CONSERVATIVE SPDES AS FLUCTUATING MEAN FIELD LIMITS OF STOCHASTIC GRADIENT DESCENT

BENJAMIN GESS[†‡], RISHABH S. GVALANI[‡], AND VITALII KONAROVSKYI[†§]

ABSTRACT. The convergence of stochastic interacting particle systems in the mean-field limit to solutions to conservative stochastic partial differential equations is shown, with optimal rate of convergence. As a second main result, a quantitative central limit theorem for such SPDEs is derived, again with optimal rate of convergence.

The results apply in particular to the convergence in the mean-field scaling of stochastic gradient descent dynamics in overparametrized, shallow neural networks to solutions to SPDEs. It is shown that the inclusion of fluctuations in the limiting SPDE improves the rate of convergence, and retains information about the fluctuations of stochastic gradient descent in the continuum limit.

1. INTRODUCTION

The analysis of machine learning algorithms is confronted with algorithms in high dimension, with a large number of degrees of freedom (parameters), huge data-sets, and high computing capacities. This motivates the analysis of scaling limits, corresponding to the asymptotic regimes in which these parameters become large, leading to a variety of relative scaling regimes. A particularly relevant one is the so-called overparametrised regime, which corresponds to the case in which the number of parameters M is much larger than the (large) number of training samples (data) N. Indeed, a large class of real-world algorithms fall into this class, see, for example [10].

The success of such algorithms comes as a statistical surprise. Classical belief and estimates in statistics would suggest that vast overparametrization leads to overfitting [83], contradicting empirical evidence in machine learning. A systematic explanation of this observation constitutes a key challenge in the scientific understanding of machine learning. For recent progress concerning the related "double-descent" phenomenon of the error observed in machine learning we refer to [8, 40, 65, 74]. A central standing conjecture is that the (stochastic) learning algorithm employed in empirical risk minimization introduces an "implicit bias" towards minimizers that generalize well, thereby avoiding those that would lead to overfitting.

In order to prove or disprove this conjecture, universal models for machine learning are needed as the basis for the analysis of the stochastic dynamics of (stochastic) learning algorithms and their implicit bias. Motivated by this, several scaling limits of stochastic gradient descent dynamics have been analysed in the literature. In particular, the overparametrised regime, with its "mean-field" [15, 45, 78] and "lazy training" [14, 44, 84] scalings has achieved significant attention in recent years. In these works, the joint scaling limit of small learning rate $\alpha \to 0$ and overparametrization

$M \to \infty$ is considered. Roughly speaking, it is shown, for example in [66], that the empirical distribution $\nu^{M,\alpha}$ of the network parameters following stochastic gradient descent converges to the solution μ^0 of a (deterministic) gradient flow in the sense that

$$d(\nu^{M,\alpha}, \mu^0) = O(M^{-\frac{1}{2}}) + O(\alpha^{\frac{1}{2}}) \quad \text{for } \alpha \to 0, \, M \to \infty.$$

This corresponds to a law of large numbers result, since it proves the concentration of the random measures $\nu^{M,\alpha}$ onto a deterministic path. The mean behavior of stochastic gradient descent can then be analysed by considering the limiting dynamics μ^0. However, after passing to μ^0 all of the information about the inherent fluctuations of stochastic gradient descent is lost. Since there is substantial empirical evidence that stochasticity is decisive for the implicit bias of stochastic gradient descent [2, 49, 82], universal limiting models incorporating these fluctuations are needed. In this work, we rigorously identify a class of nonlinear conservative SPDEs which serve as such a fluctuating continuum model.

All previous known results rely on considering the *joint* scaling limit $M \to \infty$, $\alpha \to 0$. In contrast, in practice, the sizes of networks M are typically large, while the learning rate α is moderately small (e.g. [81]). This corresponds to the scaling limit $M \to \infty$ with $\alpha > 0$ small but fixed. The identification of a scaling limit in this regime is demanding, since it informally corresponds to the solutions of a nonlinear SPDE with challenging well-posedness properties. The development of a corresponding well-posedness framework and the rigorous treatment of the scaling regime $M \to \infty$, $\alpha > 0$ are two of the main contributions of the present work.

We now give a more precise account of the setup and results of the present work. Supervised learning starts from a given training set of data[1] $\mathcal{D} \subseteq \mathbb{R}^{n_0} \times \mathbb{R}^{k_0}$ with inputs $\Theta = \{\theta : (\theta, f(\theta)) \in \mathcal{D}\}$ and labels $\{f(\theta) : (\theta, f(\theta)) \in \mathcal{D}\}$. One then chooses a space of hypotheses. Here, we consider a fully-connected feed-forward network with one hidden layer

$$(1.1) \qquad f^M(x, \theta) = \frac{1}{M} \sum_{i=1}^{M} c_i \phi(U_i \theta + b_i) = \frac{1}{M} \sum_{i=1}^{M} \Phi(x_i, \theta)$$

with weights/parameters $x_i := (c_i, U_i, b_i) \in \mathbb{R}^{k_0} \times \mathbb{R}^{d_0 \times n_0} \times \mathbb{R}^{d_0} =: \mathbb{R}^d$, $\Phi(x_i, \theta) = c_i \phi(U_i \theta + b_i)$, and ϕ a nonlinear activation function. In practice, ϕ is often chosen as the rectified linear unit (ReLU). In fact, the results of this paper apply to more general choices of Φ, and we only restrict for simplicity to the specific choice in this introduction. Thereby, we obtain a parametrization of the space of hypotheses $\mathcal{M} = \{f^M(x, \theta) : x \in \mathbb{R}^{Md}\}$. The aim of risk minimization then is to select a suitable model $f^M(x, \cdot) \in \mathcal{M}$ minimizing the risk

$$(1.2) \qquad L(x) = \int_{\mathcal{D}} |f(\theta) - f^M(x, \theta)|^2 d\vartheta(\theta),$$

where we concentrate on square loss $l(x, y) = \frac{1}{2}|x - y|^2$, and ϑ is some measure on the data set $\mathcal{D}$. For example, $\mathcal{D}$ finite with ϑ the uniform distribution corresponds to empirical risk minimization. This leads to the optimization problem $\mathrm{Err} :=^! \min_x L(x)$. In machine learning, this optimization is approximated by variants of the stochastic gradient descent algorithm [73], corresponding to a random choice of the direction of descent. More precisely, the optimization dynamics are specified via

$$(1.3) \qquad x(n+1) = x(n) - \frac{\alpha}{P} \sum_{p=1}^{P} \nabla_x l(f(x(n), \theta_p), f(\theta_p)),$$

---

[1]For simplicity we assume that the ground-truth is given by a function f.

where θ_p are i.i.d. samples drawn from ϑ, α is the learning rate, P is the mini-batch size, and $x(0)$ is initialized i.i.d from a measure μ_0. Stochastic gradient descent corresponds to $P = 1$ and gradient descent to $P = N$. Note that, since $\mathbb{E}_\vartheta \nabla_x l(f(\cdot), f^M(x, \cdot)) = \nabla_x L(x)$, this is an unbiased estimator of gradient descent. The convergence of these (stochastic) optimization algorithms depends crucially on the properties of the empirical risk L. In machine learning, this risk landscape is typically non-convex, non-smooth, and degenerate, making the rigorous analysis of the convergence of stochastic gradient descent challenging.

In a series of works, see, e.g. [30, 61], it has been shown that for small learning rate $\alpha > 0$, the time discrete dynamics (1.3) can be approximated up to first order by the following SDE

$$(1.4) \qquad dX_t^{M,\alpha} = U(X_t^{M,\alpha}, \mu_t^{M,\alpha})dt + \sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}(X_t^{M,\alpha})dB(t),$$

where $\mu_t^{M,\alpha} := \frac{1}{M} \sum_i \delta_{(X_t^{M,\alpha})_i}$ is the empirical measure of the above system, B is a Brownian motion in $\mathbb{R}^{Md}$, $U(x, \mu) = (V(x_i, \mu))_{i \in [M]}$, $\Sigma(x, \mu) = (\tilde{A}(x_i, x_j, \mu))_{i,j \in [M]}$ for $x = (x_i)_{i \in [M]} \in \mathbb{R}^{Md}$, and

$$V(x_i, \mu) = \nabla F(x_i) - \langle \nabla_{x_i} K(x_i, \cdot), \mu \rangle,$$

$$(1.5) \qquad \begin{aligned} G(x_i, \mu, \theta) &= \left(f(\theta) - \int \Phi(y, \theta) \mu(dy) \right) \nabla_{x_i} \Phi(x_i, \theta) \\ &\quad - \mathbb{E}_\vartheta \left[\left(f(\theta) - \int \Phi(y, \theta) \mu(dy) \right) \nabla_{x_i} \Phi(x_i, \theta) \right], \end{aligned}$$

$$\tilde{A}(x_i, x_j, \mu) = \mathbb{E}_\vartheta \left[G(x_i, \mu, \theta) \otimes G(x_j, \mu, \theta) \right]$$

and $\sigma = \frac{\alpha}{P}$ is the fluctuation intensity. Notably, compared to plain gradient descent, (1.5) retains information on the fluctuations in (1.3), and offers a higher order of approximation.

In the case of shallow networks (1.1) and square loss (1.2) we observe, following [12, 74, 76], that we can represent the square loss as follows

$$L(x) = C_f - \frac{1}{M} \sum_{i=1}^{M} F(x_i) + \frac{1}{2M^2} \sum_{i,j=1}^{M} K(x_i, x_j)$$

where $C_f = \mathbb{E}_\vartheta |f(\theta)|^2$, $F(x_i) = \mathbb{E}_\vartheta [f(\theta)\Phi(x_i, \theta)]$, $K(x_i, x_j) = \mathbb{E}_\vartheta [\Phi(x_i, \theta)\Phi(x_j, \theta)]$, $i, j \in [M]$. This representation of the loss reveals intricate relations to statistical physics by interpreting the parameters as particles interacting via the interaction potential given by the risk (see [76]): The empirical distribution $\mu^{M,\alpha}$ of the parameter dynamics (1.4) can be identified as a solution to the martingale problem

$$(1.6) \quad d\mu_t^{M,\alpha} = -\nabla \cdot (V(\cdot, \mu_t^{M,\alpha})\mu_t^{M,\alpha})dt + \frac{\sigma}{2}D^2 : (A(\cdot, \mu_t^{M,\alpha})\mu_t^{M,\alpha})dt + \sigma^{\frac{1}{2}}\nabla \cdot (d\mathcal{M}_t^{M,\alpha})$$

with initial datum $\mu_0^M = \frac{1}{M} \sum_{i=1}^M \delta_{(X_0^{M,\alpha})_i}$ and where $\mathcal{M}^{M,\alpha}$ is a continuous martingale satisfying

$$[\langle \psi, \mathcal{M}_\cdot^{M,\alpha} \rangle]_t = \int_0^t \int \int \psi(x) \otimes \psi(y) : \tilde{A}(x, y, \mu_s^{M,\alpha})\mu_s^{M,\alpha}(dx)\mu_s^{M,\alpha}(dy)ds$$

and $A(x, \mu) = \tilde{A}(x, x, \mu)$. Based on this, in several works [12, 74, 76] it has been informally suggested that in the overparametrised limit ($M \to \infty$) the dynamics will converge to solutions of the martingale problem

$$(1.7) \qquad d\mu_t^\alpha = -\nabla \cdot (V(\cdot, \mu_t^\alpha)\mu_t^\alpha)dt + \frac{\sigma}{2}D^2 : (A(\cdot, \mu_t^\alpha)\mu_t^\alpha)dt + \sigma^{\frac{1}{2}}\nabla \cdot (d\mathcal{M}_t^\alpha),$$

where $\mathcal{M}$ is a continuous martingale satisfying

$$(1.8) \qquad [\langle \psi, \mathcal{M}_\cdot^\alpha \rangle]_t = \int_0^t \int \int \psi(x) \otimes \psi(y) : \tilde{A}(x, y, \mu_s^\alpha) \mu_s^\alpha(dx) \mu_s^\alpha(dy) ds.$$

The proof of this conjecture is one of the main results of this work. Precisely, we prove that in the $M \to \infty$ scaling limit, the empirical measure $\mu^{M,\alpha}$ converges to a solution μ^α of the conservative SPDE (1.9), with the optimal rate of convergence

$$d(\mu^{M,\alpha}, \mu^\alpha) = O(M^{-1/2}) \quad \text{for } M \to \infty.$$

The rigorous proof relies on the development of a well-posedness framework for (1.7), that is, to an infinite dimensional martingale problem with nonlocal coefficients and degenerate ellipticity. We approach this by analyzing instead the probabilistically strong well-posedness of SPDEs that have (1.7) as their martingale problem. However, the naive guess for such an SPDE, corresponding to informally taking the square-root of the coefficients in the quadratic variation (1.8), leads to an SPDE with irregular diffusion coefficients, for which the validity of strong uniqueness appears unclear. Instead, as one of the first main ideas of this work, we introduce an alternative "coupling" SPDE associated to (1.7), by

$$(1.9) \quad d\mu_t^\alpha = -\nabla \cdot (V(\cdot, \mu_t^\alpha)\mu_t^\alpha)dt + \frac{\sigma}{2}D^2 : (A(\cdot, \mu_t^\alpha)\mu_t^\alpha)dt + \sigma^{\frac{1}{2}}\nabla \cdot \int_\Theta G(\cdot, \mu_t, \theta)\, W(d\theta, dt),$$

which we will call the *Stochastic Mean-Field Equation*, where W is a cylindrical Wiener process in $L_2(\Theta, \vartheta)$. We then prove the (probabilistically strong) well-posedness of this SPDE. Since (1.9) is an SPDE with nonlocal coefficients, and degenerate coercivity, this is a challenging task. We next state a paraphrased version of our main results on the well-posedness of (1.9). We refer the reader to Theorems 2.16, 2.23, and 2.26 for the precise results under varying assumptions on the coefficients and the initial data.

Theorem 1.1 (See Theorems 2.16, 2.23, and 2.26)**.** *Given sufficiently nice initial data* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *and sufficiently nice coefficients* $V(\cdot), A(\cdot), G(\cdot, \cdot)$, *the stochastic mean-field equation* (1.9) *has a unique superposition solution in the sense of Definition 2.6.*

The proof of this result relies on establishing a superposition principle for the stochastic mean-field equation (1.9), i.e. the proof that each solution to (1.9) is given as a superposition of solutions to the SDE with interaction

$$(1.10) \qquad dX(u, t) = V(X(u, t), \mu_t)\, dt + \int_\Theta G(X(u, t), \mu_t, \theta)W(d\theta, dt),$$

where $\mu_t = \mu_0 \circ X(\cdot, t)^{-1}$. While the superposition principle for deterministic PDE has been well-established in a series of ground-breaking works [3, 5, 23, 34, 79], even in infinite dimensions [71], the few existing results for the case of SPDEs [16, 35] rely on restrictive assumptions on the regularity of the coefficients or initial data, which are not satisfied in the present case. Therefore, a new argument for the case of measure-valued, discrete initial data corresponding to (1.6) and Lipschitz continuous coefficients is developed in the present work, as well as a new proof for measure-valued initial data, relaxing the regularity assumptions on the coefficients from [16].

We then prove the uniqueness of solutions to this Lagrangian system, which by the superposition principle implies the uniqueness of solutions to (1.9). Based on the superposition principle, we next establish the convergence of the empirical measures $\mu^{M,\alpha}$ to the solution μ^α of the stochastic mean-field equation (1.9) with optimal rate $M^{-\frac{1}{2}}$.

$$\mu^{M,\alpha} \xrightarrow{\quad M \to \infty \quad} \text{SMFE } (1.9)_{\mu_0}$$

$$\alpha \to 0 \Big\downarrow \qquad\qquad\qquad \Big\downarrow \alpha \to 0$$

$$\text{TE } (1.11)_{\mu_0^M} \xrightarrow{\quad M \to \infty \quad} \text{TE } (1.11)_{\mu_0}$$

FIGURE 1. The limits of small learning rate and large parameter size for the empirical measure $\mu^{M,\alpha}$. Here, $(1.11)_\nu$ (resp. $(1.9)_\nu$) denotes a solution of (1.11) (resp. (1.9)) with initial datum ν.

Theorem 1.2 (See Theorem 2.14). *Let $\mu^{M,\alpha}$ be the superposition empirical measure associated to the SDE (1.4) started from independent samples of μ_0. Then, for fixed $\alpha > 0$ and as $M \to \infty$, we have the estimate*

$$\mathbb{E} \sup_{t \in [0,T]} \mathcal{W}_2^2(\mu_t^{M,\alpha}, \mu_t^\alpha) \lesssim \mathbb{E}\mathcal{W}_2^2(\mu_0^M, \mu_0) \lesssim M^{-1},$$

where μ^α is a superposition solution of the stochastic mean-field equation (1.9) with $\sigma = \frac{\alpha}{P}$ and initial datum μ_0, $\mathcal{W}_p$ denotes the p-Wasserstein distance, and the implicit constants are independent of α.

We next analyze the law of large numbers behavior of solutions μ^α to (1.9) in the limit of small learning rate $\alpha \to 0$, proving an optimal rate of convergence to the deterministic transport equation

$$(1.11) \qquad\qquad d\mu_t^0 = -\nabla \cdot \left(V(t, \cdot, \mu_t^0)\mu_t^0 \right) dt.$$

Theorem 1.3 (See Theorem 3.1). *If μ^α is a superposition solution of (1.9) with initial datum μ_0, then in the limit $\alpha \to 0$, we have the estimate*

$$(1.12) \qquad\qquad \mathbb{E} \sup_{t \in [0,T]} \mathcal{W}_2^2(\mu_t^\alpha, \mu_t^0) \lesssim \alpha,$$

where μ^0 is the solution of the transport equation (1.11) with initial datum μ_0 and the implicit constant is independent of $\alpha > 0$.

Combining Theorems 1.2 and 1.3, we conclude

$$(1.13) \qquad\qquad \mathbb{E} \sup_{t \in [0,T]} \mathcal{W}_2^2(\mu_t^{M,\alpha}, \mu_t^0) \lesssim \alpha + M^{-1},$$

which implies that the limits $M \to \infty$ and $\alpha \to 0$ of $\mu^{M,\alpha}$ commute. In addition, using the results of the present work the intermediate limits can be characterized: Indeed, since $\mu^{M,\alpha}$ is itself shown to be a superposition solution to the stochastic mean-field equation (1.9), (1.12) implies that taking the limit $\alpha \to 0$ for $M > 0$ fixed yields a solution μ^M to the transport equation (1.11) with initial datum μ_0^M. Subsequently, considering the limit $M \to \infty$ and applying (1.13) implies that μ^M converges to a solution to the transport equation (1.11) with initial datum μ_0. Taking the limits in the opposite order follows in an analogous manner, see Figure 1.

Having established the law of large numbers behavior of $\mu^{M,\alpha}$, we next turn to its asymptotic fluctuations and prove a quantified central limit theorem for the stochastic mean-field equation (1.9), again providing optimal bounds on the rate of convergence. As discussed in [76], there

are two sources of fluctuations, one due to the sampling from the initial measure μ_0 and one due to the dynamical fluctuations in SGD. We define the corresponding fluctuation field

$$\eta^{M,\alpha} := \min\{\alpha^{-1/2}, M^{1/2}\}(\mu^{M,\alpha} - \mu^0).$$

As is done in [76, Section 4.4], we will focus on the case where the dynamical fluctuations dominate, i.e. α decreases slowly than M^{-1}. Precisely, assuming that $M = M(\alpha)$ with $\lim_{\alpha\to0} \alpha^{-1}M^{-1}(\alpha) < \infty$, we prove that the fluctuation field $\eta^\alpha = \alpha^{-1/2}(\mu^{M(\alpha),\alpha} - \mu^0)$ for $\alpha \to 0$ converges to a solution of the linear SPDE

$$\begin{aligned}(1.14) \qquad d\eta_t = &-\nabla \cdot \left(V(\cdot, \mu_t^0)\eta_t + \langle \tilde{V}(x,\cdot), \eta_t \rangle_0 \mu_t^0(dx) \right) dt \\ &- P^{-\frac{1}{2}} \int_\Theta \nabla \cdot \left(G(\cdot, \mu_t^0, \theta)\mu_t^0 \right) W(d\theta, dt),\end{aligned}$$

and prove an optimal rate of convergence. This generalizes the result obtained in [76, Section 4.4], and proves that (1.9) correctly reproduces the central limit fluctuations of the stochastic gradient descent. The proof of this optimal rate of convergence relies on a careful estimation of the error terms, including a new stopping time argument.

Theorem 1.4 (See Theorem 3.7). *Given sufficiently nice initial data $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and sufficiently nice coefficients V, A, G, consider the fluctuation field η^α as defined earlier. Then, η^α converges to a weak solution η of (1.14) with initial datum $\eta_0 = a\tilde{\eta}_0$, where $\tilde{\eta}_0$ is a centred Gaussian random variable with covariance*

$$\mathbb{E}\langle \varphi, \tilde{\eta}_0 \rangle \langle \psi, \tilde{\eta}_0 \rangle = \int \left(\varphi(x) - \langle \varphi, \mu_0 \rangle \right) \left(\psi(x) - \langle \psi, \mu_0 \rangle \right) d\mu_0(x),$$

for any smooth φ, ψ and $a = \lim_{\alpha\to0} \alpha^{-\frac{1}{2}} M^{-\frac{1}{2}}(\alpha)$. Furthermore, η^α satisfies

$$\mathbb{E} \sup_{t\in[0,T]} \|\eta_t^\alpha - \eta_t\|_{H^{-J}}^2 \lesssim \alpha$$

for some appropriately chosen $J \in \mathbb{N}$.

Remark 1.5. For the case that the fluctuations arising from the sampling from the initial measure dominate, that is, if $M = M(\alpha)$ with $\lim_{\alpha\to0} \alpha^{-1}M^{-1}(\alpha) = +\infty$, then the same arguments as in the proof of Theorem 1.4 imply that $\eta^M = M^{1/2}(\mu^{M(\alpha),\alpha} - \mu^0)$ converges to a solution η to the linear PDE (1.14) with $G = 0$ started from the centered Gaussian random variable $\tilde{\eta}_0$ defined in Theorem 1.4 and yield the optimal rate of convergence

$$\mathbb{E} \sup_{t\in[0,T]} \|\eta_t^M - \eta_t\|_{H^{-J}}^2 \lesssim M^{-1}.$$

We next show that the central limit Theorem 1.4 in particular implies that the stochastic mean-field equation offers a higher order approximation of the SGD dynamics than the deterministic mean field equation. We define the empirical distribution of SGD by

$$(1.15) \qquad \nu_t^{M,\frac{1}{M}} = \frac{1}{M} \sum_{i=1}^M \delta_{x_i(\lfloor Mt \rfloor)}, \quad t \geq 0,$$

where $x(k) = (x_i(k))_{i\in[M]}$, $k \in \mathbb{N}_0$, is defined by (1.3) with $\alpha = \frac{1}{M}$ and $P = 1$. The central limit theorem obtained in [78] gives

$$\nu^{M,\frac{1}{M}} = \mu^0 + M^{-\frac{1}{2}}\eta + o(M^{-\frac{1}{2}}),$$

with η as in Theorem 1.4. Moreover, for $\mu^{\frac{1}{M}}$ we have by Theorem 1.4

$$\mu^{\frac{1}{M}} = \mu^0 + M^{-\frac{1}{2}}\eta + O(M^{-1}).$$

This indicates that the solutions $\mu^{\frac{1}{M}}$ to (1.9) provide a higher order approximation to the SGD dynamics $\nu^{M,\frac{1}{M}}$, in the sense that

$$\nu^{M,\frac{1}{M}} - \mu^{\frac{1}{M}} = o(M^{-\frac{1}{2}}),$$

which supersedes the order of approximation by the non-fluctuating limit $d(\nu^{M,\frac{1}{M}}, \mu^0) \approx M^{-1/2}$.

Theorem 1.6. *(See Theorem 4.1) Let $\mu^{\frac{1}{M}}$ be a superposition solution to the stochastic mean-field equation (1.9) with $\alpha = \frac{1}{M}$. Let also $\nu^{M,\frac{1}{M}}$, be the empirical process associated to the SGD (1.3), which is defined by (1.15). Then, for every $p \in [1, 2)$,*

$$\mathcal{W}_p(\mathrm{Law}(\mu^{\frac{1}{M}}), \mathrm{Law}(\nu^{M,\frac{1}{M}})) = o(M^{-\frac{1}{2}}).$$

1.1. **Overview of the literature.** To first order, in the small learning rate limit $\alpha \to 0$, stochastic gradient descent converges to deterministic gradient descent. As argued above, this law of large numbers scaling limit does not incorporate information on the fluctuations of stochastic gradient descent. However, considering higher order approximations, stochastic gradient descent can be shown to converge to solutions to so-called stochastic modified equations

$$(1.16) \qquad dX_t = -\nabla\big(L(X_t) + \tfrac{1}{4}\alpha|\nabla L(X_t)|^2\big)dt + (\alpha\Sigma(X_t))^{\frac{1}{2}}dW_t$$

with Σ given in terms of the variance of the stochastic sampling of the empirical loss, see [30, 60]. For the validity of this limit for moderately large learning rates see [62]. A discussion of (1.16) with jump noise can be found in [67].

The effect of the randomness inherent to stochastic gradient descent on the implicit bias and on implicit regularization has been analyzed in [2, 49, 82].

Overparameterised limits of shallow networks in the mean-field training regime have received considerable attention in recent years. In [15, 45, 66, 75, 78] the convergence of gradient descent to a Wasserstein gradient flow has been shown and analysed. Notably, this limit is different from the "lazy training" regime which can be treated in terms of a linearisation around initialization, see [29, 44]. An instructive comparison of the scaling regimes is given in [14] and of their performance in [13, 38, 70].

Linear SPDE have been rigorously identified in the context of central limit fluctuations in stochastic gradient descent in [75, 77]. A fluctuating, nonlinear mean-field limit, incorporating the fluctuations of stochastic gradient descent was, informally, suggested in [12, 74, 76], taking the form of the conservative SPDE (2.1) below. The rigorous derivation of this conservative SPDE as well as a proof of a quantified central limit theorem remained open problems in the literature. These are solved in the present work.

For an overview of the literature on conservative SPDEs we refer to [33, Section 1.1]. We here concentrate on nonlocal conservative SPDEs. In [20] nonlocal, nonlinear stochastic Fokker-Planck equations have been considered, proving the uniqueness of solutions by several methods, e.g. by duality arguments, coupling arguments, and the Krylov–Rozovskii variational framework. Under less restrictive assumptions on the coefficients and solutions, this has been extended in [58, 25, p. 115]. The case of measure-valued solutions has been treated in [16]. Additionally, motivated by applications to fluid dynamics, signed measure-valued solutions to nonlocal, nonlinear stochastic Fokker-Planck equations have been considered in the literature, see, for example, e.g. [6, 55, 56, 72] and the references therein.

The convergence of interacting particle systems to solutions of nonlocal, nonlinear Fokker-Planck equations, and the closely related phenomenon of propagation of chaos has been considered in [16], [54] and [59, Theorem 2.3] and the references therein. SDE with interaction have been analyzed in [9, 25, 26, 26, 28, 68], and their relation to SPDEs has been considered in [27, 54].

The SPDE considered in this work bears some similarity with the so-called Dean–Kawasaki equation introduced in [22, 48], and which corresponds to (1.6) for independent particles. In contrast to the Dean–Kawasaki equation, the noise caused by SGD is spatially correlated, which allows the development of a full mathematical treatment introduced in the present work. The more singular case of the Dean–Kawasaki equation has attracted considerable interest in the literature, yielding the construction of (renormalized) solutions [7, 50, 53, 80], negative results on the existence of non-trivial solutions [51, 52], and regularized models [18, 19].

Central limit theorems for conservative, local SPDEs have been analyzed in [24]. The case of linear transport noise has been analyzed in [36]. For central limit theorems for parabolic SPDEs with multiplicative, semilinear noise we refer to [11, 41, 42] and the references therein. Higher order approximations of interacting particle systems by conservative, local SPDEs have been shown in [24], and for non-interacting particle systems up to arbitrary order in [17]. The authors are not aware of any previous results on central limit theorems for nonlocal conservative SPDEs.

1.2. **Outline of the paper.**

In Section 2, well-posedness results for the SDE with interaction (1.10) is shown assuming Lipschitz continuity of its coefficients, and the existence of a superposition solution to the stochastic mean-field equation (1.9) is established. This allows us to connect the uniqueness with the superposition principle in Corollary 2.12. The well-posedness of (1.10) and the continuous dependence of its solutions with respect to the initial particle distribution is obtained in Section 2.1. Section 2.2 is devoted to the proof of the uniqueness for the stochastic mean-field equation (1.9). The limit Theorems 3.1 and 3.7 are proved in Section 3. In Section 4, the higher order approximation of the SGD dynamics by solutions to the stochastic mean-field equation is obtained.

1.3. **Basic notation.**

Let $d \in \mathbb{N}$ be fixed. For $m \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ the space of m-times continuously differentiable functions from an open set $\Gamma \subset \mathbb{R}^d$ to $\mathbb{R}$ is denoted by $\mathrm{C}^m(\Gamma)$. The subspace of $\mathrm{C}^m(\Gamma)$ of all bounded together with their derivatives (resp. compactly supported) functions is denoted by $\mathrm{C}_b^m(\Gamma)$ (resp. $\mathrm{C}_c^m(\Gamma)$). We write $\mathrm{C}(\Gamma)$, $\mathrm{C}_b(\Gamma)$ and $\mathrm{C}_c(\Gamma)$ for $\mathrm{C}^0(\Gamma)$, $\mathrm{C}_b^0(\Gamma)$ and $\mathrm{C}_c^0(\Gamma)$, respectively. Let $\varphi, f_i \in \mathrm{C}^m(\Gamma)$, $i \in [n] := \{1, \ldots, n\}$. We set $\partial_i \varphi = \frac{\partial}{\partial x_i} \varphi$ and $\partial_{i,j}^2 \varphi = \frac{\partial^2}{\partial x_i x_j} \varphi$. For $f = (f_i)_{i \in [n]}$, we write ∇f for the matrix with rows $(\partial_j f_i)_{j \in [d]}$, $i \in [n]$, and $\nabla \cdot f = \sum_{i=1}^d \partial_i f_i$ if $n = d$. We also set $D^2 \varphi = (\partial_{i,j} \varphi)_{i,j \in [d]}$. The supremum norm in $\mathrm{C}_b^m(\Gamma)$ will be denoted by $\| \cdot \|_{\mathrm{C}_b^m}$, that is,

$$\|f\|_{\mathrm{C}_b^m} = \sum_{|\alpha| \leq m} \sup_{x \in \Gamma} |D^\alpha f(x)|,$$

where $D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$ and $|\alpha| = \alpha_1 + \cdots + \alpha_d$ for $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$

For vectors $a, b \in \mathbb{R}^d$ and matrices $A, B \in \mathbb{R}^{d \times d}$ we will use the notation $a \cdot b = \sum_{i=1}^d a_i b_i$, $|a| = \sqrt{a \cdot a}$, $a \otimes b = (a_i b_j)_{i,j \in [d]}$, $A : B = \sum_{i,j=1}^d a_{i,j} b_{i,j}$, $A \cdot b = \left(\sum_{j=1}^d a_{i,j} b_j \right)_{i \in [d]}$ and $|A| = \sqrt{A : A}$. In particular, $(\nabla f) \cdot g = \left(\sum_{j=1}^d (\partial_j f_i) g_j \right)_{i \in [n]}$, where $f = (f_i)_{i \in [n]}$ and $g = (g_i)_{i \in [d]}$.

For $p \geq 0$ we also introduce the subset $\mathrm{C}_p^2(\mathbb{R}^d)$ of all functions f from $\mathrm{C}^2(\mathbb{R}^d)$ such that $|f(x)| + (1 + |x|)|\nabla f(x)| + (1 + |x|^2)|D^2 f(x)| \leq C(1 + |x|^p)$, $x \in \mathbb{R}^d$, for some $C > 0$.

For every $m \in \mathbb{N}_0$ and $\delta \in (0, 1)$ denote by $\mathrm{C}^{m,\delta}(\mathbb{R}^d)$ the subset of all functions from $\mathrm{C}^m(\mathbb{R}^d)$ whose m-th derivatives are locally δ-Hölder continuous, that is, a function $f \in \mathrm{C}^m(\mathbb{R}^d)$ belongs to $\mathrm{C}^{m+\delta}(\mathbb{R}^d)$ if for every $R > 0$ there exists a constant C such that

$$|D^\alpha f(x) - D^\alpha f(y)| \le C|x - y|^\delta$$

for all $x, y \in \mathbb{R}^d$ with $|x - y| \le R$ and $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = m$. We equip the space $\mathrm{C}^{m,\delta}(\mathbb{R}^d)$ with the Fréchet topology generated by the following seminorms

$$\|f\|_{m+\delta,K} = \sup_{x \in K} \frac{|f(x)|}{1 + |x|} + \sum_{1 \le |\alpha| \le m} \sup_{x \in K} |D^\alpha f(x)| + \sum_{|\alpha| = m} \sup_{\substack{x, y \in K \\ x \ne y}} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x - y|^\delta}$$

for all compact sets $K \subset \mathbb{R}^d$. Set $\mathrm{C}_{lb}^{m,\delta} = \{f \in \mathrm{C}^{m,\delta} : \|f\|_{m+\delta} := \|f\|_{m+\delta,\mathbb{R}^d} < \infty\}$.

We also denote by $\tilde{\mathrm{C}}^{m,\delta}$ the set of all functions $f \in \mathrm{C}^m(\mathbb{R}^d \times \mathbb{R}^d)$ whose mixed m-th derivatives are locally δ-Hölder continuous, that is, for every $R > 0$ there exists a constant $C > 0$ such that

$$|D_x^\alpha D_y^\alpha f(x, y) - D_{x'}^\alpha D_y^\alpha f(x', y) - D_x^\alpha D_{y'}^\alpha f(x, y') + D_{x'}^\alpha D_{y'}^\alpha f(x', y')| \le C|x - x'|^\delta |y - y'|^\delta$$

for all $x, y \in \mathbb{R}^d$ with $|x - y| < R$ and $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = m$. Similarly to $\mathrm{C}^{m,\delta}(\mathbb{R}^d)$, the space $\tilde{\mathrm{C}}^{m,\delta}(\mathbb{R}^d)$ will be equipped with the Fréchet topology generated by the seminorms

$$\|f\|_{m+\delta,K}^\sim = \sup_{x,y \in K} \frac{|f(x, y)|}{(1 + |x|)(1 + |y|)} + \sum_{1 \le |\alpha| \le m} \sup_{x, y \in K} |D_x^\alpha D_y^\alpha f(x, y)|$$

$$+ \sum_{|\alpha| = m} \sup_{\substack{x, x', y, y' \in K \\ x \ne x', y \ne y'}} \frac{|D_x^\alpha D_y^\alpha f(x, y) - D_{x'}^\alpha D_y^\alpha f(x', y) - D_x^\alpha D_{y'}^\alpha f(x, y') + D_{x'}^\alpha D_{y'}^\alpha f(x', y')|}{|x - x'|^\delta |y - y'|^\delta}.$$

for all compact sets $K \subset \mathbb{R}^d$. Set $\tilde{\mathrm{C}}_{lb}^{m,\delta}(\mathbb{R}^d) = \{f \in \tilde{\mathrm{C}}^{m,\delta}(\mathbb{R}^d) : \|f\|_{m+\delta}^\sim := \|f\|_{m+\delta,\mathbb{R}^d}^\sim < \infty\}$.

Let $L_2(\mathbb{R}^d)$ be the Hilbert space of all 2-integrable functions on $\mathbb{R}^d$ with respect to the Lebesgue measure with the usual L_2-norm $\|\cdot\|_{L_2}$ and inner product $\langle \cdot, \cdot \rangle_{L_2}$.

For $J \in \mathbb{N}_0$ and an open domain $\Gamma \subset \mathbb{R}^d$ we denote the complete extension of the space $\mathrm{C}_c^\infty(\Gamma)$ with respect to the norm defined by

$$\|\varphi\|_{J,\Gamma}^2 = \sum_{|\alpha| \le J} \int_\Gamma |D^\alpha \varphi(x)|^2 dx$$

by $H^J(\Gamma)$. It is well-known that $H^J(\Gamma)$ is a separable Hilbert space with the inner product

$$\langle \varphi_1, \varphi_2 \rangle_{J,\Gamma} = \sum_{|\alpha| \le J} \int_\Gamma D^\alpha \varphi_1(x) D^\alpha \varphi_2(x) dx.$$

The dual space to $H^J(\Gamma)$ equipped with the norm

$$\|f\|_{-J,\Gamma} = \sup_{\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{\langle \varphi, f \rangle_{0,\Gamma}}{\|\varphi\|_J}$$

will be denoted by $H^{-J}(\Gamma)$. It is also a separable Hilbert space with the inner product denoted by $\langle \cdot, \cdot \rangle_{-J,\Gamma}$. According to the Riesz representation theorem, there exists the isometry between $H^J(\Gamma)$ and $H^{-J}(\Gamma)$ denoted by $L_{J,\Gamma}$. We will often drop Γ from the notation of the inner product and the norm on a Sobolev space, if it does not lead to the confusion. For more details about the Sobolev spaces $H^J(\Gamma)$ and $H^{-J}(\Gamma)$ we refer the reader to, e.g., [1].

