Theory of Stochastic Processes
Vol. 29 (45), no. 1, 2025, pp. 52–64

K. M. KUSTAROVA

DIFFERENCE APPROXIMATION FOR EQUATIONS WITH INTERACTION

This paper investigates stochastic differential equations with interaction, introduced by Dorogovtsev the model of the evolution of large systems of interacting particles in random environments. The study emphasizes the difference approximation scheme for these equations, which involve approximating solutions in an infinite-dimensional, nonlinear space of measures. The key contributions include the formulation of approximation schemes for compactly supported initial measures, the derivation of Wasserstein distance-based estimates, and spatial discretization techniques.

1. Introduction

This paper is devoted to stochastic differential equations with interaction. Such equations were introduced by Dorogovtsev in order to describe the evolution of the large systems of interacting particles in random media [2]. The main equation has the following form:

$$\mathrm{d}x(u,t) = a(x(u,t),\mu_t)\,\mathrm{d}t + \int_{\mathbb{R}^d} b(x(u,t),\mu_t,p)\,W(\mathrm{d}p,\mathrm{d}t), \tag{1}$$

where

$$x(u,0)=u, \quad u\in\mathbb{R}^d, \quad a:\mathbb{R}^d\times\mathfrak{M}_2(\mathbb{R}^d)\to\mathbb{R}^d, \quad b:\mathbb{R}^d\times\mathfrak{M}_2(\mathbb{R}^d)\to\mathbb{R}^{d\times d}$$

and

$$\mu_t = \mu_0\circ x(\cdot,t)^{-1}.$$

Here the probability measure μ_0 plays a role of the initial mass distribution. For every $u\in\mathbb{R}^d$ the random process $x(u,t), t\geq 0$, describes the trajectory of a particle which starts from the point u. W is a Wiener sheet on $\mathbb{R}^d\times[0,+\infty)$ i.e. the centered Gaussian measure with independent values on disjoint sets and Lebesgue measure as a structure measure. W is responsible for the influence of the random media. The presence of μ in the coefficients of equation reflects the fact that the trajectory of each particle depends on the mass distribution of the whole system. In order to guarantee the existence of the solution, the coefficients of the equation are assumed to be Lipschitz continuous with respect to the measure-valued argument in the Wasserstein distance of some order.

In recent years, there has been growing interest in stochastic differential equations with interaction, because they provide a natural way to describe the behavior of large systems of particles influenced by randomness. Such models are not limited to single trajectories but describe whole stochastic flows and the associated measure-valued processes. This viewpoint is important both for theoretical reasons and for practical applications such as simulation of particle systems. A number of works have developed this direction. For example, Gess, Kassing, and Konarovskyi [3] studied stochastic gradient descent (SGD) using the language of interacting particle systems. They showed that SGD can be approximated by conservative stochastic partial differential equations (SPDEs) in the

2000 *Mathematics Subject Classification.* 60H10 · 60H35 · 65C30.

Key words and phrases. Stochastic differential equations · Interacting particle systems · Difference approximation · Euler–Maruyama scheme · Wasserstein distance · Spatial discretization.

mean-field limit, and that fluctuations around the limit are essential for accuracy. In the same paper, the authors introduced Stochastic Modified Flows (SMFs), which are stochastic differential equations with additional regularizing terms. These flows allow precise control of statistics and describe SGD in regimes with small learning rates.

This line of research builds on earlier contributions. Dorogovtsev [2] developed a framework for stochastic flows with interaction and measure-valued processes. Kunita [4] studied stochastic flows in detail and proved continuity results for random fields. Related ideas can also be found in the works of Dawson [1] and Sznitman [7] on measure-valued limits and propagation of chaos, and in Villani [8] on Wasserstein geometry. These works provide the background for the present paper. The focus here is on equations with interaction of Dorogovtsev type. Such equations describe both the flow $x(\cdot, t), t \geq 0$ and the induced process of measures μ_t. Because the dynamics of each particle depend on the whole distribution, the state space is infinite-dimensional and nonlinear. This makes the construction of approximation schemes more complicated than in the classical case.

