Lemma A.5 applied to each $\Phi_l, l = 1, \dots, d$.

To see $\tilde{\chi}, \partial_{y_i} \tilde{\chi}, \partial_{\bar{y}_i} \tilde{\chi}, \partial_{y_i} \partial_{y_j} \tilde{\chi}, \partial_{\bar{y}_i} \partial_{y_j} \tilde{\chi}, \partial_{\bar{y}_i} \partial_{\bar{y}_j} \tilde{\chi} \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, we note that, as per Remark A.7, Lemmas A.1 and A.5 hold with $\mathcal{L}_{x,\mu}^2$ in the place of $\mathcal{L}_{x,\mu}$ and $\bar{\pi}$ from Equation (26) in the place of π . By product rule and triangle inequality that $b, \Phi \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ implies $G : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times d}$ defined by $G(\tilde{x}, \tilde{y}, \mu) = \Phi(\bar{x}, \bar{y}, \mu) \otimes b(x, y, \mu)$ where $x, \bar{x}, y, \bar{y} \in \mathbb{R}^d$ are such that $\tilde{x} = (x, \bar{x}), \tilde{y} = (y, \bar{y})$ satisfies $G \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, and under Assumption A5), $\tilde{f} \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{2d})$, and $\tilde{a} \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{2d \times 2d})$ (using the notation of Remark A.7). As noted after (27), we also have G integrates against $\bar{\pi}$ to 0, so we get the result holds via Lemma A.5 applied to each $\tilde{\chi}_{k,l}$ for $k, l = 1, \dots, d$.

The fact that $\chi, \partial_{y_i} \chi, \partial_{\bar{y}_i} \chi, \partial_{y_i} \partial_{y_j} \chi, \partial_{\bar{y}_i} \partial_{y_j} \chi, \partial_{\bar{y}_i} \partial_{\bar{y}_j} \chi \in \mathcal{M}_p^{\hat{\zeta}_2}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ holds similarly. Here we see that, again by product rule and triangle inequality, $b, \Phi \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ implies $G : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ defined by $G(\tilde{x}, \tilde{y}, \mu) = \partial_\mu \Phi(\bar{x}, \bar{y}, \mu)[x]b(x, y, \mu)$ where $x, \bar{x}, y, \bar{y} \in \mathbb{R}^d$ are such that $\tilde{x} = (x, \bar{x}), \tilde{y} = (y, \bar{y})$ satisfies $G \in \mathcal{M}_p^{\hat{\zeta}_2}(\mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, where here one must track how many derivatives in x, μ , and z of $\partial_\mu \Phi(x, y, \mu)[z]$ one gets by the fact that $\Phi \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$. This is what results in the smaller collection of multi-indices $\hat{\zeta}_2$. The result then follows again from Lemma A.5 applied to the coordinate functions of χ .

For $\Xi, \partial_{y_i} \Xi, \partial_{y_i} \partial_{y_j} \Xi \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ for $F = \gamma$ and $\Xi, \partial_{y_i} \Xi, \partial_{y_i} \partial_{y_j} \Xi \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$ for $F = D$, we first establish that $\gamma \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $D \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$. By assumption all the coefficients which appear in the definition of γ and D in Equation (6) are already assumed to be in $\mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ for the appropriate choices of k . Thus, to conclude that γ and D are in $\mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, we need only show that $\Phi, \partial_{y_i} \Phi, \partial_{x_i} \Phi$, and $\partial_{x_i} \partial_{y_j} \Phi$ are in $\mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ for $i, j = 1, \dots, d$. Since $\hat{\zeta} \subset \hat{\zeta}_1$, we already know this for Φ and $\partial_{y_i} \Phi$. For $\partial_{x_i} \Phi$ and $\partial_{x_i} \partial_{y_j} \Phi$, we note that $\hat{\zeta}_1$ is constructed so that $G \in \mathcal{M}_p^{\hat{\zeta}_1}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ implies $\partial_{x_i} G \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d), i = 1, \dots, d$. So indeed $\gamma \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $D \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$.