For $I = [0, T]$ or $I = [0, \infty)$ the space of all càdlàg functions from I to a metric space E equipped with the Skorohod topology will be denoted by $D(I, E)$. The set $C(I, E)$ of all continuous functions from I to E is a closed subset of $D(I, E)$ and the induced topology on $C(I, E)$ is equivalent to the topology of uniform convergence on compacts.

The space of all probability measures (resp. signed measures with finite total variations) on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ equipped with the topology of weak convergence will be denoted by $\mathcal{P}(\mathbb{R}^d)$ (resp. by $\mathcal{M}(\mathbb{R}^d)$). Let $\mathcal{P}_p(\mathbb{R}^d)$ denote the subset of $\mathcal{P}(\mathbb{R}^d)$ of all probability measures with finite p-moment for $p \geq 1$, that is,

$$\mathcal{P}_p(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \ \langle \phi_p, \mu \rangle < \infty \right\},$$

where $\phi_p(x) = |x|^p$, $x \in \mathbb{R}^d$, and $\langle \varphi, \mu \rangle$ (and also $\langle \varphi(x), \mu(dx) \rangle$) is the integration of φ with respect to μ. It is well-know that $\mathcal{P}_p(\mathbb{R}^d)$ is a Polish space with the Wasserstein distance given for each $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ by

$$\mathcal{W}_p(\mu, \nu) = \inf \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x - y|^p \chi(dx, dy) \right)^{\frac{1}{p}},$$

where the infimum is taken over all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν.

We will fix a measure space $(\Theta, \mathcal{G}, \vartheta)$ such that ϑ is a finite measure and the space $L_2(\Theta, \vartheta) := L_2(\Theta, \mathcal{G}, \vartheta)$, which consists of all 2-integrable with respect to ϑ functions (more precisely, equivalence classes) from Θ to $\mathbb{R}$, is separable. The usual inner product and norm on $L_2(\Theta, \vartheta)$ are denoted by $\langle \cdot, \cdot \rangle_\vartheta$ and $\| \cdot \|_\vartheta$, respectively. Let W_t, $t \geq 0$, be a cylindrical Wiener process on $L_2(\Theta, \vartheta)$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\mathcal{F}_t)_{t \geq 0}$ be the right-continuous complete extension of the filtration generated by W, which exists according to [46, Lemma 7.8]. We recall that for an $(\mathcal{F}_t)$-progressively measurable $L_2(\Theta, \vartheta)$-valued process g_t, $t \geq 0$, satisfying

$$\int_0^t \int_\Theta g_t^2(\theta) \vartheta(d\theta) < \infty \quad \text{a.s.}$$

for every $t > 0$, the integral[2] defined by

$$\int_0^t g_s(\theta) W(d\theta, ds) := \int_0^t G_s dW_s, \quad t \geq 0,$$

is a continuous local $(\mathcal{F}_t)$-martingale with quadratic variation

$$(1.17) \qquad \left[\int_0^\cdot g_s(\theta) W(d\theta, ds) \right]_t = \int_0^t \int_\Theta g_s^2(\theta) \vartheta(d\theta), \quad t \geq 0,$$

where $G_t = \langle g_t, \cdot \rangle_\vartheta$, $t \geq 0$, is an $(\mathcal{F}_t)$-progressively measurable process on the space of Hilbert–Schmidt operators on $L_2(\Theta, \vartheta)$. Denote the space of all Hilbert–Schmidt operators from $L_2(\Theta, m)$ to a Hilber space H by $\mathcal{L}_2(L_2(\Theta, m); H)$ and $\| \cdot \|_{\mathrm{HS}, H}$ be the Hilbert–Schmidt norm on that space. In particular, $\|G_t\|_{\mathrm{HS}, \mathbb{R}} = \|g_t\|_\vartheta$.

---

[2]For the definition of the stochastic integral with respect to a cylindrical Wiener process see, e.g., [37, Section 2.2.4]. The equality (1.17) holds due to Theorem 2.4 [37] and the fact that $G_t G_t^* = \|g\|_\vartheta^2$ for $G_t = \langle g_t, \cdot \rangle_\vartheta$.

2. Well-posedness and superposition principle

In this section, we establish the well-posedness and the superposition principle for the stochastic mean-field equation

$$
d\mu_t = \frac{1}{2} D^2 : (A(t, \cdot, \mu_t)\mu_t)\, dt - \nabla \cdot (V(t, \cdot, \mu_t)\mu_t)\, dt
$$

(2.1)
$$
- \int_\Theta \nabla \cdot (G(t, \cdot, \mu_t, \theta)\mu_t)\, W(d\theta, dt),
$$

where W is a cylindrical Wiener process in $L_2(\Theta, \vartheta)$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the functions $V : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \Omega \to \mathbb{R}^d$ and $G : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \Omega \to (L_2(\Theta, \vartheta))^d$, $A : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \Omega \to \mathbb{R}^{d \times d}$ satisfy the following Assumption 2.1.

Assumption 2.1. *The functions V and G are $\mathcal{B}([0, \infty)) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathcal{P}_2(\mathbb{R}^d)) \otimes \mathcal{F}$-measurable and bounded (in (t, x, μ)) on every compact subset of $[0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ a.s., the restrictions of V and G to the time interval $[0, t]$ are $\mathcal{B}([0, t]) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathcal{P}_2(\mathbb{R}^d)) \otimes \mathcal{F}_t$-measurable, and*

$$
A(t, x, \mu) = \big(\langle G_i(t, x, \mu, \cdot), G_j(t, x, \mu, \cdot)\rangle_\vartheta\big)_{i,j \in [d]}
$$

for all $t \geq 0$, $x \in \mathbb{R}^d$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Definition 2.2. Let $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$. A continuous $(\mathcal{F}_t)$-adapted process μ_t, $t \geq 0$, in $\mathcal{P}_2(\mathbb{R}^d)$ is a *(strong) solution* to the stochastic mean-field equation (2.1) started from μ_0 if for every $\varphi \in \mathrm{C}_c^2(\mathbb{R}^d)$ a.s. the equality

$$
\langle \varphi, \mu_t \rangle = \langle \varphi, \mu_0 \rangle + \frac{1}{2} \int_0^t \langle D^2\varphi : A(s, \cdot, \mu_s), \mu_s \rangle\, ds
$$

(2.2)
$$
+ \int_0^t \langle \nabla\varphi \cdot V(s, \cdot, \mu_s), \mu_s \rangle\, ds + \int_0^t \int_\Theta \langle \nabla\varphi \cdot G(s, \cdot, \mu_s, \theta), \mu_s \rangle\, W(d\theta, ds)
$$

holds for every $t \geq 0$.

All integrals in the definition above are well-defined due to the a.s. boundedness of the functions inside the integral $\langle \cdot, \mu_s \rangle$.

Remark 2.3. If for every $T > 0$ and compact set K in $\mathcal{P}_2(\mathbb{R}^d)$ the coefficients V and G are a.s. bounded on $[0, T] \times \mathbb{R}^d \times K$, that is, there exists a (random) constant $C > 0$ such that

$$
|V(t, x, \mu)| + \|\|G(t, x, \mu, \cdot)\|\|_m \leq C, \quad t \in [0, T],\ x \in \mathbb{R}^d,\ \mu \in K,
$$

and μ_t, $t \geq 0$, is a solution to (2.1), then the integral equality (2.2) in Definition 2.2 holds for every $\varphi \in \mathrm{C}_b^2(\mathbb{R}^d)$. This follows from the dominated convergence theorem and the compactness of $\{\mu_t,\ t \in [0, T]\}$ in $\mathcal{P}_2(\mathbb{R}^d)$ due to the continuity of μ_t, $t \geq 0$.

Remark 2.4. Let μ_t, $t \geq 0$, be a solution to the equation (2.1). Then for every $\varphi \in \mathrm{C}_c^2(\mathbb{R}^d)$ the process $\langle \varphi, \mu_t \rangle$, $t \geq 0$, is a continuous local $(\mathcal{F}_t)$-semimartingale with quadratic variation

$$
[\langle \varphi, \mu_\cdot \rangle]_t = \int_0^t \int_\Theta \langle \nabla\varphi \cdot G(s, \cdot, \mu_s, \theta), \mu_s \rangle \langle \nabla\varphi \cdot G(s, \cdot, \mu_s, \theta), \mu_s \rangle\, \vartheta(d\theta)ds
$$

$$
= \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\nabla\varphi(x) \otimes \nabla\varphi(y)) : \tilde{A}(s, x, y, \mu_s)\mu_s(dx)\mu_s(dy)ds,
$$

where $\tilde{A}(s, x, y, \mu) = (\langle G_i(s, x, \mu, \cdot), G_j(s, y, \mu, \cdot)\rangle_\vartheta)_{i,j \in [d]}$. The expression for the quadratic variation of $\langle \varphi, \mu_t \rangle$, $t \geq 0$, directly follows from (1.17).

Together with the stochastic mean-field equation (2.1), we will consider the following associated SDE with interaction

$$dX(u,t) = V(t, X(u,t), \bar\mu_t)dt + \int_\Theta G(t, X(u,t), \bar\mu_t, \theta)W(d\theta, dt),$$

(2.3)

$$X(u,0) = u, \quad \bar\mu_t = \mu_0 \circ X^{-1}(\cdot, t), \quad u \in \mathbb{R}^d, \ t \geq 0.$$

This type of equation was introduced and studied by Dorogovtsev in [26, Section 2]. We next give the definition of a solution to (2.3), following [26, Definition 2.1.1].

Definition 2.5. A family of continuous processes $\{X(u,t), \ t \geq 0\}$, $u \in \mathbb{R}^d$, is called a *(strong) solution* to the SDE with interaction (2.3) if the restriction of X to the time interval $[0,t]$ is $\mathcal{B}([0,t]) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{F}_t$-measurable, $\bar\mu_t = \mu_0 \circ X^{-1}(\cdot, t) \in \mathcal{P}_2(\mathbb{R}^d)$ a.s. for all $t \geq 0$ and for every $u \in \mathbb{R}^d$ a.s.

$$X(u,t) = u + \int_0^t V(s, X(u,s), \bar\mu_s)ds + \int_0^t \int_\Theta G(s, X(u,s), \bar\mu_s, \theta)W(d\theta, ds)$$

for all $t \geq 0$.

We remark that $\bar\mu_t$, $t \geq 0$, is an $(\mathcal{F}_t)$-progressively measurable process in $\mathcal{P}_2(\mathbb{R}^d)$. Moreover, due to Fubini's theorem it does not depend on the version of $X(u, \cdot)$, $u \in \mathbb{R}^d$, that is, if $Y(u,t)$, satisfies the same measurablity conditions from Definition 2.5 as X and for every $u \in \mathbb{R}^d$ a.s. $Y(u, \cdot) = X(u, \cdot)$, then $\bar\mu_t = \mu_0 \circ Y^{-1}(\cdot, t)$, $t \geq 0$, a.s.

The key tool in the investigation of solutions to the stochastic mean-field equation (2.1) is the fact that it satisfies the (strong) superposition principle.

Definition 2.6. A continuous process μ_t, $t \geq 0$, in $\mathcal{P}_2(\mathbb{R}^d)$ started from $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ is a *strong superpositon solution*[3] to the stochastic mean-field equation (2.1) or *satisfies the superposition principle* if there exists a solution $X(u,t)$, $t \geq 0$, $u \in \mathbb{R}^d$, to the SDE with interaction (2.3) such that $\mu_t = \mu_0 \circ X^{-1}(\cdot, t)$, $t \geq 0$, a.s.

In order to build a strong superposition solution to the equation (2.1), we will need the Lipschitz continuity assumption on the coefficients, which will guarantee the well-posedness of the SDE (2.3).

Assumption 2.7. *The coefficients V and G are Lipschitz continuous with respect to x and μ, that is, for every $T > 0$ there exists $L > 0$ such that a.s. for every $t \in [0,T]$, $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$*

$$|V(t,x,\mu) - V(t,y,\nu)| + \||G(t,x,\mu,\cdot) - G(t,y,\nu,\cdot)\||_\vartheta \leq L \left(|x-y| + \mathcal{W}_2(\mu,\nu)\right).$$

and

$$|V(t,0,\delta_0)| + \||G(t,0,\delta_0,\cdot)\||_\vartheta \leq L,$$

where δ_0 denotes the δ-measure at 0 on $\mathbb{R}^d$.

Remark 2.8. Assumption 2 implies linear growth of V and G, that is, a.s.

$$|V(t,x,\mu)| + \||G(t,x,\mu,\cdot)\||_\vartheta \leq L(1 + |x| + \mathcal{W}_2(\mu, \delta_0))$$

for all $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

---

[3]The notion of the strong superposiltion solution considered in this work is close to one introduced by Flandoli (see [35, Definition 5]). Since we will only work with equation (2.3) which has a unique solution, we avoid more general definitions, like in [3, 35, 79], which needs the introduction of distributions on the path space.

In the next section, using a standard approach, we will prove the well-posedness of the SDE with interaction (2.3) stated in the following theorem.

Theorem 2.9 (Well-posedness of the SDE with interaction)**.** *Under Assumptions 2.1 and 2.7, the SDE with interaction* (2.3) *has a unique strong solution* $X(u,t)$, $t \geq 0$, $u \in \mathbb{R}^d$, *for every* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$*. Moreover, for every* $T > 0$ *and* $p \geq 2$ *there exists a constant* $C > 0$ *(only depending of* L, p, d *and* T*) such that*

$$\mathbb{E} \sup_{t \in [0,T]} |X(u,t)|^p \leq C(1 + \langle \phi_p, \mu_0 \rangle + |u|^p)$$

for all $u \in \mathbb{R}^d$*, where* $\phi_p(x) = |x|^p$, $x \in \mathbb{R}^d$*.*

Corollary 2.10 (Moment preservation property)**.** *Under Assumptions 2.1 and 2.7, the measure-valued process* $\bar{\mu}_t = \mu_0 \circ X^{-1}(\cdot, t)$, $t \geq 0$, *is moment preserving, that is, for every* $T > 0$ *and* $p \geq 2$

$$\mathbb{E} \sup_{t \in [0,T]} \langle \phi_p, \bar{\mu}_t \rangle \leq C \left(1 + \langle \phi_p, \mu_0 \rangle \right).$$

The corollary directly follows from the inequality in Theorem 2.9 by its integration with respect to μ_0.

Next, using Itô's formula, one can prove the existence of a superposition solution to (2.1).

Theorem 2.11 (Existence of solutions)**.** *Let* V, G, A *satisfy Assumptions 2.1 and 2.7. Then for every* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *there exists a solution* μ_t, $t \geq 0$, *to the stochastic mean-field equation* (2.1) *started from* μ_0 *that satisfies the superposition principle and is moment preserving.*

The proof of this theorem directly follows from Lemma 2.13 below. We note that the uniqueness of the stochastic mean-field equation is closely related to the superposition principle. Indeed, the well-posedness of the SDE with interaction (see Theorem 2.9) and Theorem 2.11 immediately imply the following corollary.

Corollary 2.12. *Let* V, G, A *satisfy Assumptions 2.1, 2.7. The stochastic mean-field equation* (2.1) *has a unique solution if and only if every its solution in the sense of Definition 2.2 is a strong superposition solution in the sense of Definition 2.6.*

In Section 2.2, we will prove the uniqueness to the stochastic mean-field equation making further assumptions on the initial condition and its coefficients.

2.1. **SDE with interaction.** We start this section with the proof that any solution to the SDE with interaction (2.3) provides a solution to the equation (2.1) in the sense of Definition 2.2.

Lemma 2.13. *Let* V, G, A *satisfy Assumption 2.1,* $X(u,t)$, $t \geq 0$, $u \in \mathbb{R}^d$, *be a solution to the SDE with interaction* (2.3) *with* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *and* $\bar{\mu}_t = \mu_0 \circ X^{-1}(\cdot, t)$, $t \geq 0$, *be continuous in* $\mathcal{P}_2(\mathbb{R}^d)$*. Then* $\bar{\mu}_t$, $t \geq 0$, *is a solution to the stochastic mean-field equation* (2.1) *started from* μ_0*.*

Proof. We note that for every $u \in \mathbb{R}^d$ the process $X(u,t)$, $t \geq 0$, is a continuous semimartingale in $\mathbb{R}^d$ with quadratic variation

$$[X(u, \cdot)]_t = \int_0^t A(s, X(u,s), \bar{\mu}_s) ds, \quad t \geq 0.$$

Indeed, according to (1.17) and the polarisation equality, one gets

$$[X_i(u, \cdot), X_j(u, \cdot)]_t = \int_0^t \langle G_i(s, X(u,s), \bar{\mu}_s, \cdot), G_j(s, X(u,s), \bar{\mu}_s, \cdot) \rangle_\vartheta ds$$

$$= \int_0^t A_{i,j}(s, X(u,s), \bar{\mu}_s)ds, \quad t \geq 0,$$

for all $i, j \in [d]$.

Next, taking $\varphi \in C_c^2(\mathbb{R}^d)$ and applying Itô's formula to $\varphi(X(u,t))$ for every $u \in \mathbb{R}^d$, we get a.s.

$$\varphi(X(u,t)) = \varphi(u) + \int_0^t \nabla\varphi(X(u,s)) \cdot V(s, X(u,s), \bar{\mu}_s)ds$$

$$+ \frac{1}{2}\int_0^t D^2\varphi(X(u,s)) : A(s, X(u,s), \bar{\mu}_s)ds$$

$$+ \int_0^t \nabla\varphi(X(u,s)) \cdot dN(u,s), \quad t \geq 0.$$

where $N(u,t) = \int_0^t \int_\Theta G(s, X(u,s), \bar{\mu}_s, \theta)W(d\theta, ds), t \geq 0$. Using the definition of the stochastic integral with respect to a cylindrical Wiener process, it is easily to see that

$$\varphi(X(u,t)) = \varphi(u) + \int_0^t \nabla\varphi(X(u,s)) \cdot V(s, X(u,s), \bar{\mu}_s)ds$$

$$+ \frac{1}{2}\int_0^t D^2\varphi(X(u,s)) : A(s, X(u,s), \bar{\mu}_s)ds$$

$$+ \int_0^t \int_\Theta \nabla\varphi(X(u,s)) \cdot G(s, X(u,s), \bar{\mu}_s, \theta)W(d\theta, ds), \quad t \geq 0.$$

Since φ has a compact support and the set $\{\bar{\mu}_s, \ s \in [0,t]\}$ is compact a.s. as the image of the compact set $[0,t]$ under the continuous map $s \mapsto \bar{\mu}_s$, the functions $\nabla\varphi(x) \cdot V(s, x, \bar{\mu}_s)$, $D^2\varphi(x) : A(s, x, \bar{\mu}_s)$ and $\|\nabla\varphi(x) \cdot G(s, x, \bar{\mu}_s, \cdot)\|_\vartheta^2, s \in [0,t], x \in \mathbb{R}^d$ are bounded a.s., by Assumption 2.1. Hence, we may integrate the above expression with respect to μ_0 and use Fubini's theorem and the equality $\int_{\mathbb{R}^d} \psi(X(u,t))\mu_0(du) = \langle \psi, \bar{\mu}_t \rangle$ for $\psi \in C_b(\mathbb{R}^d)$ to get that $\bar{\mu}_t, t \geq 0$, satisfies (2.2). This ends the proof of the lemma. $\square$

We next prove Theorem 2.9. Since its proof is similar to the proofs of [26, Theorem 2.2.1], we will only provide a sketch.

Proof of Theorem 2.9. Let $\bar{\mu}_t^0 = \mu_0, t \geq 0$. We define inductively for every $n \geq 1$ the family of continuous processes $\{X_n(u,t), \ t \geq 0\}, u \in \mathbb{R}^d$, as solutions to the usual SDEs

$$(2.4) \qquad dX_n(u,t) = V\left(s, X_n(u,t), \bar{\mu}_s^{n-1}\right)dt + \int_\Theta G\left(t, X_n(u,t), \bar{\mu}_s^{n-1}, \theta\right)W(d\theta, dt),$$

$$X_n(u,0) = u$$

and

$$(2.5) \qquad\qquad\qquad \bar{\mu}_n = \mu_0 \circ X_n^{-1}(\cdot, t), \quad t \geq 0.$$

Using Assumptions 2.1 and 2.7, it is easy to see that SDE (2.4) has a unique solution for every $u \in \mathbb{R}^d$ and for every $p \geq 2, n \geq 1$ and $T > 0$ there exists a constant $C > 0$, independent of u, such that

$$(2.6) \qquad\qquad\qquad \mathbb{E}\sup_{t \in [0,T]} |X_n(u,t)|^p \leq C(1 + |u|^p)$$

for all $n \geq 1$ and $u \in \mathbb{R}^d$.

We will first show that X_n satisfies the measurability assumptions from Definition 2.5. For this we will prove that X_n has a continuous version in u. Taking $u, v \in \mathbb{R}^d$, $T > 0$, $p \geq 2$, and using Hölder's inequality and the Burkholder–Davis–Gundy inequality, we estimate for each $t \in [0, T]$

$$\mathbb{E} \sup_{s \in [0,t]} |X_n(u, s) - X_n(v, s)|^p \leq C_1 |u - v|^p$$

$$+ C_1 \mathbb{E} \int_0^t \left| V\left(s, X_n(u, s), \bar{\mu}_s^{n-1}\right) - V\left(s, X_n(v, s), \bar{\mu}_s^{n-1}\right) \right|^p ds$$

$$+ C_1 \mathbb{E} \int_0^t \left\| |G(s, X_n(u, s), \bar{\mu}_s^{n-1}, \cdot) - G(s, X_n(v, s), \bar{\mu}_s^{n-1}, \cdot) | \right\|_\vartheta^p ds,$$

where C_1 is a constant that depends only on p, T and d. By Assumption 2.7 and Gronwall's lemma, we get

$$(2.7) \qquad \mathbb{E} \sup_{t \in [0,T]} |X_n(u, t) - X_n(v, t)|^p \leq C |u - v|^p,$$

where C also depends only on p, T, L and d. Therefore, we can conclude from [46, Theorem 3.23] that $X_n(u, \cdot)$, $u \in \mathbb{R}^d$, has a continuous version as a $C([0, \infty))$-valued process, which will be also denoted by X_n. Note that the choice of different version of $X_n(u, \cdot)$, $u \in \mathbb{R}^d$, does not change the fact that $X_n(u, \cdot)$ solves equation (2.4) for every $u \in \mathbb{R}^d$. Thus, the desired measurability of X_n follows from the continuity of $(u, t) \mapsto X_n(u, t)$ a.s. We also note that $\bar{\mu}^n$, defined by (2.5), is an $(\mathcal{F}_t)$-progressively measurable continuous process in $\mathcal{P}_2(\mathbb{R}^d)$, where the continuity follows from bound (2.6) and de la Vallée-Poussin [63, Theorem 1.8]. Note that $\bar{\mu}.$ does not depend on the choice of a version of $X_n(u, \cdot)$, $u \in \mathbb{R}^d$.

Next, using Assumption 2.7 again, we can estimate for every $T > 0$, $t \in [0, T]$, $u \in \mathbb{R}^d$ and $n \geq 1$

$$\mathbb{E} \sup_{s \in [0,t]} |X_{n+1}(u, s) - X_n(u, s)|^2 \leq C \int_0^t \mathbb{E} \sup_{r \in [0,s]} |X_{n+1}(u, r) - X_n(u, r)|^2 ds$$

$$+ C \int_0^t \mathbb{E} \sup_{r \in [0,s]} \mathcal{W}_2^2\left(\bar{\mu}_r^n, \bar{\mu}_r^{n-1}\right) ds,$$

where C is independent of u, t and n. By Gronwall's lemma, we have

$$(2.8) \qquad \mathbb{E} \sup_{s \in [0,t]} |X_{n+1}(u, s) - X_n(u, s)|^2 \leq C e^{CT} \int_0^t \mathbb{E} \sup_{r \in [0,s]} \mathcal{W}_2^2\left(\bar{\mu}_r^n, \bar{\mu}_r^{n-1}\right) ds.$$

Using the definition of the Wasserstein distance, we further estimate for $t \in [0, T]$ and $n \geq 2$

$$\mathbb{E} \sup_{s \in [0,t]} \mathcal{W}_2^2\left(\bar{\mu}_s^n, \bar{\mu}_s^{n-1}\right) \leq \mathbb{E} \sup_{s \in [0,t]} \int_{\mathbb{R}^d} |X_n(u, s) - X_{n-1}(u, s)|^2 \mu_0(du)$$

$$\leq \int_{\mathbb{R}^d} \mathbb{E} \sup_{s \in [0,t]} |X_n(u, s) - X_{n-1}(u, s)|^2 \mu_0(du)$$

$$\leq C e^{CT} \int_0^t \mathbb{E} \sup_{r \in [0,s]} \mathcal{W}_2^2\left(\bar{\mu}_r^{n-1}, \bar{\mu}_r^{n-2}\right) ds,$$

where we have used the equality $\mu_0(\mathbb{R}^d) = 1$ and (2.8) in the last step. Iterating the above inequality $n - 2$ times, we get

$$\mathbb{E} \sup_{t \in [0,T]} \mathcal{W}_2^2\left(\bar{\mu}_t^n, \bar{\mu}_t^{n-1}\right) \leq C^2 e^{2CT} \int_0^t \int_0^{s_1} \mathbb{E} \sup_{r \in [0,s_2]} \mathcal{W}_2^2\left(\bar{\mu}_r^{n-2}, \bar{\mu}_r^{n-3}\right) ds_2 ds_1 \leq \dots$$

$$\leq C^{n-1}e^{(n-1)CT}\int_0^t\int_0^{s_1}\cdots\int_0^{s_{n-2}}\mathbb{E}\sup_{r\in[0,s_{n-1}]}\mathcal{W}_2^2\left(\bar{\mu}_r^1,\bar{\mu}_r^0\right)ds_{n-1}\ldots ds_2ds_1$$

$$\leq\frac{C^{n-1}e^{(n-1)CT}T^{n-1}}{(n-1)!}\mathbb{E}\sup_{t\in[0,T]}\mathcal{W}_2^2\left(\bar{\mu}_t^1,\bar{\mu}_t^0\right).$$

The finiteness of the expectation on the right hand side of the above inequality follows from (2.6). By (2.8),

$$\mathbb{E}\sup_{t\in[0,T]}|X_{n+1}(u,t)-X_n(u,t)|^2\leq\frac{C^ne^{nCT}T^n}{(n-1)!}\mathbb{E}\sup_{t\in[0,T]}\mathcal{W}_2^2\left(\bar{\mu}_t^1,\bar{\mu}_t^0\right).$$

Next, using the Borel–Cantelli lemma, it is easily seen that there exist continuous processes $\bar{\mu}_t$, $t\geq 0$, in $\mathcal{P}_2(\mathbb{R}^d)$ and $X(u,t)$, $t\geq 0$, in $\mathbb{R}^d$, $u\in\mathbb{R}^d$, such that for every $T>0$ and $u\in\mathbb{R}^d$

$$\sup_{t\in[0,T]}|X_n(u,t)-X(u,t)|\to 0\quad\text{and}\quad\sup_{t\in[0,T]}\mathcal{W}_2\left(\bar{\mu}_t^n,\bar{\mu}_t\right)\to 0\quad\text{a.s.}$$

as $n\to\infty$. Moreover, for every $u\in\mathbb{R}^d$ a.s.

$$X(u,t)=u+\int_0^t V(s,X(u,s),\bar{\mu}_s)ds+\int_0^t\int_\Theta G(s,X(u,s),\bar{\mu}_s,\theta)W(d\theta,ds)$$

for all $t\geq 0$, by Assumption 2.7. Since the constant in inequality (2.7) does not depend on u,v and n, the inequality remains true for X_n replaced by X, by Fatou's lemma. Therefore, the $C([0,\infty))$-valued random field $X(u,\cdot)$, $u\in\mathbb{R}^d$, has a continuous version, which is also denoted by X. This implies that X satisfies the measurability assumptions of Definition 2.5. We also remark that for every $\varphi\in C_b(\mathbb{R}^d)$ and $t\geq 0$ one has a.s.

$$\langle\varphi,\bar{\mu}_t\rangle=\lim_{n\to\infty}\langle\varphi,\bar{\mu}_t^n\rangle=\lim_{n\to\infty}\int_{\mathbb{R}^d}\varphi(X_n(u,t))\mu_0(du)=\int_{\mathbb{R}^d}\varphi(X(u,t))\mu_0(du),$$

by the dominated convergence theorem. This completes the proof of the well-posedness of SDE (2.3).

We next show the finiteness of moments of the solution X. For each $u\in\mathbb{R}^d$ and $n,m\in\mathbb{N}$ we define the stopping times

$$\tau_u^n=\inf\left\{t\geq 0:\ |X(u,t)|\geq n\right\}\quad\text{and}\quad\sigma^n=\inf\left\{t\geq 0:\ \mathcal{W}_2(\bar{\mu}_t,\delta_0)\geq n\right\}.$$

Let also $\tau_u^{n,m}=\tau_u^n\wedge\sigma^m$. Then, using Hölder's inequality, the Burkholder–Davis–Gundy inequality (see, e.g., Theorem 3.28 [47]) and Remark 2.8, we estimate for $p\geq 2$ and every $n,m\in\mathbb{N}$, $t\in[0,T]$

$$\mathbb{E}\sup_{s\in[0,t]}|X(u,s\wedge\tau_u^{n,m})|^p\leq C|u|^p+C\mathbb{E}\sup_{s\in[0,t]}\left|\int_0^{s\wedge\tau_u^{n,m}}V(r,X(u,r),\bar{\mu}_r)dr\right|^p$$

$$+C\mathbb{E}\sup_{s\in[0,t]}\left|\int_0^{s\wedge\tau_u^{n,m}}G(r,X(u,s),\bar{\mu}_r,\theta)W(d\theta,dr)\right|^p$$

$$\leq C|u|^p+C\mathbb{E}\int_0^{t\wedge\tau_u^{n,m}}|V(r,X(u,r),\bar{\mu}_r)|^p\,ds$$

$$+C\mathbb{E}\left(\int_0^{t\wedge\tau_u^{n,m}}\||G(s,X(u,s),\bar{\mu}_s,\cdot)\||_\vartheta^2\,ds\right)^{\frac{p}{2}}$$

$$\leq C|u|^p+C\mathbb{E}\int_0^{t\wedge\tau_u^{n,m}}\left(1+|X(u,s)|^p+\mathcal{W}_2^p(\bar{\mu}_s,\delta_0)\right)ds$$

$$\leq C(1 + |u|^p) + C\mathbb{E} \int_0^t |X(u, s \wedge \tau_u^{n,m})|^p ds$$

$$+ C\mathbb{E} \int_0^t \mathcal{W}_2^p(\bar{\mu}_{s \wedge \tau_u^{n,m}}, \delta_0) ds$$

$$\leq C(1 + |u|^p) + C\mathbb{E} \int_0^t \sup_{r \in [0,s]} |X(u, r \wedge \tau_u^{n,m})|^p ds$$

$$+ C \int_0^t \mathbb{E} \mathcal{W}_2^p(\bar{\mu}_{s \wedge \tau_u^{n,m}}, \delta_0) ds,$$

where the constant C depends only on L, p, d and T. By Gronwall's lemma, we get

$$\mathbb{E} \sup_{s \in [0,t]} |X(u, s \wedge \tau_u^{n,m})|^p \leq Ce^{CT} \left(1 + |u|^p + \int_0^t \mathbb{E} \mathcal{W}_2^p(\bar{\mu}_{s \wedge \tau_u^{n,m}}, \delta_0) ds \right).$$

Making $n \to \infty$ and using Fatou's lemma, we obtain

$$(2.9) \qquad \mathbb{E} \sup_{s \in [0,t]} |X(u, s \wedge \sigma^m)|^p \leq Ce^{CT} \left(1 + |u|^p + \int_0^t \mathbb{E} \mathcal{W}_2^p(\bar{\mu}_{s \wedge \sigma^m}, \delta_0) ds \right)$$

for all $t \in [0, T]$, $u \in \mathbb{R}^d$ and $m \geq 1$. In order to bound the integral in the inequality above, we first estimate

$$\mathbb{E} \sup_{s \in [0,t]} \mathcal{W}_2^p(\bar{\mu}_{s \wedge \sigma^m}, \delta_0) = \mathbb{E} \left(\sup_{s \in [0,t]} \int_{\mathbb{R}^d} |X(u, s \wedge \sigma^m)|^2 \mu_0(du) \right)^{\frac{p}{2}}$$

$$\leq \mathbb{E} \int_{\mathbb{R}^d} \sup_{s \in [0,t]} |X(u, s \wedge \sigma^m)|^p \mu_0(du)$$

$$\leq Ce^{CT} \int_{\mathbb{R}^d} \left(1 + |u|^p + \int_0^t \mathbb{E} \mathcal{W}_2^p(\bar{\mu}_{s \wedge \sigma^m}, \delta_0) ds \right) \mu_0(du)$$

$$= C_1 (1 + \langle \phi_p, \mu_0 \rangle) + C_1 \int_0^t \mathbb{E} \mathcal{W}_2^p(\bar{\mu}_{s \wedge \sigma^m}, \delta_0) ds$$

$$= C_1 (1 + \langle \phi_p, \mu_0 \rangle) + C_1 \int_0^t \mathbb{E} \sup_{r \in [0,s]} \mathcal{W}_2^p(\bar{\mu}_{r \wedge \sigma^m}, \delta_0) ds,$$

where $C_1 = Ce^{CT}$. Hence, Gronwall's lemma yields

$$\mathbb{E} \sup_{s \in [0,t]} \mathcal{W}_2^p(\bar{\mu}_{s \wedge \sigma^m}, \delta_0) \leq C_1 e^{C_1 T} (1 + \langle \phi_p, \mu_0 \rangle)$$

for all $t \in [0, T]$ and $m \geq 1$. Combining the inequality above with (2.9), we obtain

$$\mathbb{E} \sup_{s \in [0,t]} |X(u, s \wedge \sigma^m)|^p \leq C (1 + \langle \phi_p, \mu_0 \rangle + |u|^p),$$

where the constant C depends only on L, p, d and T. Now, making $m \to \infty$ and using Fatou's lemma, we complete the proof of the theorem. $\square$

We will further prove the continuous dependence of $\bar{\mu}.$ on the initial condition. This is known for SDEs with interaction with non-random coefficients driven by a Brownian sheet (see [26, Exercise 5.3.1]).

Theorem 2.14. *Let the coefficients of SDE* (2.3) *satisfy Assumptions* 2.1, 2.7 *and let* X_i, $i \in [2]$, *be solutions to* (2.3) *with* $\mu_0 = \mu_0^i \in \mathcal{P}_2(\mathbb{R}^d)$. *Then, for every* $T > 0$, *there exists a constant* $C > 0$ *such that*

$$(2.10) \qquad \mathbb{E} \sup_{t \in [0,T]} |X_1(u,t) - X_2(v,t)|^2 \leq C \left(|u-v|^2 + \mathcal{W}_2^2(\mu_0^1, \mu_0^2) \right), \quad u, v \in \mathbb{R}^d,$$

and

$$(2.11) \qquad \mathbb{E} \sup_{t \in [0,T]} \mathcal{W}_2^2 \left(\bar{\mu}_t^1, \bar{\mu}_t^2 \right) \leq C \mathcal{W}_2^2 \left(\mu_0^1, \mu_0^2 \right),$$

where $\bar{\mu}_t^i = \mu_0^i \circ X_i(\cdot, t)$.