2. Euler-Maruyama scheme for compactly supported initial measure μ_0

The goal of this paper is to build a rigorous difference approximation scheme for these equations. We do this in several steps. First, we consider compactly supported initial measures μ_0 and construct an Euler–Maruyama type scheme [5]. Next, we study stability of solutions with respect to different initial measures, using Wasserstein distance estimates. Finally, we perform spatial discretization and prove quantitative convergence rates.

The main tools are the Burkholder–Davis–Gundy inequality [6], the Kolmogorov–Totoki continuity theorem [4], and stability estimates for interacting stochastic flows. In particular, the Kolmogorov–Totoki theorem guarantees the existence of Hölder-continuous modifications of the solution, which allows us to control the supremum norm of the error over compact sets.

The main contribution of this work is to give explicit convergence rates for the proposed approximation scheme and to provide a mathematical justification for numerical methods in interacting particle systems. In this section, we will develop the Euler-Maruyama scheme for equations with interaction, specifically for compactly supported initial measures μ_0. We will demonstrate the convergence of the scheme, and provide error estimates to support its practical implementation.

We consider the equation (1) with the additional condition

$$\operatorname{supp} \mu_0 \subset [-\mathbb{R}, \mathbb{R}]^d = K$$

for some positive R. Since the measure μ_t is an image of μ_0 under the mapping $x(\cdot, t)$, then it is natural to think that the restriction of x on K plays important role in construction of the solution and must be approximated first. Let us recall the definitions of Wasserstein distances. For two probability measures μ_1, μ_2 on $\mathbb{R}^d$ (all measures are supposed to be defined on the Borel σ-field of corresponding space) define $C(\mu_1, \mu_2)$ as a set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ which have μ_1 and μ_2 as its marginal projections. Now denote by $\mathfrak{M}_2$ the set of probability measures on $\mathbb{R}^d$ which have finite n-th moment. In particular $\mathfrak{M}_0$ is the set of all probability measures.

The Wasserstein distances $\gamma_0(\mu_1, \mu_2)$ and $\gamma_2(\mu_1, \mu_2)$ are defined as follows:

$$\gamma_0(\mu_1, \mu_2) = \inf_{\zeta \in C(\mu_1, \mu_2)} \iint_{\mathbb{R}^d} \frac{|u - v|}{1 + |u - v|} \, \zeta(\mathrm{d}u, \mathrm{d}v), \tag{2}$$

$$\gamma_2(\mu_1, \mu_2) = \inf_{\zeta \in C(\mu_1, \mu_2)} \left(\iint_{\mathbb{R}^d} |u - v|^2 \, \zeta(\mathrm{d}u, \mathrm{d}v) \right)^{\frac{1}{2}}. \tag{3}$$

It is well known that the space $(\mathfrak{M}_2, \gamma_2)$ is a complete separable metric space [8]. As shown in Theorem 2.3.1 of [2], if the coefficients of the equation with interaction are Lipschitz continuous with respect to their arguments (in the γ_2-metric on the space of measures), then the solution exists and is unique (see also Theorem 2.1.2 in [2]).

The functions

$$a : \mathbb{R}^d \times \mathfrak{M}_2 \to \mathbb{R}^d, \quad b : \mathbb{R}^d \times \mathfrak{M}_2 \times \mathbb{R}^d \to \mathbb{R}^d$$

are said to satisfy the Lipschitz condition if there exist constants $L > 0$ such that for all $x, y \in \mathbb{R}^d$, $\mu^1, \mu^2 \in \mathfrak{M}_2$, the following inequalities hold:

$$|a(x,\mu^1) - a(y,\mu^2)| \le L\left(|x-y| + \gamma_2(\mu^1,\mu^2)\right), \tag{4}$$

$$\int_{\mathbb{R}^d} |b(x,\mu^1,p) - b(y,\mu^2,p)|^2 \,\mathrm{d}p \le L^2\left(|x-y|^2 + \gamma_2^2(\mu^1,\mu^2)\right). \tag{5}$$