Now, we note that applying Lemma A.6 to each coordinate function of γ and D , this implies that $\bar{\gamma} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $\bar{D} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$. By triangle inequality, this implies $\gamma - \bar{\gamma} \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $D - \bar{D} \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$ so that Lemma A.5 yields $\Xi, \partial_{y_i} \Xi, \partial_{y_i} \partial_{y_j} \Xi \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ with $k = d$ for $F = \gamma$ and $k = d \times d$ for $F = D$.

Lastly, to gain $\bar{D}^{1/2} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$ from $\bar{D} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$ under Assumption A6), we use that mapping which takes a positive-definite matrix to its unique positive-definite square root is Fréchet differentiable up to arbitrary order, with all derivatives being bounded on sets of uniformly bounded, uniformly positive definite matrices (see Equation (6) in [18]). Thus, by chain rule and the fact that $\bar{D} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, we can see indeed that $\bar{D}^{1/2} \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$. See also [14] Lemma A.7 for how the growth of derivatives of $\bar{D}^{1/2}$ can be controlled in terms of $\bar{\lambda}_-$ and the derivatives of \bar{D} . \square

APPENDIX B. ON DIFFERENTIATION OF FUNCTIONS ON SPACES OF MEASURES

We will need the following two definitions from [8]:

Definition B.1. Given a function $u : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, we may define a lifting of u to $\tilde{u} : L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d) \rightarrow \mathbb{R}$ via $\tilde{u}(X) = u(\mathcal{L}(X))$ for $X \in L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$. Here we assume $\tilde{\Omega}$ is a Polish space, $\tilde{\mathcal{F}}$ its Borel σ -field, and $\tilde{\mathbb{P}}$ is an atomless probability measure (since $\tilde{\Omega}$ is Polish, this is equivalent to every singleton having zero measure).

Here, denoting by $\mu(| \cdot |^r) := \int_{\mathbb{R}^d} |x|^r \mu(dx)$ for $r > 0$,

$$\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d) : \mu(| \cdot |^2) = \int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty\}.$$

$\mathcal{P}_2(\mathbb{R}^d)$ is a Polish space under the L^2 -Wasserstein distance

$$\mathbb{W}_2(\mu_1, \mu_2) := \inf_{\pi \in \mathcal{C}_{\mu_1, \mu_2}} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dx, dy) \right]^{1/2},$$

where $\mathcal{C}_{\mu_1, \mu_2}$ denotes the set of all couplings of μ_1, μ_2 .

We say u is **L-differentiable** or **Lions-differentiable** at $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists a random variable X_0 on some $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ satisfying the above assumptions such that $\mathcal{L}(X_0) = \mu_0$ and \tilde{u} is Fréchet differentiable at X_0 .

The Fréchet derivative of \tilde{u} can be viewed as an element of $L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$ by identifying $L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$ and its dual. From this, one can find that if u is L-differentiable at $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, there is a deterministic measurable function $\xi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $D\tilde{u}(X_0) = \xi(X_0)$, and that ξ is uniquely defined μ_0 -almost everywhere on \mathbb{R}^d . We denote this equivalence class of $\xi \in L^2(\mathbb{R}^d, \mu_0; \mathbb{R}^d)$ by $\partial_\mu u(\mu_0)$ and call $\partial_\mu u(\mu_0)[\cdot] : \mathbb{R}^d \rightarrow \mathbb{R}^d$ the **Lions derivative** of u at μ_0 . Note that this definition is independent of the choice of X_0 and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. See [8] Section 5.2.

To avoid confusion when u depends on more variables than just μ , if $\partial_\mu u(\mu_0)$ is differentiable at $z_0 \in \mathbb{R}^d$, we denote its derivative at v_0 by $\partial_z \partial_\mu u(\mu_0)[z_0]$.