Proof. Using Hölder's inequality, the Burkholder–Davis–Gundy inequality and Assumption 2.7, we estimate for each $t \in [0, T]$ and $u, v \in \mathbb{R}^d$

$$\mathbb{E} \sup_{s \in [0,t]} |X_1(u,s) - X_2(v,s)|^2 \leq 3|u-v|^2$$

$$+ 3t\mathbb{E} \int_0^t \left| V(s, X_1(u,s), \bar{\mu}_s^1) - V(s, X_2(v,s), \bar{\mu}_s^2) \right|^2 ds$$

$$+ 3\mathbb{E} \sup_{s \in [0,t]} \left| \int_0^s \int_\Theta \left(G(r, X_1(u,r), \bar{\mu}_r^1, \theta) - G(r, X_2(v,r), \bar{\mu}_r^2, \theta) \right) W(d\theta, dr) \right|^2$$

$$\leq 3|u-v|^2 + C \int_0^t \left(\mathbb{E}|X_1(u,s) - X_2(v,s)|^2 + \mathbb{E}\mathcal{W}_2^2 \left(\bar{\mu}_s^1, \bar{\mu}_s^2 \right) \right) ds$$

$$\leq 3|u-v|^2 + C \int_0^t \mathbb{E} \sup_{r \in [0,s]} |X_1(u,r) - X_2(v,r)|^2 ds + C \int_0^t \mathbb{E}\mathcal{W}_2^2 \left(\bar{\mu}_s^1, \bar{\mu}_s^2 \right) ds,$$

where C is independent of u, v and t. By Gronwall's lemma, we get for each $t \in [0, T]$

$$(2.12) \qquad \mathbb{E} \sup_{s \in [0,t]} |X_1(u,s) - X_2(v,s)|^2 \leq C|u-v|^2 + C \int_0^t \mathbb{E}\mathcal{W}_2^2 \left(\bar{\mu}_s^1, \bar{\mu}_s^2 \right) ds.$$

We next take an arbitrary probability measure χ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_0^i, $i \in [2]$. Then the probability measure χ_t defined by $\chi_s(B) = \chi \left(\{ (u,v) : (X_1(u,s), X_2(v,s)) \in B \} \right)$, $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$, that is the pushforward of χ under the map $(u,v) \mapsto (X_1(u,s), X_2(v,s))$, has the marginals $\bar{\mu}_s^i$, $i \in [2]$, for each $t \in [0, T]$. Thus, by the estimate above, we obtain

$$\mathbb{E} \sup_{s \in [0,t]} \mathcal{W}_2^2(\bar{\mu}_s^1, \bar{\mu}_s^2) \leq \mathbb{E} \sup_{s \in [0,t]} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x-y|^2 \chi_s(dx, dy)$$

$$\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E} \sup_{s \in [0,t]} |X_1(u,s) - X_2(v,s)|^2 \chi(du, dv)$$

$$\leq C \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |u-v|^2 \chi(du, dv) + C \int_0^t \mathbb{E}\mathcal{W}_2^2(\bar{\mu}_s^1, \bar{\mu}_s^2) ds.$$

Taking infimum over all χ with marginals μ_0^i, $i \in [2]$, we get the inequality

$$\mathbb{E} \sup_{s \in [0,t]} \mathcal{W}_2^2(\bar{\mu}_s^1, \bar{\mu}_s^2) \leq C\mathcal{W}_2^2(\mu_0^1, \mu_0^2) + C \int_0^t \mathbb{E}\mathcal{W}_2^2(\bar{\mu}_s^1, \bar{\mu}_s^2) ds$$

$$\leq C\mathcal{W}_2^2(\mu_0^1, \mu_0^2) + C \int_0^t \mathbb{E} \sup_{r \in [0,s]} \mathcal{W}_2^2(\bar{\mu}_r^1, \bar{\mu}_r^2) ds,$$

where C depends only on T, L and d. Next, Gronwall's lemma implies

$$\mathbb{E} \sup_{t \in [0,T]} \mathcal{W}_2^2(\bar{\mu}_s^1, \bar{\mu}_s^2) \le C e^{CT} \mathcal{W}_2^2(\mu_0^1, \mu_0^2).$$

This completes the proof of (2.11).

The estimate (2.10) now directly follows from (2.11) and (2.12). This concludes the proof of the proposition. $\qquad\square$

We will also need to use the fact that two paths of particles described by an SDE with interaction never meet.

Lemma 2.15. *Let V and G satisfy Assumptions 2.1, 2.7 and let X be a solution to (2.3). Then for every $u, v \in \mathbb{R}^d$, $u \ne v$, a.s. $X(u,t) \ne X(v,t)$ for all $t \ge 0$.*

Proof. We set $X(u,v,t) = X(u,t) - X(v,t)$, $t \ge 0$, and introduce the following $(\mathcal{F}_t)$-stopping times

$$\sigma_n = \inf \left\{ t \ge 0 : \ |X(u,v,t)| \le \frac{1}{n} \right\}, \quad n \ge 1.$$

Then by Itô's formula and Assumption 2.7, we can estimate for every $T > 0$ and $t \in [0,T]$

$$\mathbb{E} \frac{1}{|X(u,v,t \wedge \sigma_n)|} \le \frac{1}{|u-v|} + C \int_0^t \mathbb{E} \frac{1}{|X(u,v,s \wedge \sigma_n)|} ds,$$

for some constant C independent of t and n. By Gronwall's lemma, $\mathbb{E} \frac{1}{|X(u,v,t \wedge \sigma_n)|} < \frac{e^{CT}}{|u-v|}$. Passing to the limit as $n \to \infty$ and using Fatou's lemma, one gets $\mathbb{E} \frac{1}{|X(u,v,t \wedge \sigma)|} < \frac{e^{CT}}{|u-v|}$, $t \in [0,T]$, where σ is defined similarly to σ_n with $\frac{1}{n}$ replaced by 0. Therefore, $X(u,v,t \wedge \sigma) > 0$ a.s. for all $t \ge 0$. This implies $t < \sigma$ a.s. for all $t \ge 0$. Hence, $\sigma = +\infty$ a.s. which completes the proof of the lemma. $\qquad\square$

2.2. **Uniqueness and superposition principle.** The main goal of this section is to prove the uniqueness of solutions to the stochastic mean-field equation. We will consider separately a few types of initial particle distributions: atomic, with L_2-density, and with a finite second moment. Depending on the type of initial conditions, we will need additional assumptions on the coefficients of the equation, which will appear in corresponding sections.

2.2.1. *Atomic initial conditions.* We remind the reader that $\phi_m(x) = |x|^m$, $x \in \mathbb{R}^d$, and set

$$\mathcal{N}_n = \left\{ \mu \in \mathcal{P}_2(\mathbb{R}^d) : \ \mu = \sum_{l=1}^n \alpha_l \delta_{x^l} \text{ for some } \alpha_l \ge 0 \text{ and } x^l \in \mathbb{R}^d \right\}, \quad n \ge 1.$$

It is easily seen that $\mathcal{N}_n$ is a closed subspace of $\mathcal{P}_2(\mathbb{R}^d)$ for each $n \ge 1$.

Theorem 2.16. *Let V, G, A satisfy Assumptions 2.1, 2.7, and let $\mu_0 \in \mathcal{N}_n$ for some $n \ge 1$. Then there exists a unique solution μ_t, $t \ge 0$, to (2.1) started from μ_0 and satisfying*

$$(2.13) \qquad \qquad \sup_{t \in [0,T]} \langle \phi_p, \mu_t \rangle < \infty \quad \text{a.s.}$$

for $p = n^2 + n$ and all $T > 0$. Moreover, μ_t, $t \ge 0$, is a superposition solution to (2.1).

We define the following function $F_n : \mathcal{P}_{n(n-1)}(\mathbb{R}^d) \to [0, \infty)$ by

$$(2.14) \qquad F_n(\mu) = \int_{\mathbb{R}^{dn}} \prod_{i,j : i \ne j} |z^i - z^j| \prod_{i=1}^n \mu(dz^i) = \int_{\mathbb{R}^{dn}} \prod_{i,j : i < j} |z^i - z^j|^2 \prod_{i=1}^n \mu(dz^i).$$

Then

(2.15) $$F_n^{-1}(\{0\}) = \mathcal{N}_{n-1}$$

for every $n \geq 2$.

Remark 2.17. The function F_n can be written as a polynomial of the maps

$$\mu \mapsto \int_{\mathbb{R}^d} z_1^{k_1} \dots z_d^{k_d} \mu(dz)$$

with $k_1 + \dots + k_d \leq n(n-1)$, $k_i \in \mathbb{N} \cup \{0\}$, $i \in [d]$.

To prove Theorem 2.16, we will apply Itô's formula to the semimartingale $F(\mu_t)$, $t \geq 0$. Hence, we will need an analog of Itô's formula for functions of the form

(2.16) $$G(\mu) = f(\langle \varphi_1, \mu \rangle, \dots, \langle \varphi_m, \mu \rangle), \quad \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

for some smooth functions f and φ_l, $l \in m$, involving functional derivatives [21, Section 2]

$$\frac{\delta G(\mu)}{\delta \mu}(x) = \frac{\partial}{\partial \varepsilon} G(\mu + \varepsilon \delta_x)|_{\varepsilon=0}$$

and

$$\frac{\delta^2 G(\mu)}{\delta \mu^2}(x, y) = \frac{\partial^2}{\partial \varepsilon_1 \partial \varepsilon_2} G(\mu + \varepsilon_1 \delta_x + \varepsilon_2 \delta_y)|_{\varepsilon_1=\varepsilon_2=0}$$

in order to simplify computations.

We will often work with the linear stochastic mean-field equation obtained by freezing a solution μ_t, $t \geq 0$, in coefficients. In particular, we will write $v(t, x)$, $a(t, x)$ and $g(t, x, \theta)$ for $V(t, x, \mu_t)$, $A(t, x, \mu_t)$ and $G(t, x, \mu_t, \theta)$.

Proposition 2.18. *Let V, A, G satisfy Assumptions 2.1, 2.7, $p \geq 2$ and ν_t, $t \geq 0$, be a continuous $(\mathcal{F}_t)$-adapted process in $\mathcal{P}_2(\mathbb{R}^d)$ satisfying for every $\varphi \in \mathrm{C}_c^2(\mathbb{R}^d)$*

(2.17)
$$\langle \varphi, \nu_t \rangle = \langle \varphi, \nu_0 \rangle + \int_0^t \langle \nabla \varphi \cdot v(s, \cdot), \nu_s \rangle ds + \frac{1}{2} \int_0^t \langle D^2 \varphi : a(s, \cdot), \nu_s \rangle ds$$
$$+ \int_0^t \int_\Theta \langle \nabla \varphi \cdot g(s, \cdot, \theta), \nu_s \rangle W(d\theta, ds).$$

If there exist an $(\mathcal{F}_t)$-stopping time σ such that $\sup_{t \in [0, \sigma]} \langle \phi_p, \nu_t \rangle < \infty$ a.s., then

(2.18)
$$G(\nu_{t \wedge \sigma}) = G(\nu_0) + \int_0^{t \wedge \sigma} \left\langle \nabla \frac{\delta G(\nu_s)}{\delta \nu_s} \cdot v(s, \cdot), \nu_s \right\rangle ds$$
$$+ \frac{1}{2} \int_0^{t \wedge \sigma} \left\langle D^2 \frac{\delta G(\nu_s)}{\delta \nu_s} : a(s, \cdot), \nu_s \right\rangle ds$$
$$+ \frac{1}{2} \int_0^{t \wedge \sigma} \left\langle \nabla \otimes \nabla \frac{\delta^2 G(\nu_s)}{\delta \nu_s^2} : \tilde{a}(s, \cdot), \nu_s \otimes \nu_s \right\rangle ds$$
$$+ \int_0^{t \wedge \sigma} \int_\Theta \left\langle \nabla \frac{\delta G(\nu_s)}{\delta \nu_s} \cdot g(s, \cdot, \theta), \nu_s \right\rangle W(d\theta, ds),$$

for each function G defined by (2.16) with $\varphi_l \in \mathrm{C}_p^2(\mathbb{R}^d)$, $l \in [m]$, and $f \in \mathrm{C}^2(\mathbb{R}^m)$, where $\tilde{a}_{i,j}(t, x, y) = \langle g_i(t, x, \cdot), g_j(t, y, \cdot) \rangle_\vartheta$, $i, j \in [d]$.

Proof. We will first show that equality (2.17) holds for every $\varphi \in C^2_p(\mathbb{R}^d)$ and t replaced by $t \wedge \sigma$. We fix $g \in C^2_b(\mathbb{R})$ such that $g(x) = 0$, $x \leq 0$, and $g(x) = 1$, $x \geq 1$. Set for $n \geq 1$

$$\kappa_n(x) := \begin{cases} g\left(n - \ln|x|\right), & \text{if } x \neq 0, \\ 1, & \text{if } x = 0, \end{cases} \quad x \in \mathbb{R}^d.$$

It is easy to see that for each $x \in \mathbb{R}^d$ $\kappa_n(x) \to 1$ as $n \to \infty$ and κ_n, $n \geq 1$, belongs uniformly to $C^2_0(\mathbb{R}^d)$, that is, there exists a constant $C > 0$ such that

$$|\kappa_n(x)| + (1 + |x|)|\nabla\kappa_n(x)| + (1 + |x|^2)|D^2\kappa_n(x)| \leq C$$

for all $x \in \mathbb{R}^d$ and $n \geq 1$. Let $\varphi \in C^2_p(\mathbb{R}^d)$ and $\varphi_n = \varphi\kappa_n$, $n \geq 1$. Note that $\varphi_n \in C^2_c(\mathbb{R}^d)$ for all $n \geq 1$. Therefore, for every $t \geq 0$

$$\langle \varphi_n, \nu_t \rangle = \langle \varphi_n, \nu_0 \rangle + \int_0^t \langle \nabla\varphi_n \cdot v(s, \cdot), \nu_s \rangle ds + \frac{1}{2} \int_0^t \langle D^2\varphi_n : a(s, \cdot), \nu_s \rangle ds$$

$$+ \int_0^t \int_\Theta \langle \nabla\varphi_n \cdot g(s, \cdot, \theta), \nu_s \rangle W(d\theta, ds).$$

We will next pass to the limit as $n \to \infty$. A simple computation gives that $\varphi_n(x) \to \varphi(x)$, $\nabla\varphi_n(x) \to \nabla\varphi(x)$ and $D^2\varphi_n(x) \to D^2\varphi(x)$ as $n \to \infty$ for all $x \in \mathbb{R}^d$. Define for $k \geq 1$

$$\sigma_k = \inf\left\{ t \geq 0 : \langle \phi_p, \nu_t \rangle > k, \, \mathcal{W}_2(\nu_t, \delta_0) > k \right\} \wedge \sigma.$$

Since the set $\{\mu \in \mathcal{P}_2(\mathbb{R}^d) : \langle \phi_p, \mu \rangle \leq k\}$ is closed in $\mathcal{P}_2(\mathbb{R}^d)$ and ν_t, $t \geq 0$, is a continuous process, σ_k is an $(\mathcal{F}_t)$-stopping time, by [32, Proposition 2.1.5 (a)]. Let $k \geq 1$ and $t \geq 0$ be fixed. Since $\sup_{s \in [0, \sigma_k]} \langle \phi_p, \nu_s \rangle \leq k$ and $\varphi_n(x) \leq C(1 + |x|^p)$, $x \in \mathbb{R}^d$, $n \geq 1$, for some $C > 0$, the dominated convergence theorem implies that

$$\langle \varphi_n, \nu_{t \wedge \sigma_k} \rangle \to \langle \varphi, \nu_{t \wedge \sigma_k} \rangle \quad \text{a.s.}$$

as $n \to \infty$. By Remark 2.8, for every $x \in \mathbb{R}^d$ and $s \in [0, t \wedge \sigma_k]$ one has

$$\begin{aligned}
(2.19) \quad |\nabla\varphi_n \cdot v(s, x)| &\leq (|\nabla\varphi(x)||\kappa_n(x)| + |\varphi(x)||\nabla\kappa_n(x)|) |V(s, x, \mu_s)| \\
&\leq L\left(|\nabla\varphi(x)||\kappa_n(x)| + |\varphi(x)||\nabla\kappa_n(x)|\right) (1 + |x| + \mathcal{W}_2(\mu_s, \delta_0)) \\
&\leq L\left(|\nabla\varphi(x)||\kappa_n(x)| + |\varphi(x)||\nabla\kappa_n(x)|\right) (1 + |x| + k) \\
&\leq C_1(1 + |x|^p),
\end{aligned}$$

where C_1 is a constant which is independent of s and x. Since $\int_0^{t \wedge \sigma_k} C_1(1 + \langle \phi_p, \nu_s \rangle) ds \leq C_1(1 + k)t < \infty$, the dominated convergence theorem implies

$$\int_0^{t \wedge \sigma_k} \langle \nabla\varphi_n \cdot v(s, \cdot), \nu_s \rangle \, ds \to \int_0^{t \wedge \sigma_k} \langle \nabla\varphi \cdot v(s, \cdot), \nu_s \rangle \, ds \quad \text{a.s.}$$

as $n \to \infty$. Similarly, by Remark 2.8, for every $s \in [0, t \wedge \sigma_k]$

$$\begin{aligned}
|D^2\varphi_n : a(s, x)| &\leq |\kappa_n(x)D^2\varphi(x) : a(s, x)| + |(\nabla\varphi(x) \otimes \nabla\kappa_n(x)) : a(s, x)| \\
&\quad + |(\nabla\kappa_n(x) \otimes \nabla\varphi(x)) : a(s, x)| + |\varphi(x)D^2\kappa_n(x) : a(s, x)| \\
&\leq \left(|\kappa_n(x)||D^2\varphi(x)| + 2|\nabla\varphi(x)||\nabla\kappa_n(x)| + |\varphi(x)||D^2\kappa_n(x)|\right) |a(s, x)| \\
&\leq \left(|\kappa_n(x)||D^2\varphi(x)| + 2|\nabla\varphi(x)||\nabla\kappa_n(x)| \right. \\
&\quad \left. + C|\varphi(x)||D^2\kappa_n(x)|\right) (1 + |x| + \mathcal{W}_2(\mu_s, \delta_0))^2 \leq C_1(1 + |x|^p),
\end{aligned}$$

where C_1 is also a constant independent of x and s. Therefore, using the dominated convergence theorem again, we get

$$\int_0^{t \wedge \sigma_k} \left\langle D^2 \varphi_n : a(x, \cdot), \nu_s \right\rangle ds \to \int_0^{t \wedge \sigma_k} \left\langle D^2 \varphi : a(x, \cdot), \nu_s \right\rangle ds \quad \text{a.s.}$$

as $n \to \infty$. It remains only to show the convergence of the stochastic integrals. We consider

$$\mathbb{E} \left(\int_0^{t \wedge \sigma_k} \left\langle (\nabla \varphi - \nabla \varphi_n) \cdot g(s, \cdot, \theta), \nu_s \right\rangle W(d\theta, ds) \right)^2$$

$$= \mathbb{E} \int_0^{t \wedge \sigma_k} \int_\Theta \left\langle (\nabla \varphi - \nabla \varphi_n) \cdot g(s, \cdot, \theta), \nu_s \right\rangle^2 \vartheta(d\theta) ds$$

$$\leq \mathbb{E} \int_0^{t \wedge \sigma_k} \int_\Theta \left\langle |\nabla \varphi - \nabla \varphi_n|^2 |g(s, \cdot, \theta)|^2, \nu_s \right\rangle \vartheta(d\theta) ds$$

$$\leq \mathbb{E} \int_0^{t \wedge \sigma_k} \left\langle |\nabla \varphi - \nabla \varphi_n|^2 \|g(s, \cdot, \theta)\|_\vartheta^2, \nu_s \right\rangle ds \to 0$$

as $n \to \infty$. In the last step, we have used the dominated convergence theorem, since $|\nabla \varphi - \nabla \varphi_n|^2 \|g(s, x)\|_\vartheta^2 \leq C(1 + |x|^p)$ on $[0, t \wedge \sigma_k]$ with a (non-random) constant C that does not depend on s and x, where the estimate can be obtained in a similar manner to (2.19). Consequently,

$$\int_0^{t \wedge \sigma_k} \int_\Theta \left\langle \nabla \varphi_n \cdot g(s, \cdot, \theta), \nu_s \right\rangle W(d\theta, ds) \to \int_0^{t \wedge \sigma_k} \int_\Theta \left\langle \nabla \varphi \cdot g(s, \cdot, \theta), \nu_s \right\rangle W(d\theta, ds)$$

in $L_2(\Omega)$ as $n \to \infty$.

Summarizing obtained convergence results, we can conclude that for each $t \geq 0$ and $k \geq 0$ we have a.s.

$$\langle \varphi, \nu_{t \wedge \sigma_k} \rangle = \langle \varphi, \nu_0 \rangle + \int_0^{t \wedge \sigma_k} \langle \nabla \varphi \cdot v(s, \cdot), \nu_s \rangle ds + \frac{1}{2} \int_0^{t \wedge \sigma_k} \langle D^2 \varphi : a(s, \cdot), \nu_s \rangle ds$$

$$+ \int_0^{t \wedge \sigma_k} \int_\Theta \langle \nabla \varphi \cdot g(s, \cdot, \theta), \nu_s \rangle W(d\theta, ds).$$

Now we pass to the limit as $k \to \infty$. Remark that the map $t \mapsto \langle \varphi, \nu_t \rangle$ is not continuous a.s. in the Euclidean topology on $[0, \infty)$ because φ is not bounded in general. However, $\sigma_k \to \sigma$ in the discrete topology as $k \to \infty$, i.e. a.s. there exists $\tilde{k}$ such that $\sigma_k = \sigma$ for all $k \geq \tilde{k}$. This allows to pass to the limit as $k \to \infty$. This gives (2.17) with t replaced by $t \wedge \sigma$. We can conclude from this equality that the process $\langle \varphi, \nu_{t \wedge \sigma} \rangle$ has a continuous version and is an $(\mathcal{F}_t)$-semimartingale.

We next set $\langle \varphi, \nu_t \rangle = (\langle \varphi_1, \nu_t \rangle, \ldots, \langle \varphi_m, \nu_t \rangle)$. Applying Itô's formula, one obtains

$$G(\nu_{t \wedge \sigma}) = f \left(\langle \varphi_1, \nu_{t \wedge \sigma} \rangle, \ldots, \langle \varphi_m, \nu_{t \wedge \sigma} \rangle \right) = f \left(\langle \varphi, \nu_0 \rangle \right)$$

$$= \sum_{l=1}^m \int_0^{t \wedge \sigma} \partial_l f(\langle \varphi, \nu_s \rangle) d\langle \varphi_l, \nu_s \rangle$$

$$+ \frac{1}{2} \sum_{l,k=1}^m \int_0^{t \wedge \sigma} \partial_l \partial_k f(\langle \varphi, \nu_s \rangle) d\left[\langle \varphi_l, \nu_\cdot \rangle, \langle \varphi_k, \nu_\cdot \rangle \right]_s$$

$$= f \left(\langle \varphi, \nu_0 \rangle \right) + \sum_{l=1}^m \int_0^{t \wedge \sigma} \partial_l f(\langle \varphi, \nu_s \rangle) \langle \nabla \varphi_l \cdot v(s, \cdot), \nu_s \rangle ds$$

$$+ \frac{1}{2} \sum_{l=1}^m \int_0^{t \wedge \sigma} \partial_l f(\langle \varphi, \nu_s \rangle) \langle D^2 \varphi_l : a(s, \cdot), \nu_s \rangle ds$$

$$+ \sum_{l=1}^{m} \int_{0}^{t \wedge \sigma} \int_{\Theta} \partial_l f(\langle \varphi, \nu_s \rangle) \langle \nabla \varphi_l \cdot g(s, \cdot, \theta), \nu_s \rangle W(d\theta, ds)$$

$$+ \frac{1}{2} \sum_{l,k=1}^{m} \int_{0}^{t \wedge \sigma} \int_{\Theta} \partial_l \partial_k f(\langle \varphi, \nu_s \rangle) \langle \nabla \varphi_l \cdot g_l(s, \cdot, \theta), \nu_s \rangle \langle \nabla \varphi_k \cdot g_k(s, \cdot, \theta), \nu_s \rangle \vartheta(d\theta) ds.$$

Using the equalities

$$\frac{\delta G(\mu)}{\delta \mu}(x) = \sum_{l=1}^{m} \partial_l f(\langle \varphi, \mu \rangle) \varphi_l(x)$$

and

$$\frac{\delta^2 G(\mu)}{\delta \mu^2}(x, y) = \sum_{l,k=1}^{m} \partial_l \partial_k f(\langle \varphi, \mu \rangle) \varphi_l(x) \varphi_k(y),$$

we get (2.18) that completes the proof of the proposition. □

Remark 2.19. One can extend the obtained Itô's formula to any, e.g., bounded twice continuously differentiable function G on $\mathcal{P}_2(\mathbb{R}^d)$, using an approximation analog to Bernstein polynomials similarly as it was done in the proof of [51, Theorem 2]. We do not consider this extension here since the obtained Itô formula is needed only for the proof of Theorem 2.16, where it will be applied to the function F_n, that is defined by (2.14) and satisfies the assumptions of Proposition 2.18 .

Corollary 2.20. *Let μ_t, $t \geq 0$, be a solution to the stochastic mean-field equation (2.1) whose coefficients satisfy Assumptions 2.1 and 2.7. Then equality (2.2) holds for every $\varphi \in \mathrm{C}_b^2(\mathbb{R}^d)$.*

Proof. We note that μ_t, $t \geq 0$, is a continuous process in $\mathcal{P}_2(\mathbb{R}^d)$. Hence, $\sup_{s \in [0,t]} \langle \phi_2, \mu_s \rangle < \infty$ a.s. for all $t \geq 0$. Consequently, the corollary directly follows from Proposition 2.18 with $f(x) = x$, $x \in \mathbb{R}$, and the inclusion $\mathrm{C}_b^2(\mathbb{R}^d) \subset \mathrm{C}_2^2(\mathbb{R}^d)$. □

The following lemma gives the key property of solutions to the stochastic mean-field equation started from atomic initial condition that allows to prove their uniqueness.

Lemma 2.21. *Let V, A, G satisfy Assumptions 2.1, 2.7, $n \geq 2$, and ν_t, $t \geq 0$, be defined in Proposition 2.18 with $\mu_0 \in \mathcal{N}_{n-1}$. Assume that there exists an $(\mathcal{F}_t)$-stopping time σ such that $\sup_{t \in [0,\sigma]} \langle \phi_p, \nu_t \rangle < \infty$ a.s. for $p = n^2 - n$. Then a.s. $\nu_{t \wedge \sigma} \in \mathcal{N}_{n-1}$, $t \geq 0$.*

Proof. To prove the lemma, it is enough to show that a.s. $F_n(\nu_{t \wedge \sigma}) = 0$, $t \geq 0$, by observation (2.15). For this we will apply Itô's formula to the function $F := F_n$, which satisfies assumptions of Proposition 2.18, according to Remark 2.17. We first compute the derivatives of F which appear in (2.18). For $x \in \mathbb{R}^d$ we have

$$\frac{\delta F(\mu)}{\delta \mu}(x) = \frac{\partial}{\partial \varepsilon} \int_{\mathbb{R}^{dn}} \prod_{i,j:i \neq j} |z^i - z^j| \prod_{i=1}^{n} \left(\mu(dz^i) + \varepsilon \delta_x \right) \Big|_{\varepsilon=0}$$

$$= \sum_{i_1=1}^{n} \int_{\mathbb{R}^{dn}} \prod_{i,j:i \neq j} |z^i - z^j| \delta_x(dz^{i_1}) \prod_{i \neq i_1} \mu(dz^i).$$

Writing $\{i,j\} \neq \{i_1, i_2\}$ for the set $\{(i,j) \in [n]^2 : i \neq j, (i,j) \notin \{(i_1, i_2), (i_2, i_1)\}\}$, we get for $k, l \in [d]$

$$\frac{\partial}{\partial x_k} \frac{\delta F(\mu)}{\delta \mu}(x) = 2 \sum_{i_1, i_2: i_1 \neq i_2} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_2}) \prod_{\{i,j\} \neq \{i_1, i_2\}} |z^i - z^j| \delta_x(z^{i_1}) \prod_{i: i \neq i_1} \mu(dz^i)$$

and

$$\frac{\partial^2}{\partial x_k \partial x_l} \frac{\delta F(\mu)}{\delta \mu}(x) = 4 \sum_{\substack{i_1, i_2, i_3: \\ i_1 \neq i_2 \neq i_3}} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_2})(z_l^{i_1} - z_l^{i_3})$$

$$\prod_{\substack{\{i,j\} \neq \{i_1, i_2\} \\ \{i,j\} \neq \{i_1, i_3\}}} |z^i - z^j| \delta_x(z^{i_1}) \prod_{i: i \neq i_1} \mu(dz^i)$$

$$+ 2\delta_{kl} \sum_{i_1, i_2: i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \prod_{\{i,j\} \neq \{i_1, i_2\}} |z^i - z^j| \delta_x(z^{i_1}) \prod_{i: i \neq i_1} \mu(dz^i)$$

$$=: I^{k,l}(x, \mu) + \delta_{kl} \tilde{I}^k(x, \mu),$$

where $\delta_{kl} = \mathbb{I}_{\{k\}}(l)$ denotes the Kronecker delta. Similarly, for $x, y \in \mathbb{R}^d$ and $k, l \in [d]$

$$\frac{\delta^2 F}{\delta \mu^2}(x, y) = \sum_{i_1, i_2: i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \prod_{i \neq j} |z^i - z^j| \delta_x(dz^{i_1}) \delta_y(dz^{i_2}) \prod_{i \neq i_1, i_2} \mu(dz^i),$$

$$\frac{\partial}{\partial x_k} \frac{\delta^2 F}{\delta \mu^2}(x, y) = 2 \sum_{i_1, i_2: i_1 \neq i_2} \sum_{i_3: i_3 \neq i_1} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_3})$$

$$\prod_{\{i,j\} \neq \{i_1, i_3\}} |z^i - z^j| \delta_x(dz^{i_1}) \delta_y(dz^{i_2}) \prod_{i: i \neq i_1, i_2} \mu(dz^i)$$

and

$$\frac{\partial^2}{\partial x_k \partial y_l} \frac{\delta^2 F}{\delta \mu^2}(x, y) = 4 \sum_{i_1, i_2: i_1 \neq i_2} \sum_{i_3: i_3 \neq i_1} \sum_{i_4: i_4 \neq i_1, i_2} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_3})(z_l^{i_2} - z_l^{i_4})$$

$$\prod_{\substack{\{i,j\} \neq \{i_1, i_3\} \\ \{i,j\} \neq \{i_2, i_4\}}} |z^i - z^j| \delta_x(dz^{i_1}) \delta_y(dz^{i_2}) \prod_{i: i \neq i_1, i_2} \mu(dz^i)$$

$$+ 4 \sum_{i_1, i_2: i_1 \neq i_2} \sum_{i_3: i_3 \neq i_1, i_2} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_3})(z_l^{i_2} - z_l^{i_1})$$

$$\prod_{\substack{\{i,j\} \neq \{i_1, i_3\} \\ \{i,j\} \neq \{i_2, i_1\}}} |z^i - z^j| \delta_x(dz^{i_1}) \delta_y(dz^{i_2}) \prod_{i: i \neq i_1, i_2} \mu(dz^i)$$

$$- 2\delta_{kl} \sum_{i_1, i_2: i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \prod_{\{i,j\} \neq \{i_1, i_2\}} |z^i - z^j| \delta_x(dz^{i_1}) \delta_y(dz^{i_2}) \prod_{i: i \neq i_1, i_2} \mu(dz^i)$$

$$=: J_1^{k,l}(x, y, \mu) + J_2^{k,l}(x, y, \mu) + \delta_{kl} \tilde{J}^k(x, y, \mu).$$

We next estimate the terms which appear after applying Itô's formula from Proposition 2.18 to the semimartingale $F(\nu_t)$, $t \geq 0$. For the first term, we get

$$\left\langle \nabla \frac{\delta F(\nu_s)}{\delta \nu_s} \cdot v(s, \cdot), \nu_s \right\rangle = 2 \sum_{i_1, i_2 : i_1 \neq i_2} \int_{\mathbb{R}^{dn}} (z^{i_1} - z^{i_2}) \cdot v(s, z^{i_1}) \prod_{\{i,j\} \neq \{i_1, i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i).$$

Interchanging i_1 and i_2, the expression above can be rewritten as

$$\sum_{i_1, i_2 : i_1 \neq i_2} \int_{\mathbb{R}^{dn}} (z^{i_1} - z^{i_2}) \cdot v(s, z^{i_1}) \prod_{\{i,j\} \neq \{i_1, i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i)$$

$$+ \sum_{i_1, i_2 : i_1 \neq i_2} \int_{\mathbb{R}^{dn}} (z^{i_2} - z^{i_1}) \cdot v(s, z^{i_2}) \prod_{\{i,j\} \neq \{i_2, i_1\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i)$$

$$= \sum_{i_1, i_2 : i_1 \neq i_2} \int_{\mathbb{R}^{dn}} (z^{i_1} - z^{i_2}) \cdot \left(v(s, z^{i_1}) - v(s, z^{i_2}) \right) \prod_{\{i,j\} \neq \{i_1, i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i).$$

Therefore, using the Lipschitz continuity of $v(s, \cdot)$ from Assumption 2.7, we can estimate

$$\left| \left\langle \nabla \frac{\delta F(\nu_s)}{\delta \nu_s} \cdot v_s, \nu_s \right\rangle \right| \leq L \sum_{i_1, i_2 : i_1 \neq i_2} \int_{\mathbb{R}^{dn}} |z^{i_1} - z^{i_2}|^2 \prod_{\{i,j\} \neq \{i_1, i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i)$$

$$= L \sum_{i_1, i_2 : i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \prod_{i,j : i \neq j} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i) = Ln(n-1)F(\nu_s).$$

In order to estimate

$$K := \left\langle D^2 \frac{\delta F(\nu_s)}{\delta \nu_s} : a(s, \cdot), \nu_s \right\rangle + \left\langle \nabla \otimes \nabla \frac{\delta^2 F(\nu_s)}{\delta \nu_s^2} : \tilde{a}(s, \cdot), \nu_s \otimes \nu_s \right\rangle,$$

we first split $J_1^{k,l}$ into three terms $J_{1,1}^{k,l}, J_{1,2}^{k,l}, J_{1,3}^{k,l}$ defined by

$$J_{1,1}^{k,l}(x, y, \mu) = 4 \sum_{\substack{i_1, i_2, i_3, i_4 : \\ i_1 \neq i_2 \neq i_3 \neq i_4}} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_3})(z_l^{i_2} - z_l^{i_4})$$

$$\prod_{\substack{\{i,j\} \neq \{i_1, i_3\} \\ \{i,j\} \neq \{i_2, i_4\}}} |z^i - z^j| \delta_x(dz^{i_1}) \delta_y(dz^{i_2}) \prod_{i : i \neq i_1, i_2} \mu(dz^i),$$

$$J_{1,2}^{k,l}(x, y, \mu) = 4 \sum_{\substack{i_1, i_2, i_4 : \\ i_1 \neq i_2 \neq i_4}} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_2})(z_l^{i_2} - z_l^{i_4})$$

$$\prod_{\substack{\{i,j\} \neq \{i_1, i_2\} \\ \{i,j\} \neq \{i_2, i_4\}}} |z^i - z^j| \delta_x(dz^{i_1}) \delta_y(dz^{i_2}) \prod_{i : i \neq i_1, i_2} \mu(dz^i)$$

and

$$J_{1,3}^{k,l}(x, y, \mu) = 4 \sum_{\substack{i_1, i_2, i_4 : \\ i_1 \neq i_2 \neq i_4}} \int_{\mathbb{R}^{dn}} (z_k^{i_1} - z_k^{i_4})(z_l^{i_2} - z_l^{i_4})$$

$$\prod_{\substack{\{i,j\}\neq\{i_1,i_4\}\\\{i,j\}\neq\{i_2,i_4\}}}|z^i-z^j|\delta_x(dz^{i_1})\delta_y(dz^{i_2})\prod_{i:i\neq i_1,i_2}\mu(dz^i),$$

which are obtained by summing over $\{i_3 : i_3 \neq i_2, i_4\}$, $\{i_3 : i_3 = i_2, i_3 \neq i_4\}$ and $\{i_3 : i_3 = i_4, i_3 \neq i_2\}$ in the second sum, respectively. We remark that $J_{1,1}^{k,l}$ only appears for $n \geq 4$ and $J_{1,2}^{k,l}$, $J_{1,3}^{k,l}$ appear for $n \geq 3$. We rewrite K as the sum $K_1 + K_2 + K_3$, where

$$K_1 = \langle J_{1,1}(\nu_s) : \tilde{a}(s,\cdot), \nu_s \otimes \nu_s \rangle,$$
$$K_2 = \langle I(\nu_s) : a(s,\cdot), \nu_s \rangle + \langle J_{1,2}(\nu_s) : \tilde{a}(s,\cdot), \nu_s \otimes \nu_s \rangle$$
$$+ \langle J_{1,3}(\nu_s) : \tilde{a}(s,\cdot), \nu_s \otimes \nu_s \rangle + \langle J_2(\nu_s) : \tilde{a}(s,\cdot), \nu_s \otimes \nu_s \rangle$$

and

$$K_3 = \left\langle \tilde{I}(\nu_s) \cdot b(s,\cdot), \nu_s \right\rangle + \left\langle \tilde{J}(\nu_s) : \tilde{b}(s,\cdot), \nu_s \otimes \nu_s \right\rangle.$$