In what follows, we assume this Lipschitz condition holds unless explicitly specified otherwise. For $n > 1$, define firstly the values of the approximating sequence in the points of the uniform partition $t_k = \frac{k}{n}$, where $k = 0, 1, \ldots, n$. For $t \in [t_k, t_{k+1})$ we define the approximation:

$$\begin{aligned} x_n(t,u) = {} & x_n(t_k,u) + a\big(x_n(t_k,u), \mu^n_{t_k}\big)(t - t_k) \\ & + \int_{t_k}^t \int_{\mathbb{R}^d} b\big(x_n(t_k,u), \mu^n_{t_k}, p\big) W(\mathrm{d}p, \mathrm{d}s), \qquad x_n(0,u) = u, \end{aligned} \tag{6}$$

where $\mu^n_{t_k}$ denotes the empirical measure of the approximation at time t_k, i.e.

$$\mu^n_{t_k} = \mu_0 \circ \big(x_n(\cdot, t_k)\big)^{-1}.$$

We are going to prove the convergence of this scheme to the solution and estimate the speed of convergence. To do this we will use Kolmogorov condition of continuity. This condition can guarantee Hölder continuity and estimate the tails of distribution of Hölder coefficient. So, let us start from the moments estimation.

Lemma 2.1. *For fixed $q \in \mathbb{N}$, there exists a constant $C_q > 0$ such that*

$$\forall n \ge 1 \ \forall u \in K \quad \mathbb{E} \max_{t \in [0,1]} \|x_n(t,u)\|^{2q} \le C_q.$$

Proof. Let us denote $t_k = \frac{k}{n}$, where $k = 0, 1, \ldots, n$, and n is the partition size. Note that

$$\begin{aligned} x_n(t,u) = {} & u + \int_0^t \sum_{k=0}^{n-1} a\big(x_n(t_k,u), \mu^n_{t_k}\big)\, 1_{[t_k,t_{k+1})}(s)\,\mathrm{d}s \\ & + \int_0^t \sum_{k=0}^{n-1} \int_{\mathbb{R}^d} b\big(x_n(t_k,u), \mu^n_{t_k}, p\big)\, 1_{[t_k,t_{k+1})}(s)\, W(\mathrm{d}p, \mathrm{d}s). \end{aligned} \tag{7}$$

Hence,

$$\begin{aligned} \mathbb{E} \sup_{t\in[0,1]} \|x_n(t,u)\|^{2q} \le C_q^1 \Bigg(& \|u\|^{2q} + \mathbb{E} \sup_{t\in[0,1]} \left\| \int_0^t \sum_{k=0}^{n-1} a\big(x_n(t_k,u), \mu^n_{t_k}\big)\, \mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s \right\|^{2q} \\ & + \mathbb{E} \sup_{t\in[0,1]} \left\| \int_0^t \int_{\mathbb{R}^d} \sum_{k=0}^{n-1} b\big(x_n(t_k,u), \mu^n_{t_k}, p\big)\, \mathbf{1}_{[t_k,t_{k+1})}(s)\, W(\mathrm{d}p, \mathrm{d}s) \right\|^{2q} \Bigg). \end{aligned} \tag{8}$$

Using the Burkholder–Davis–Gundy inequality, we obtain:

$$\begin{aligned}&\mathbb{E}\sup_{t\in[0,1]}\left\|\int_0^t\int_{\mathbb{R}^d}\sum_{k=0}^{n-1}b\big(x_n(t_k,u),\mu_{t_k}^n,p\big)\,\mathbf{1}_{[t_k,t_{k+1})}(s)\,W(\mathrm{d}p,\mathrm{d}s)\right\|^{2q}\\ &\qquad\le\tilde{C}_q\,\mathbb{E}\left(\int_0^1\int_{\mathbb{R}^d}\left\|\sum_{k=0}^{n-1}b\big(x_n(t_k,u),\mu_{t_k}^n,p\big)\,\mathbf{1}_{[t_k,t_{k+1})}(s)\right\|^2\mathrm{d}p\,\mathrm{d}s\right)^q.\end{aligned}\tag{9}$$