Definition B.2. ([8] Definition 5.83) We say $u : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is **Fully C²** if the following conditions are satisfied:

- (1) u is C^1 in the sense of L-differentiation, and its first derivative has a jointly continuous version $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (\mu, z) \mapsto \partial_\mu u(\mu)[z] \in \mathbb{R}^d$.
- (2) For each fixed $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, the version of $\mathbb{R}^d \ni z \mapsto \partial_\mu u(\mu)[z] \in \mathbb{R}^d$ from the first condition is differentiable on \mathbb{R}^d in the classical sense and its derivative is given by a jointly continuous function $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (\mu, z) \mapsto \partial_z \partial_\mu u(\mu)[z] \in \mathbb{R}^{d \times d}$.
- (3) For each fixed $z \in \mathbb{R}^d$, the version of $\mathcal{P}_2(\mathbb{R}^d) \ni \mu \mapsto \partial_\mu u(\mu)[z] \in \mathbb{R}^d$ in the first condition is continuously L-differentiable component-by-component, with a derivative given by a function $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \ni (\mu, z, \bar{z}) \mapsto \partial_\mu^2 u(\mu)[z][\bar{z}] \in \mathbb{R}^{d \times d}$ such that for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $X \in L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$ with $\mathcal{L}(X) = \mu$, $\partial_\mu^2 u(\mu)[z][X]$ gives the Fréchet derivative at X of $L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d) \ni X' \mapsto \partial_\mu u(\mathcal{L}(X'))[z]$ for every $z \in \mathbb{R}^d$. Denoting $\partial_\mu^2 u(\mu)[z][\bar{z}]$ by $\partial_\mu^2 u(\mu)[z, \bar{z}]$, the map $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \ni (\mu, z, \bar{z}) \mapsto \partial_\mu^2 u(\mu)[z, \bar{z}]$ is also assumed to be continuous in the product topology.

Remark B.3. Conditions 1) and 2) from Definition B.2 along with local boundedness of $\partial_\mu u$ and $\partial_z \partial_\mu u$ is sufficient to apply Itô's formula for measure-dependent functions as used in the proofs Section 4 and the proof of Theorem 3.1 - see Section 5.6.4 in [8].

Remark B.4. In this paper we will in fact also look at functions $u : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ which are required to have 3 Lions Derivatives. We will assume such functions are **Fully C²**, and satisfy:

- (4) For each fixed $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ the version of $\mathbb{R}^d \times \mathbb{R}^d \ni (z_1, z_2) \mapsto \partial_\mu^2 u(\mu)[z_1, z_2] \in \mathbb{R}^{d \times d}$ in the third condition is differentiable on \mathbb{R}^{2d} in the classical sense and its derivative is given by a jointly continuous function $\partial_z \partial_\mu^2 u(\mu)[z_1, z_2] = (\partial_{z_1} \partial_\mu^2 u(\mu)[z_1, z_2], \partial_{z_2} \partial_\mu^2 u(\mu)[z_1, z_2]) \in \mathbb{R}^{d \times d \times d} \times \mathbb{R}^{d \times d \times d}$ for $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \ni (\mu, z_1, z_2)$.
- (5) For each fixed $(z_1, z_2) \in \mathbb{R}^{2d}$, the version of $\mathcal{P}_2(\mathbb{R}^d) \ni \mu \mapsto \partial_\mu^2 u(\mu)[z_1, z_2] \in \mathbb{R}^{d \times d}$ in the third condition is continuously L-differentiable component-by-component, with a derivative given by a function $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \ni (\mu, z_1, z_2, z_3) \mapsto \partial_\mu^3 u(\mu)[z_1, z_2][z_3] \in \mathbb{R}^{d \times d \times d}$ such that for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $X \in L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d)$ with $\mathcal{L}(X) = \mu$, $\partial_\mu^3 u(\mu)[z_1, z_2][X]$ gives the Fréchet derivative at X of $L^2(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}; \mathbb{R}^d) \ni X' \mapsto \partial_\mu^2 u(\mathcal{L}(X'))[z_1, z_2]$ for every $(z_1, z_2) \in \mathbb{R}^{2d}$. Denoting $\partial_\mu^3 u(\mu)[z_1, z_2][z_3]$

by $\partial_\mu^2 u(\mu)[z_1, z_2, z_3]$, the map $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \ni (\mu, z_1, z_2, z_3) \mapsto \partial_\mu^3 u(\mu)[z_1, z_2, z_3]$ is also assumed to be continuous in the product topology.

Though we don't require higher than 3 Lions derivatives in this paper, when we state general results for higher Lions derivatives in terms of the spaces from Definition 2.4, we assume the analogous higher continuity.