In the above equality the functions b and $\tilde{b}$ are defined by $b_k(s,\cdot) = a_{k,k}(s,\cdot)$ and $\tilde{b}_k(s,\cdot) = \tilde{a}_{k,k}(s,\cdot)$, $k \in [d]$. We next compute

$$K_1 = 4 \sum_{\substack{i_1,i_2,i_3,i_4:\\i_1\neq i_2\neq i_3\neq i_4}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1}-z^{i_3})\otimes(z^{i_2}-z^{i_4})\right]:\tilde{a}(s,z^{i_1},z^{i_2}) \prod_{\substack{\{i,j\}\neq\{i_1,i_3\}\\\{i,j\}\neq\{i_2,i_4\}}}|z^i-z^j|\prod_{i=1}^{n}\nu_s(dz^i)$$

$$= \sum_{\substack{i_1,i_2,i_3,i_4:\\i_1\neq i_2\neq i_3\neq i_4}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1}-z^{i_3})\otimes(z^{i_2}-z^{i_4})\right]$$

$$: \left\langle g(s,z^{i_1},\cdot)-g(s,z^{i_3},\cdot), g(s,z^{i_2},\cdot)-g(s,z^{i_4},\cdot)\right\rangle_\vartheta \prod_{\substack{\{i,j\}\neq\{i_1,i_3\}\\\{i,j\}\neq\{i_2,i_4\}}}|z^i-z^j|\prod_{i=1}^{n}\nu_s(dz^i)$$

$$= \sum_{\substack{i_1,i_2,i_3,i_4:\\i_1\neq i_2\neq i_3\neq i_4}} \int_{\mathbb{R}^{dn}} \left\langle (z^{i_1}-z^{i_3})\cdot(g(s,z^{i_1},\cdot)-g(s,z^{i_3},\cdot)),\right.$$

$$\left.(z^{i_2}-z^{i_4})\cdot(g(s,z^{i_2},\cdot)-g(s,z^{i_4},\cdot))\right\rangle_\vartheta \prod_{\substack{\{i,j\}\neq\{i_1,i_3\}\\\{i,j\}\neq\{i_2,i_4\}}}|z^i-z^j|\prod_{i=1}^{n}\nu_s(dz^i).$$

Therefore, using the Cauchy-Schwarz inequality and then the Lipschitz continuity of $g(s,\cdot)$, we estimate

$$|K_1| \leq \sum_{\substack{i_1,i_2,i_3,i_4:\\i_1\neq i_2\neq i_3\neq i_4}} \int_{\mathbb{R}^{dn}} |z^{i_1}-z^{i_3}|\|g(s,z^{i_1},\cdot)-g(s,z^{i_3},\cdot)\|_\vartheta|z^{i_2}-z^{i_4}|\|g(s,z^{i_2},\cdot)-g(s,z^{i_4},\cdot)\|_\vartheta$$

$$\prod_{\substack{\{i,j\}\neq\{i_1,i_3\}\\\{i,j\}\neq\{i_2,i_4\}}}|z^i-z^j|\prod_{i=1}^{n}\nu_s(dz^i)$$

$$\leq L^2 \sum_{\substack{i_1,i_2,i_3,i_4:\\i_1\neq i_2\neq i_3\neq i_4}} \int_{\mathbb{R}^{dn}} |z^{i_1}-z^{i_3}|^2|z^{i_2}-z^{i_4}|^2 \prod_{\substack{\{i,j\}\neq\{i_1,i_3\}\\\{i,j\}\neq\{i_2,i_4\}}}|z^i-z^j|\prod_{i=1}^{n}\nu_s(dz^i)$$

$$= L^2 \sum_{\substack{i_1,i_2,i_3,i_4: \\ i_1 \neq i_2 \neq i_3 \neq i_4}} \int_{\mathbb{R}^{dn}} \prod_{i,j:i \neq j} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i) = L^2 \frac{n!}{(n-4)!} F(\nu_s).$$

We rewrite K_2 in similar way as K_1:

$$K_2 = 4 \sum_{\substack{i_1,i_2,i_3: \\ i_1 \neq i_2 \neq i_3}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1} - z^{i_2}) \otimes (z^{i_1} - z^{i_3}) \right] : a(s, z^{i_1}) \prod_{\substack{\{i,j\} \neq \{i_1,i_2\} \\ \{i,j\} \neq \{i_1,i_3\}}} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i)$$

$$+ 4 \sum_{\substack{i_1,i_2,i_4: \\ i_1 \neq i_2 \neq i_4}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1} - z^{i_2}) \otimes (z^{i_2} - z^{i_4}) \right] : \tilde{a}(s, z^{i_1}, z^{i_2}) \prod_{\substack{\{i,j\} \neq \{i_1,i_2\} \\ \{i,j\} \neq \{i_2,i_4\}}} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i)$$

$$+ 4 \sum_{\substack{i_1,i_2,i_4: \\ i_1 \neq i_2 \neq i_4}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1} - z^{i_4}) \otimes (z^{i_2} - z^{i_4}) \right] : \tilde{a}(s, z^{i_1}, z^{i_2}) \prod_{\substack{\{i,j\} \neq \{i_1,i_4\} \\ \{i,j\} \neq \{i_2,i_4\}}} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i)$$

$$+ 4 \sum_{\substack{i_1,i_2,i_3: \\ i_1 \neq i_2 \neq i_3}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1} - z^{i_3}) \otimes (z^{i_2} - z^{i_1}) \right] : \tilde{a}(s, z^{i_1}, z^{i_2}) \prod_{\substack{\{i,j\} \neq \{i_1,i_3\} \\ \{i,j\} \neq \{i_2,i_1\}}} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i).$$

Interchanging the indexes of summations in second, third and fourth terms in the following way $[i_1 \mapsto i_2, i_2 \mapsto i_1, i_4 \mapsto i_3]$, $[i_1 \mapsto i_2, i_2 \mapsto i_3, i_4 \mapsto i_1]$ and $[i_2 \mapsto i_3, i_3 \mapsto i_2]$, respectively, and using the equalities $a(s,x) = \langle g(s,x,\cdot), g(s,x,\cdot) \rangle_\vartheta$, $\tilde{a}(s,x,y) = \langle g(s,x,\cdot), g(s,y,\cdot) \rangle$, we get

$$K_2 = 4 \sum_{\substack{i_1,i_2,i_3: \\ i_1 \neq i_2 \neq i_3}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1} - z^{i_2}) \otimes (z^{i_3} - z^{i_1}) \right] :$$

$$\left[- \langle g(s, z^{i_1}, \cdot), g(s, z^{i_1}, \cdot) \rangle_\vartheta + \langle g(s, z^{i_2}, \cdot), g(s, z^{i_1}, \cdot) \rangle_\vartheta \right.$$

$$\left. - \langle g(s, z^{i_2}, \cdot), g(s, z^{i_3}, \cdot) \rangle_\vartheta + \langle g(s, z^{i_1}, \cdot), g(s, z^{i_3}, \cdot) \rangle_\vartheta \right] \prod_{\substack{\{i,j\} \neq \{i_1,i_2\} \\ \{i,j\} \neq \{i_1,i_3\}}} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i)$$

$$= 4 \sum_{\substack{i_1,i_2,i_3: \\ i_1 \neq i_2 \neq i_3}} \int_{\mathbb{R}^{dn}} \left[(z^{i_1} - z^{i_2}) \otimes (z^{i_3} - z^{i_1}) \right]$$

$$: \langle g(s, z^{i_1}, \cdot) - g(s, z^{i_2}, \cdot), g(s, z^{i_3}, \cdot) - g(s, z^{i_1}, \cdot) \rangle_\vartheta \prod_{\substack{\{i,j\} \neq \{i_1,i_2\} \\ \{i,j\} \neq \{i_1,i_3\}}} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i).$$

Similarly as before, we get

$$K_2 \leq 4L^2 \frac{n!}{(n-3)!} F(\nu_s).$$

We now compute

$$K_3 = 2 \sum_{i_1,i_2:i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \left[\sum_{k=1}^d a_{k,k}(s, z^{i_1}) \right] \prod_{\{i,j\} \neq \{i_1,i_2\}} |z^i - z^j| \prod_{i=1}^n \nu_s(dz^i)$$

$$- 2 \sum_{i_1,i_2:i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \left[\sum_{k=1}^{d} \tilde{a}_{k,k}(s, z^{i_1}, z^{i_2}) \right] \prod_{\{i,j\} \neq \{i_1,i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i)$$

$$= \sum_{i_1,i_2:i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \left[\sum_{k=1}^{d} \left(\langle g_k(s, z^{i_1}, \cdot), g_k(s, z^{i_1}, \cdot) \rangle_\vartheta + \langle g_k(s, z^{i_2}, \cdot), g_k(s, z^{i_2}, \cdot) \rangle_\vartheta \right. \right.$$

$$\left. \left. - 2 \langle g_k(s, z^{i_1}, \cdot), g_k(s, z^{i_2}, \cdot) \rangle_\vartheta \right) \right] \prod_{\{i,j\} \neq \{i_1,i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i)$$

$$= \sum_{i_1,i_2:i_1 \neq i_2} \int_{\mathbb{R}^{dn}} \sum_{k=1}^{d} \|g_k(s, z^{i_1}, \cdot) - g_k(s, z^{i_2}, \cdot)\|_\vartheta^2 \prod_{\{i,j\} \neq \{i_1,i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i).$$

Hence, using the Lipschitz continuity of $g(s, \cdot)$ again, we get

$$|K_3| \leq L^2 \sum_{i_1,i_2:i_1 \neq i_2} \int_{\mathbb{R}^{dn}} |z^{i_1} - z^{i_2}|^2 \prod_{\{i,j\} \neq \{i_1,i_2\}} |z^i - z^j| \prod_{i=1}^{n} \nu_s(dz^i) = L^2 n(n-1) F(\nu_s).$$

Combining obtained estimates for K_i, $i \in [3]$, we can see that $|K| \leq C_n L^2 F(\nu_s)$.

We next define for $p = n^2 - n$ and every $k \geq 1$ the $(\mathcal{F}_t)$-stopping time σ_k as follows

$$\sigma_k = \inf \{t \geq 0 : \langle \phi_p, \nu_t \rangle \geq k\} \wedge \sigma.$$

Then there exists a (non-random) constant $C > 0$ such that $F(\mu_t) \leq C$ for all $t \in [0, \sigma_k]$, by Remark 2.17. Applying Itô's formula to $F(\mu_t \wedge \sigma_k)$, $t \geq 0$, taking the expectations and using estimates obtained above, we have for every $k \geq 1$

$$\mathbb{E} F(\mu_{t \wedge \sigma_k}) \leq Ln(n-1)\mathbb{E} \int_0^{t \wedge \sigma_k} F(\mu_s) ds + L^2 C_n \mathbb{E} \int_0^{t \wedge \sigma_k} F(\mu_s) ds$$

$$\leq Ln(n-1) \int_0^t \mathbb{E} F(\mu_{s \wedge \sigma_k}) ds + L^2 C_n \int_0^t \mathbb{E} F(\mu_{s \wedge \sigma_k}) ds, \quad t \geq 0.$$

By Gronwall's lemma, $\mathbb{E}\left[F(\mu_{t \wedge \sigma_k})\right] = 0$ for all $t \geq 0$. This implies that for every $t \geq 0$ a.s. $\mu_{t \wedge \sigma_k} \in \mathcal{N}_{n-1}$. Since $\mathcal{N}_{n-1}$ is closed and μ_t, $t \geq 0$, is continuous, we obtain that a.s. $\mu_{t \wedge \sigma_k} \in \mathcal{N}_{n-1}$ for all $t \geq 0$ and $k \geq 1$. Making $k \to \infty$, we can conclude that with probability 1 $\mu_{t \wedge \sigma} \in \mathcal{N}_{n-1}$, $t \geq 0$, that ends the proof of the proposition. $\square$

We next prove the main statement of this section.

Proof of Theorem 2.16. Let μ_t, $t \geq 0$, be a solution to the stochastic mean-field equation (2.1) started from $\mu_0 \in \mathcal{N}_n$ and satisfying (2.13). We will show that μ_t is a superposition solution. Let $\mu_0 = \sum_{i=1}^{n} \alpha_i \delta_{x^i}$ for some $\alpha_i \geq 0$ and $x^i \in \mathbb{R}^d$, $i \in [n]$. Without loss of generality, we may assume that α_i, $i \in [n]$, are strictly positive and x^i, $i \in [n]$, are distinct. Applying Lemma 2.21, we can conclude that a.s. $\mu_t \in \mathcal{N}_n$ for all $t \geq 0$. By the continuity of $\mu_.$, there exist $(\mathcal{F}_t)$-adapted continuous processes $\alpha_i(t)$, $Y^i(t)$, $t \geq 0$, $i \in [n]$, such that $a_i(0) = \alpha_i$, $Y_i(0) = x_i$, $i \in [n]$, and

$$\mu_t = \sum_{i=1}^{n} \alpha_i(t) \delta_{Y^i(t)}, \quad t \geq 0.$$

For every $i \in [n]$ we define the stopping time

$$\tau_0 := \inf \left\{ t \geq 0 : |Y^i(t) - x^i| \geq \varepsilon \text{ or } |Y^j(t) - x^i| \leq 2\varepsilon \text{ for some } j \neq i \right\},$$

where $\varepsilon = \frac{1}{3} \min_{i \neq j} |x^i - x^j|$. For each $i \in [n]$ we next consider functions $\varphi, \psi_i \in C^2_c(\mathbb{R}^d)$ such that $\varphi_i(x) = 1$, $\psi_i(x) = x^i$ for $|x - x^i| < \varepsilon$ and $\varphi_i(x) = \psi_i(x) = 0$ for $|x - x^i| \geq 2\varepsilon$. Then, using the definition of a solution to the stochastic mean-field equation, we obtain

$$\alpha_i(t \wedge \tau_0) = \langle \varphi_i, \mu_{t \wedge \tau_0} \rangle = \langle \varphi_i, \mu_0 \rangle = \alpha_i, \quad t \geq 0,$$

since $\nabla \varphi_i(x) = 0$ and $D^2 \varphi_i(x) = 0$ for $|x - x^i| < \varepsilon$ and for $|x - x^i| > 2\varepsilon$. Similarly, using additionally the previous observation, for every $t \geq 0$

$$\begin{aligned}
\alpha_i Y_i(t \wedge \tau_0) = \alpha_i(t \wedge \tau_0) Y_i(t \wedge \tau_0) &= \langle \psi_i, \mu_{t \wedge \tau_0} \rangle \\
&= \alpha_i Y_i(0) + \int_0^{t \wedge \tau_0} \alpha_i(s) v(s, Y_i(s)) ds + \int_0^{t \wedge \tau_0} \int_\Theta \alpha_i(s) g(s, Y_i(s), \theta) W(d\theta, ds) \\
&= \alpha_i Y_i(0) + \int_0^{t \wedge \tau_0} \alpha_i v(s, Y_i(s)) ds + \int_0^{t \wedge \tau_0} \int_\Theta \alpha_i g(s, Y_i(s), \theta) W(d\theta, ds)
\end{aligned}$$

holds. Hence, Y_i is a solution to the usual SDE

$$(2.20) \qquad dY_i(t) = v(t, Y_i(t)) dt + \int_\Theta g(t, Y_i(t), \theta) W(d\theta, dt), \quad Y_i(0) = x^i$$

on the interval $[0, \tau_0]$. Note that (2.20) has at most one solution due to the Lipschitz continuity of its coefficients. Let $X(u, t)$, $t \geq 0$, $u \in \mathbb{R}^d$, be a solution to SDE with interaction (2.3). Since $X(x^i, t)$, $t \geq 0$, also solves equation (2.20), one has $Y_i(t) = X(x^i, t)$, $t \in [0, \tau_0]$, for each $i \in [n]$. Next, considering the process $\mu^1_t := \mu_{t+\tau_0}$ conditioning to the σ-algebra $\mathcal{F}_{\tau_0}$, we get that, μ^1_t, $t \geq 0$, satisfies the same stochastic mean-field equation with the initial condition μ_{τ_0}. Applying our argument above again to μ^1_t, $t \geq 0$, we get that $\alpha_i(t + \tau_0) = \alpha_i$ and $Y_i(t + \tau_0) = X(x^i, t + \tau_0)$ for $t \in [0, \tau_1]$ and $i \in [n]$, where

$$\begin{aligned}
\tau_1 := \inf \big\{ t \geq 0 : {}& |Y_i(t + \tau_0) - Y_i(\tau_0)| \geq \varepsilon \\
& \text{or } |Y_j(t + \tau_0) - Y_i(\tau_0)| \leq 2\varepsilon \text{ for some } j \neq i \big\},
\end{aligned}$$

and $\varepsilon = \frac{1}{3} \min_{i \neq j} |Y_i(\tau_0) - Y_j(\tau_0)|$. Hence, $\alpha_i(t) = \alpha_i$ and $Y_i(t) = X(x^i, t)$, $t \in [0, \tau_0 + \tau_1]$, $i \in [n]$. Repining our argument infinitely many times and using the uniform continuity of Y_i, $i \in [n]$, a.s. on any compact time interval, we get that $\alpha_i(t) = \alpha_i$ and $Y_i(t) = X(x^i, t)$, $t \in [0, \tau]$, for

$$\tau = \inf \{ t \geq 0 : Y_i(t) = Y_j(t) \text{ for some } j \neq i \}.$$

Since $X(x^i, \cdot)$ and $X(x^j, \cdot)$ never meet for distinct x^i, x^j, by Lemma 2.15, we can conclude that $\tau = +\infty$. Consequently, $\mu_t = \mu_0 \circ X^{-1}(\cdot, t)$, $t \geq 0$. This ends the proof of the uniqueness of the superposition principle for the stochastic mean-field equation. $\square$

2.2.2. *Initial condition with L_2-density.* In this section, we adapt the method from [58] in order to prove the uniqueness for the stochastic mean-field equation (2.1) if the initial condition has an L_2-density with respect to the Lebesgue measure and the coefficients are bounded a.s. We will need the following assumption.

Assumption 2.22. *For every $T > 0$ there exists a (non-random) constant $L > 0$ such that a.s. for every $x \in \mathbb{R}^d$, $t \in [0, T]$, and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$*

$$|V(t, x, \mu)| + \|\|G(t, x, \mu, \cdot)\|\|_\vartheta \leq L$$

holds.

Using the notation $\frac{d\mu}{dx}$ for the density of a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ with respect to the Lebesgue measure, we state the main result of this section.

Theorem 2.23. *Let the coefficients of stochastic mean-field equation* (2.1) *satisfy Assumptions 2.1, 2.7 and 2.22, and* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *be absolutely continuous with respect to Lebesgue measure with* $\frac{d\mu_0}{dx} \in L_2(\mathbb{R}^d)$. *Then* (2.1) *has a unique solution* μ_t, $t \geq 0$, *started from* μ_0. *Moreover,* μ_t, $t \geq 0$, *is a superposition solution which is absolutely continuous with respect to the Lebesgue measure with*

$$(2.21) \qquad \mathbb{E}\left\|\frac{d\mu_t}{dx}\right\|_{L_2}^2 \leq e^{Ct}\left\|\frac{d\mu_0}{dx}\right\|_{L_2}^2$$

for all $t \geq 0$ *and some constant* $C > 0$.

Proof. The proof of the theorem is similar to the proof of the uniqueness result for the nonlinear SPDE in [58, Section 3], which differs from the equation (2.1) by non-random and homogeneous (independent of time) coefficients, on one side, and a more general structure on the other side. Therefore, we just describe the main steps omitting details. Let μ_t, $t \geq 0$, be a solution to (2.1) started from μ_0. As in the previous section, we freeze μ_t in the coefficients. As before, we set $v(t,x) = V(t,x,\mu_t)$, $a(t,x) = A(t,x,\mu_t)$ and $g(t,x,\theta) = G(t,x,\mu_t,\theta)$. We will consider μ_t, $t \geq 0$, as a solution to the following linear SPDE

$$(2.22) \qquad d\mu_t = \frac{1}{2}D^2 : (a(t,\cdot)\mu_t)dt - \nabla \cdot (v(t,\cdot)\mu_t)dt - \int_{\Theta} \nabla \cdot (g(t,\cdot,\theta)\mu_t)W(d\theta,dt).$$

This means that equality (2.17) holds for the process μ_t, $t \geq 0$, where test functions φ can be taken in $\mathrm{C}_b^2(\mathbb{R}^d)$, by Corollary 2.20 and Assumption 2.7.

Our goal is to show that (2.22) has a unique solution. But now we will assume that a solution ν_t, $t \geq 0$, can take values in the space $\mathcal{M}(\mathbb{R}^d)$ of all signed measures on $\mathbb{R}^d$ with finite total variations and $\nu_0 \in \mathcal{M}(\mathbb{R}^d)$. Thus, let ν_t, $t \geq 0$, be a continuous $(\mathcal{F}_t)$-adapted process which satisfies (2.17) for all $\varphi \in \mathrm{C}_b(\mathbb{R}^d)$ and ν_0 is absolutely continuous with respect to the Lebesgue measure with $\frac{d\nu_0}{dx}$ from $L_2(\mathbb{R}^d)$.

For any $\rho \in \mathcal{M}(\mathbb{R}^d)$ and $\varepsilon > 0$, we define

$$P_\varepsilon \rho(x) := \int_{\mathbb{R}^d} p_\varepsilon(x - y)\rho(dy),$$

where p_ε is the heat kernel defined by $p_\varepsilon(x) = \frac{1}{(2\pi\varepsilon)^{-d/2}}e^{-\frac{|x|^2}{2\varepsilon}}$. For every $\varphi \in \mathrm{C}_b(\mathbb{R}^d)$, we will also write $P_\varepsilon\varphi(x)$ for $\int_{\mathbb{R}^d} p_\varepsilon(x-y)\varphi(y)dy$ and $P_\varepsilon(\varphi\rho)(x)$ for $\int_{\mathbb{R}^d} p_\varepsilon(x-y)\varphi(y)\rho(dy)$. Setting $\nu_t^\varepsilon := P_\varepsilon\nu_t$, $t \geq 0$, and following the computations from [58, Section 3], based on the application of Itô's formula, we get

$$\mathbb{E}\|\nu_t^\varepsilon\|_{L_2}^2 = \|\nu_0^\varepsilon\|_{L_2}^2 - \sum_{i=1}^d \mathbb{E}\int_0^t 2\left\langle \nu_s^\varepsilon, \partial_{x_i} P_\varepsilon(v_i(s,\cdot)\nu_s)\right\rangle_{L_2} ds$$

$$+ \sum_{i,j=1}^d \mathbb{E}\int_0^t \left\langle \nu_s^\varepsilon, \partial_{x_i}\partial_{x_j} P_\varepsilon(a_{i,j}(s,\cdot)\nu_s)\right\rangle_{L_2} ds$$

$$+ \mathbb{E}\int_0^t \int_{\Theta}\left\|\sum_{i=1}^d \partial_{x_i} P_\varepsilon(g_i(s,\cdot,\theta)\nu_s)\right\|_{L_2}^2 \vartheta(d\theta)ds, \quad t \geq 0.$$

Using Assumptions 2.7, 2.22 and [58, Lemma 3.2], one can immediately conclude that for every $T > 0$ there exists a constant C such that

$$(2.23) \qquad \mathbb{E} \int_0^t 2 \left\langle \nu_s^\varepsilon, \partial_{x_i} P_\varepsilon(v_i(s, \cdot)\nu_s) \right\rangle_{L_2} ds \le C \int_0^t \mathbb{E} \left\| P_\varepsilon(|\nu_s|) \right\|_{L_2}^2 ds, \quad t \in [0, T],$$

where $|\nu_s|$ is the total variation of ν_s. Using the integration by parts in the L_2-norm, the sum of the last two integrals in the expansion of $\mathbb{E}\|\nu_t^\varepsilon\|_{L_2}^2$ can be rewritten as

$$\frac{1}{2} \sum_{i,j=1}^d \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\frac{(x_i - y_i)(x_j - y_j)}{4\varepsilon^2} - \frac{1}{2\varepsilon} \mathbb{I}_{\{i=j\}} \right) p_{2\varepsilon}(x - y)$$
$$\cdot \left\langle g_i(s, y, \theta) - g_i(s, x, \theta), g_j(s, y, \theta) - g_j(s, x, \theta) \right\rangle_\vartheta \nu_s(dx)\nu_s(dy).$$

This leads to the same estimate (2.23), see the proof of [58, Lemma 3.3] for more details. Hence, for every $T > 0$, there exists a constant $C > 0$ such that

$$\mathbb{E}\|\nu_t^\varepsilon\|_{L_2}^2 \le \|\nu_0^\varepsilon\|_{L_2}^2 + C \int_0^t \mathbb{E} \left\| P_\varepsilon(|\nu_s|) \right\|_{L_2}^2 ds.$$

Following proofs of Corollary 3.1 and Theorem 3.3 from [58], we can conclude that every solution to the linear SPDE (2.22) with L_2-density is unique. Moreover, if its solution ν_t, $t \ge 0$, is a measure-valued process, then ν_t is absolutely continuous with respect to the Lebesgue measure with density satisfying (2.21) for every $t \ge 0$. This immediately implies that the process μ_t, $t \ge 0$, is a unique solution to (2.22) started from μ_0 and is absolutely continuous with respect to the Lebesgue measure that satisfies (2.21).

In order to show that μ_t, $t \ge 0$, is superposition solution, consider a solution $Y(u, t)$, $t \ge 0$, $u \in \mathbb{R}^d$, to the equation

$$(2.24) \qquad \begin{aligned} dY(u, t) &= v(t, Y(u, t))dt + \int_\Theta g(t, Y(u, t), \theta)W(d\theta, dt), \\ Y(u, 0) &= u, \quad \bar{\nu}_t = \mu_0 \circ Y(\cdot, t)^{-1}, \end{aligned}$$

which exists and is unique due to Theorem 2.9, since its coefficients v, g satisfy Assumptions 2.1, 2.7. By Lemma 2.13, $\bar{\nu}_t$, $t \ge 0$, is a solution to the (2.22) started from μ_0. The uniqueness result, stated above, yields that $\bar{\nu}_t = \mu_t$, $t \ge 0$. On the other hand, the unique solution X to (2.3) solves also (2.24). Hence, Y must coincide with X. This completes the proof of the superposition principle.

Then, the uniqueness for the stochastic mean-field equation (2.1) directly follows from Corollary 2.12. $\qquad \square$

2.2.3. *Initial condition with finite second moment.* We will now obtain the well-posedness and superposition principle for the stochastic mean-field equation for general initial conditions. This case includes both types of initial conditions considered before, however, stronger assumptions on the regularity of the coefficients will be needed. Our main idea is to transform a solution to the stochastic mean-field equation by a smooth (in space) stochastic flow to get a solution to a continuity equation with random coefficients for which the superposition principle can be easily obtain, e.g., by a duality method. This will imply that the original equation has only superposition solutions which will yield the uniqueness result by Corollary 2.12. For the construction of the transformation flow, we will mainly use results from [57].

We first introduce the following assumption on the coefficients of the equation and formulate the main result of this section.

Assumption 2.24. *There exists* $\delta \in (0, 1)$ *such that* $V_i(t, \cdot, \mu) \in C_{lb}^{1,\delta}(\mathbb{R}^d)$, $\tilde{A}_{i,j}(t, \cdot, \mu) \in \tilde{C}_{lb}^{3,\delta}(\mathbb{R}^d)$ *a.s. for all* $t \geq 0$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $i, j \in [d]$, *and for every* $T > 0$ *and a compact set* $K \subset \mathcal{P}_2(\mathbb{R}^d)$ *a.s.*

$$\sup_{t \in [0,T], \mu \in K} \left(\|V_i(t, \cdot, \mu)\|_{1+\delta} + \|\tilde{A}_{i,j}(t, \cdot, \mu)\|_{3+\delta}^{\sim} \right) < \infty, \quad i, j \in [d],$$

where $\tilde{A}$ *was defined in Remark* 2.4.

Remark 2.25. Let $\mu_t, t \geq 0$, be an arbitrary continuous process in $\mathcal{P}_2(\mathbb{R}^d)$. Then Assumption 2.24 implies that for each $i, j \in [d]$ the processes $\int_0^t V_i(s, \cdot, \mu_s)ds, t \geq 0$, and $\int_0^t \tilde{A}_{i,j}(s, \cdot, \mu_s)ds, t \geq 0$, are a.s. continuous in $C^{1,\delta}(\mathbb{R}^d)$ and $\tilde{C}^{3,\delta}(\mathbb{R}^d)$, respectively. Indeed, due to the continuity of the process $\mu_t, t \geq 0$, the set $K_T^\mu := \{\mu_t : t \in [0, T]\}$ is compact in $\mathcal{P}_2(\mathbb{R}^d)$ for every $T > 0$. Then the direct computation shows that for every $T > 0$, $t, t' \in [0, T]$ and compact $K \subset \mathbb{R}^d$

$$\left\| \int_0^t V_i(s, \cdot, \mu_s)ds - \int_0^{t'} V_i(s, \cdot, \mu_s)ds \right\|_{1+\delta, K} \leq \sup_{\substack{s \in [0,T], \\ \mu \in K_T^\mu}} \|V_i(s, \cdot, \mu)\|_{1+\delta} |t - t'|$$

and

$$\left\| \int_0^t \tilde{A}_{i,j}(s, \cdot, \mu_s)ds - \int_0^{t'} \tilde{A}_{i,j}(s, \cdot, \mu_s)ds \right\|_{3+\delta, K}^{\sim} \leq \sup_{\substack{s \in [0,T], \\ \mu \in K_T^\mu}} \left\| \tilde{A}_{i,j}(s, \cdot, \mu) \right\|_{3+\delta} |t - t'|$$

for all $i, j \in [d]$, that guarantees the continuity.

Theorem 2.26. *Let the coefficients of the stochastic mean-field equation* (2.1) *satisfy Assumptions* 2.1, 2.7 *and* 2.24. *Then for every* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *the equation* (2.1) *has a unique solution started from* μ_0. *Moreover, it is a superposition solution.*

In order to prove the theorem we will state a few auxiliary statements. Let $\mu_t, t \geq 0$, be a solution to the stochastic mean-field equation (2.1) whose coefficients satisfy Assumptions 2.1, 2.7 and 2.24. As before, we will freeze μ_t in the coefficients, setting $v(t, x) = V(t, x, \mu_t)$, $a(t, x) = A(t, x, \mu_t)$, $\tilde{a}(t, x, y) = \tilde{A}(t, x, y, \mu_t)$ and $g(t, x, \theta) = G(t, x, \mu_t, \theta)$. We consider the following field of local martingales

$$M(x, t) = \int_0^t \int_\Theta g(s, x, \theta)W(d\theta, ds), \quad x \in \mathbb{R}^d, \ t \geq 0.$$

Note that its local quadratic variation

$$[M_i(x, \cdot), M_j(y, \cdot)]_t = \int_0^t \tilde{a}_{i,j}(s, x, y)ds, \quad t \geq 0, \ x, y \in \mathbb{R}^d, \ i, j \in [d],$$

is $\tilde{C}^{3,\delta}(\mathbb{R}^d)$-valued continuous process, by Remark 2.25. Consequently, there exists a version of M which is a $C^{3,\delta'}$-valued continuous process for every $\delta' \in (0, \delta)$, by [57, Theorem 3.1.2]. Moreover, for each α with $|\alpha| \leq 3$, $D^\alpha M(x, t), t \geq 0, x \in \mathbb{R}^d$, is a family of continuous local martingales with quadratic variation

$$\left[D^\alpha M_i(x, \cdot), D^\beta M_j(y, \cdot) \right]_t = \int_0^t D_x^\alpha D_y^\beta \tilde{a}_{i,j}(s, x, y)ds, \quad t \geq 0,$$

for any $x, y \in \mathbb{R}^d$, $i, j \in [d]$ and α, β with $|\alpha| \leq 3$, $|\beta| \leq 3$, by [57, Theorem 3.1.3].

We consider the following family of linear stochastic transport equations written in integral form

$$(2.25) \qquad \psi_k(t,x) = x_k - \int_0^t \nabla\psi_k(s,x) \cdot M(x, \circ ds), \quad t \geq 0, \ x \in \mathbb{R}^d, \ k \in [d],$$

where the Stratonovich integral was defined in [57, Section 3.2]. Using the connection between the Itô and Stratonovich integrals (see [57, Theorem 3.2.5]), one gets

$$
\int_0^t \nabla\psi_k(s,x) \cdot M(x, \circ ds) = \int_0^t \nabla\psi_k(s,x) \cdot M(x, ds) + \frac{1}{2}\sum_{i=1}^d \left[\int_0^\cdot M_i(x, ds), \partial_i \psi_k(\cdot, x) \right]_t
$$

$$
= \int_0^t \nabla\psi_k(s,x) \cdot g(s,x,\theta) W(d\theta, ds)
$$

$$
- \frac{1}{2}\sum_{i=1}^d \left[M_i(x, \cdot), \partial_i \int_0^\cdot \nabla\psi_k(s,x) \cdot M(x, ds) \right]_t.
$$

In order to compute the quadratic variation on the right hand side of the above expression, we will use [57, Theorem 3.1.3]. Thus, for $x, y \in \mathbb{R}^d$

$$
\left[M_i(y, \cdot), \partial_i \int_0^\cdot \nabla\psi_k(s,x) \cdot M(x, ds) \right]_t = \frac{\partial}{\partial x_i} \left[M_i(y, \cdot), \int_0^\cdot \nabla\psi_k(s,x) \cdot M(x, ds) \right]
$$

$$
= \frac{\partial}{\partial x_i} \sum_{j=1}^d \int_0^t \partial_j \psi_k(s,x) \int_\Theta g_j(s,x,\theta) g_j(s,y,\theta) \vartheta(d\theta) ds
$$

$$
= \sum_{j=1}^d \int_0^t \frac{\partial}{\partial x_i} \left(\partial_j \psi_k(s,x) \tilde{a}_{i,j}(s,x,y) \right) ds.
$$

Therefore, equality (2.25) can be rewritten in Itô's form as follows

$$(2.26) \qquad
\begin{aligned}
\psi_k(t,x) = x_k &- \int_0^t \int_\Theta \nabla\psi_k(s,x) \cdot g(s,x,\theta) W(d\theta, ds) \\
&+ \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} \nabla_x \cdot (\tilde{a}(s,x,y) \cdot \nabla\psi_k(s,x)) \delta_x(dy) ds
\end{aligned}
$$

for any $t \geq 0$, $x \in \mathbb{R}^d$ and $k \in [d]$.

Proposition 2.27. *Under the assumptions of Theorem 2.26, for each $k \in [d]$ there exist $\delta' \in (0, \delta)$ and an $(\mathcal{F}_t)$-adapted continuous $C^{3,\delta'}$-valued process $\psi_k(t, \cdot)$, $t \geq 0$, that satisfies (2.25) (and, therefore, (2.26)). Moreover, a.s. for every $t \geq 0$ the map $\psi(t, \cdot) = (\psi_1(t, \cdot), \ldots, \psi_d(t, \cdot)) : \mathbb{R}^d \to \mathbb{R}^d$ is invertible and $\varphi(t, \cdot) := \psi^{-1}(t, \cdot)$ is an $(\mathcal{F}_t)$-adapted continuous $C^{3,\delta'}(\mathbb{R}^d)$-valued stochastic process that satisfies the equation*

$$
\varphi(t,x) = x + \int_0^t M(\varphi(s,x), \circ ds)
$$

for every $t \geq 0$ and $x \in \mathbb{R}^d$.

We note that [57, Theorem 3.2.5] yields that the inverse flow $\varphi(t, \cdot)$, $t \geq 0$, to $\psi(t, \cdot)$, $t \geq 0$, solves the equation

$$
\varphi(t, x) = x + \int_0^t \int_\Theta g(s, \varphi(s, x), \theta) W(d\theta, ds)
$$
(2.27)
$$
+ \frac{1}{2} \int_0^t (\nabla_x \cdot \tilde{a})(s, \varphi(s, x), \varphi(s, x)) ds, \quad t \geq 0, \ x \in \mathbb{R}^d,
$$

in Itô form, where $\nabla_x \cdot \tilde{a}(s, x, y) = \left(\sum_{j=1}^d \frac{\partial}{\partial x_j} \tilde{a}_{i,j}(s, x, y) \right)_{i \in [d]}$. Indeed,

$$
\int_0^t M_k(\varphi(s, x), \circ ds) = \int_0^t M_k(\varphi(s, x), ds) + \frac{1}{2} \sum_{i=1}^d \left[\partial_i \int_0^\cdot M_k(\varphi(s, x), ds), \varphi_i(\cdot, x) \right]_t.
$$

By [57, Theorem 3.1.3], we can compute for $x, y \in \mathbb{R}^d$

$$
\left[\frac{\partial}{\partial x_i} \int_0^\cdot M_k(\varphi(s, x), ds), \varphi_i(\cdot, y) \right]_t = \frac{\partial}{\partial x_i} \int_0^t \int_\Theta g_k(s, x, \theta) g_i(s, y, \theta) \vartheta(d\theta) ds
$$
$$
= \int_0^t \frac{\partial}{\partial x_i} \tilde{a}_{k,i}(s, x, y) ds.
$$

This implies the equivalence between the equations in Itô and Stratonovich form.