Applying Jensen's inequality:

$$\begin{aligned}&\mathbb{E}\sup_{t\in[0,1]}\left\|\int_0^t\sum_{k=0}^{n-1}a\big(x_n(t_k,u),\mu_{t_k}^n\big)\,\mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s\right\|^{2q}\\ &\qquad\le\tilde{C}_q^2\,\mathbb{E}\left(\int_0^1\left\|\sum_{k=0}^{n-1}a\big(x_n(t_k,u),\mu_{t_k}^n\big)\,\mathbf{1}_{[t_k,t_{k+1})}(s)\right\|\mathrm{d}s\right)^{2q}.\end{aligned}\tag{10}$$

Using the Lipschitz condition and the boundedness of the functions a and b, we obtain the following upper bound:

$$\begin{aligned}&C_q^1\tilde{C}_q^2\,\mathbb{E}\sup_{t\in[0,1]}\left\|\int_0^t\sum_{k=0}^{n-1}a\big(x_n(t_k,u),\mu_{t_k}^n\big)\,\mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s\right\|^{2q}\\ &\quad+C_q^1\tilde{C}_q^1\,\mathbb{E}\left(\int_0^1\int_{\mathbb{R}^d}\left\|\sum_{k=0}^{n-1}b\big(x_n(t_k,u),\mu_{t_k}^n,p\big)\,\mathbf{1}_{[t_k,t_{k+1})}(s)\right\|^2\mathrm{d}p\,\mathrm{d}s\right)^q\le C_q,\end{aligned}\tag{11}$$

where $C_q>0$ depends only on q, d, the compact set K, and the Lipschitz constants of a and b, but is independent of n and u.

Finally, we obtain

$$\mathbb{E}\sup_{t\in[0,1]}\|x_n(t,u)\|^{2q}\le C_q^1\|u\|^{2q}+C_q.\tag{12}$$

Since $u\in K$, the norm $\|u\|$ is uniformly bounded, so the right-hand side is bounded by a constant $\widetilde{C}_q$ independent of n and u. This completes the proof. □

The next lemma estimates the differences between the values of $x_n(t)$ at different points of K.

Lemma 2.2. *For every $p\in\mathbb{N}$ there exists a positive constant C_p such that*

$$\mathbb{E}\sup_{n\ge1}\sup_{t\in[0,1]}\|x_n(t,u_2)-x_n(t,u_1)\|^{2p}\le C_p\|u_2-u_1\|^{2p}.$$

Proof. Using the same integral representation as in the proof of the previous lemma and the Lipschitz continuity 2 of the functions, we have:

$$\begin{aligned}x_n(t,u_2)-x_n(t,u_1)&=u_2-u_1\\ &\quad+\int_0^t\sum_{k=0}^{n-1}\big[a(x_n(t_k,u_2),\mu_{t_k}^n)-a(x_n(t_k,u_1),\mu_{t_k}^n)\big]\mathbf{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}s\\ &\quad+\int_0^t\int_{\mathbb{R}^d}\sum_{k=0}^{n-1}\big[b(x_n(t_k,u_2),\mu_{t_k}^n,p)-b(x_n(t_k,u_1),\mu_{t_k}^n,p)\big]\\ &\qquad\times\mathbf{1}_{[t_k,t_{k+1})}(s)\,W(\mathrm{d}p,\mathrm{d}s).\end{aligned}$$