REFERENCES

- [1] D. Benedetto, E. Caglioti, J. A. Carrillo, and M. Pulvirenti, *A Non-Maxwellian Steady Distribution for One-Dimensional Granular Media*, Journal of Statistical Physics **91** (1998), 979–990.
- [2] A. Bensoussan, J. L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, North Holland, Amsterdam, 1978.
- [3] Z. Bezemek and K. Spiliopoulos, *Large deviations for interacting multiscale particle systems* (2020), available at [arXiv:2011.03032 \[math.PR\]](#).
- [4] ———, *Moderate deviations for fully coupled multiscale weakly interacting particle systems* (2022), available at [arXiv:2202.08403 \[math.PR\]](#).
- [5] V. Borkar and V. Gaitsgory, *Averaging of singularly perturbed controlled stochastic differential equations*, Applied Mathematics and Optimization **56** (2007), no. 2, 169–209.
- [6] R. Buckdahn, J. Li, S. Peng, and C. Rainer, *Mean-field stochastic differential equations and associated PDEs*, Ann. Probab. **45** (2017), no. 2, 824–878.
- [7] P. Cardaliaguet, *Notes on mean field games (from P. L. Lions' lectures at Collège de France)*, 2013, <https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf>, unpublished.
- [8] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I*, Springer, NY, 2018.
- [9] S. Cerrai, *Second Order PDE's in Finite and Infinite Dimension: A Probabilistic Approach*, Springer, NY, 2001.
- [10] J.F. Chassagneux, D. Crisan, and F. Delarue, *A Probabilistic approach to classical solutions of the master equation for large population equilibria* (2015), available at [arXiv:1411.3009 \[math.PR\]](#).
- [11] J.F. Chassagneux, L. Szpruch, and A. Tse, *Weak quantitative propagation of chaos via differential calculus on the space of measures*, Ann. Appl. Probab. **32** (2022), no. 3, 1929–1969.
- [12] P. E. Chaudru de Raynal and N. Frikha, *From the backward Kolmogorov PDE on the Wasserstein space to propagation of chaos for McKean-Vlasov SDEs*, Journal de Mathématiques Pures et Appliquées **156** (2021), no. 2.
- [13] ———, *Well-posedness for some non-linear diffusion processes and related PDE on the Wasserstein space* (2018), available at [arXiv:1811.06904 \[math.CA\]](#).
- [14] P. Chen, J. Lu, and L. Xu, *Approximation to stochastic variance reduced gradient Langevin dynamics by stochastic delay differential equations*, Applied Mathematics & Optimization **85** (2022), no. 15.
- [15] D. Crisan and E. McMurray, *Smoothing properties of McKean-Vlasov SDEs*, Probability Theory and Related Fields **171** (2018), no. 2, 97–148.
- [16] G. Da prato, S. Kwapien, and J. Zabczyk, *Regularity of solutions of linear stochastic equations in hilbert spaces*, Stochastics **23** (1988), 1–23.
- [17] D. A. Dawson, *Critical dynamics and fluctuations for a mean-field model of cooperative behavior*, J. Stat. Phys. **31** (1983), 29–85.
- [18] P. Del Moral and A. Niclas, *A Taylor expansion of the square root matrix function*, Journal of Mathematical Analysis and Applications **465** (2018), no. 1, 259–266.
- [19] F. Delarue, D. Lacker, and K. Ramanan, *From the master equation to mean field game limit theory: a central limit theorem*, Electron. J. Probab. **24** (2019), 1–54.
- [20] M. G. Delgadino, R. S. Gvalani, and G. A. Pavliotis, *On the diffusive-mean field limit for weakly interacting diffusions exhibiting phase transitions*, Archive for Rational Mechanics and Analysis **241** (2021), 91–148.
- [21] P. Drábek and J. Milota, *Methods of Nonlinear Analysis: Applications to Differential Equations*, 2nd ed., Birkhäuser, Basel, 2015.
- [22] A. B. Duncan and G. A. Pavliotis, *Brownian Motion in an N-scale periodic Potential* (2016), available at [arXiv:1605.05854 \[math.ph\]](#).
- [23] S. D. Eidel'man, *Parabolic Systems*, Nauka, Moscow, 1965 (Russian); English transl. in *Parabolic Systems*, North-Holland Publishing, Amsterdam, 1969.
- [24] B. Fernandez and S. Méléard, *A Hilbertian approach for fluctuations on the McKean-Vlasov model*, Stochastic Processes and their Applications **71** (1997), 33–53.
- [25] A. Friedman, *Partial Differential Equations of Parabolic Type*, R. E. Krieger Publishing, Malabar, Florida, 1983.
- [26] A. Ganguly and P. Sundar, *Inhomogeneous functionals and approximations of invariant distributions of ergodic diffusions: Central limit theorem and moderate deviation asymptotics*, Stochastic Processes and their Applications **133** (2021), 74–110.
- [27] J. Garnier, G. Papanicolaou, and T. W. Yang, *Large deviations for a mean field model of systemic risk*, SIAM Journal of financial mathematics **4** (2013), no. 1, 151–184.
- [28] ———, *Consensus convergence with stochastic effects*, Vietnam Journal of mathematics **45** (2017), no. 1-2, 51–75.
- [29] W. R.P. Hammersley, D. Šiška, and L. Szpruch, *McKean-Vlasov SDEs under measure dependent Lyapunov conditions*, Ann. Inst. H. Poincaré Probab. Statist. **57** (2021), no. 2, 1032–1057.
- [30] M. Hitsuda and I. Mitoma, *Tightness problem and stochastic evolution equation arising from fluctuation phenomena for interacting diffusions*, Journal of Multivariate Analysis **19** (1986), no. 2, 311–328.