Proof of Proposition 2.27. The existence of ψ_k, $k \in [d]$, follows from [57, Theorem 6.1.8]. We observe that

$$
\varphi(t, x) = x + \int_0^t M(\varphi(s, x), \circ ds), \quad t \geq 0, \ x \in \mathbb{R}^d,
$$

is the stochastic characteristic equation for the SPDE (2.25). By (2.27) and [57, Theorems 3.4.6, 4.7.3], it has a unique continuous $C^{3, \delta'}(\mathbb{R}^d)$-valued solution which is also a stochastic flow of C^3-diffeomorphisms. Using [57, Theorem 6.1.2], we can conclude that $\psi_k(t, \cdot)$ is the k-th coordinate of $\varphi^{-1}(t, \cdot)$ for any $t \geq 0$. This immediately implies the equality $\psi^{-1}(t, \cdot) = \varphi(t, \cdot)$ which ends the proof of the proposition. $\qquad\square$

We will next consider for every $t \geq 0$ the following probability measure $\rho_t = \mu_t \circ \psi^{-1}(t, \cdot)$ on $\mathbb{R}^d$. It is easy to see that ρ_t, $t \geq 0$, is a continuous process in $\mathcal{P}(\mathbb{R}^d)$. Let us show that this process (locally) satisfies a continuity equation with random coefficients.

Lemma 2.28. *Let the coefficients of* (2.1) *satisfy Assumptions* 2.1, 2.7 *and* 2.24. *Then the measure-valued process* ρ_t, $t \geq 0$, *defined above, is a solution to the equation*

(2.28) $d\rho_t = -\nabla(b(t, \cdot)\rho_t)dt, \quad \rho_0 = \mu_0,$

that is, for every $\varphi \in C_c^2(\mathbb{R}^d)$ *a.s. the equality*

$$
\langle \rho_t, \varphi \rangle = \langle \mu_0, \varphi \rangle + \int_0^t \langle \nabla \varphi \cdot b(s, \cdot), \rho_s \rangle ds, \quad t \geq 0,
$$

holds, where $b(t, x) = \tilde{b}(t, \psi^{-1}(t, x))$ *and*

$$
\tilde{b}_k(t, x) = \nabla \psi_k(t, x) \cdot v(t, x) - \frac{1}{2} \int_{\mathbb{R}^d} \nabla_x \cdot (\tilde{a}(t, x, y) \cdot \nabla \psi_k(t, x)) \, \delta_x(dy)
$$

for all $t \geq 0$, $x \in \mathbb{R}^d$, $k \in [d]$.

Proof. Let $\varphi \in C_c^2(\mathbb{R}^d)$. We note that $\langle \varphi, \rho_t \rangle = \langle \varphi \circ \psi(t, \cdot), \mu_t \rangle$ for every $t \geq 0$, by the definition of the measure ρ_t. Hence, we first apply Itô's formula to $\varphi \circ \psi(t, x)$, $t \geq 0$, for each $x \in \mathbb{R}^d$. One gets for $t \geq 0$

$$\varphi \circ \psi(t, x) = \varphi(\psi(t, x)) = \varphi(x)$$

$$- \sum_{k=1}^d \int_0^t \int_\Theta (\partial_k \varphi \circ \psi)(s, x) \left[\nabla \psi_k(s, x) \cdot g(s, x, \theta) \right] W(d\theta, ds)$$

$$+ \frac{1}{2} \sum_{k=1}^d \int_0^t (\partial_k \varphi \circ \psi)(s, x) \int_{\mathbb{R}^d} \nabla_x \cdot (\tilde{a}(t, x, y) \cdot \nabla \psi_k(t, x)) \, \delta_x(dy) ds$$

$$+ \frac{1}{2} \sum_{k,l=1}^d \int_0^t (\partial_{k,l}^2 \varphi \circ \psi)(s, x) \left[(\nabla \psi_k(s, x) \otimes \nabla \psi_l(s, x)) : a(s, x) \right] ds.$$

Next, we consider a non-negative function $\kappa \in C_c^2(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} \kappa(x) dx = 1$, and set $\delta_x^\varepsilon(y) = \frac{1}{\varepsilon^d} \kappa \left(\frac{1}{\varepsilon}(x - y) \right)$, $y \in \mathbb{R}^d$, for each $x \in \mathbb{R}^d$ and $\varepsilon > 0$. Integrating δ_x^ε by μ_t and denoting $\mu_t^\varepsilon(x) = \langle \delta_x^\varepsilon, \mu_t \rangle$, we get for every $x \in \mathbb{R}^d$ a.s.

$$\mu_t^\varepsilon(x) = \mu_0^\varepsilon(x) + \frac{1}{2} \int_0^t \langle D^2 \delta_x^\varepsilon : a(s, \cdot), \mu_s \rangle \, ds$$

$$+ \int_0^t \langle \nabla \delta_x^\varepsilon \cdot v(s, \cdot), \mu_s \rangle \, ds + \int_0^t \int_\Theta \langle \nabla \delta_x^\varepsilon \cdot g(s, \cdot, \theta), \mu_s \rangle \, W(d\theta, ds)$$

for all $t \geq 0$. Using now the expressions for $\varphi \circ \psi(t, x)$ and $\mu_t^\varepsilon(x)$ and Itô's formula, we obtain for every $x \in \mathbb{R}^d$ a.s.

$$\varphi \circ \psi(t, x) \mu_t^\varepsilon(x) = \varphi(x) \mu_0^\varepsilon(x)$$

$$- \sum_{k=1}^d \int_0^t \int_\Theta \mu_s^\varepsilon(x) (\partial_k \varphi \circ \psi)(s, x) \left[\nabla \psi_k(s, x) \cdot g(s, x, \theta) \right] W(d\theta, ds)$$

$$+ \frac{1}{2} \sum_{k=1}^d \int_0^t \mu_s^\varepsilon(x) (\partial_k \varphi \circ \psi)(s, x) \int_{\mathbb{R}^d} \nabla_x \cdot (\tilde{a}(s, x, y) \cdot \nabla \psi_k(s, x)) \, \delta_x(dy) ds$$

$$+ \frac{1}{2} \sum_{k,l=1}^d \int_0^t \mu_s^\varepsilon(x) (\partial_{k,l}^2 \varphi \circ \psi)(s, x) \left[(\nabla \psi_k(s, x) \otimes \nabla \psi_l(s, x)) : a(s, x) \right] ds$$

$$+ \frac{1}{2} \int_0^t \varphi \circ \psi(s, x) \langle D^2 \delta_x^\varepsilon : a(s, \cdot), \mu_s \rangle \, ds + \int_0^t \varphi \circ \psi(s, x) \langle \nabla \delta_x^\varepsilon \cdot v(s, \cdot), \mu_s \rangle \, ds$$

$$+ \int_0^t \int_\Theta \varphi \circ \psi(s, x) \langle \nabla \delta_x^\varepsilon \cdot g(s, \cdot, \theta), \mu_s \rangle \, W(d\theta, ds)$$

$$- \sum_{k=1}^d \int_0^t \int_\Theta \langle \nabla \delta_x^\varepsilon \cdot g(s, \cdot, \theta), \mu_s \rangle (\partial_k \varphi \circ \psi)(s, x) \left[\nabla \psi_k(s, x) \cdot g(s, x, \theta) \right] \vartheta(d\theta) ds$$

for all $t \geq 0$. Note that for every $t \geq 0$ a.s. there exists a compact $K \subset \mathbb{R}^d$ such that $\operatorname{supp} \varphi \circ \psi(s, \cdot) \in K$ for any $s \in [0, t]$. Indeed, one can take $K = \{ \psi^{-1}(s, x) : s \in [0, t], x \in \operatorname{supp} \varphi \}$, which is a compact set as the image of $[0, t] \times \operatorname{supp} \varphi$ under the continuous map $(s, x) \mapsto \psi^{-1}(s, x)$ (for the continuity of ψ^{-1} see Proposition 2.27). Therefore, we may integrate the above

expression with respect to the Lebesgue measure dx over $\mathbb{R}^d$. Then taking $\varepsilon \to 0$, we get

$$\int_{\mathbb{R}^d} \varphi \circ \psi(t,x) \mu_t^\varepsilon(x) dx \to \langle \varphi \circ \psi(t,\cdot), \mu_t \rangle \quad \text{a.s.}$$

and

$$\int_{\mathbb{R}^d} \varphi(x) \mu_0^\varepsilon(x) dx \to \langle \varphi, \mu_0 \rangle \quad \text{a.s.}$$

Next using the equalities $\nabla_y \delta_x^\varepsilon(y) = -\nabla_x \delta_x^\varepsilon(y)$, $D_y^2 \delta_x^\varepsilon(y) = D_x^2 \delta_x^\varepsilon(y)$, the integration by parts formula and Fubini's theorem, we can conclude that the sum of all terms consisting of the integrals $\int_0^t (\dots) ds$ in the expression for $\int_{\mathbb{R}^d} \varphi \circ \psi(t,x) \mu_t^\varepsilon(x) dx$ converges a.s. to

$$I_1 := \frac{1}{2} \sum_{k=1}^d \int_0^t \left\langle (\partial_k \varphi \circ \psi)(s,x) \int_{\mathbb{R}^d} \nabla_x \cdot (\tilde{a}(s,x,y) \cdot \nabla \psi_k(s,x)) \delta_x(dy), \mu_s(dx) \right\rangle ds$$

$$+ \frac{1}{2} \sum_{k,l=1}^d \int_0^t \left\langle (\partial_{k,l}^2 \varphi \circ \psi)(s,x) \left[(\nabla \psi_k(s,x) \otimes \nabla \psi_l(s,x)) : a(s,x) \right], \mu_s(dx) \right\rangle ds$$

$$+ \frac{1}{2} \int_0^t \left\langle D^2(\varphi \circ \psi)(s,x) : a(s,\cdot), \mu_s \right\rangle ds + \int_0^t \left\langle \nabla(\varphi \circ \psi)(s,x) \cdot v(s,\cdot), \mu_s \right\rangle ds$$

$$- \sum_{k,i,j=1}^d \int_0^t \int_{\mathbb{R}^d} \left\langle \frac{\partial}{\partial x_i} \left[(\partial_k \varphi \circ \psi)(s,x) \partial_j \psi_k(s,x) \tilde{a}_{i,j}(s,x,y) \right], \mu_s(dx) \right\rangle \delta_x(dy) ds.$$

Taking into account that the second and the third integrals are the same and that the last term can be rewritten as

$$- \sum_{k,l=1}^d \int_0^t \left\langle (\partial_{k,l}^2 \varphi \circ \psi)(s,x) \partial_j \left[(\nabla \psi_k(s,x) \otimes \nabla \psi_l(s,x)) : \tilde{a}_{i,j}(s,x,x) \right], \mu_s(dx) \right\rangle ds$$

$$- \sum_{k=1}^d \int_0^t \left\langle (\partial_k \varphi \circ \psi)(s,x) \int_{\mathbb{R}^d} \nabla_x \cdot (\tilde{a}(s,x,y) \cdot \nabla \psi_k(s,x)) \delta_x(dy), \mu_s(dx) \right\rangle ds,$$

we obtain

$$I_1 = -\frac{1}{2} \sum_{k=1}^d \int_0^t \left\langle (\partial_k \varphi \circ \psi)(s,x) \int_{\mathbb{R}^d} \nabla_x \cdot (\tilde{a}(s,x,y) \cdot \nabla \psi_k(s,x)) \delta_x(dy), \mu_s(dx) \right\rangle ds$$

$$+ \sum_{k=1}^d \int_0^t \left\langle (\partial_k \varphi \circ \psi)(s,\cdot) \nabla \psi_k(s,\cdot) \cdot v(s,\cdot), \mu_s \right\rangle ds$$

$$= \int_0^t \left\langle (\nabla \varphi \circ \psi)(s,\cdot) \cdot b(s, \psi(s,\cdot)), \mu_s \right\rangle ds.$$

We now show that the difference of stochastic integrals converges to zero. Set

$$F_\varepsilon(s,x,\theta) = \varphi \circ \psi(s,x) \left\langle \nabla \delta_x^\varepsilon \cdot g(s,\cdot,\theta), \mu_s \right\rangle$$

$$- \sum_{k=1}^d \mu_s^\varepsilon(x) (\partial_k \varphi \circ \psi)(s,x) \left[\nabla \psi_k(s,x) \cdot g(s,x,\theta) \right]$$

$$= \int_{\mathbb{R}^d} \varphi \circ \psi(s,x) \nabla_y \delta_x^\varepsilon(y) \cdot g(s,y,\theta) \mu_s(dy)$$

$$- \int_{\mathbb{R}^d} \delta_x^\varepsilon(y) \nabla(\varphi \circ \psi)(s,x) \cdot g(s,x,\theta)\mu_s(dy)$$

$$= \int_{\mathbb{R}^d} \left(\varphi \circ \psi(s,x)\nabla_y \delta_x^\varepsilon(y) \cdot g(s,y,\theta) - \delta_x^\varepsilon(y)\nabla(\varphi \circ \psi)(s,x) \cdot g(s,x,\theta) \right)\mu_s(dy)$$

for all $s \geq 0$, $x \in \mathbb{R}^d$ and $\theta \in \Theta$. Using stochastic Fubini's theorem, one can see that

$$(2.29) \qquad \int_{\mathbb{R}^d} \left(\int_0^t \int_\Theta F_\varepsilon(s,x,\theta)W(d\theta,ds) \right)dx = \int_0^t \int_\Theta \left(\int_{\mathbb{R}^d} F_\varepsilon(s,x,\theta)dx \right)W(d\theta,ds).$$

Moreover,

$$\int_0^t \int_\Theta \left(\int_{\mathbb{R}^d} F_\varepsilon(s,x,\theta)dx \right)^2 \vartheta(d\theta)ds$$

$$= \int_0^t \int_{\mathbb{R}^{4d}} \varphi \circ \psi(s,x)\varphi \circ \psi(s,\tilde{x})\nabla_y \delta_x^\varepsilon(y) \otimes \nabla_{\tilde{y}}\delta_{\tilde{x}}^\varepsilon(\tilde{y}) : \tilde{a}(s,y,\tilde{y})\mu_s(dy)\mu_s(d\tilde{y})dxd\tilde{x}ds$$

$$+ \int_0^t \int_{\mathbb{R}^{4d}} \delta_x^\varepsilon(y)\delta_{\tilde{x}}^\varepsilon(\tilde{y})\nabla(\varphi \circ \psi)(s,x) \otimes \nabla(\varphi \circ \psi)(s,\tilde{x}) : \tilde{a}(s,x,\tilde{x})\mu_s(dy)\mu_s(d\tilde{y})dxd\tilde{x}ds$$

$$- 2\int_0^t \int_{\mathbb{R}^{4d}} \varphi \circ \psi(s,x)\delta_{\tilde{x}}^\varepsilon(\tilde{y})\nabla_y \delta_x^\varepsilon(y) \otimes \nabla\left(\varphi \circ \psi(s,\tilde{x}) : \tilde{a}(s,y,\tilde{x})\right)\mu_s(dy)\mu_s(d\tilde{y})dxd\tilde{x}ds$$

for every $t \geq 0$. After the integration by parts, similarly as before, we obtain

$$\int_0^t \int_\Theta \left(\int_{\mathbb{R}^d} F_\varepsilon(s,x,\theta)dx \right)^2 \vartheta(d\theta)ds \to 0 \quad \text{a.s.}$$

as $\varepsilon \to 0$. This simply implies the convergence of the right hand side of (2.29) to zero in probability as $\varepsilon \to 0$, by, e.g., [43, Theorem II.7.2']. Summarizing our computations, we get that for every $t \geq 0$ a.s.

$$\langle \varphi \circ \psi(t,\cdot), \mu_t \rangle = \langle \varphi, \mu_0 \rangle + \int_0^t \langle (\nabla\varphi \circ \psi)(s,\cdot) \cdot b(s,\psi(s,\cdot)), \mu_s \rangle\,ds.$$

This directly implies that the process ρ_t, $t \geq 0$, solves the equation (2.28). $\qquad \square$

Note that we cannot control the growth of the coefficient b in the PDE (2.28) in the spatial variable. Therefore, the known superposition principle, e.g. from [4, 23, 79], cannot be applied. However, we can construct a superposition solution to (2.28) and a solution to its dual equation precisely, and then use the duality principle to prove that only the superposition solution solves (2.28).

Let $Y(u,t)$, $u \in \mathbb{R}^d$, $t \geq 0$, be a unique solution to (2.3) with the frozen coefficients v and g. In particular, for every $u \in \mathbb{R}^d$ a.s.

$$(2.30) \qquad Y(u,t) = u + \int_0^t v(s,Y(u,s))ds + \int_0^t \int_\Theta g(s,Y(u,s),\theta)W(d\theta,ds), \quad t \geq 0.$$

By Assumptions 2.7, 2.24 and [57, Theorem 4.6.5], Y has a modification (also denoted by Y) which is a continuous process in the space of $C^{1,\delta'}(\mathbb{R}^d)$-diffeomorphisms for every $\delta' \in (0,\delta)$, i.e., a.s. $Y(\cdot,t)$, $t \geq 0$, is a continuous $C^{1,\delta'}(\mathbb{R}^d)$-valued process and a.s. for every $t \geq 0$ there exists the inverse map $Y^{-1}(\cdot,t) : \mathbb{R}^d \to \mathbb{R}^d$ to $Y(\cdot,t)$ which is a $C^{1,\delta'}$-valued continuous process. We also set

$$(2.31) \qquad Z(\cdot,t) = \psi(t,Y(\cdot,t)), \quad t \geq 0,$$

which is a continuous process in the space of $C^{1,\delta'}$-diffeomeorphisms for some $\delta \in (0,\delta)$, according to Proposition 2.27.

Lemma 2.29. *Under the assumptions of Lemma* 2.28, *the process* $\bar{\rho}_t = \mu_0 \circ Z^{-1}(\cdot, t)$, $t \geq 0$, *is the unique solution to* (2.28).

Proof. We first apply the generalised Itô's formula from [57, Theorem 3.3.1] to $Z(u,t) = \psi(t, Y(u,t))$, $t \geq 0$. We get for every $u \in \mathbb{R}^d$ a.s.

$$Z(u,t) = u + \int_0^t \tilde{b}(s, Y(u,s))ds = u + \int_0^t b(s, Z(u,s))ds, \quad t \geq 0.$$

where $\tilde{b}$ and b were defined in Lemma 2.28. Let $\varphi \in C_c^2(\mathbb{R}^d)$. By the usual chain-rule, one obtains for each $u \in \mathbb{R}^d$ a.s.

$$\varphi(Z(u,t)) = \varphi(u) + \int_0^t (\nabla \varphi)(Z(u,s)) \cdot b(s, Z(u,s))ds, \quad t \geq 0.$$

Then, integrating the obtained expression with respect to μ_0 yields that the process $\bar{\rho}_t$, $t \geq 0$, satisfies the equality

$$\langle \varphi, \bar{\rho}_t \rangle = \langle \varphi, \mu_0 \rangle + \int_0^t \langle \nabla \varphi \cdot b(s, \cdot), \bar{\rho}_s \rangle ds, \quad t \geq 0,$$

a.s. This implies that $\bar{\rho}_t$, $t \geq 0$, is a solution to (2.28).

We next prove the uniqueness of solutions to (2.28). Let ρ_t, $t \geq 0$, be a solution to (2.28). For a fixed function $\varphi \in C_c^2(\mathbb{R}^d)$ define

$$\gamma(t, \cdot) = \varphi(Z^{-1}(\cdot, t)), \quad t \geq 0,$$

which is a $C^{3,\delta'}(\mathbb{R}^d)$-valued continuous process. Then similarly to the proof of [57, Lemma 6.1.1], one can show that for every $x \in \mathbb{R}^d$ a.s.

$$\gamma(t,x) = \varphi(x) - \int_0^t \nabla \gamma(s,x) \cdot b(s,x)ds, \quad t \geq 0.$$

Repeating now the computation from the proof of Lemma 2.28, we get a.s.

$$\langle \gamma(t,\cdot), \rho_t \rangle = \langle \gamma(0,\cdot), \varphi \rangle + \int_0^t \langle \nabla \gamma(s,\cdot) \cdot b(s,\cdot), \rho_s \rangle ds - \int_0^t \langle \nabla \gamma(s,\cdot) \cdot b(s,x), \rho_s \rangle ds$$
$$= \langle \mu_0, \varphi \rangle, \quad t \geq 0.$$

Thus, from the definition of γ it follows that a.s.

$$\langle \varphi, \mu_0 \rangle = \int_{\mathbb{R}^d} \gamma(t,x)\rho_t(dx) = \int_{\mathbb{R}^d} \varphi(Z^{-1}(x,t))\rho_t(dx) = \int_{\mathbb{R}^d} \varphi(u)\tilde{\rho}_t(du),$$

where $\tilde{\rho}_t(A) = \rho_t\{Z(u,t) : u \in A\}$, $A \in \mathcal{B}(\mathbb{R}^d)$. It is easy to see that $\tilde{\rho}_t$, $t \geq 0$, is a continuous process in $\mathcal{P}(\mathbb{R}^d)$. Consequently, the above expression, which holds for every $\varphi \in C_c^2(\mathbb{R}^d)$, yields that $\tilde{\rho}_t = \mu_0$, $t \geq 0$, a.s. Hence, $\rho_t = \mu_0 \circ Z^{-1}(\cdot, t) = \bar{\rho}_t$, $t \geq 0$, a.s. This completes the proof of the lemma. $\square$

We now ready to prove the well-posedness of the stochastic mean-field equation for arbitrary initial conditions.

Proof of Theorem 2.26. Let μ_t, $t \geq 0$, be an arbitrary solution to the stochastic mean-field equation (2.1). Let also ψ, ρ, Y and Z be defined by (2.25), (2.28), (2.30), and (2.31), respectively, where the process μ_t, $t \geq 0$, is frozen in the coefficients V, A, $\tilde{A}$ and G, i.e, $v(t,x) = V(t,x,\mu_t)$, $a(t,x) = A(t,x,\mu_t)$, $\tilde{a}(t,x,y) = \tilde{A}(t,x,y,\mu_t)$ and $g(t,x,\theta) = G(t,x,\mu_t,\theta)$. We recall that $\rho_t = \mu_t \circ \psi^{-1}(t, \cdot)$, $t \geq 0$. Then by Lemmas 2.28 and 2.29, we obtain that a.s.

$$\rho_t = \mu_0 \circ Z^{-1}(\cdot, t) = \left(\mu_0 \circ Y^{-1}(\cdot, t)\right) \circ \psi^{-1}(t, \cdot), \quad t \geq 0.$$

Using that $\psi(t, \cdot)$ is a bijection, we conclude that $\mu_t = \mu_0 \circ Y^{-1}(\cdot, t)$, $t \geq 0$, a.s. Hence, Y is a solution to the SDE with interaction (2.3) due to the definition of Y and the fact that $\bar{\mu}_t = \mu_0 \circ Y^{-1}(\cdot, t) = \mu_t$ for all $t \geq 0$. Moreover, Y is also the unique solution, according to Theorem 2.9. This yields that μ_t, $t \geq 0$, is a superposition solution to (2.1). This completes the proof of the theorem since the superposition principle implies the uniqueness of the stochastic mean-field equation, by Corollary 2.12. $\square$

3. Limit theorems for the stochastic mean-field equation

The aim of this section is to prove an analog of the law of large numbers (LLN) and the central limit theorem (CLT) for solutions to the stochastic mean-field equation (2.1). More precisely, we provide a rate of convergence of the superposition solution μ_t^ε, $t \geq 0$, to

$$
(3.1) \quad
\begin{aligned}
d\mu_t^\varepsilon = {}& \frac{\varepsilon}{2} D^2 : (A(t, \cdot, \mu_t^\varepsilon)\mu_t^\varepsilon) dt - \nabla \cdot (V(t, \cdot, \mu_t^\varepsilon)\mu_t^\varepsilon) \, dt \\
& - \sqrt{\varepsilon} \int_\Theta \nabla \cdot (G(t, \cdot, \mu_t^\varepsilon, \theta)\mu_t^\varepsilon) \, W(d\theta, dt)
\end{aligned}
$$

started from $\mu_0^\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$ to the superposition solution μ_t^0, $t \geq 0$, to the PDE

$$
(3.2) \quad d\mu_t^0 = -\nabla \cdot \left(V(t, \cdot, \mu_t^0)\mu_t^0 \right) dt
$$

started from $\mu_0^0 \in \mathcal{P}_2(\mathbb{R}^d)$. We also show that the fluctuation field

$$
(3.3) \quad \eta_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}} \left(\mu_t^\varepsilon - \mu_t^0 \right), \quad t \geq 0,
$$

converges to a Gaussian process η_t^0, $t \geq 0$, which is a solution to the linear SPDE

$$
(3.4) \quad
\begin{aligned}
d\eta_t^0 = {}& -\nabla \cdot \left(V(t, \cdot, \mu_t^0)\eta_t^0 + \langle \tilde{V}(t, x, \cdot), \eta_t^0 \rangle_0 \mu_t^0(dx) \right) dt \\
& - \int_\Theta \nabla \cdot \left(G(t, \cdot, \mu_t^0, \theta)\mu_t^0 \right) W(d\theta, dt),
\end{aligned}
$$

and estimate the speed of the convergence.

3.1. Law of large numbers. In this section, we will prove the following theorem.

Theorem 3.1 (LLN for the stochastic mean-field equation). *Let the coefficients A, V, G of the stochastic mean-field equation (3.1) satisfy Assumptions 2.1 and 2.7. Furthermore, let μ_t^ε, $t \geq 0$, be a superposition solution to the SPDE (3.1) started from $\mu_0^\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$ for each $\varepsilon > 0$ and μ_t^0, $t \geq 0$, be a superposition solution to the PDE (3.2) started from $\mu_0^0 \in \mathcal{P}_2(\mathbb{R}^d)$. Then, for every $T > 0$ there exists a constant $C > 0$ such that*

$$
(3.5) \quad \mathbb{E} \sup_{t \in [0,T]} \mathcal{W}_2^2(\mu_t^\varepsilon, \mu_t^0) \leq C \left(\varepsilon \left(1 + \langle \phi_2, \mu_0^\varepsilon \rangle \right) + \mathcal{W}_2^2(\mu_0^\varepsilon, \mu_0^0) \right),
$$

where $\phi_2(x) = |x|^2$, $x \in \mathbb{R}^d$.

Proof. The proof is similar to the proof of Theorem 2.14. Since the processes μ_t^ε, $t \geq 0$, and μ_t^0, $t \geq 0$, are superposition solutions to (3.1) and (3.2), respectively, we have $\mu_t^\varepsilon = \mu_0^\varepsilon \circ X_\varepsilon(t, \cdot)$, $t \geq 0$, $\varepsilon \geq 0$, where X_ε are solutions to the corresponding SDEs with interaction (2.3). Using the Burkholder–Davis–Gundy inequality, Assumption 2.7 and Remark 2.8, we get for each $t \in [0, T]$ and $u, v \in \mathbb{R}^d$

$$
\mathbb{E} \sup_{s \in [0,t]} |X_\varepsilon(u, s) - X_0(v, s)|^2 \leq 3|u - v|^2
$$

$$+ 3T\mathbb{E} \int_0^t \left| V(s, X_\varepsilon(u, r), \mu_s^\varepsilon) - V(s, X_0(v, s), \mu_s^0) \right|^2 ds$$

$$+ 3\varepsilon\mathbb{E} \sup_{s \in [0,t]} \left| \int_0^s \int_\Theta G(r, X_\varepsilon(u, r), \mu_r^\varepsilon, \theta) W(d\theta, dr) \right|^2$$

$$\leq 3|u - v|^2 + C \int_0^t \left(\mathbb{E}|X_\varepsilon(u, s) - X_0(v, s)|^2 + \mathbb{E}\mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) \right) ds$$

$$+ \varepsilon C \mathbb{E} \int_0^t \|G(s, X_\varepsilon(u, s), \mu_s^\varepsilon)\|_\vartheta^2 \, ds$$

$$\leq 3|u - v|^2 + C \int_0^t \mathbb{E} \sup_{r \in [0,s]} |X_\varepsilon(u, r) - X_0(v, r)|^2 ds + C \int_0^t \mathbb{E}\mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) ds$$

$$+ \varepsilon 3CL^2 \int_0^t \mathbb{E} \left(1 + |X_\varepsilon(u, s)|^2 + \mathcal{W}_2^2(\mu_s^\varepsilon, \delta_0) \right) ds,$$

where C is independent of u, v, t and ε. We note that

$$\mathcal{W}_2^2(\mu_s^\varepsilon, \delta_0) = \int_{\mathbb{R}^d} |x|^2 \mu_s^\varepsilon(dx) = \int_{\mathbb{R}^d} |X_\varepsilon(\tilde{u}, s)|^2 \mu_0^\varepsilon(d\tilde{u}).$$

Hence, by Gronwall's lemma, we conclude

$$\mathbb{E} \sup_{s \in [0,t]} |X_\varepsilon(u, s) - X_0(v, s)|^2 \leq C|u - v|^2 + C \int_0^t \mathbb{E}\mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) ds$$

$$+ \varepsilon C \int_0^t \mathbb{E} \left(1 + |X_\varepsilon(u, s)|^2 + \int_{\mathbb{R}^d} |X_\varepsilon(\tilde{u}, s)|^2 \mu_0^\varepsilon(d\tilde{u}) \right) ds$$

$$\leq C|u - v|^2 + C \int_0^t \mathbb{E}\mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) ds + \varepsilon C \left(1 + |u|^2 + \langle \phi_2, \mu_0^\varepsilon \rangle \right) ds$$

for all $t \in [0, T]$, $u, v \in \mathbb{R}^d$ and $\varepsilon > 0$, where C depends only on L and T. Note that in the last step of the inequality, we have used Theorem 2.9.

For fixed $\varepsilon > 0$, let χ be an arbitrary probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_0^ε and μ_0^0. Let also $\chi_s(B) = \chi\{(u, v) : (X_\varepsilon(u, s), X_0(v, s)) \in B\}$, $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$. Then, for $t \in [0, T]$, we have the following

$$\mathbb{E} \sup_{s \in [0,t]} \mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) \leq \mathbb{E} \sup_{s \in [0,t]} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x - y|^2 \chi_s(dx, dy)$$

$$\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E} \sup_{s \in [0,t]} |X_\varepsilon(u, s) - X_0(v, s)|^2 \chi(du, dv)$$

$$\leq C \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |u - v|^2 \chi(du, dv) + C \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_0^t \mathbb{E}\mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) ds \chi(du, dv)$$

$$+ \varepsilon C \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E} \left(1 + |u|^2 + \langle \phi_2, \mu_0^\varepsilon \rangle \right) \chi(du, ds)$$

$$= C \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |u - v|^2 \chi(du, dv) + C \int_0^t \mathbb{E}\mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) ds$$

$$+ \varepsilon C \left(1 + 2\langle \phi_2, \mu_0^\varepsilon \rangle \right)$$

$$\leq C \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |u - v|^2 \chi(du, dv) + C \int_0^t \mathbb{E} \sup_{r \in [0,s]} \mathcal{W}_2^2(\mu_r^\varepsilon, \mu_r^0) ds$$

$$+ \varepsilon C \left(1 + 2\langle\phi_2, \mu_0^\varepsilon\rangle\right).$$

Taking the infimum over all probability measures χ with marginals μ_0^ε and μ_0^0, we obtain

$$\mathbb{E} \sup_{s\in[0,t]} \mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0) \leq C_1 \mathcal{W}_2^2(\mu_0^\varepsilon, \mu_0^0) + C_1 \int_0^t \mathbb{E} \sup_{r\in[0,s]} \mathcal{W}_2^2(\mu_r^\varepsilon, \mu_r^0) ds$$
$$+ \varepsilon C_2 \left(1 + 2\langle\phi_2, \mu_0^\varepsilon\rangle\right)$$

for all $t \in [0, T]$. Using Gronwall's lemma again, we get the required inequality. This completes the proof of the theorem. $\square$

Remark 3.2. Combining the estimate for $\mathbb{E} \sup_{s\in[0,t]} |X_\varepsilon(u, s) - X_0(v, s)|^2$ from the proof of Theorem 3.1 and the inequality (3.5) from the statement of Theorem 3.1, one gets for every $t \in [0, T]$ and $\varepsilon > 0$

$$\mathbb{E} \sup_{s\in[0,t]} |X_\varepsilon(u, s) - X_0(v, s)|^2 \leq C \left(|u - v|^2 + \mathcal{W}_2^2(\mu_0^\varepsilon, \mu_0^0) + \varepsilon \left(1 + |u|^2 + \langle\phi_2, \mu_0^\varepsilon\rangle\right)\right),$$

where the constant C depends only on T, L and d.

3.2. **Central limit theorem.** We note that Theorem 3.1 implies that $\mu^\varepsilon \to \mu^0$ as $\mu_0^\varepsilon \to \mu_0^0$ and $\varepsilon \to 0$. It this section, we will consider the fluctuations $\eta_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}} \left(\mu_t^\varepsilon - \mu_t^0\right)$ of μ_t^ε, $t \geq 0$, around μ_t^0, $t \geq 0$, as $\varepsilon \to 0$. Note that the process η_t^ε, $t \geq 0$, takes values in the space $\mathcal{M}(\mathbb{R}^d)$ of all signed measures on $\mathbb{R}^d$ with finite total variation. Since the Sobolev embedding theorem (see [1, Theorem 4.12]) tells us that the space $H^J(\mathbb{R}^d)$ is continuously embedded into $C_b^m(\mathbb{R}^d)$ for any $J > \frac{d}{2} + m$, for every $J > \frac{d}{2}$ and $\varepsilon > 0$ the process η_t^ε, $t \geq 0$, can be considered as a continuous process in $H^{-J}(\mathbb{R}^d)$, by the identification $\langle\varphi, \eta_t^\varepsilon\rangle_0 := \langle\varphi, \eta_t^\varepsilon\rangle$, $\varphi \in H^J(\mathbb{R}^d)$, where $\langle\cdot, \cdot\rangle_0$ denotes the dualization between $H^J(\mathbb{R}^d)$ and $H^{-J}(\mathbb{R}^d)$. Moreover, a simple computation shows that

$$d\eta_t^\varepsilon = \frac{\sqrt{\varepsilon}}{2} D^2 : (A(t, \cdot, \mu_t^\varepsilon)\mu_t^\varepsilon) - \nabla \cdot \left(V(t, \cdot, \mu_t^\varepsilon)\eta_t^\varepsilon + \langle\tilde{V}(t, x, \cdot), \eta_t^\varepsilon\rangle_0 \mu_t^0(dx)\right) dt$$
$$- \int_\Theta \nabla \cdot (G(t, \cdot, \mu_t^\varepsilon, \theta)\mu_t^\varepsilon) W(d\theta, dt),$$

where we assume that $V(t, x, \mu) = \bar{V}(t, x) + \langle\tilde{V}(t, x, \cdot), \mu\rangle$. Passing formally to the limit as $\varepsilon \to 0$, we expect that the limit of η_t^ε, $t \geq 0$, is a solution to the same equation with $\varepsilon = 0$.

Therefore, the main goal of this section is to show that η_t^ε, $t \geq 0$, converges to the solution η_t^0, $t \geq 0$, to the linear SPDE (3.4) and to estimate the speed of convergence. We first prove some auxiliary statements and study the well posedness of the linear SPDE (3.4). We set $\Gamma_n = (-n, n)^d$ for $n \in \bar{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$. We start with an auxiliary technical lemma that will prove useful in the proofs of the later results.

Lemma 3.3. Let $J \in \mathbb{N}$, $n \in \bar{\mathbb{N}}$ and $v = (v_1, \ldots, v_d)$ with $v_i \in C_b^J(\Gamma_n)$, $i \in [d]$. Then the map $\mathcal{G} : H^{-J+1}(\Gamma_n) \to H^{-J}(\Gamma_n)$ defined by

$$\langle\varphi, \mathcal{G}(f)\rangle_{0,\Gamma_n} = \langle\nabla\varphi \cdot v, f\rangle_{0,\Gamma_n}, \quad f \in H^{-J+1}(\Gamma_n), \ \varphi \in C_c^\infty(\Gamma_n),$$

satisfies

(3.6) $$|\langle\mathcal{G}(f), f\rangle_{-J,\Gamma_n}| \leq C \max_{i\in[d]} \|v_i\|_{C_b^J} \|f\|_{-J,\Gamma_n}^2$$

for all $f \in H^{-J+1}(\Gamma_n)$, where the constant C depends only on J and d.

Proof. We first remark that $\mathcal{G}$ maps $H^{-J+1}(\Gamma_n)$ to $H^{-J}(\Gamma_n)$. Indeed, we have

$$\|\mathcal{G}(f)\|_{H^{-J}} = \sup_{\varphi \in C_c^\infty(\Gamma_n)} \frac{\langle \nabla\varphi \cdot v, f\rangle_{0,\Gamma_n}}{\|\varphi\|_J}$$

$$\leq \sup_{\varphi \in C_c^\infty(\Gamma_n)} \frac{\|\nabla\varphi \cdot v\|_{J-1}}{\|\varphi\|_J}\|f\|_{-J+1} \leq C \max_{i\in[d]} \|v_i\|_{C_b^{J-1}}\|f\|_{-J+1}$$

for all $f \in H^{-J+1}$.