Applying the Burkholder–Davis–Gundy inequality, we obtain:

$$\mathbb{E}\sup_{t\in[0,1]}\|x_n(t,u_2)-x_n(t,u_1)\|^{2p}\le C_p\|u_2-u_1\|^{2p}$$
$$+C_pL^{2p}\int_0^1\mathbb{E}\sup_{z\in[0,s]}\|x_n(z,u_2)-x_n(z,u_1)\|^{2p}\,\mathrm{d}s.$$

Here, L denotes the Lipschitz constant from Definition 2, shared by both a and b. Applying the Gronwall–Bellman inequality, we obtain:

$$\mathbb{E}\sup_{t\in[0,1]}\|x_n(t,u_2)-x_n(t,u_1)\|^{2p}\le C_p\|u_2-u_1\|^{2p},$$

which concludes the proof. □

By Lemma 2.2 and Kolmogorov's continuity theorem, the process x_n admits a Hölder-continuous modification with respect to both variables.

We recall that Theorem 2.1 is a version of the Kolmogorov–Totoki continuity criterion, originally stated by Kunita [4] (see Lemmas 1.8.1 and 1.8.2, p. 41), which provides sufficient conditions for the existence of a Hölder-continuous modification of a stochastic field.

Theorem 2.1. *Let the random $\mathbb{R}^d$-valued field y on K satisfies condition from Kolmogorov's criteria, i.e., there exist the positive constants γ, α, C such that $\alpha>d$ and*

$$\forall u,v\in K\quad \mathbb{E}\|y(u)-y(v)\|^{\gamma}\le C\|u-v\|^{\alpha}.$$

Then for some absolute constants (depending on K) R

$$\big(\mathbb{E}\sup_{u\in K}\|y(u)\|^{\gamma}\big)^{1/\gamma}\le R\left(\mathbb{E}\|y(0)\|^{\gamma}\right)^{1/\gamma}+R\cdot C^{1/\gamma}.$$

As a consequence of this statement, and by combining the moment bounds from Lemma 2.2 with the continuity criterion provided by Theorem 2.1, we obtain an explicit estimate for the rate of convergence of the proposed approximation scheme for the equation with interaction (1).

3. Spatial discretization

In this section, we aim to establish the convergence of the numerical scheme under consideration. Specifically, we will derive error estimates for the proposed approximation, analyze the bounds for the difference between the exact and numerical solutions, and validate the method's applicability through theoretical results.

Let $K\subset\mathbb{R}^d$ represent the spatial domain. To approximate initial measure in this domain, we select N^d grid points $\{u_{i_1,\dots,i_d}\}$, where $i_1,i_2,\dots,i_d$ are indices defining the grid points in each dimension of the domain for spatial discretization.

In the case where $d=1$, i.e., for a one-dimensional domain, partitioning is straightforward. We divide the interval into N equal subintervals, resulting in grid points u_i spaced evenly by a step size $\Delta u=\frac{R-(-R)}{N}$, assuming the domain is $K=[-R,R]$.

For the case of a domain with dimensionality $d>1$, the partitioning of the domain $K\subset\mathbb{R}^d$ involves evenly dividing it along each coordinate axis. The domain K is divided into N^d grid cells, where N represents the number of divisions in each dimension. Each grid cell corresponds to a rectangle formed by the Cartesian product of intervals along different coordinate axes:

$$K=\prod_{i=1}^{d}[a_i,b_i)\,,$$

where, for each coordinate i, the interval is divided into N equal subintervals:

$$[a_i, b_i] = \bigcup_{k=1}^{N} [a_i + (k-1)\Delta x_i, a_i + k\Delta x_i), \quad \Delta x_i = \frac{b_i - a_i}{N},$$

where Δx_i is the step size along the i-th coordinate. Therefore, each grid cell can be represented as

$$\prod_{i=1}^{d} [a_i + (k_i - 1)\Delta x_i, a_i + k_i \Delta x_i), \quad k_i = 1, \dots, N.$$