- [31] W. Hong, S. Li, and W. Liu, *Strong Convergence Rates in Averaging Principle for Slow-Fast McKean-Vlasov SPDEs* (2021), available at [arXiv:2107.14401 \[math.PR\]](#).
- [32] W. Hong, S. Li, and X. Sun, *Diffusion Approximation for Multi-Scale McKean-Vlasov SDEs Through Different Methods* (2022), available at [arXiv:2206.01928 \[math.PR\]](#).
- [33] W. Hong, S. Li, W. Liu, and X. Sun, *Central Limit Type Theorem and Large Deviations for Multi-Scale McKean-Vlasov SDEs* (2021), available at [arXiv:2112.08203 \[math.PR\]](#).
- [34] X. Huang and F.Y. Wang, *Derivative Estimates on Distributions of McKean-Vlasov SDEs* (2020), available at [arXiv:2006.16731 \[math.PR\]](#).
- [35] ———, *Distribution dependent SDEs with singular coefficients*, *Stochastic Processes and their Applications* **129** (2019), no. 11.
- [36] X. Huang, S. Wang, and F.F. Yang, *Weak Solution and Invariant Probability Measure for McKean-Vlasov SDEs with Integrable Drifts* (2021), available at [arXiv:2108.05802 \[math.PR\]](#).
- [37] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, NY, 1998.
- [38] T. Kolokolnikov, A. Bertozzi, R. Fetecau, and M. Lewis, *Emergent Behaviour in Multi-particle Systems with Non-local Interactions*, *Physica D* **260** (2013), 1-4.
- [39] M. Kong, Y. Shi, and X. Sun, *Well-posedness and averaging principle of McKean-Vlasov SPDEs driven by cylindrical α -stable process* (2021), available at [arXiv:2106.05561 \[math.PR\]](#).
- [40] T. G. Kurtz and J. Xiong, *A stochastic evolution equation arising from the fluctuations of a class of interacting particle systems*, *Communications in Mathematical Sciences* **2** (2004), no. 3, 325–358.
- [41] D. Lacker, *Limit theory for controlled McKean-Vlasov dynamics*, *SIAM Journal on Control and Optimization* **55** (2017), no. 3, 1641–1672.
- [42] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ural'ceva, *Linear and Quasi-linear Equations of Parabolic Type*, AMS, NY, 1968.
- [43] Y. Li, F. Wu, and L. Xie, *Poisson equation on Wasserstein space and diffusion approximations for McKean-Vlasov equation* (2022), available at [arXiv:2203.12796 \[math.PR\]](#).
- [44] S. Mehri and W. Stannat, *Weak solutions to Vlasov–McKean equations under Lyapunov-type conditions*, *Stochastics and Dynamics* **19** (2019), no. 9.
- [45] A. Muntean and F. Toschi (eds.), *Collective dynamics from bacteria to crowds: An excursion through modeling, analysis and simulation*, CISM International Centre for Mechanical Sciences. Courses and Lectures, vol. 533, Springer, Vienna, 2014.
- [46] M. R. Morse and K. Spiliopoulos, *Moderate deviations principle for systems of slow-fast diffusions*, *Asymptotic Analysis* **105** (2017), no. 3–4, 97–135.