Since $H^{-J}(\Gamma_n)$ is the dual space of $H^J(\Gamma_n)$, for every $f \in H^{-J+1}(\Gamma_n) \subset H^{-J}(\Gamma_n)$ there exists a unique function $\tilde{f} = L_{J,\Gamma_n}^{-1}f \in H^J(\Gamma_n)$ such that $\langle\varphi, f\rangle_{0,\Gamma_n} = \langle\varphi, \tilde{f}\rangle_{J,\Gamma_n}$ for any $\varphi \in C_c^\infty(\Gamma_n)$. We set

$$R := \left\{ f \in H^{-J}(\Gamma_n) : \ \tilde{f} \in C_c^\infty(\Gamma_n) \right\} \subset H^{-J+1}(\Gamma_n)$$

and compute

$$\langle\mathcal{G}(f), f\rangle_{-J,\Gamma_n} = \langle\tilde{f}, \mathcal{G}(f)\rangle_{0,\Gamma_n} = \langle\nabla\tilde{f}\cdot v, f\rangle_{0,\Gamma_n} = \langle\nabla\tilde{f}\cdot v, \tilde{f}\rangle_{J,\Gamma_n}$$

$$= \sum_{|\alpha|\leq J} \int_{\Gamma_n} D^\alpha\left(\nabla\tilde{f}(x)\cdot v(x)\right) D^\alpha\tilde{f}(x)dx$$

$$= \sum_{i=1}^d \sum_{|\alpha|\leq J} \int_{\Gamma_n} D^\alpha\left(\partial_i\tilde{f}(x)v_i(x)\right) D^\alpha\tilde{f}(x)dx.$$

For $\beta, \alpha \in \mathbb{N}_0^d$ we say $\beta \leq \alpha$ if $\beta_i \leq \alpha_i$ for all $i \in [d]$. We also set $\binom{\alpha}{\beta} = \binom{\alpha_1}{\beta_1}\ldots\binom{\alpha_d}{\beta_d}$ for $\beta \leq \alpha$. Then, for every multi-index α with $|\alpha| = J$, we compute

$$D^\alpha\left(\partial_i\tilde{f}(x)v_i(x)\right) = \sum_{\beta\leq\alpha} \binom{\alpha}{\beta} D^\beta\left(\partial_i\tilde{f}(x)\right) D^{\alpha-\beta}v_i(x)$$

$$= \sum_{\beta<\alpha} \binom{\alpha}{\beta} D^\beta\left(\partial_i\tilde{f}(x)\right) D^{\alpha-\beta}v_i(x) + v_i(x)D^\alpha\partial_i\tilde{f}(x),$$

where $\beta < \alpha$ means $\beta \leq \alpha$ and $|\beta| < |\alpha|$. Hence,

$$\langle\mathcal{G}(f), f\rangle_{-J,\Gamma_n} = \sum_{i=1}^d \sum_{|\alpha|<J} \sum_{\beta\leq\alpha} \binom{\alpha}{\beta} \int_{\Gamma_n} D^\beta\left(\partial_i\tilde{f}(x)\right) D^{\alpha-\beta}v_i(x)D^\alpha\tilde{f}(x)dx$$

$$(3.7) \qquad + \sum_{i=1}^d \sum_{|\alpha|=J} \sum_{\beta<\alpha} \binom{\alpha}{\beta} \int_{\Gamma_n} D^\beta\left(\partial_i\tilde{f}(x)\right) D^{\alpha-\beta}v_i(x)D^\alpha\tilde{f}(x)dx$$

$$+ \sum_{i=1}^d \sum_{|\alpha|=J} \int_{\Gamma_n} v_i(x)D^\alpha\partial_i\tilde{f}(x)D^\alpha\tilde{f}(x)dx.$$

Using the fact that $\tilde{f} \in C_c^\infty(\Gamma_n)$ and the integration by parts, the last integral in the equality above can be rewritten as follows

$$\int_{\Gamma_n} v_i(x)D^\alpha\partial_i\tilde{f}(x)D^\alpha\tilde{f}(x)dx = \frac{1}{2}\int_{\Gamma_n} v_i(x)\partial_i\left(D^\alpha\tilde{f}(x)\right)^2 dx$$

$$(3.8) \qquad = -\frac{1}{2}\int_{\Gamma_n} \partial_i v_i(x)\left(D^\alpha\tilde{f}(x)\right)^2 dx.$$

Thus, Young's inequality and the equalities (3.7), (3.8) yield

$$|\langle \mathcal{G}(f), f\rangle_{-J,\Gamma_n}| \leq C \max_{i\in[d]} \|v_i\|_{\mathrm{C}_b^J(\Gamma_n)} \|\tilde{f}\|_{J,\Gamma_n}^2 = C \max_{i\in[d]} \|v_i\|_{\mathrm{C}_b^J(\Gamma_n)} \|f\|_{-J,\Gamma_n}^2,$$

for every $f \in R$ and some constant depending only on d and J. We now prove the estimate (3.6) for all $f \in H^{-J+1}$. Since the set R is dense in $H^{-J+1}(\Gamma_n)$, there exists a sequence $\{f_m,\ m \geq 1\} \subset R$ which converges to f in $H^{-J+1}(\Gamma_n)$. Since $\mathcal{G}$ is a bounded linear operator, $\mathcal{G}(f_m) \to \mathcal{G}(f)$ in $H^{-J}(\Gamma_n)$. Moreover, $f_m \to f$ in $H^{-J}(\Gamma_n)$ due to the continuous embedding of $H^{-J+1}(\Gamma_n)$ into $H^{-J}(\Gamma_n)$. Hence,

$$|\langle \mathcal{G}(f), f\rangle_{-J,\Gamma_n}| = \lim_{m\to\infty} |\langle \mathcal{G}(f_m), f_m\rangle_{-J,\Gamma_n}| \leq \lim_{m\to\infty} C \max_{i\in[d]} \|v_i\|_{\mathrm{C}_b^J(\Gamma_n)} \|f_m\|_{-J,\Gamma_n}^2$$

$$= C \max_{i\in[d]} \|v_i\|_{\mathrm{C}_b^J(\Gamma_n)} \|f\|_{-J,\Gamma_n}^2.$$

This completes the proof of the lemma. $\qquad\square$

We now prove the well-posedness of the linear SPDE (3.4).

Proposition 3.4. *Let $J \geq \frac{d}{2} + 3$ and $\bar{v} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\tilde{v} : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, $g : [0,T] \times \mathbb{R}^d \to L_2(\Theta, \vartheta)$ be measurable functions such that for every $t \in [0,T]$, $\bar{v}(t,\cdot) \in \mathrm{C}^J(\mathbb{R}^d)$, $\tilde{v}(t,\cdot,\cdot) \in \mathrm{C}^J(\mathbb{R}^d \times \mathbb{R}^d)$, and*

(3.9)
$$\sup_{t\in[0,T]} \|\bar{v}(t,\cdot)\|_{\mathrm{C}_b^J} + \sup_{t\in[0,T],y\in\mathbb{R}^d} \|\tilde{v}(t,\cdot,y)\|_{\mathrm{C}_b^J}$$
$$+ \sup_{t\in[0,T],x\in\mathbb{R}^d} \|\tilde{v}(t,x,\cdot)\|_J + \sup_{t\in[0,T],x\in\mathbb{R}^d} \frac{\|g(t,x,\cdot)\|_\vartheta}{1+|x|} < \infty.$$

Let also μ_t, $t \in [0,T]$, be a continuous curve in $\mathcal{P}_2(\mathbb{R}^d)$, $\varrho \in H^{-J+1}(\mathbb{R}^d)$, and $v(t,x) = \bar{v}(t,x) + \langle \tilde{v}(t,x,\cdot), \mu_t\rangle$. Then, there exists a unique $(\mathcal{F}_t)$-adapted continuous process in $H^{-J}(\mathbb{R}^d)$ such that for every $\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)$ we have a.s.

(3.10)

$$\langle \varphi, \eta_t\rangle_0 = \langle \varphi, \varrho\rangle_0 + \int_0^t \left(\langle \nabla\varphi \cdot v(s,\cdot), \eta_s\rangle_{0,\mathbb{R}^d} + \langle \nabla\varphi \cdot \langle \tilde{v}(s,x,\cdot), \eta_s\rangle_{0,\mathbb{R}^d}, \mu_s(dx)\rangle \right) ds$$

$$+ \int_0^t \int_\Theta \langle \nabla\varphi \cdot g(s,\cdot,\theta), \mu_s\rangle\, W(d\theta, ds), \quad t \in [0,T].$$

Moreover, η_t, $t \geq 0$, is a Gaussian process in $H^{-J}(\mathbb{R}^d)$ and

(3.11)
$$\mathbb{E} \sup_{t\in[0,T]} \|\eta_t\|_{-J,\mathbb{R}^d}^2 < \infty.$$

Proof. To prove the well-posedness of (3.10), we will use the general theory developed in [31, 64], where one needs to check that the coefficients of the equation satisfy some conditions (see (H0)-(H4) below for more details). Unfortunately, working in the Sobolev space $H^{-J}(\mathbb{R}^d)$, we will not be able to verify (H0). In order to overcome this problem, we will first construct solutions to cut-off versions of the SPDE in the Sobolev spaces $H^{-J}(\Gamma_n)$ for all $G_n = (-n,n)^d$, $n \in \mathbb{N}$, and then pass to the limit as $n \to \infty$.

Since for every $n \in \bar{\mathbb{N}}$ the space $\mathcal{V}_n := H^{-J+1}(\Gamma_n)$ is continuously and densely embedded into the Hilbert space $\mathcal{H}_n := H^{-J}(\Gamma_n)$,

$$\mathcal{V}_n \subset \mathcal{H}_n \cong \mathcal{H}_n^* \subset \mathcal{V}_n^*$$

is a Gelfand triple, where $\mathcal{V}_n^*$ is the dual space of $\mathcal{V}_n$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_n} = \langle \cdot, \cdot \rangle_{-J, \Gamma_n}$. In particular,

$$\langle f, \varphi \rangle_{\mathcal{V}_n^* \mathcal{V}_n} = \langle f, \varphi \rangle_{\mathcal{H}_n}, \quad f \in \mathcal{H}_n, \ \varphi \in \mathcal{V}_n.$$

Let $\mathcal{L}_2 \left(L_2(\Theta, \vartheta); \mathcal{V}_n \right)$ denote the space of all Hilbert–Schmidt operators from $L_2(\Theta, \vartheta)$ to $\mathcal{V}_n$ and $\| \cdot \|_{\mathrm{HS}, \mathcal{V}_n}$ be the Hilbert–Schmidt norm on that space. For the cylindrical Brownian motion W_t, $t \geq 0$, in $L_2(\Theta, \vartheta)$ and fixed $n \in \bar{\mathbb{N}}$, we consider the following equation

$$(3.12) \qquad d\xi_t^n = U_n(t, \xi_t^n)dt + B_n(t)dW_t,$$

where $U_n = \bar{U}_n + \tilde{U}_n : [0, T] \times \mathcal{V}_n \to \mathcal{H}_n \subset \mathcal{V}_n^*$ and $B_n : [0, T] \to \mathcal{L}_2 \left(L_2(\Theta, \vartheta); \mathcal{V}_n \right)$ are measurable functions defined by

$$\langle \bar{U}_n(t, f), \varphi \rangle_{0, \Gamma_n} = \langle \nabla \varphi \cdot v(t, \cdot), f \rangle_{0, \Gamma_n},$$
$$\langle \tilde{U}_n(t, f), \varphi \rangle_{0, \Gamma_n} = \langle \nabla \varphi(x) \cdot \langle \kappa_n \tilde{v}(t, x, \cdot), f \rangle_{0, \Gamma_n}, \mu_t(dx) \rangle$$

and

$$\langle B_n(t)h, \varphi \rangle_{0, \Gamma_n} = \langle h, \langle \nabla \varphi(x) \cdot g(t, x, \cdot), \mu_t(dx) \rangle \rangle_\vartheta$$

for all $\varphi \in \mathrm{C}_c^\infty(\Gamma_n)$, $f \in \mathcal{V}_n$ and $h \in L_2(\Theta, \vartheta)$, where we extend the function φ to the whole space $\mathbb{R}^d$ by zero in the integrals with respect to μ_t, the function $\kappa_n \in \mathrm{C}_c^\infty(\mathbb{R}^d)$ has a compact support in Γ_n and $\kappa_n = 1$ on Γ_{n-1} for every $n \in \mathbb{N}$, $\kappa_\infty = 1$ on $\mathbb{R}^d$, and $\sup_{n \geq 1} \|\kappa_n\|_{\mathrm{C}_b^{J-1}} < \infty$. The fact that U_n and $B_n(t)h$ take values in $\mathcal{H}_n$ and $\mathcal{V}_n$, respectively, follows from the continuous embedding of the Sobolev space $H^{-J}(\Gamma_n)$ into $\mathrm{C}_b^1(\Gamma_n)$.

To prove the well-posedness of the equation (3.12), we need to check the following conditions for all $n \in \mathbb{N}$.

(H0) There exists an orthogonal set $\{e_1, e_2, \dots\}$ in $\mathcal{V}_n$ such that it constitutes an orthonormal basis of $\mathcal{H}_n$.

(H1) (Hemicontinuity) The map $s \mapsto {}_{\mathcal{V}_n^*}\langle U_n(t, f_1 + sf_2), f \rangle_{\mathcal{V}_n}$ is continuous on $\mathbb{R}$ for any $f_1, f_2, f \in \mathcal{V}_n$ and $t \in [0, T]$.

(H2) (Monotonicity) There exists a constant $C > 0$ such that

$${}_{\mathcal{V}_n^*}\langle U_n(t, f_1) - U_n(t, f_2), f_1 - f_2 \rangle_{\mathcal{V}_n} \leq C\|f_1 - f_2\|_{\mathcal{H}_n}^2$$

for all $t \in [0, T]$ and $f_1, f_2 \in \mathcal{V}_n$.

(H3) (One-side linear growth) For any $k \in \mathbb{N}$ the operator $U_n(t, \cdot)$ maps the set $E^k := \mathrm{span}\{e_1, \dots, e_k\}$ into $\mathcal{V}_n$, and there exists a constant $C > 0$ such that

$$\langle U_n(t, f), f \rangle_{\mathcal{V}_n} \leq C(1 + \|f\|_{\mathcal{V}_n}^2)$$

for all $f \in E^k$ and some constant independent of k and t.

(H4) (Growth) There exists a constant $C > 0$ such that

$$\|U_n(t, f)\|_{\mathcal{V}_n^*} \leq C(1 + \|f\|_{\mathcal{V}_n}), \quad \|B_n(t)\|_{\mathrm{HS}, \mathcal{V}_n}^2 \leq C$$

for all $t \in [0, T]$ and $f \in \mathcal{V}_n$.

The condition (H0) follows from the compact embedding of the space $\mathcal{V}_n$ into $\mathcal{H}_n$. Indeed, due to the fact that the inner product $\langle \cdot, \cdot \rangle_{\mathcal{V}_n}$ induces a closed quadratic form on $\mathcal{H}_n$, there exists a densely defined self-adjoint operator $L : \mathcal{H}_n \supset D(L) \to \mathcal{H}_n$ on $\mathcal{H}_n$ such that $\langle f_1, f_2 \rangle_{\mathcal{V}_n} = \langle f_1, Lf_2 \rangle_{\mathcal{H}_n}$ for all $f_1 \in \mathcal{V}_n$ and $f_2 \in D(L)$. By the compact embedding of $\mathcal{V}_n$ into $\mathcal{H}_n$, the operator L has a discrete spectrum (see also [31, Section 2.1]) $\{\lambda_k, \ k \geq 1\}$ with corresponding eigenbasis $\{e_k, \ k \geq 1\}$ in $\mathcal{H}_n$. It is easy to see that $\{e_k, \ k \geq 1\}$ is an orthogonal system in $\mathcal{V}_n$ and $e_k \in H^{-J+2}(\Gamma_n)$ for every $k \geq 1$.

The condition (H1) holds due to the linearity of $U_n(t, \cdot)$ for every $t \in [0, T]$.

In order to check the monotonicity of U_n, we estimate for $f \in \mathcal{V}_n$ and $t \in [0, T]$

$$_{\mathcal{V}_n^*}\langle U_n(t, f), f \rangle_{\mathcal{V}_n} = \langle U_n(t, f), f \rangle_{\mathcal{H}_n} = \langle \bar{U}_n(t, f), f \rangle_{\mathcal{H}_n} + \langle \tilde{U}_n(t, f), f \rangle_{\mathcal{H}_n}.$$

By Lemma 3.3,

$$\langle \bar{U}_n(t, f), f \rangle_{\mathcal{H}_n} = \langle \bar{U}_n(t, f), f \rangle_{-J, \Gamma_n} \leq C \max_{i \in [d]} \sup_{t \in [0, T]} \|v_i(t, \cdot)\|_{\mathrm{C}_b^J} \|f\|_{-J, \Gamma_n}^2,$$

where the constant C is independent of n. We remark that $\sup_{t \in [0,T]} \|v_i(t, \cdot)\|_{\mathrm{C}_b^J}$ is finite according to (3.9). For every $f \in \mathcal{V}_n \subset H^{-J}(\Gamma_n)$ we define $\tilde{f} = L_{J, \Gamma_n}^{-1} f \in H^J(\Gamma_n)$ and note that $\langle \tilde{f}, \varphi \rangle_{J, \Gamma_n} = \langle f, \varphi \rangle_{0, \Gamma_n}$ for all $\varphi \in \mathrm{C}_c^\infty(\Gamma_n)$. By [1, Theorem 5.29], the extension by zero of $\tilde{f}$ on $\mathbb{R}^d$, which we also denote $\tilde{f}$, belongs to $H^J(\mathbb{R}^d)$. We next estimate

$$
\begin{aligned}
\langle \tilde{U}_n(t, f), f \rangle_{\mathcal{H}_n} = \langle \tilde{U}_n(t, f), f \rangle_{-J, \Gamma_n} &= \langle \tilde{U}_n(t, f), \tilde{f} \rangle_{0, \Gamma_n} \\
&= \left\langle \nabla \tilde{f}(x) \cdot \langle \kappa_n \tilde{v}(t, x, \cdot), f \rangle_{0, \Gamma_n}, \mu_t(dx) \right\rangle \\
&\leq \sum_{i=1}^d \int_{\mathbb{R}^d} \left| \partial_i \tilde{f}(x) \right| \|\kappa_n \tilde{v}_i(t, x, \cdot)\|_{J, \Gamma_n} \|f\|_{-J, \Gamma_n} \mu_t(dx) \\
&\leq \|\tilde{f}\|_{\mathrm{C}_b^1} \|f\|_{-J, \Gamma_n} \sum_{i=1}^d \sup_{t \in [0,T], x \in \mathbb{R}^d} \|\tilde{v}_i(t, x, \cdot)\|_{J, \mathbb{R}^d}.
\end{aligned}
$$

(3.13)

By the continuous embedding of $H^J(\mathbb{R}^d)$ into $\mathrm{C}_b^1(\mathbb{R}^d)$, we obtain

$$\|\tilde{f}\|_{\mathrm{C}_b^1} \leq C \|\tilde{f}\|_{J, \mathbb{R}^d} = C \|\tilde{f}\|_{J, \Gamma_n} = C \|f\|_{-J, \Gamma_n}.$$

Thus, the linearity of $U_n(t, \cdot)$ and the estimates above imply the monotonicity of U_n with the constant C independent of n. This establishes condition (H2).

The function $U_n(t, \cdot)$ maps $E^k = \mathrm{span}\{e_1, \ldots, e_k\}$ to $\mathcal{V}_n$ due to the fact that $e_k \in H^{-J+2}(\Gamma_n)$ for all $k \geq 1$. Moreover, Lemma 3.3 and a similar computation to (3.13) imply

$$\langle U_n(t, f), f \rangle_{\mathcal{V}_n} \leq C \max_{i \in [d]} \sup_{t \in [0,T], x \in \mathbb{R}^d} \left(\|v_i(t, \cdot)\|_{\mathrm{C}_b^{J-1}} + \|\tilde{v}_i(t, x, \cdot)\|_{J-1, \mathbb{R}^d} \right) \|f\|_{\mathcal{V}_n}^2$$

for all $f \in E^k$, where the constant C is independent of n and k. Thus, condition (H3) holds.

We separately estimate the norms of $\bar{U}_n$ and $\tilde{U}_n$ in $\mathcal{H}_n$ as follows

$$
\begin{aligned}
\|\bar{U}_n(t, f)\|_{\mathcal{H}_n} &= \sup_{\varphi \in \mathrm{C}_c^\infty(\Gamma_n)} \frac{1}{\|\varphi\|_{J, \Gamma_n}} \langle \nabla \varphi \cdot v(t, \cdot), f \rangle_{0, \Gamma_n} \\
&\leq \sup_{\varphi \in \mathrm{C}_c^\infty(\Gamma_n)} \frac{\|\nabla \varphi \cdot v(t, \cdot)\|_{J-1, \Gamma_n}}{\|\varphi\|_{J, \Gamma_n}} \|f\|_{\mathcal{V}_n} \\
&\leq C \max_{i \in [d]} \sup_{t \in [0,T]} \|v_i(t, \cdot)\|_{\mathrm{C}_b^{J-1}} \|f\|_{\mathcal{V}_n},
\end{aligned}
$$

for all $f \in \mathcal{V}_n$ and $t \in [0, T]$, by the continuous embedding of $H^J(\mathbb{R}^d)$ into $\mathrm{C}_b^1(\mathbb{R}^d)$. Similarly,

$$
\begin{aligned}
\|\tilde{U}_n(t, f)\|_{\mathcal{H}_n} &= \sup_{\varphi \in \mathrm{C}_c^\infty(\Gamma_n)} \frac{1}{\|\varphi\|_{J, \Gamma_n}} \langle \nabla \varphi(x) \cdot \langle \kappa_n \tilde{v}(t, x, \cdot), f \rangle_{0, \Gamma_n}, \mu_t(dx) \rangle \\
&\leq C \max_{i \in [d]} \sup_{t \in [0,T], x \in \mathbb{R}^d} \|\kappa_n \tilde{v}_i(t, x, \cdot)\|_{J-1, \mathbb{R}^d} \|f\|_{\mathcal{V}_n} \\
&\leq C \|\kappa_n\|_{\mathrm{C}_b^{J-1}} \max_{i \in [d]} \sup_{t \in [0,T], x \in \mathbb{R}^d} \|\tilde{v}_i(t, x, \cdot)\|_{J-1, \mathbb{R}^d} \|f\|_{\mathcal{V}_n}
\end{aligned}
$$

$$\leq C \max_{i \in [d]} \sup_{t \in [0,T], x \in \mathbb{R}^d} \|\tilde{v}_i(t, x, \cdot)\|_{J-1, \mathbb{R}^d} \|f\|_{\mathcal{V}_n}$$

for all $f \in \mathcal{V}_n$ and $t \in [0, T]$, where the constant C is independent of n due to the specific choice of κ_n.

To estimate the Hilbert–Schmidt norm of $B_n(t)$, we consider an orthonormal basis $\{h_k, \ k \geq 1\}$ in $L_2(\Theta, \vartheta)$ and compute for every $k \geq 1$ and $\varphi \in \mathrm{C}_c^\infty(\Gamma_n)$

$$\langle B_n(t) h_k, \varphi \rangle_{0, \Gamma_n} = \langle h_k, \langle \nabla \varphi(x) \cdot g(t, x, \cdot), \mu_t(dx) \rangle \rangle_\vartheta = \left\langle \nabla \varphi \cdot g^k(t, \cdot), \mu_t \right\rangle,$$

where $g^k(t, x) = \langle g(t, x, \cdot), h_k \rangle_\vartheta$. Hence, by the continuous embedding of $H^{J-1}(\mathbb{R}^d)$ into $\mathrm{C}_b^1(\mathbb{R}^d)$ due to $J \geq \frac{d}{2} + 3$, we obtain

$$\|B_n(t) h_k\|_{\mathcal{V}_n} = \sup_{\varphi \in \mathrm{C}_c^\infty(\Gamma_n)} \frac{\left\langle \nabla \varphi \cdot g^k(t, \cdot), \mu_t \right\rangle}{\|\varphi\|_{J-1, \Gamma_n}} \leq C \langle |g^k(t, \cdot)|, \mu_t \rangle,$$

where the constant C is independent of n. Consequently, by Jensen's inequality and Fubini's theorem,

$$\|B_n(t)\|_{\mathrm{HS}, \mathcal{V}_n}^2 = \sum_{k=1}^\infty \|B_n(t) h_k\|_{\mathcal{V}_n}^2 \leq C^2 \sum_{k=1}^\infty \langle |g^k(t, \cdot)|, \mu_t \rangle^2$$

$$\leq C^2 \sum_{i=1}^d \left\langle \sum_{k=1}^\infty (g_i^k(t, \cdot))^2, \mu_t \right\rangle = C^2 \sum_{i=1}^d \left\langle \|g_i(t, x, \cdot)\|_\vartheta^2, \mu_t(dx) \right\rangle.$$

We note that $\sup_{t \in [0,T]} \langle \|g_i(t, x, \cdot)\|_\vartheta^2, \mu_t(dx) \rangle < \infty$ due to (3.9). Thus, condition (H4) holds true with a constant C that is independent of n.

Now we can apply [64, Theorem 1.1], to obtain that, for every $n \geq 1$ and $\varrho \in H^{-J+1} \subset \mathcal{V}_n$, there exists a continuous $\mathcal{H}_n$-valued $(\mathcal{F}_t)$-adapted process η_t^n, $t \in [0, T]$, which has a $dt \times \mathbb{P}$-version $\bar{\eta}^n$ from $L^2([0, T] \times \Omega, dt \times \mathbb{P}; V)$ and $\mathbb{P}$-a.s.

$$\eta_t^n = \varrho + \int_0^t U_n(s, \bar{\eta}_s^n) ds + \int_0^t B_n(s) dW_s, \quad t \in [0, T].$$

Using [64, Lemma 2.1], one can see that

$$\mathbb{E} \sup_{t \in [0,T]} \|\eta_t^n\|_{\mathcal{H}_n}^2 + \mathbb{E} \int_0^T \|\bar{\eta}_t^n\|_{\mathcal{V}_n}^2 dt \leq C(1 + \|\varrho\|_{\mathcal{V}_n}^2),$$

where the constant C is independent of n because all constants in the conditions (H1)-(H4) are independent of n. Since for every $m \geq n$ the relationships $\mathcal{H}_m \subset \mathcal{H}_n$, $\mathcal{V}_m \subset \mathcal{H}_n$, and $\|\cdot\|_{\mathcal{H}_n} \leq \|\cdot\|_{\mathcal{H}_m}$, $\|\cdot\|_{\mathcal{V}_n} \leq \|\cdot\|_{\mathcal{V}_m}$ hold, we get

$$\mathbb{E} \sup_{t \in [0,T]} \|\eta_t^m\|_{\mathcal{H}_n}^2 + \mathbb{E} \int_0^T \|\bar{\eta}_t^m\|_{\mathcal{V}_n}^2 dt \leq C(1 + \mathbb{E}\|\varrho\|_{\mathcal{V}_m}^2) \leq C(1 + \|\varrho\|_{\mathcal{V}_\infty}^2)$$

for all $m \geq n$. Hence the sequence $\{\eta^m, \ m \geq n\}$ is relatively compact in the weak topologies of $L_2(\mathcal{H}_n) := L_2([0, T] \times \Omega, dt \times \mathbb{P}; \mathcal{H}_n)$, and $L_2(\mathcal{V}_n) := L_2([0, T] \times \Omega, dt \times \mathbb{P}; \mathcal{V}_n)$ for every $n \geq 1$. Using a diagonal argument, there exists a subsequence $\{m_k, \ k \geq 1\}$ such that $\eta^{m_k} \to \bar{\eta}$ in the weak topologies of $L_2(\mathcal{H}_n)$ and $L_2(\mathcal{V}_n)$ for every $n \geq 1$. In particular,

$$\mathbb{E} \int_0^T \|\bar{\eta}_t\|_{\mathcal{V}_n}^2 dt \leq C(1 + \|\varrho\|_{\mathcal{V}_\infty}^2)$$

for all $n \geq 1$. Using Fatou's lemma and the fact that the sequence $\|\bar{\eta}_t\|_{\mathcal{V}_n}$, $n \geq 1$, increases, we get

$$\mathbb{E} \int_0^T \lim_{n \to \infty} \|\bar{\eta}_t\|_{\mathcal{V}_n}^2 dt \leq C(1 + \|\varrho\|_{-J+1,\mathbb{R}^d}^2).$$

This yields that $\bar{\eta} \in L_2([0,T] \times \Omega, dt \times \mathbb{P}; H^{-J+1}(\mathbb{R}^d))$ and

$$\mathbb{E} \int_0^T \|\bar{\eta}_t\|_{-J+1,\mathbb{R}^d}^2 dt \leq C(1 + \|\varrho\|_{-J+1,\mathbb{R}^d}^2).$$

Next, let $\gamma \in L_\infty([0,T] \times \Omega, dt \times \mathbb{P}; \mathbb{R})$ and $\varphi \in \mathrm{C}_c(\mathbb{R}^d)$. We take $n \in \mathbb{N}$ satisfying $\mathrm{supp}\,\varphi \subset \Gamma_n$ and compute

$$\mathbb{E} \int_0^T \gamma(t) \langle \bar{\eta}_t, \varphi \rangle_{0,\mathbb{R}^d} dt = \mathbb{E} \int_0^T \gamma(t) \langle \bar{\eta}_t, \varphi \rangle_{0,\Gamma_n} dt = \lim_{k \to \infty} \mathbb{E} \int_0^T \gamma(t) \langle \eta_t^{m_k}, \varphi \rangle_{0,\Gamma_n} dt$$

$$= \mathbb{E} \int_0^T \gamma(t) \langle \varrho, \varphi \rangle_{0,\Gamma_n} dt + \lim_{k \to \infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \langle U_{m_k}(s, \bar{\eta}_t^{m_k}), \varphi \rangle_{0,\Gamma_n} ds\, dt$$

$$+ \lim_{k \to \infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \langle B_{m_k}(s), \varphi \rangle_{0,\Gamma_n} dW_s\, dt$$

$$= \mathbb{E} \int_0^T \gamma(t) \langle \varrho, \varphi \rangle_{0,\mathbb{R}^d} dt + \lim_{k \to \infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \langle \nabla \varphi \cdot v(s, \cdot), \eta_s^{m_k} \rangle_{0,\Gamma_n} ds\, dt$$

$$+ \lim_{k \to \infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \left\langle \nabla \varphi(x) \cdot \langle \kappa_{m_k} \tilde{v}(s, x, \cdot), \eta_s^{m_k} \rangle_{0,\Gamma_{m_k}}, \mu_s(dx) \right\rangle ds\, dt$$

$$+ \int_0^T \gamma(t) \int_0^t \int_\Theta \langle \nabla \varphi \cdot g(s, \cdot, \theta), \mu_s \rangle W(d\theta, ds)\, dt.$$

Using the convergence of η^{m_k} in the weak topology of $L_2(\mathcal{V}_n)$, we get

$$\lim_{k \to \infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \langle \nabla \varphi \cdot v(s, \cdot), \eta_s^{m_k} \rangle_{0,\Gamma_n} ds\, dt = \mathbb{E} \int_0^T \gamma(t) \int_0^t \langle \nabla \varphi \cdot v(s, \cdot), \bar{\eta}_s \rangle_{0,\Gamma_n} ds\, dt$$

$$= \mathbb{E} \int_0^T \gamma(t) \int_0^t \langle \nabla \varphi \cdot v(s, \cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d} ds\, dt.$$

Next, we rewrite for every $k, l \geq 1$

$$\langle \kappa_{m_k} \tilde{v}(s, x, \cdot), \eta_s^{m_k} \rangle_{0,\Gamma_{m_k}} - \langle \tilde{v}(s, x, \cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d} = \langle (\kappa_{m_k} - 1) \tilde{v}(s, x, \cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d}$$
$$+ \langle (\kappa_{m_k} - \kappa_l) \tilde{v}(s, x, \cdot), \eta_s^{m_k} - \bar{\eta}_s \rangle_{0,\Gamma_{m_k}} + \langle \kappa_l \tilde{v}(s, x, \cdot), \eta_s^{m_k} - \bar{\eta}_s \rangle_{0,\Gamma_l}.$$

Since

$$|\langle (1 - \kappa_{m_k}) \tilde{v}(s, x, \cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d}| \leq \|(\kappa_{m_k} - 1) \tilde{v}(s, x, \cdot)\|_{J-1,\mathbb{R}^d} \|\bar{\eta}_s\|_{-J+1,\mathbb{R}^d},$$
$$|\langle (\kappa_{m_k} - \kappa_l) \tilde{v}(s, x, \cdot), \eta_s^{m_k} - \bar{\eta}_s \rangle_{0,\Gamma_{m_k}}| \leq \|(\kappa_{m_k} - \kappa_l) \tilde{v}(s, x, \cdot)\|_{J-1,\Gamma_{m_k}} \|\eta_s^{m_k} - \bar{\eta}_s\|_{\mathcal{V}_{m_k}},$$
$$\leq C\|(1 - \kappa_l) \tilde{v}(s, x, \cdot)\|_{J-1,\mathbb{R}^d} \|\eta_s^{m_k} - \bar{\eta}_s\|_{\mathcal{V}_{m_k}}$$

and $\eta^{m_k} \to \bar{\eta}$ in the weak topology of $L_2(\mathcal{V}_l)$ for every l, it is easily seen that

$$\lim_{k \to \infty} \left(\mathbb{E} \int_0^T \gamma(t) \int_0^t \left\langle \nabla \varphi(x) \cdot \langle \kappa_{m_k} \tilde{v}(s, x, \cdot), \eta_s^{m_k} \rangle_{0,\Gamma_{m_k}}, \mu_s(dx) \right\rangle ds\, dt \right.$$

$$\left. - \mathbb{E} \int_0^T \gamma(t) \int_0^t \left\langle \nabla \varphi(x) \cdot \langle \tilde{v}(s, x, \cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d}, \mu_s(dx) \right\rangle ds\, dt \right)$$

$$= \lim_{k\to\infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \left\langle \nabla\varphi(x) \cdot \langle (\kappa_{m_k} - 1)\tilde{v}(s,x,\cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d}, \mu_s(dx) \right\rangle ds dt$$

$$+ \lim_{l\to\infty} \lim_{k\to\infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \left\langle \nabla\varphi(x) \cdot \langle (\kappa_{m_k} - \kappa_l)\tilde{v}(s,x,\cdot), \eta_s^{m_k} - \bar{\eta}_s \rangle_{0,\Gamma_{m_k}}, \mu_s(dx) \right\rangle ds dt$$

$$+ \lim_{l\to\infty} \lim_{k\to\infty} \mathbb{E} \int_0^T \gamma(t) \int_0^t \left\langle \nabla\varphi(x) \cdot \langle \kappa_l \tilde{v}(s,x,\cdot), \eta_s^{m_k} - \bar{\eta}_s \rangle_{0,\Gamma_l}, \mu_s(dx) \right\rangle ds dt = 0.$$

This implies that for that for a.e. $t \in [0,T]$ and all $\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)$

$$\langle \varphi, \bar{\eta}_t \rangle_{0,\mathbb{R}^d} = \langle \varphi, \varrho \rangle_{0,\mathbb{R}^d} + \int_0^t \left(\langle \nabla\varphi \cdot v(s,\cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d} + \langle \nabla\varphi \cdot \langle \tilde{v}(s,x,\cdot), \bar{\eta}_s \rangle_{0,\mathbb{R}^d}, \mu_s(dx) \rangle \right) ds$$

$$+ \int_0^t \int_\Theta \langle \nabla\varphi \cdot g(s,\cdot,\theta), \mu_s \rangle W(d\theta, ds), \quad t \in [0,T].$$

Taking

$$\eta_t = \varrho + \int_0^t U_\infty(s, \bar{\eta}_s) ds + \int_0^t B_\infty(s) dW_s, \quad t \in [0,T],$$

and using the continuity of the right hand side of the expression above, we get the existence of solutions to (3.4).

Itô's formula from [69, Theorem 4.2.4] and the condition (H2), which also holds true for $n = \infty$, imply the uniqueness of solutions to (3.10) and the bound (3.11). The fact that η_t, $t \in [0,T]$, is a Gaussian process in $H^{-J}(\mathbb{R}^d)$ follows from the linearity of the equation. This completes the proof of the proposition. $\qquad\square$

Lemma 3.5. *Let the coefficients A, V, G satisfy Assumptions 2.1, 2.7. Let also μ_t^ε, $t \geq 0$, be a superposition solution to the stochastic mean-field equation (3.1) started from $\mu_0^\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$, μ_t, $t \geq 0$, be a superposition solution to the PDE (3.2) started from $\mu_0^0 \in \mathcal{P}_2(\mathbb{R}^d)$ and the process η_t^ε, $t \geq 0$, be defined by (3.3). Then for every $J > \frac{d}{2} + 1$ and $T > 0$ there exists a constant $C > 0$ such that*

$$\mathbb{E} \sup_{t\in[0,T]} \|\eta_t^\varepsilon\|_{-J}^2 \leq C \left(1 + \langle \phi_2, \mu_0^\varepsilon \rangle \right) + \frac{C}{\varepsilon} \mathcal{W}_2^2 \left(\mu_0^\varepsilon, \mu_0^0 \right).$$

for all $\varepsilon > 0$.