Thus, there are N^d cells in total, each of which is a d-dimensional rectangle. To approximate the initial measure μ_0 over the domain, we define an empirical measure μ_0^N that is concentrated at these grid points. This empirical measure is constructed as a weighted sum of Dirac delta functions, each centered at one of the grid points $u_{i_1,\dots,i_d}$:

$$\mu_0^N = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_d=1}^{N_d} \alpha_{i_1,i_2,\dots,i_d} \delta_{u_{i_1,i_2,\dots,i_d}},$$

The weight $\alpha_{i_1,i_2,\dots,i_d}$ for each rectangle is calculated based on the measure μ_0 concentrated within the grid cell. Each weight corresponds to the measure of a subregion of K, defined by the product of intervals in each coordinate direction:

$$\alpha_{i_1,i_2,\dots,i_d} = \mu_0 \left(\prod_{k=1}^{d} \left[\frac{i_k}{n}, \frac{i_k + 1}{n} \right) \right),$$

where $\prod_{k=1}^{d} \left[\frac{i_k}{n}, \frac{i_k+1}{n} \right)$ represents the hyperrectangle in $\mathbb{R}^d$ corresponding to the grid point $u_{i_1,i_2,\dots,i_d}$.

To calculate the Wasserstein distance, we define a piecewise function $f : \mathbb{R}^d \to \mathbb{R}^d$ as follows:

$$f(\vec{u}) = \frac{\vec{k}}{n}, \quad \vec{u} = \left\{ u_i : u_i \in \left[\frac{k_i}{n}, \frac{k_i + 1}{n} \right) \right\}$$

where $\vec{k} = (k_1, k_2, \dots, k_d)$.

Using this piecewise function f, we can estimate the Wasserstein distance $\gamma_2(\mu_0, \mu_0^N)$ between the initial and discrete measures as:

$$\gamma_2^2(\mu_0, \mu_0^N) \le \int_{\mathbb{R}^d} ||f(\vec{u}) - \vec{u}||^2 \mu_0(d\vec{u}),$$

In this context, μ_0 represents a probability measure defined on $\mathbb{R}^d$. It describes the distribution of the random variable u, which is used to compute the expected squared distance in the 2-Wasserstein metric 2. Specifically, μ governs the weight assigned to each point in $\mathbb{R}^d$ during integration, which can be further simplified to:

$$\gamma_2^2(\mu_0, \mu_0^N) \le \int_{\mathbb{R}^d} ||f(\vec{u}) - \vec{u}||^2 \mu_0(d\vec{u}) \le \frac{C}{n^2} \int_{\mathbb{R}^d} \mu_0(\mathrm{d}u) = \frac{C}{n^2}.$$

4. Main convergence results

In this section we establish the main convergence results for the proposed difference approximation scheme. Building on the moment estimates of Section 2 and the spatial discretization framework of Section 3, we derive stability bounds for solutions corresponding to different initial measures and prove quantitative convergence rates in the Wasserstein metric.

Lemma 4.1. *Let $x_1(t,u)$ and $x_2(t,u)$ be two stochastic processes satisfying the stochastic differential equations (1). Here μ_0^1 and μ_0^2 denote the probability measures of the initial conditions $x_1(0,u)$ and $x_2(0,u)$, respectively. Then, for every $p \geq 2$, there exists a constant $C_p > 0$ such that*

$$\mathbb{E} \sup_{t\in[0,1]} \|x_1(t,u) - x_2(t,u)\|^{2p} \leq C_p \cdot \gamma_2^{2p}(\mu_0^1, \mu_0^2),$$

where γ_2 denotes the 2-Wasserstein distance as defined in Definition 2.