- [47] E. Pardoux and A. Y. Veretennikov, *On Poisson equation and diffusion approximation I*, *Annals of Probability* **29** (2001), no. 3, 1061–1085.
- [48] ———, *On Poisson equation and diffusion approximation II*, *Annals of Probability* **31** (2003), no. 3, 1166–1192.
- [49] G. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Springer, NY, 2014.
- [50] G. Pavliotis and G. A. Stuart, *Multiscale Methods*, Springer, NY, 2008.
- [51] H. Qiao and W. Wei, *Efficient filtering for multiscale McKean-Vlasov Stochastic differential equations* (2022), available at [arXiv:2206.05037 \[math.PR\]](#).
- [52] P. Ren, *Singular McKean-Vlasov SDEs: Well-Posedness, Regularities and Wangs Harnack Inequality* (2021), available at [arXiv:2110.08846 \[math.PR\]](#).
- [53] M. Röckner and L. Xie, *Diffusion approximation for fully coupled stochastic differential equations*, *The Annals of Probability* **49** (2021), no. 3, 101–122.
- [54] M. Röckner, X. Sun, and Y. Xie, *Strong convergence order for slow-fast McKean-Vlasov stochastic differential equations*, *Ann. Inst. H. Poincaré Probab. Statist.* **57** (2021), no. 1, 547–576.
- [55] ———, *Strong and weak convergence in the averaging principle for SDEs with Hölder coefficients* (2019), available at [arXiv:1907.09256 \[math.PR\]](#).
- [56] M. Röckner, L. Xie, and L. Yang, *Asymptotic behavior of multiscale stochastic partial differential equations* (2020), available at [arXiv:2010.14897 \[math.PR\]](#).
- [57] K. Spiliopoulos, *Fluctuation analysis and short time asymptotics for multiple scales diffusion processes*, *Stochastics and Dynamics* **14** (2014), no. 3, 1350026.
- [58] K. Spiliopoulos, J. A. Sirignano, and K. Giesecke, *Fluctuation Analysis for the Loss From Default*, *Stochastic Processes and their Applications* **124** (2014), no. 7, 2322–2362.
- [59] A. Tse, *Higher order regularity of nonlinear Fokker-Planck PDEs with respect to the measure component*, *Journal de Mathématiques Pures et Appliquées* **150** (2021), 134–180.
- [60] A. Yu. Veretennikov, *Bounds for the mixing rates in the theory of stochastic equations*, *Theory Probab. Appl.* **32** (1987), 273–281.
- [61] F. Y. Wang, *Distribution dependent SDEs for Landau type equations*, *Stochastic Processes and their Applications* **128** (2017), no. 2, 595–621.
- [62] J. Xu, J. Liu, J. Liu, and Y. Miao, *Strong Averaging Principle for Two-Time-Scale Stochastic McKean-Vlasov Equations*, *Applied Mathematics and Optimization* **84** (2021), 837–867.
- [63] R. Zwanzig, *Diffusion in a rough potential*, *Proc. Natl. Acad. Sci.* **85** (1988), 2029–2030.

BOSTON UNIVERSITY, DEPARTMENT OF MATHEMATICS AND STATISTICS, 111 CUMMINGTON MALL, BOSTON, MA 02215, USA
Email address, Zachary William Bezemek: **bezemek@bu.edu**
Email address, Konstantinos Spiliopoulos: **kspiliop@bu.edu**