Proof. Since for each $\varepsilon \geq 0$ the process μ_t^ε, $t \geq 0$, is a superposition solution to the corresponding equation, $\mu_t^\varepsilon = \mu_0^\varepsilon \circ X_\varepsilon(t,\cdot)$, $t \geq 0$, where X_ε is a solution to the SDE with interaction (2.3) with G replaced by $\sqrt{\varepsilon}G$. Similarly to the proof of Theorem 3.1, we fix $\varepsilon > 0$ and consider an arbitrary probability measure χ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_0^ε and μ_0^0. Using Remark 3.2 and the continuous embedding of $H_0^J(\mathbb{R}^d)$ into $\mathrm{C}_b^1(\mathbb{R}^d)$, we get

$$\mathbb{E} \sup_{t\in[0,T]} \|\eta_t^\varepsilon\|_{-J}^2 = \frac{1}{\varepsilon} \mathbb{E} \sup_{t\in[0,T]} \sup_{\varphi\in\mathrm{C}_c^\infty(\Gamma)} \frac{1}{\|\varphi\|_J^2} \langle \varphi, \mu_t^\varepsilon - \mu_t^0 \rangle^2$$

$$= \frac{1}{\varepsilon} \mathbb{E} \sup_{t\in[0,T]} \sup_{\varphi\in\mathrm{C}_c^\infty(\Gamma)} \frac{1}{\|\varphi\|_J^2} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\varphi(X_\varepsilon(u,t)) - \varphi(X_0(v,t))) \chi(du, dv) \right)^2$$

$$\leq \frac{1}{\varepsilon} \mathbb{E} \sup_{t\in[0,T]} \sup_{\varphi\in\mathrm{C}_c^\infty(\Gamma)} \frac{1}{\|\varphi\|_J^2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\varphi(X_\varepsilon(u,t)) - \varphi(X_0(v,t)))^2 \chi(du, dv)$$

$$\leq \frac{1}{\varepsilon} \mathbb{E} \sup_{t\in[0,T]} \sup_{\varphi\in\mathrm{C}_c^\infty(\Gamma)} \frac{\sup_{x\in\Gamma} |\nabla\varphi(x)|^2}{\|\varphi\|_J^2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |X_\varepsilon(u,t) - X_0(v,t)|^2 \chi(du, dv)$$

$$\leq \frac{C}{\varepsilon} \mathbb{E} \sup_{t \in [0,T]} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |X_\varepsilon(u,t) - X_0(v,t)|^2 \, \chi(du, dv)$$

$$\leq \frac{C}{\varepsilon} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E} \sup_{t \in [0,T]} |X_\varepsilon(u,t) - X_0(v,t)|^2 \, \chi(du, dv)$$

$$\leq \frac{C_1}{\varepsilon} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(|u-v|^2 + \mathcal{W}_2^2(\mu_0^\varepsilon, \mu_0^0) + \varepsilon \left(1 + |u|^2 + \langle \phi_2, \mu_0^\varepsilon \rangle \right) \right) \chi(du, dv)$$

$$= \frac{C_1}{\varepsilon} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |u-v|^2 \chi(du, dv) + \mathcal{W}_2^2(\mu_0^\varepsilon, \mu_0^0) + \varepsilon \left(1 + 2 \langle \phi_2, \mu_0^\varepsilon \rangle \right) \right).$$

Taking infimum over all χ with marginals μ_0^ε and μ_0^0, we obtain the needed estimate, which completes the proof of the lemma. $\qquad\square$

To prove the convergence of the fluctuation field η^ε, we will need the following assumptions on the coefficients of the stochastic mean-field equation (3.1).

Assumption 3.6.

(i) *The coefficients A, V, G do not depend on $\omega \in \Omega$ and there exist Borel measurable functions $\bar{V} : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{V} : [0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ such that*

$$V(t, x, \mu) = \bar{V}(t, x) + \langle \tilde{V}(t, x, \cdot), \mu \rangle$$

for all $t \in [0, \infty)$, $x \in \mathbb{R}^d$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

(ii) *For all $t \geq 0$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $x, y \in \mathbb{R}^d$, and some $J \geq \frac{d}{2} + 4$ one has that $\bar{V}(t, \cdot) \in C_b^J(\mathbb{R}^d)$, $\tilde{V}(t, \cdot, \cdot) \in C^J(\mathbb{R}^d \times \mathbb{R}^d)$, and for every compact set $K \in \mathcal{P}_2(\mathbb{R}^d)$ and $i \in [d]$*

$$\sup_{t \in [0,T]} \|\bar{V}_i(t, \cdot)\|_{C_b^J} + \sup_{t \in [0,T]} \|\tilde{V}_i(t, \cdot, \cdot)\|_{C_b^J \times H^J} + \sup_{t \in [0,T], \mu \in K} \|A_{i,i}(t, \cdot, \mu)\|_{C_b} < \infty,$$

where

$$\|f\|_{C_b^m \times H^J}^2 = \sum_{|\alpha| \leq m} \sum_{|\beta| \leq J} \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} \left(D_x^\alpha D_y^\beta f(x, y) \right)^2 dy.$$

Now we can formulate the main result of this section.

Theorem 3.7 (Quantified CLT for the stochastic mean field equation). *Let the coefficients A, V, G of the equations (3.1), (3.2) satisfy Assumptions 2.1, 2.7, 3.6 for some $J \geq \frac{d}{2} + 4$. Let μ_t^ε, $t \geq 0$, be a superposition solution to the stochastic mean-field equation (3.1) started from $\mu_0^\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$ for each $\varepsilon > 0$, μ_t, $t \geq 0$, be a superposition solution to the PDE (3.2) started from $\mu_0^0 \in \mathcal{P}_2(\mathbb{R}^d)$ and the continuous Gaussian process η_t^0, $t \geq 0$, be a solution to the linear SPDE (3.4) started from $\varrho \in H^{-J+1}(\mathbb{R}^d)$. Furthermore, we assume that $\mathcal{W}_2^2(\mu_0^\varepsilon, \mu_0^0) \leq C\varepsilon$, $\|\eta_0^\varepsilon - \varrho\|_{-J}^2 \leq C\varepsilon$ and $\langle \phi_4, \mu_0^\varepsilon \rangle \leq C$ for some $C > 0$ and all $\varepsilon \in (0, 1]$, and define the continuous process $\eta^\varepsilon = \frac{1}{\sqrt{\varepsilon}} \left(\mu_t^\varepsilon - \mu_t^0 \right)$, $t \geq 0$. Then for every $T > 0$ there exists a constant $C > 0$ such that the inequality*

$$\mathbb{E} \sup_{t \in [0,T]} \|\eta_t^\varepsilon - \eta_t^0\|_{-J}^2 \leq C\varepsilon$$

holds.

Proof. The existence of a unique solution η_t^0, $t \in [0, T]$, to the linear SPDE (3.4) follows from Proposition 3.4. Moreover, it is a.s. a continuous $H^{-J+1}(\mathbb{R}^d)$-valued process such that

$$\mathbb{E} \sup_{t \in [0,T]} \|\eta_t^0\|_{-J+1}^2 < \infty.$$

We also remark that η_t^ε, $t \in [0,T]$, is a.s. a continuous $H^{-J+1}(\mathbb{R}^d)$-valued process for every $\varepsilon > 0$.

Define the functions $Q = U + \tilde{U} : [0,T] \times H^{-J+1}(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \to H^{-J}(\mathbb{R}^d)$, $R : [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \to H^{-J}(\mathbb{R}^d)$, and $B : [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathcal{L}_2(L_2(\Theta,\vartheta), H^{-J}(\mathbb{R}^d))$ as follows

$$\langle U(t,f,\mu),\varphi\rangle_0 := \langle \nabla\varphi \cdot V(t,\cdot,\mu), f\rangle_0,$$

$$\langle \tilde{U}(t,f,\mu),\varphi\rangle_0 := \left\langle \nabla\varphi(x) \cdot \langle \tilde{V}(t,x,\cdot), f\rangle_0, \mu(dx)\right\rangle$$

$$\langle R(t,\mu),\varphi\rangle_0 := \langle D^2\varphi : A(t,\cdot,\mu), \mu\rangle$$

and

$$\langle B(t)h,\varphi\rangle_0 = \langle h, \langle \nabla\varphi(x) \cdot G(t,x,\mu,\cdot), \mu(dx)\rangle\rangle_\vartheta$$

for all $\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)$, $f \in H^{-J+1}(\mathbb{R}^d)$, and $h \in L_2(\Theta,\vartheta)$. Then

$$\eta_t^\varepsilon = \eta_0^\varepsilon + \frac{\sqrt{\varepsilon}}{2}\int_0^t R(s,\mu_s^\varepsilon)ds + \int_0^t Q(s,\eta_s^\varepsilon,\mu_s^\varepsilon)ds + \int_0^t B(s,\mu_s^\varepsilon)dW_s, \quad t \in [0,T],$$

for all $\varepsilon \geq 0$.

Let $\zeta^\varepsilon = \eta_t^\varepsilon - \eta_t^0$, $t \in [0,T]$. Then

$$\zeta_t^\varepsilon = \zeta_0^\varepsilon + \frac{\sqrt{\varepsilon}}{2}\int_0^t R(s,\mu_s^\varepsilon)ds + \int_0^t \left(Q(s,\eta_s^\varepsilon,\mu_s^\varepsilon) - Q(s,\eta_s^0,\mu_s^0)\right)ds$$

$$+ \int_0^t \left(B(s,\mu_s^\varepsilon) - B(s,\mu_s^0)\right)dW_s, \quad t \in [0,T].$$

Using Itô's formula from [69, Theorem 4.2.4], we get

$$\|\zeta_t^\varepsilon\|_{-J}^2 = \|\zeta_0^\varepsilon\|_{-J}^2 + \sqrt{\varepsilon}\int_0^t \langle R(s,\mu_s^\varepsilon), \zeta_s^\varepsilon\rangle_{-J}ds + 2\int_0^t \langle Q(s,\eta_s^\varepsilon,\mu_s^\varepsilon) - Q(s,\eta_s^0,\mu_s^0), \zeta_s^\varepsilon\rangle_{-J}ds$$

$$+ \int_0^t \|B(s,\mu_s^\varepsilon) - B(s,\mu_s^0)\|_{\mathrm{HS},H^{-J}}^2 ds + \int_0^t \langle \zeta_s^\varepsilon, (B(s,\mu_s^\varepsilon) - B(s,\mu_s^0))dW_s\rangle_{-J}$$

for all $t \in [0,T]$. We fix an $(\mathcal{F}_t)$-stopping time τ and apply the Burkholder–Davis–Gundy inequality to estimate $\mathbb{E}\sup_{s\in[0,\tau]}\|\zeta_s^\varepsilon\|_{-J}^2 = \mathbb{E}\sup_{s\in[0,t]}\|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2$. We get

$$\mathbb{E}\sup_{s\in[0,t]}\|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2 \leq \|\zeta_0^\varepsilon\|_{-J}^2 + \sqrt{\varepsilon}\mathbb{E}\sup_{s\in[0,t]}\left|\int_0^{s\wedge\tau}\langle R(r,\mu_r^\varepsilon), \zeta_r^\varepsilon\rangle_{-J}dr\right|$$

$$+ 2\mathbb{E}\sup_{s\in[0,t]}\left|\int_0^{s\wedge\tau}\langle Q(r,\eta_r^\varepsilon,\mu_r^\varepsilon) - Q(r,\eta_r^0,\mu_r^0), \zeta_r^\varepsilon\rangle_{-J}dr\right|$$

$$+ \mathbb{E}\sup_{s\in[0,t]}\int_0^{s\wedge\tau}\|B(r,\mu_r^\varepsilon) - B(r,\mu_r^0)\|_{\mathrm{HS},H^{-J}}^2 dr$$

$$+ \mathbb{E}\sup_{s\in[0,t]}\left|\int_0^{s\wedge\tau}\langle \zeta_r^\varepsilon, (B(r,\mu_r^\varepsilon) - B(r,\mu_r^0))dW_r\rangle_{-J}\right|$$

$$\leq \|\zeta_0^\varepsilon\|_{-J}^2 + \sqrt{\varepsilon}\mathbb{E}\int_0^t |\langle R(s\wedge\tau,\mu_{s\wedge\tau}^\varepsilon), \zeta_{s\wedge\tau}^\varepsilon\rangle_{-J}|\,ds$$

$$+ 2\mathbb{E}\int_0^t \left|\langle Q(s\wedge\tau,\eta_{s\wedge\tau}^\varepsilon,\mu_{s\wedge\tau}^\varepsilon) - Q(s\wedge\tau,\eta_{s\wedge\tau}^0,\mu_{s\wedge\tau}^0), \zeta_{s\wedge\tau}^\varepsilon\rangle_{-J}\right|\,ds$$

$$+ \mathbb{E}\int_0^t \|B(s\wedge\tau,\mu_{s\wedge\tau}^\varepsilon) - B(s\wedge\tau,\mu_{s\wedge\tau}^0)\|_{\mathrm{HS},H^{-J}}^2 ds$$

$$+ C\mathbb{E}\left(\int_0^t \|\hat{B}^\varepsilon(s \wedge \tau)\|^2_{\mathrm{HS},\mathbb{R}} ds\right)^{\frac{1}{2}} =: \sum_{i=1}^5 I_i(t),$$

where $\hat{B}^\varepsilon : [0,T] \to \mathcal{L}_2(L_2(\Theta,\vartheta),\mathbb{R})$ is defined as follows

$$\hat{B}^\varepsilon(s)h := \langle \zeta_s^\varepsilon, B(s,\mu_s^\varepsilon)h - B(s,\mu_s^0)h \rangle_{-J}$$

for all $s \in [0,T]$ and $h \in L_2(\Theta,\vartheta)$, and C is the universal constant from the Burkholder–Davis–Gundy inequality. We next estimate every term in the inequality above.

Estimate of I_2. Using the Cauchy–Schwarz inequality and Young's inequality, we get

$$\sqrt{\varepsilon}|\langle R(s,\mu_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}| \leq \sqrt{\varepsilon}\|R(s,\mu_s^\varepsilon)\|_{-J}\|\zeta_s^\varepsilon\|_{-J} \leq \frac{\varepsilon}{2}\|R(s,\mu_s^\varepsilon)\|^2_{-J} + \frac{1}{2}\|\zeta_s^\varepsilon\|^2_{-J}.$$

Then, by the continuity of the embedding of $H^J(\mathbb{R}^d)$ into $\mathrm{C}_b^2(\mathbb{R}^d)$, Assumptions 2.1, 2.7 and Remark 2.8, we have

$$
\begin{aligned}
\|R(s,\mu_s^\varepsilon)\|^2_{-J} &= \sup_{\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|^2_J}\langle D^2\varphi : A(s,\cdot,\mu_s^\varepsilon), \mu_s^\varepsilon \rangle^2 \\
&\leq \sup_{\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{\|\varphi\|^2_{\mathrm{C}_b^2}}{\|\varphi\|^2_J} \sup_{t \in [0,T]} \langle |A(t,\cdot,\mu_t^\varepsilon)|, \mu_t^\varepsilon \rangle^2 \\
&\leq C \sup_{t \in [0,T]} \langle \||G(t,x,\mu_t^\varepsilon,\cdot)|\|^2_\vartheta, \mu_t^\varepsilon(dx) \rangle^2 \\
&\leq C \sup_{t \in [0,T]} \langle 1 + |x|^2 + \mathcal{W}_2^2(\mu_t^\varepsilon, \delta_0), \mu_t^\varepsilon(dx) \rangle^2 \\
&\leq C\left(1 + 4 \sup_{t \in [0,T]} \mathcal{W}_2^4(\mu_t^\varepsilon, \delta_0)\right).
\end{aligned}
$$

Using Corollary 2.10 and the fact that $\langle \phi_4, \mu_0^\varepsilon \rangle < C$ for some constant $C > 0$ and each $\varepsilon \in (0,1]$, we get

$$\mathbb{E}\sup_{t \in [0,T]} \mathcal{W}_2^4(\mu_t^\varepsilon, \delta_0) = \mathbb{E}\sup_{t \in [0,T]} \langle \phi_2, \mu_t^\varepsilon \rangle^2 \leq \mathbb{E}\sup_{t \in [0,T]} \langle \phi_4, \mu_t^\varepsilon \rangle \leq C$$

for all $\varepsilon \in (0,1]$. Hence, for every $\varepsilon \in (0,1]$, we have

$$I_2(t) \leq C\varepsilon + C\int_0^t \|\zeta_{s\wedge\tau}^\varepsilon\|^2_{-J} ds \leq C\varepsilon + C\int_0^t \sup_{r \in [0,s]} \|\zeta_{r\wedge\tau}^\varepsilon\|^2_{-J} ds, \quad t \in [0,T].$$

Estimate of I_3. We estimate I_3 as follows

$$(3.14) \quad
\begin{aligned}
\left|\langle U(s,\eta_s^\varepsilon,\mu_s^\varepsilon) - U(s,\eta_s^0,\mu_s^0), \zeta_s^\varepsilon \rangle_{-J}\right| &\leq \left|\langle U(s,\eta_s^\varepsilon,\mu_s^\varepsilon) - U(s,\eta_s^0,\mu_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}\right| \\
&\quad + \left|\langle U(s,\eta_s^0,\mu_s^\varepsilon) - U(s,\eta_s^0,\mu_s^0), \zeta_s^\varepsilon \rangle_{-J}\right|.
\end{aligned}
$$

By Lemma 3.3,

$$
\begin{aligned}
\left|\langle U(s,\eta_s^\varepsilon,\mu_s^\varepsilon) - U(s,\eta_s^0,\mu_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}\right| &= \left|\langle U(s,\zeta_s^\varepsilon,\mu_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}\right| \\
&\leq C\max_{i \in [d]} \sup_{t \in [0,T]} \|V_i(t,\cdot,\mu_t^\varepsilon)\|_{\mathrm{C}_b^J} \|\zeta_s^\varepsilon\|^2_{-J}.
\end{aligned}
$$

We note that

$$
\begin{aligned}
\sup_{t \in [0,T]} \|V_i(t,\cdot,\mu_t^\varepsilon)\|_{\mathrm{C}_b^J} &\leq \sup_{t \in [0,T]} \|\bar{V}_i(t,\cdot)\|_{\mathrm{C}_b^J} + \sup_{t \in [0,T]} \|\langle \tilde{V}_i(t,\cdot,y), \mu_t^\varepsilon(dy) \rangle\|_{\mathrm{C}_b^J} \\
&\leq \sup_{t \in [0,T]} \|\bar{V}_i(t,\cdot)\|_{\mathrm{C}_b^J} + \sup_{t \in [0,T], y \in \mathbb{R}^d} \|\tilde{V}_i(t,\cdot,y)\|_{\mathrm{C}_b^J} \leq C,
\end{aligned}
$$

where C is a non-random constant, according to Assumption 3.6 and the continuous embedding of $H^J(\mathbb{R}^d)$ into $C_b(\mathbb{R}^d)$. In order to estimate the second term of the right hand side of (3.14), we consider the following bilinear map $\hat{U} : [0,T] \times H^{-J+1}(\mathbb{R}^d) \times H^{-J}(\mathbb{R}^d) \to H^{-J}(\mathbb{R}^d)$ defined by

$$(3.15) \qquad \langle \hat{U}(t, f_1, f_2), \varphi \rangle_0 := \left\langle \nabla \varphi(x) \cdot \langle \tilde{V}(t, x, \cdot), f_1 \rangle_0, f_2(dx) \right\rangle_0$$

for all $\varphi \in C_c^\infty(\mathbb{R}^d)$, which coincides with $\tilde{U}$ on $[0,T] \times H^{-J+1}(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$. Then, using Lemma 3.3, we obtain

$$
\begin{aligned}
\left| \langle U(s, \eta_s^0, \mu_s^\varepsilon) - U(s, \eta_s^0, \mu_s^0), \zeta_s^\varepsilon \rangle_{-J} \right| &= \sqrt{\varepsilon} |\langle \hat{U}(s, \eta_s^0, \eta_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}| \\
&\leq \sqrt{\varepsilon} |\langle \hat{U}(s, \eta_s^0, \zeta_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}| + \sqrt{\varepsilon} |\langle \hat{U}(s, \eta_s^0, \eta_s^0), \zeta_s^\varepsilon \rangle_{-J}| \\
&\leq \sqrt{\varepsilon} C \max_{i \in [d]} \|\langle \tilde{V}_i(s, \cdot, x), \eta_s^0(dx) \rangle_0\|_{C_b^J} \|\zeta_s^\varepsilon\|_{-J}^2 \\
&\quad + \sqrt{\varepsilon} \|\hat{U}(s, \eta_s^0, \eta_s^0)\|_{-J} \|\zeta_s^\varepsilon\|_{-J}.
\end{aligned}
$$
(3.16)

We estimate separately

$$\|\langle \tilde{V}_i(s, \cdot, x), \eta_s^0(dx) \rangle\|_{C_b^J} \leq \|\tilde{V}_i(s, \cdot, \cdot)\|_{C_b^J \times H^J} \|\eta_s^0\|_{-J}$$

and

$$
\begin{aligned}
\|\hat{U}(s, \eta_s^0, \eta_s^0)\|_{-J} &= \sup_{\varphi \in C_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J} \langle \nabla \varphi(x) \cdot \langle \tilde{V}(t, x, \cdot), \eta_s^0 \rangle_0, \eta_s^0(dx) \rangle_0 \\
&\leq \sup_{\varphi \in C_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J} \|\nabla \varphi\|_{J-1} \|\tilde{V}(s, \cdot, \cdot)\|_{C_b^{J-1} \times H^{J-1}} \|\eta_s^0\|_{-J+1}^2 \\
&\leq C \sup_{t \in [0,T]} \|\tilde{V}(t, \cdot, \cdot)\|_{C_b^{J-1} \times H^{J-1}} \|\eta_s^0\|_{-J+1}^2.
\end{aligned}
$$

Therefore, by Assumption (3.6) and Young's inequality,

$$
\begin{aligned}
\left| \langle U(s, \eta_s^0, \mu_s^\varepsilon) - U(s, \eta_s^0, \mu_s^0), \zeta_s^\varepsilon \rangle_{-J} \right| &\leq \sqrt{\varepsilon} C \|\eta_s^0\|_{-J} \|\zeta_s^\varepsilon\|_{-J}^2 + \sqrt{\varepsilon} C \|\eta_s^0\|_{-J+1}^2 \|\zeta_s^\varepsilon\|_{-J} \\
&\leq \sqrt{\varepsilon} C \|\eta_s^0\|_{-J} \|\zeta_s^\varepsilon\|_{-J}^2 + C \varepsilon \|\eta_s^0\|_{-J+1}^4 + C \|\zeta_s^\varepsilon\|_{-J}^2.
\end{aligned}
$$
(3.17)

We next estimate

$$
\begin{aligned}
|\langle \tilde{U}(s, \eta_s^\varepsilon, \mu_s^\varepsilon) &- \tilde{U}(s, \eta_s^0, \mu_s^0), \zeta_s^\varepsilon \rangle_{-J}| \\
&\leq |\langle \tilde{U}(s, \eta_s^\varepsilon, \mu_s^\varepsilon) - \tilde{U}(s, \eta_s^0, \mu_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}| + |\langle \tilde{U}(s, \eta_s^0, \mu_s^\varepsilon) - \tilde{U}(s, \eta_s^0, \mu_s^0), \zeta_s^\varepsilon \rangle_{-J}| \\
&= |\langle \tilde{U}(s, \zeta_s^\varepsilon, \mu_s^\varepsilon), \zeta_s^\varepsilon \rangle| + \sqrt{\varepsilon} |\langle \hat{U}(s, \eta_s^0, \eta_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}|
\end{aligned}
$$

where $\hat{U}$ is defined by (3.15). Then,

$$
\begin{aligned}
|\langle \tilde{U}(s, \zeta_s^\varepsilon, \mu_s^\varepsilon), \zeta_s^\varepsilon \rangle| &\leq \|\tilde{U}(s, \zeta_s^\varepsilon, \mu_s^\varepsilon)\|_{-J} \|\zeta_s^\varepsilon\|_{-J} \\
&\leq \|\zeta_s^\varepsilon\|_{-J} \sup_{\varphi \in C_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J} \left\langle \nabla \varphi \cdot \langle \tilde{V}(s, x, \cdot), \zeta_s^\varepsilon \rangle_0, \mu_s^\varepsilon(dx) \right\rangle \\
&\leq \|\zeta_s^\varepsilon\|_{-J}^2 \sup_{\varphi \in C_c^\infty(\mathbb{R}^d)} \frac{\|\varphi\|_{C_b^1}}{\|\varphi\|_J} \|\tilde{V}(s, \cdot, \cdot)\|_{C_b \times H^J} \leq C \sup_{t \in [0,T]} \|\tilde{V}(s, \cdot, \cdot)\|_{C_b \times H^J} \|\zeta_s^\varepsilon\|_{-J}^2.
\end{aligned}
$$

Using the first equality in the estimate (3.16) and then (3.17), we can conclude that

$$\sqrt{\varepsilon} |\langle \hat{U}(s, \eta_s^0, \eta_s^\varepsilon), \zeta_s^\varepsilon \rangle_{-J}| \leq \sqrt{\varepsilon} C \|\eta_s^0\|_{-J} \|\zeta_s^\varepsilon\|_{-J}^2 + C \varepsilon \|\eta_s^0\|_{-J+1}^4 + C \|\zeta_s^\varepsilon\|_{-J}^2.$$

Thus, we have shown that there exists a constant $C > 0$ such that

$$I_3(t) \le C \int_0^t \mathbb{E}\|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2 ds + C\sqrt{\varepsilon} \int_0^t \mathbb{E}\|\eta_{s\wedge\tau}^0\|_{-J}\|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2 ds + C\varepsilon \int_0^t \mathbb{E}\|\eta_{s\wedge\tau}^0\|_{-J+1}^4 ds$$

$$\le C \int_0^t \mathbb{E} \sup_{r\in[0,s]} \|\zeta_{r\wedge\tau}^\varepsilon\|_{-J}^2 ds + C\sqrt{\varepsilon} \int_0^t \mathbb{E} \left[\|\eta_{s\wedge\tau}^0\|_{-J} \sup_{r\in[0,s]} \|\zeta_{r\wedge\tau}^\varepsilon\|_{-J}^2 \right] ds$$

$$+ C\varepsilon\mathbb{E} \sup_{s\in[0,T]} \|\eta_s^0\|_{-J+1}^4$$

for all $t \in [0, T]$.

Estimate of I_4. Let $\{h_k, \ k \ge 1\}$ be an orthonormal basis in $L_2(\Theta, \vartheta)$ and $\bar{G}^k(t, x, \mu) = \langle G(t, x, \mu, \cdot), h_k \rangle, k \ge 1$. Then

$$\|B(s, \mu_s^\varepsilon) - B(s, \mu_s^0)\|_{\mathrm{HS}, H^{-J}}^2 = \sum_{k=1}^\infty \|B(s, \mu_s^\varepsilon)h_k - B(s, \mu_s^0)h_k\|_{-J}^2$$

$$= \sum_{k=1}^\infty \sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J^2} \left(\langle \nabla\varphi \cdot \bar{G}^k(s, \cdot, \mu_s^\varepsilon), \mu_s^\varepsilon \rangle - \langle \nabla\varphi \cdot \bar{G}^k(s, \cdot, \mu_s^0), \mu_s^0 \rangle \right)^2$$

(3.18)
$$\le 2 \sum_{k=1}^\infty \sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J^2} \left(\langle \nabla\varphi \cdot \bar{G}^k(s, \cdot, \mu_s^\varepsilon), \mu_s^\varepsilon \rangle - \langle \nabla\varphi \cdot \bar{G}^k(s, \cdot, \mu_s^0), \mu_s^\varepsilon \rangle \right)^2$$

$$+ 2 \sum_{k=1}^\infty \sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J^2} \left(\langle \nabla\varphi \cdot \bar{G}^k(s, \cdot, \mu_s^0), \mu_s^\varepsilon \rangle - \langle \nabla\varphi \cdot \bar{G}^k(s, \cdot, \mu_s^0), \mu_s^0 \rangle \right)^2$$

$$= 2 \sum_{k=1}^\infty \sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J^2} \left\langle \nabla\varphi \cdot (\bar{G}^k(s, \cdot, \mu_s^\varepsilon) - \bar{G}^k(s, \cdot, \mu_s^0)), \mu_s^\varepsilon \right\rangle^2$$

$$+ 2 \sum_{k=1}^\infty \sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J^2} \left\langle \nabla\varphi \cdot \bar{G}^k(s, \cdot, \mu_s^0), \mu_s^\varepsilon - \mu_s^0 \right\rangle^2.$$

Using Jensen's inequality, Fubini's theorem, Parseval's identity, and Assumption 2.7, we get

$$\sum_{k=1}^\infty \sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J^2} \left\langle \nabla\varphi \cdot (\bar{G}^k(s, \cdot, \mu_s^\varepsilon) - \bar{G}^k(s, \cdot, \mu_s^0)), \mu_s^\varepsilon \right\rangle^2$$

$$\le \sum_{k=1}^\infty \sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)} \frac{1}{\|\varphi\|_J^2} \|\varphi\|_{\mathrm{C}_b^1}^2 \left\langle |\bar{G}^k(s, \cdot, \mu_s^\varepsilon) - \bar{G}^k(s, \cdot, \mu_s^0)|^2, \mu_s^\varepsilon \right\rangle$$

$$\le C \left\langle \sum_{k=1}^\infty |\bar{G}^k(s, \cdot, \mu_s^\varepsilon) - \bar{G}^k(s, \cdot, \mu_s^0)|^2, \mu_s^\varepsilon \right\rangle$$

$$= C \sum_{i=1}^d \left\langle \|G_i(s, \cdot, \mu_s^\varepsilon) - G_i(s, \cdot, \mu_s^0)\|_\vartheta^2, \mu_s^\varepsilon \right\rangle \le CL\mathcal{W}_2^2(\mu_s^\varepsilon, \mu_s^0).$$

In order to estimate the second term in the right hand side of (3.18), we use the fact that μ_t^ε, $t \ge 0$, is a superposition solution for every $\varepsilon \ge 0$. Therefore, for every $\varepsilon \ge 0$ there exists a solution X_ε to the corresponding SDE with interaction (2.3) such that $\mu_t^\varepsilon = \mu_0^\varepsilon \circ X_\varepsilon(t, \cdot)$, $t \ge 0$. Let χ be an arbitrary probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_0^ε and μ_0^0. Then, by Jensen's inequality

and the mean value theorem, we have

$$
\begin{aligned}
\langle \nabla\varphi \cdot \bar{G}^k(s,\cdot,\mu_s^0), \mu_s^\varepsilon - \mu_s^0\rangle^2 = \bigg(& \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \big(\nabla\varphi(X_\varepsilon(u,s)) \cdot \bar{G}^k(s,X_\varepsilon(u,s),\mu_s^0) \\
& - \nabla\varphi(X_0(v,s)) \cdot \bar{G}^k(s,X_0(v,s),\mu_s^0)\big)\chi(du,dv) \bigg)^2
\end{aligned}
$$

$$
\begin{aligned}
\leq\ & 2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left[\nabla\varphi(X_\varepsilon(u,s))\cdot\left(\bar{G}^k(s,X_\varepsilon(u,s),\mu_s^0) - \bar{G}^k(s,X_0(v,s),\mu_s^0)\right)\right]^2\chi(du,dv) \\
& + 2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left[(\nabla\varphi(X_\varepsilon(u,s)) - \nabla\varphi(X_0(v,s)))\cdot\bar{G}^k(s,X_0(v,s),\mu_s^0)\right]^2\chi(du,dv) \\
\leq\ & 2\|\varphi\|_{\mathrm{C}_b^1}^2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|\bar{G}^k(s,X_\varepsilon(u,s),\mu_s^0) - \bar{G}^k(s,X_0(v,s),\mu_s^0)\right|^2\chi(du,dv) \\
& + 2\|\varphi\|_{\mathrm{C}_b^2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|X_\varepsilon(u,s) - X_0(v,s)|^2|\bar{G}^k(s,X_0(v,s),\mu_s^0)|^2\chi(du,dv).
\end{aligned}
$$

Thus, using the continuous embedding of $H^J(\mathbb{R}^d)$ into $\mathrm{C}_b^2(\mathbb{R}^d)$, Fubini's theorem, and Parseval's identity, the second term of the right hand side of (3.18) can be estimated as follows

$$
\begin{aligned}
\sum_{k=1}^\infty &\sup_{\varphi\in\mathrm{C}_c^\infty(\mathbb{R}^d)}\frac{1}{\|\varphi\|_J^2}\langle\nabla\varphi\cdot\bar{G}^k(s,\cdot,\mu_s^0),\mu_s^\varepsilon-\mu_s^0\rangle^2 \\
&\leq C\sum_{k=1}^\infty\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|\bar{G}^k(s,X_\varepsilon(u,s),\mu_s^0) - \bar{G}^k(s,X_0(v,s),\mu_s^0)\right|^2\chi(du,dv) \\
&\quad + C\sum_{k=1}^\infty\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|X_\varepsilon(u,s) - X_0(v,s)|^2|\bar{G}^k(s,X_0(v,s),\mu_s^0)|^2\chi(du,dv) \\
&= C\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|\|G(s,X_\varepsilon(u,s),\mu_s^0) - G(s,X_0(u,s),\mu_s^0)\|_\vartheta\right|^2\chi(du,dv) \\
&\quad + C\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|X_\varepsilon(u,s) - X_0(v,s)|^2\left|\|G(s,X_0(v,s),\mu_s^0)\|_\vartheta\right|^2\chi(du,dv).
\end{aligned}
$$

Using Assumption 2.7 and the expression $\|G_i(t,x,\mu)\|_\vartheta^2 = A_{i,i}(t,x,\mu)$ for all $i\in[d]$, we can bound the expression above by

$$
C\left(L + \sum_{i=1}^d\|A_{i,i}(s,\cdot,\mu_s^0)\|_{\mathrm{C}_b}^2\right)\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|X_\varepsilon(u,s) - X_0(v,s)|^2\chi(du,dv).
$$

Using Assumption 3.6, Theorem 3.1, Remark 3.2, and the estimate above, we get

$$
\begin{aligned}
I_4(t) \leq\ & C\int_0^t\mathbb{E}\mathcal{W}_2^2(\mu_{s\wedge\tau}^\varepsilon,\mu_{s\wedge\tau}^0)ds + C\int_0^t\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}|X_\varepsilon(u,s\wedge\tau) - X_0(v,s\wedge\tau)|^2\chi(du,dv)ds \\
\leq\ & CT\mathbb{E}\sup_{s\in[0,T]}\mathcal{W}_2^2(\mu_s^\varepsilon,\mu_s^0) + CT\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}\sup_{s\in[0,T]}|X_\varepsilon(u,s) - X_0(v,s)|^2\chi(du,dv) \\
\leq\ & C\mathcal{W}_2^2(\mu_0^\varepsilon,\mu_0^0) + \varepsilon C(1 + \langle\phi_2,\mu_0^\varepsilon\rangle) + C\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|u - v|^2\chi(du,dv).
\end{aligned}
$$

Taking infimum over all measures χ on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals μ_0^ε and μ_0^0, the last term on the right hand side of the above inequality will give $\mathcal{W}_2^2(\mu_0^\varepsilon,\mu_0^0)$. Hence, by the inequality

$$
\langle\phi_2,\mu_0^\varepsilon\rangle = \mathcal{W}_2^2(\mu_0^\varepsilon,\delta_0) \leq 2\mathcal{W}_2^2(\mu_0^\varepsilon,\mu_0^0) + 2\mathcal{W}_2^2(\mu_0^0,\delta_0) \leq C
$$

and the assumption of the theorem, we obtain

$$I_4(t) \le C\varepsilon$$

for all $t \in [0, T]$.