Proof. Subtracting the equations, we write

$$\begin{aligned} x_2(t,u) - x_1(t,u) = &\int_0^t \big(a(x_2(s,u),\mu_s^2) - a(x_1(s,u),\mu_s^1)\big)\, ds \\ &+ \int_0^t \int_{\mathbb{R}^d} \big(b(x_2(s,u),\mu_s^2,p) - b(x_1(s,u),\mu_s^1,p)\big)\, W(\mathrm{d}p,\mathrm{d}s). \end{aligned}$$

Define

$$\begin{aligned} I_t &= \int_0^t \big(a(x_2(s,u),\mu_s^2) - a(x_1(s,u),\mu_s^1)\big)\, \mathrm{d}s, \\ II_t &= \int_0^t \int_{\mathbb{R}^d} \big(b(x_2(s,u),\mu_s^2,p) - b(x_1(s,u),\mu_s^1,p)\big)\, W(\mathrm{d}p,\mathrm{d}s). \end{aligned}$$

Then

$$\mathbb{E} \sup_{t\in[0,1]} \|x_2(t,u) - x_1(t,u)\|^{2p} \leq C_p \left(\mathbb{E} \sup_{t\in[0,1]} \|I_t\|^{2p} + \mathbb{E} \sup_{t\in[0,1]} \|II_t\|^{2p} \right).$$

Using the Lipschitz condition 2 on a, we get:

$$\|a(x_2(s,u),\mu_s^2) - a(x_1(s,u),\mu_s^1)\| \leq L_a(\|x_2(s,u) - x_1(s,u)\| + \gamma_2(\mu_s^1,\mu_s^2)).$$

Thus the inequality,

$$\|I_t\| \leq \int_0^t L_a(\|x_2(s,u) - x_1(s,u)\| + \gamma_2(\mu_s^1,\mu_s^2))ds.$$

Consequently,

$$\mathbb{E} \sup_{t\in[0,1]} \|I_t\|^{2p} \leq C_p \int_0^1 \mathbb{E}\left[\|x_2(s,u) - x_1(s,u)\|^{2p} + \gamma_2^{2p}(\mu_s^1,\mu_s^2)\right] ds.$$

For I_t, applying the Burkholder–Davis–Gundy inequality and the Lipschitz continuity of b, we obtain

$$\mathbb{E} \sup_{t\in[0,1]} \|II_t\|^{2p} \leq C_p \int_0^1 \mathbb{E}\left[\|x_2(s,u) - x_1(s,u)\|^{2p} + \gamma_2^{2p}(\mu_s^1,\mu_s^2)\right] \mathrm{d}s.$$

Let $g(t) = \mathbb{E}\sup_{r\in[0,t]} \|x_2(r,u) - x_1(r,u)\|^{2p}$, and $\varphi(t) = \mathbb{E}[\gamma_2^{2p}(\mu_t^1,\mu_t^2)]$. Then

$$g(t) \leq C_p \int_0^t g(s)\mathrm{d}s + C_p \int_0^t \varphi(s)\mathrm{d}s.$$

To estimate $\varphi(t)$, fix $\kappa \in C(\mu_0^1,\mu_0^2)$. Then

$$\gamma_2^2(\mu_t^1,\mu_t^2) \leq \int \|x_1(t,u) - x_2(t,v)\|^2\, \kappa(\mathrm{d}u,\mathrm{d}v).$$

Raising both sides to the power p, we obtain

$$\gamma_2^{2p}(\mu_t^1,\mu_t^2) \leq \left(\int \|x_1(t,u) - x_2(t,v)\|^2\, \kappa(\mathrm{d}u,\mathrm{d}v)\right)^p.$$

Applying the elementary inequality

$$\|x_1(t,u)-x_2(t,v)\|^2 \le 2\|x_1(t,u)-x_1(t,v)\|^2 + 2\|x_1(t,v)-x_2(t,v)\|^2,$$

and then raising both sides to the power p and using a triangle-type inequality, we obtain:

$$\gamma_2^{2p}(\mu_t^1,\mu_t^2) \le C_p \int \left[\|x_1(t,u)-x_1(t,v)\|^{2p} + \|x_1(t,v)-x_2(t,v)\|^{2p}\right] \kappa(\mathrm{d}u,\mathrm{d}v).$$