For the final term I_5, we have

$$\|\hat{B}^\varepsilon(s)\|_{\text{HS},\mathbb{R}}^2 = \sum_{k=1}^\infty \langle \zeta_s^\varepsilon, B(s, \mu_s^\varepsilon)h_k - B(s, \mu_s^0)h_k \rangle_{-J}^2$$

$$\le \|\zeta_s^\varepsilon\|_{-J}^2 \sum_{k=1}^\infty \|B(s, \mu_s^\varepsilon)h_k - B(s, \mu_s^0)h_k\|_{-J}^2$$

$$\le \|\zeta_s^\varepsilon\|_{-J}^2 \|B(s, \mu_s^\varepsilon) - B(s, \mu_s^0)\|_{\text{HS},H^{-J}}^2.$$

Therefore, for every $t \in [0, T]$

$$I_5(t) \le C\mathbb{E}\left(\int_0^t \|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2 \|B(s \wedge \tau, \mu_{s\wedge\tau}^\varepsilon) - B(s \wedge \tau, \mu_{s\wedge\tau}^0)\|_{\text{HS},H^{-J}}^2 ds \right)^{\frac{1}{2}}$$

$$\le C\mathbb{E}\left[\sup_{s\in[0,t]} \|\zeta_{s\wedge\tau}^\varepsilon\|_{-J} \left(\int_0^t \|B(s \wedge \tau, \mu_{s\wedge\tau}^\varepsilon) - B(s \wedge \tau, \mu_{s\wedge\tau}^0)\|_{\text{HS},H^{-J}}^2 ds \right)^{\frac{1}{2}} \right]$$

$$\le \frac{1}{2}\mathbb{E} \sup_{s\in[0,t]} \|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2 + \frac{C^2}{2} \int_0^t \mathbb{E}\|B(s \wedge \tau, \mu_{s\wedge\tau}^\varepsilon) - B(s \wedge \tau, \mu_{s\wedge\tau}^0)\|_{\text{HS},H^{-J}}^2 ds$$

$$= \frac{1}{2}\mathbb{E} \sup_{s\in[0,t]} \|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2 + \frac{C^2}{2} I_4(t) \le \frac{1}{2}\mathbb{E} \sup_{s\in[0,t]} \|\zeta_{s\wedge\tau}^\varepsilon\|_{-J}^2 + \frac{C^2 C}{2}\varepsilon.$$

We remark that η_t^0, $t \in [0, T]$, is a Gaussian random element in $C([0, T], H^{-J+1}(\mathbb{R}^d))$ and $\mathbb{E} \sup_{t\in[0,T]} \|\eta_t\|_{-J+1}^2 < \infty$, by Proposition 3.4. Using [39, Proposition 3.14], one can see that there exists a constant $C > 0$ such that

(3.19) $$\mathbb{E} \sup_{t\in[0,T]} \|\eta_t^0\|_{-J+1}^4 \le C \left(\mathbb{E} \sup_{t\in[0,T]} \|\eta_t^0\|_{-J+1}^2 \right)^2 < \infty.$$

Defining for each $\varepsilon > 0$ the stopping time

$$\tau^\varepsilon = \inf\left\{ t \ge 0 : \|\eta_t^0\|_{-J} \ge \frac{1}{\sqrt{\varepsilon}} \right\}$$

and combining the estimates of $I_i(t)$, $i \in [5]$, with $\tau = \tau^\varepsilon$, we get

$$\mathbb{E} \sup_{s\in[0,t]} \|\zeta_{s\wedge\tau^\varepsilon}^\varepsilon\|_{-J}^2 \le C\varepsilon + C \int_0^t \mathbb{E} \sup_{r\in[0,s]} \|\zeta_{r\wedge\tau^\varepsilon}^\varepsilon\|_{-J}^2 ds$$

for all $t \in [0, T]$, where C is independent of ε. Using Gronwall's lemma, we obtain

(3.20) $$\mathbb{E} \sup_{s\in[0,t]} \|\zeta_{s\wedge\tau^\varepsilon}^\varepsilon\|_{-J}^2 \le Ce^{TC}\varepsilon, \quad t \in [0, T].$$

Set

$$\eta^* := \sup_{t\in[0,T]} \|\eta_t^0\|_{-J}.$$

and note that

$$\sup_{t\in[0,T]} \|\eta_t^\varepsilon\|_{-J} = \frac{1}{\sqrt{\varepsilon}} \sup_{t\in[0,T]} \|\mu_t^\varepsilon - \mu_t^0\|_{-J} \le \frac{2}{\sqrt{\varepsilon}}$$

for all $\varepsilon > 0$. Thus, by (3.19), (3.20), and Chebyshev's inequality, we obtain

$$
\mathbb{E} \sup_{t \in [0,T]} \|\zeta_t^\varepsilon\|_{-J}^2 = \mathbb{E}\left[\sup_{t \in [0,T]} \|\zeta_t^\varepsilon\|_{-J}^2 \mathbb{I}_{\left\{\eta^* < \frac{1}{\sqrt{\varepsilon}}\right\}} \right] + \mathbb{E}\left[\sup_{t \in [0,T]} \|\zeta_t^\varepsilon\|_{-J}^2 \mathbb{I}_{\left\{\eta^* \geq \frac{1}{\sqrt{\varepsilon}}\right\}} \right]
$$

$$
\leq \mathbb{E} \sup_{t \in [0,T]} \|\zeta_{t \wedge \tau^\varepsilon}^\varepsilon\|_{-J}^2 + \mathbb{E}\left[\sup_{t \in [0,T]} \|\eta_t^\varepsilon - \eta_t^0\|_{-J}^2 \mathbb{I}_{\left\{\eta^* \geq \frac{1}{\sqrt{\varepsilon}}\right\}} \right]
$$

$$
\leq C\varepsilon + 4\mathbb{E}\left[\sup_{t \in [0,T]} \|\eta_t^\varepsilon\|_{-J}^2 \mathbb{I}_{\left\{\eta^* \geq \frac{1}{\sqrt{\varepsilon}}\right\}} \right] + 4\mathbb{E}\left[\sup_{t \in [0,T]} \|\eta_t^0\|_{-J}^2 \mathbb{I}_{\left\{\eta^* \geq \frac{1}{\sqrt{\varepsilon}}\right\}} \right]
$$

$$
\leq C\varepsilon + \frac{4}{\varepsilon}\mathbb{E}\mathbb{I}_{\left\{\eta^* \geq \frac{1}{\sqrt{\varepsilon}}\right\}} + 4\varepsilon\mathbb{E} \sup_{t \in [0,T]} \|\eta_t^0\|_{-J}^4
$$

$$
\leq C\varepsilon + \left(\frac{4\varepsilon^2}{\varepsilon} + 4\varepsilon\right) \mathbb{E} \sup_{t \in [0,T]} \|\eta_t^0\|_{-J}^4 \leq C\varepsilon
$$

for all $\varepsilon > 0$. This completes the proof of the theorem. $\qquad\square$

4. Mean field limit and stochastic gradient descent

In this section, we consider the one-hidden layer neural network f^M defined by (1.1) for the approximation of a function $f : \mathbb{R}^{n_0} \to \mathbb{R}$, where $c_i \in \mathbb{R}$, U_i is an $1 \times n_0$-matrix, $\theta \in \mathbb{R}^{n_0}$, $x_i = (c_i, y_i) \in \mathbb{R} \times \mathbb{R}^{n_0} =: \mathbb{R}^d$, $b_i = 0$ and $\phi \in C_b^\infty(\mathbb{R})$ is a fixed activation function. We also assume that θ is a random element in $\Theta := \mathbb{R}^{n_0}$ with distribution ϑ. To minimize the risk function L defined by (1.2), the parameter $x = (x_i)_{i \in M}$ can be estimated using the stochastic gradient descent (1.3) with $P = 1$ and the learning rate $\alpha = \frac{\beta}{M}$, where θ_k, $k \in \mathbb{N}_0$, are i.i.d. random variables with distribution ϑ and $x_i(0)$, $i \in [M]$, are i.i.d. random variables generated from a distribution μ_0. We define the empirical distribution $\nu_t^{M,\frac{\beta}{M}}$, $t \geq 0$, of the network parameters by (1.15) which is a random element in the Skorohod space $\mathrm{D}([0,\infty), \mathcal{P}(\mathbb{R}^d))$ of all càdlàg functions from $[0,\infty)$ to $\mathcal{P}(\mathbb{R}^d)$ equipped with the Skorohod topology.

We next consider the stochastic mean-field equation (3.1) with the coefficients A, V, G defined by (1.5) with $\Phi(x_i, \theta)$ replaced by $\beta\Phi(x_i, \theta)$. Using the assumptions on the activation function ϕ, it is easy to see that A, V, G satisfy Assumptions 2.1, 2.7. Thus, by Theorem 2.16, for every $M \in \mathbb{N}$ there exists a unique superposition solution $\mu_t^{\frac{1}{M}}$, $t \geq 0$, to the stochastic mean-field equation (3.1) with $\varepsilon = \frac{1}{M}$ started from $\nu_0^{M,\frac{\beta}{M}}$.

We will further consider $\mu_t^{\frac{1}{M}}$, $t \geq 0$, and $\nu_t^{M,\frac{\beta}{M}}$, $t \geq 0$, as random processes in the Sobolev space $H^{-J}(\mathbb{R}^d)$ with $J > \frac{d}{2}$. Let $\mathrm{Law}_{T,-J}(\rho)$ denote the distribution of a random element ρ_t, $t \in [0,T]$, in the Skorohod space $\mathrm{D}([0,T], H^{-J}(\mathbb{R}^d))$. The distribution of the marginals ρ_t will be denoted by $\mathrm{Law}_{-J}(\rho_t)$. The following theorem is the main result of this section.

Theorem 4.1. *Let the measure ϑ be compactly supported on Θ, f be bounded on the support of ϑ and the processes $\mu_t^{\frac{1}{M}}$, $t \geq 0$, $\nu_t^{M,\frac{\beta}{M}}$, $t \geq 0$, be defined as above. Then for every $J \geq \frac{3d}{2} + 7$, $T > 0$ and $p \in [1,2)$*

$$
\tilde{\mathcal{W}}_p\left(\mathrm{Law}_{T,-J}(\mu^{\frac{1}{M}}), \mathrm{Law}_{T,-J}(\nu^{M,\frac{\beta}{M}})\right) = o(M^{-\frac{1}{2}}),
$$

where $\tilde{\mathcal{W}}_p$ is the p-Wasserstein distance on the space $\mathrm{D}([0,T], H^{-J}(\mathbb{R}^d))$ that is equipped with the uniform norm. In particular,

$$(4.1) \qquad \sup_{t \in [0,T]} \mathcal{W}_p(\mathrm{Law}_{-J}(\mu_t^{\frac{1}{M}}), \mathrm{Law}_{-J}(\nu_t^{M,\frac{\beta}{M}})) = o(M^{-\frac{1}{2}})$$

and for every $\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)$

$$(4.2) \qquad \sup_{t \in [0,T]} |\mathbb{E}\langle \varphi, \mu_t^{\frac{1}{M}} \rangle - \mathbb{E}\langle \varphi, \nu_t^{M,\frac{\beta}{M}} \rangle| = o(M^{-\frac{1}{2}}).$$

Proof. To simplify the notation, we will write μ_t^M and ν_t^M instead of $\mu_t^{\frac{1}{M}}$ and $\nu_t^{M,\frac{\beta}{M}}$, respectively. Let $T > 0$ and $J \geq \frac{3d}{2} + 7$ be fixed. Using [78, Theorem 1.2], we get that the process ν_t^M, $t \in [0,T]$, converges in $\mathrm{D}([0,T], \mathcal{P}(\mathbb{R}^d))$ in distribution to the unique superposition solution μ_t^0, $t \geq 0$, to the PDE (3.2) started from μ_0. We next consider the fluctuation field

$$\xi_t^M = \sqrt{M}(\nu_t^M - \mu_t^0), \quad t \in [0,T].$$

According to [77, Lemma 4.3], there exists a rectangle $[-R,R]^d$ such that μ_t^0, ν_t^M and ξ_t^M are supported on it for each $t \in [0,T]$ and $M \geq 1$. We take $\Gamma = (-3\sqrt{d}R, 3\sqrt{d}R)^d$. Then, by [77, Theorem 1.5], the process ξ_t^M, $t \in [0,T]$, converges in $\mathrm{D}([0,T], H^{-J}(\Gamma))$ in distribution to the continuous Gaussian process ξ_t, $t \in [0,T]$, in $H^{-J}(\Gamma)$ satisfying for every $\varphi \in \mathrm{C}_c^\infty(\Gamma)$

$$(4.3)$$

$$\langle \varphi, \xi_t \rangle_{0,\Gamma} = \langle \varphi, \xi_0 \rangle_{0,\Gamma} + \int_0^t \left(\langle \nabla\varphi \cdot V(\cdot, \mu_s^0), \xi_s \rangle_{0,\Gamma} + \left\langle \nabla\varphi \cdot \langle \tilde{V}(x,\cdot), \xi_s \rangle_{0,\Gamma}, \mu_s^0(dx) \right\rangle \right) ds$$
$$+ M_t^\varphi, \quad t \in [0,T],$$

where M_t^φ, $t \in [0,T]$, is a mean-zero Gaussian process with variance

$$(4.4) \qquad \mathrm{Var}(M_t^\varphi) = \int_0^t \int_\Theta \langle \nabla\varphi \cdot G(\cdot, \mu_s^0, \theta), \mu_s^0 \rangle^2 \vartheta(d\theta) ds, \quad t \in [0,T].$$

Moreover, by [77, Theorem 6.2], such a process ξ_t, $t \in [0,T]$, is unique. Using the Skorohod theorem [32, Theorem 3.1.8], we may assume that the process ξ^M converges to ξ in $\mathrm{D}([0,T], H^{-J}(\Gamma))$ a.s. Since ξ_t, $t \in [0,T]$, is continuous in $H^{-J}(\Gamma)$, it is easy to see that

$$\sup_{t \in [0,T]} \|\xi_t^M - \xi_t\|_{-J,\Gamma} \to 0 \quad \text{a.s.}$$

as $M \to \infty$. Finally, we remark that

$$\sup_{M \geq 1} \mathbb{E} \sup_{t \in [0,T]} \|\xi_t^M\|_{-J,\Gamma}^2 < \infty,$$

according to [77, Lemma 4.8]. Hence, by the de la Vallée-Poussin theorem [63, Theorem 1.8], the sequence $\sup_{t \in [0,T]} \|\xi_t^M\|_{-J,\Gamma}^p$, $M \geq 1$, is uniformly integrable for every $p \in [1,2)$. Therefore,

$$(4.5) \qquad \tilde{\mathcal{W}}_p^p \left(\mathrm{Law}_{T,-J,\Gamma}(\xi^M), \mathrm{Law}_{T,-J,\Gamma}(\xi) \right) \leq \mathbb{E} \sup_{t \in [0,T]} \|\xi_t^M - \xi_t\|_{-J,\Gamma}^p \to 0$$

as $M \to \infty$ for every $p \in [1,2)$, where $\mathrm{Law}_{T,-J,\Gamma}(\rho)$ is the law of a random element ρ_t, $t \in [0,T]$, in $\mathrm{D}([0,T], H^{-J}(\Gamma))$.

We next consider the fluctuation field $\eta_t^M = \sqrt{M}(\mu_t^M - \mu_t^0)$, $t \in [0,T]$, and note that, by Theorem 3.7,

$$\mathbb{E} \sup_{t \in [0,T]} \|\eta_t^M - \eta_t\|_{-J,\mathbb{R}^d}^2 \to 0 \quad \text{as } M \to \infty,$$

where η_t, $t \geq 0$, is a unique solution to the linear SPDE (3.4). Thus,

$$(4.6) \qquad \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J,\Gamma}(\eta^M), \mathrm{Law}_{T,-J,\Gamma}(\eta) \right) \to 0 \quad \text{as } M \to \infty$$

for every $p \in [1, 2]$.

Let us make the following observations. Since μ_t^M, $t \in [0, T]$, is a superposition solution to the stochastic mean-field equation (3.1) and μ_0 is compactly supported, without loss of generality we may assume that the support of μ_t^M is contained in Γ for each $M \geq 1$ and $t \in [0, T]$, otherwise, we may choose Γ larger. Taking an arbitrary $\varphi \in \mathrm{C}_c^\infty(\Gamma)$, it is easily seen that $\langle \varphi, \eta_t \rangle$, $t \in [0, T]$, satisfies (4.3), (4.4). Hence η_t, $t \in [0, T]$, is a continuous process in $H^{-J}(\Gamma)$ and has the same distribution as ξ_t, $t \in [0, T]$. We can see also that for every $p \in [1, 2)$

$$\sqrt{M^p} \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J}(\mu_t^M), \mathrm{Law}_{T,-J}(\nu_t^M) \right)^p$$

$$= \sqrt{M^p} \inf \left\{ \mathbb{E} \sup_{t \in [0,T]} \| \tilde{\mu}_t^M - \tilde{\nu}_t^M \|_{-J,\mathbb{R}^d}^p : \tilde{\mu}^M \sim \mathrm{Law}(\mu^M), \ \tilde{\nu}^M \sim \mathrm{Law}(\nu^M) \right\}$$

$$= \sqrt{M^p} \inf \left\{ \mathbb{E} \sup_{t \in [0,T]} \| \tilde{\mu}_t^M - \tilde{\nu}_t^M \|_{-J,\Gamma}^p : \tilde{\mu}^M \sim \mathrm{Law}(\mu^M), \ \tilde{\nu}^M \sim \mathrm{Law}(\nu^M) \right\}$$

$$= \inf \left\{ \mathbb{E} \sup_{t \in [0,T]} \left\| \sqrt{M}(\tilde{\mu}_t^M - \mu_t^0) - \sqrt{M}(\tilde{\nu}_t^M - \mu_t^0) \right\|_{-J,\Gamma}^p : \right.$$

$$\left. \tilde{\mu}^M \sim \mathrm{Law}(\mu^M), \ \tilde{\nu}^M \sim \mathrm{Law}(\nu^M) \right\}$$

$$= \inf \left\{ \mathbb{E} \sup_{t \in [0,T]} \| \tilde{\eta}_t^M - \tilde{\xi}_t^M \|_{-J,\Gamma}^p : \ \tilde{\eta}^M \sim \mathrm{Law}(\eta^M), \ \tilde{\xi}^M \sim \mathrm{Law}(\xi^M) \right\}$$

$$= \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J,\Gamma}(\eta^M), \mathrm{Law}_{T,-J,\Gamma}(\xi^M) \right).$$

Therefore, using the previous observations, we can estimate for every $p \in [1, 2)$

$$\sqrt{M} \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J}(\mu^M), \mathrm{Law}_{T,-J}(\nu^M) \right) = \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J,\Gamma}(\eta^M), \mathrm{Law}_{T,-J,\Gamma}(\xi^M) \right)$$

$$\leq \tilde{\mathcal{W}}_p(\mathrm{Law}_{T,-J,\Gamma}(\eta^M), \mathrm{Law}_{T,-J,\Gamma}(\eta)) + \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J,\Gamma}(\xi), \mathrm{Law}_{T,-J,\Gamma}(\xi^M) \right)$$

$$\leq \tilde{\mathcal{W}}_p(\mathrm{Law}_{T,-J}(\eta^M), \mathrm{Law}_{T,-J}(\eta)) + \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J,\Gamma}(\xi), \mathrm{Law}_{T,-J,\Gamma}(\xi^M) \right).$$

Thus,

$$\sqrt{M} \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{T,-J}(\mu^M), \mathrm{Law}_{T,-J}(\nu^M) \right) \to 0 \quad \text{as } M \to \infty,$$

by (4.5) and (4.6). This completes the proof of the first part of the theorem.

The asymptotics (4.1), (4.2) directly follow from the inequalities

$$\sup_{t \in [0,T]} \mathcal{W}_p \left(\mathrm{Law}_{-J}(\mu_t^M), \mathrm{Law}_{-J}(\nu_t^M) \right) \leq \tilde{\mathcal{W}}_p \left(\mathrm{Law}_{-J,T}(\mu^M), \mathrm{Law}_{-J,T}(\nu^M) \right)$$

and

$$\left| \mathbb{E}\langle \varphi, \mu_t^M \rangle - \mathbb{E}\langle \varphi, \nu_t^M \rangle \right| \leq \| \varphi \|_J \mathcal{W}_p(\mathrm{Law}_{-J}(\mu_t^M), \mathrm{Law}_{-J}(\nu_t^M)),$$

for every $\varphi \in \mathrm{C}_c^\infty(\mathbb{R}^d)$ and $t \in [0, T]$, respectively. $\qquad \square$

Remark 4.2. Note that the optimal rate in the quantified central limit Theorem 3.7 implies that

$$\tilde{\mathcal{W}}_p(\mathrm{Law}_{T,-J}(\eta^M), \mathrm{Law}_{T,-J}(\eta)) = O(M^{-\frac{1}{2}}).$$

Therefore, a quantified CLT for the SGD dynamics ν^M with the same rate of convergence would imply the stronger approximation error

$$\tilde{\mathcal{W}}_p\left(\mathrm{Law}_{T,-J}(\mu^M), \mathrm{Law}_{T,-J}(\nu^M)\right) = O(M^{-1}).$$

However, proving a quantified CLT for SGD is an open problem.

References

[1] Robert A. Adams, *Sobolev spaces*, Pure and Applied Mathematics, Vol. 65, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975. MR 0450957

[2] Alnur Ali, Edgar Dobriban, and Ryan Tibshirani, *The Implicit Regularization of Stochastic Gradient Flow for Least Squares*, Proceedings of the 37th International Conference on Machine Learning, PMLR, November 2020, pp. 233–244.

[3] Luigi Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math. **158** (2004), no. 2, 227–260. MR 2096794

[4] _____, *Transport equation and Cauchy problem for non-smooth vector fields*, Calculus of variations and nonlinear partial differential equations, Lecture Notes in Math., vol. 1927, Springer, Berlin, 2008, pp. 1–41. MR 2408257

[5] Luigi Ambrosio and Gianluca Crippa, *Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields*, Transport equations and multi-D hyperbolic conservation laws, Lect. Notes Unione Mat. Ital., vol. 5, Springer, Berlin, 2008, pp. 3–57. MR 2409676

[6] Anna Amirdjanova and Jie Xiong, *Large deviation principle for a stochastic Navier-Stokes equation in its vorticity form for a two-dimensional incompressible flow*, Discrete Contin. Dyn. Syst. Ser. B **6** (2006), no. 4, 651–666. MR 2223901

[7] Sebastian Andres and Max-K. von Renesse, *Particle approximation of the Wasserstein diffusion*, J. Funct. Anal. **258** (2010), no. 11, 3879–3905. MR 2606878

[8] Mikhail Belkin, Daniel Hsu, and Ji Xu, *Two models of double descent for weak features*, SIAM J. Math. Data Sci. **2** (2020), no. 4, 1167–1180. MR 4186534

[9] M. A. Belozerova, *Asymptotic behavior of solutions to stochastic differential equations with interaction*, Theory Stoch. Process. **25** (2020), no. 2, 1–8. MR 4354470

[10] Simone Bianco, Remi Cadene, Luigi Celona, and Paolo Napoletano, *Benchmark analysis of representative deep neural network architectures*, IEEE Access **6** (2018), 64270–64277.

[11] Le Chen, Davar Khoshnevisan, David Nualart, and Fei Pu, *Central limit theorems for parabolic stochastic partial differential equations*, Ann. Inst. Henri Poincaré Probab. Stat. **58** (2022), no. 2, 1052–1077. MR 4421618

[12] Zhengdao Chen, Grant Rotskoff, Joan Bruna, and Eric Vanden-Eijnden, *A dynamical central limit theorem for shallow neural networks*, Advances in Neural Information Processing Systems (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, Curran Associates, Inc., 2020, pp. 22217–22230.

[13] Lenaic Chizat and Francis Bach, *Implicit bias of gradient descent for wide two-layer neural networks trained with the logistic loss*, Conference on Learning Theory, PMLR, 2020, pp. 1305–1338.

[14] Lénaïc Chizat, Edouard Oyallon, and Francis Bach, *On lazy training in differentiable programming*, Curran Associates Inc., Red Hook, NY, USA, 2019.

[15] Lénaïc Chizat and Francis Bach, *On the Global Convergence of Gradient Descent for Overparameterized Models using Optimal Transport*, Advances in Neural Information Processing Systems, vol. 31, Curran Associates, Inc., 2018.

[16] Michele Coghi and Benjamin Gess, *Stochastic nonlinear Fokker-Planck equations*, Nonlinear Anal. **187** (2019), 259–278. MR 3954095

[17] Federico Cornalba and Julian Fischer, *The Dean-Kawasaki equation and the structure of density fluctuations in systems of diffusing particles*, arXiv:2109.06500 (2021).

[18] Federico Cornalba, Tony Shardlow, and Johannes Zimmer, *A regularized Dean-Kawasaki model: derivation and analysis*, SIAM J. Math. Anal. **51** (2019), no. 2, 1137–1187. MR 3936895

[19] ———, *From weakly interacting particles to a regularised Dean-Kawasaki model*, Nonlinearity **33** (2020), no. 2, 864–891. MR 4055366

[20] Donald Dawson and Jean Vaillancourt, *Stochastic McKean-Vlasov equations*, NoDEA Nonlinear Differential Equations Appl. **2** (1995), no. 2, 199–229. MR 1328577

[21] Donald A. Dawson, *Measure-valued Markov processes*, École d'Été de Probabilités de Saint-Flour XXI—1991, Lecture Notes in Math., vol. 1541, Springer, Berlin, 1993, pp. 1–260. MR 1242575

[22] David S. Dean, *Langevin equation for the density of a system of interacting Langevin processes*, J. Phys. A **29** (1996), no. 24, L613–L617. MR 1446882

[23] R. J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. **98** (1989), no. 3, 511–547. MR 1022305

[24] Nicolas Dirr, Benjamin Fehrman, and Benjamin Gess, *Conservative stochastic PDE and fluctuations of the symmetric simple exclusion process*, arXiv:2012.02126 (2020).

[25] Jasmina Djordjevic and Andrey Dorogovtsev, *Clark representation formula for the solution to equation with interaction*, Theory Stoch. Process. **25** (2020), no. 2, 9–14. MR 4354471

[26] A. A. Dorogovtsev, *Meroznachnye protsessy i stokhasticheskie potoki [Measurevalued processes and stochastic flows]*, vol. 66, Proceedings of Institute of Mathematics of NAS of Ukraine. Mathematics and its Applications, Institut Matematiki, Kiev, 2007 (Russian). MR 2375817

[27] A. A. Dorogovtsev and P. Kotelenez, *Smooth stationary solutions of quasilinear stochastic partial differential equations: 1. finite mass*, Preprint No. 97-145 Dep. Math. Case Western Reserv Univ., Cleveland, OH (1997), no. 97-145.

[28] A. A. Dorogovtsev and O. V. Ostapenko, *Large deviations for flows of interacting Brownian motions*, Stoch. Dyn. **10** (2010), no. 3, 315–339. MR 2671379

[29] Simon S. Du, Xiyu Zhai, Barnabas Poczos, and Aarti Singh, *Gradient descent provably optimizes over-parameterized neural networks*, International Conference on Learning Representations, 2019.

[30] Weinan E, Chao Ma, and Lei Wu, *Machine learning from a continuous viewpoint, I*, Sci. China Math. **63** (2020), no. 11, 2233–2266. MR 4170870

[31] Abdelhadi Es-Sarhir and Max-K. von Renesse, *Ergodicity of stochastic curve shortening flow in the plane*, SIAM J. Math. Anal. **44** (2012), no. 1, 224–244. MR 2888287

[32] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, Characterization and convergence. MR 838085

[33] Benjamin Fehrman and Benjamin Gess, *Well-posedness of the Dean-Kawasaki and the nonlinear Dawson-Watanabe equation with correlated noise*, arXiv:2108.08858 (2021).

[34] Alessio Figalli, *Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients*, J. Funct. Anal. **254** (2008), no. 1, 109–153. MR 2375067

[35] Franco Flandoli, *Remarks on uniqueness and strong solutions to deterministic and stochastic differential equations*, Metrika **69** (2009), no. 2-3, 101–123. MR 2481917

[36] Lucio Galeati and Dejun Luo, *LDP and CLT for SPDEs with Transport Noise*, arXiv:2201.09697 (2022).

[37] Leszek Gawarecki and Vidyadhar Mandrekar, *Stochastic differential equations in infinite dimensions with applications to stochastic partial differential equations*, Probability and its Applications (New York), Springer, Heidelberg, 2011. MR 2560625

[38] Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz, and Andrea Montanari, *When do neural networks outperform kernel methods?*, J. Stat. Mech. Theory Exp. (2021), no. 12, Paper No. 124009, 110. MR 4412837

[39] Martin Hairer, *An introduction to stochastic PDEs*, arXiv:0907.4178 (2009).

[40] Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J. Tibshirani, *Surprises in high-dimensional ridgeless least squares interpolation*, Ann. Statist. **50** (2022), no. 2, 949–986. MR 4404925

[41] Shulan Hu, Ruinan Li, and Xinyu Wang, *Central limit theorem and moderate deviations for a class of semilinear stochastic partial differential equations*, Acta Math. Sci. Ser. B (Engl. Ed.) **40** (2020), no. 5, 1477–1494. MR 4143604

[42] Jingyu Huang, David Nualart, Lauri Viitasaari, and Guangqu Zheng, *Gaussian fluctuations for the stochastic heat equation with colored noise*, Stoch. Partial Differ. Equ. Anal. Comput. **8** (2020), no. 2, 402–421. MR 4098872

[43] Nobuyuki Ikeda and Shinzo Watanabe, *Stochastic differential equations and diffusion processes*, second ed., North-Holland Mathematical Library, vol. 24, North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, 1989. MR 1011252

[44] Arthur Jacot, Franck Gabriel, and Clément Hongler, *Neural tangent kernel: convergence and generalization in neural networks (invited paper)*, STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, [2021] ©2021, p. 6. MR 4398812

[45] Adel Javanmard, Marco Mondelli, and Andrea Montanari, *Analysis of a two-layer neural network via displacement convexity*, Ann. Statist. **48** (2020), no. 6, 3619–3642. MR 4185822

[46] Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR 1876169

[47] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR 1121940

[48] Kyozi Kawasaki, *Stochastic model of slow dynamics in supercooled liquids and dense colloidal suspensions*, Physica A: Statistical Mechanics and its Applications **208** (1994), no. 1, 35 – 64.

[49] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang, *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*, arXiv:1609.04836 (2017).

[50] Vitalii Konarovskyi, *A system of coalescing heavy diffusion particles on the real line*, Ann. Probab. **45** (2017), no. 5, 3293–3335. MR 3706744

[51] Vitalii Konarovskyi, Tobias Lehmann, and Max von Renesse, *On Dean-Kawasaki dynamics with smooth drift potential*, J. Stat. Phys. **178** (2020), no. 3, 666–681. MR 4059955

[52] Vitalii Konarovskyi, Tobias Lehmann, and Max-K. von Renesse, *Dean-Kawasaki dynamics: ill-posedness vs. triviality*, Electron. Commun. Probab. **24** (2019), Paper No. 8, 9. MR 3916340

[53] Vitalii Konarovskyi and Max-K. von Renesse, *Modified massive Arratia flow and Wasserstein diffusion*, Comm. Pure Appl. Math. **72** (2019), no. 4, 764–800. MR 3914882

[54] Peter Kotelenez, *A class of quasilinear stochastic partial differential equations of McKean-Vlasov type with mass conservation*, Probab. Theory Related Fields **102** (1995), no. 2, 159–188. MR 1337250

[55] Peter M. Kotelenez, *Stochastic flows and signed measure valued stochastic partial differential equations*, Theory Stoch. Process. **16** (2010), no. 2, 86–105. MR 2779987

[56] Peter M. Kotelenez and Bradley T. Seadler, *On the Hahn-Jordan decomposition for signed measure valued stochastic partial differential equations*, Stoch. Dyn. **12** (2012), no. 1, 1150009, 23. MR 2887921

[57] Hiroshi Kunita, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics, vol. 24, Cambridge University Press, Cambridge, 1990. MR 1070361

[58] Thomas G. Kurtz and Jie Xiong, *Particle representations for a class of nonlinear SPDEs*, Stochastic Process. Appl. **83** (1999), no. 1, 103–126. MR 1705602

[59] ———, *Numerical solutions for a class of SPDEs with application to filtering*, Stochastics in finite and infinite dimensions, Trends Math., Birkhäuser Boston, Boston, MA, 2001, pp. 233–258. MR 1797090

[60] Qianxiao Li, Cheng Tai, and Weinan E, *Stochastic modified equations and adaptive stochastic gradient algorithms*, Proceedings of the 34th International Conference on Machine Learning (Doina Precup and Yee Whye Teh, eds.), Proceedings of Machine Learning Research, vol. 70, PMLR, 06–11 Aug 2017, pp. 2101–2110.

[61] Qianxiao Li, Cheng Tai, and Weinan E, *Stochastic modified equations and dynamics of stochastic gradient algorithms I: Mathematical foundations*, Journal of Machine Learning Research (JMLR) **20** (2019), Paper No. 40, 47. MR 3948080

[62] Zhiyuan Li, Sadhika Malladi, and Sanjeev Arora, *On the validity of modeling SGD with stochastic differential equations (SDEs)*, Advances in Neural Information Processing Systems (A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, eds.), 2021.

[63] Robert S. Liptser and Albert N. Shiryaev, *Statistics of random processes. I*, expanded ed., Applications of Mathematics (New York), vol. 5, Springer-Verlag, Berlin, 2001, General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability. MR 1800857

[64] Wei Liu, *Well-posedness of stochastic partial differential equations with Lyapunov condition*, J. Differential Equations **255** (2013), no. 3, 572–592. MR 3053478

[65] Song Mei and Andrea Montanari, *The generalization error of random features regression: precise asymptotics and the double descent curve*, Comm. Pure Appl. Math. **75** (2022), no. 4, 667–766. MR 4400901

[66] Song Mei, Andrea Montanari, and Phan-Minh Nguyen, *A mean field view of the landscape of two-layer neural networks*, Proc. Natl. Acad. Sci. USA **115** (2018), no. 33, E7665–E7671. MR 3845070

[67] Thanh Huy Nguyen, Umut Simsekli, Mert Gurbuzbalaban, and Gaël Richard, *First exit time analysis of stochastic gradient descent under heavy-tailed gradient noise*, Advances in neural information processing systems **32** (2019).

[68] A. Yu. Pilipenko, *Support theorem on stochastic flows with interaction*, Theory Stoch. Process. **12** (2006), no. 1-2, 127–141. MR 2316293

[69] Claudia Prévôt and Michael Röckner, *A concise course on stochastic partial differential equations*, Lecture Notes in Mathematics, vol. 1905, Springer, Berlin, 2007. MR 2329435

[70] Maria Refinetti, Sebastian Goldt, Florent Krzakala, and Lenka Zdeborová, *Classifying high-dimensional gaussian mixtures: Where kernel methods fail and neural networks succeed*, International Conference on Machine Learning, PMLR, 2021, pp. 8936–8947.

[71] Marco Rehmeier, *Linearization and a superposition principle for deterministic and stochastic nonlinear Fokker-Planck-Kolmogorov equations*, arXiv:2012.13530 (2020).

[72] Bruno Rémillard and Jean Vaillancourt, *On signed measure valued solutions of stochastic evolution equations*, Stochastic Process. Appl. **124** (2014), no. 1, 101–122. MR 3131288

[73] Herbert Robbins and Sutton Monro, *A stochastic approximation method*, Ann. Math. Statistics **22** (1951), 400–407. MR 42668

[74] Grant Rotskoff and Eric Vanden-Eijnden, *Parameters as interacting particles: Long time convergence and asymptotic error scaling of neural networks*, Advances in Neural Information Processing Systems, vol. 31, Curran Associates, Inc., 2018.

[75] Grant M. Rotskoff and Eric Vanden-Eijnden, *Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error*, CoRR **abs/1805.00915** (2018).

[76] ———, *Trainability and accuracy of neural networks: An interacting particle system approach*, arXiv:1805.00915 (2018).

[77] Justin Sirignano and Konstantinos Spiliopoulos, *Mean field analysis of neural networks: a central limit theorem*, Stochastic Process. Appl. **130** (2020), no. 3, 1820–1852. MR 4058290

[78] ———, *Mean field analysis of neural networks: a law of large numbers*, SIAM J. Appl. Math. **80** (2020), no. 2, 725–752. MR 4074020

[79] Dario Trevisan, *Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients*, Electron. J. Probab. **21** (2016), Paper No. 22, 41. MR 3485364

[80] Max-K. von Renesse and Karl-Theodor Sturm, *Entropic measure and Wasserstein diffusion*, Ann. Probab. **37** (2009), no. 3, 1114–1191. MR 2537551

[81] Jingfeng Wu, Difan Zou, Vladimir Braverman, and Quanquan Gu, *Direction matters: On the implicit bias of stochastic gradient descent with moderate learning rate*, International Conference on Learning Representations, 2021.

[82] Lei Wu, Chao Ma, and Weinan E, *How SGD Selects the Global Minima in Over-parameterized Learning: A Dynamical Stability Perspective*, Advances in Neural Information Processing Systems, vol. 31, Curran Associates, Inc., 2018.

[83] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals, *Understanding deep learning (still) requires rethinking generalization*, Communications of the ACM **64** (2021), no. 3, 107–115.

[84] Guodong Zhang, James Martens, and Roger B Grosse, *Fast convergence of natural gradient descent for over-parameterized neural networks*, Advances in Neural Information Processing Systems (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, Curran Associates, Inc., 2019.