Taking expectation:

$$\varphi(t) \le C_p \int \mathbb{E}\|x_1(t,u)-x_1(t,v)\|^{2p}\, \kappa(\mathrm{d}u,\mathrm{d}v) + C_p g(t).$$

Applying Lemma 2.2, we get

$$\mathbb{E}\|x_1(t,u)-x_1(t,v)\|^{2p} \le C_p\|u-v\|^{2p},$$

so

$$\varphi(t) \le C_p\gamma_2^{2p}(\mu_0^1,\mu_0^2) + C_p f(t).$$

Substituting into inequality for $g(t)$:

$$g(t) \ \le\ C_p \int_0^t g(s)\,\mathrm{d}s + C_p\,\gamma_2^{2p}(\mu_0^1,\mu_0^2)\cdot t.$$

By Grönwall–Bellman inequality:

$$g(t) \ \le\ C_p\,\gamma_2^{2p}(\mu_0^1,\mu_0^2), \quad t\in[0,1].$$

Substituting back into $\varphi(t)$, we also get:

$$\varphi(t) \ \le\ C_p\,\gamma_2^{2p}(\mu_0^1,\mu_0^2).$$

This completes the proof. ☐

Theorem 4.1. *Let μ_0^1,μ_0^2 be two probability measures on $\mathbb{R}^d$ with finite second moments. Let x_1,x_2 denote the solutions of equation(1) with initial measures μ_0^1 and μ_0^2, respectively. Then, for any $p\ge 1$ and $\gamma<\frac{p-d}{p}$ (in particular, $p>d$), we have*

$$\mathbb{E}\sup_{u\in K}\sup_{t\in[0,1]}\|x_1(u,t)-x_2(u,t)\| \ \le\ C\,\left(\gamma_2^2(\mu_0^1,\mu_0^2)\right)^{\frac{\gamma}{\gamma+d}},$$

where C is a constant independent of μ_0^1,μ_0^2.

Proof. By Lemma 4.1, and assuming the Lipschitz continuity of the coefficients a and b with respect to the measure argument in the γ_2-metric (as stated in Definition 2 and Definition 2), we obtain the following moment estimate:

$$\mathbb{E}\sup_{t\in[0,1]}\|x_1(u,t)-x_2(u,t)\|^p \ \le\ C_p\,\gamma_2^{2p}(\mu_0^1,\mu_0^2),$$

where $p\ge 1$ and C_p is a constant depending only on the coefficients of the equation. In addition, by analogy with Lemma 2.2, which captures the sensitivity of the solution with respect to changes in the spatial variable, we have:

$$\mathbb{E}\sup_{t\in[0,1]}\|x_i(u_1,t)-x_i(u_2,t)\|^p \ \le\ C_p\,\|u_1-u_2\|^p,$$

for each $u_1,u_2\in K$ and $i\in\{1,2\}$. To proceed, we apply a functional version of the Kolmogorov–Totoki continuity theorem (see Kunita [4], Lemmas 1.8.1 and 1.8.2). Define the random function:

$$y(u,\cdot) \ =\ x_1(u,\cdot)-x_2(u,\cdot), \quad u\in K,$$

as a stochastic process with values in $C([0,1],\mathbb{R}^d)$. The difference process $y(u,\cdot)$ satisfies the following Hölder-type condition almost surely:

$$\|y(u_1,\cdot)-y(u_2,\cdot)\|_\infty \le \eta\|u_1-u_2\|^\gamma,$$

for all $u_1, u_2 \in K$, where $\|f\|_\infty = \sup_{t\in[0,1]} \|f(t)\|$, and η is a non-negative random variable satisfying $\mathbb{E}[\eta^p] < \infty$. Here η is independent of the choice of u_1, u_2, whose distribution depends only on the processes x_1, x_2 and the coefficients of the equation. Here $\gamma < \frac{p-d}{p}$ and d is the spatial dimension. In particular, this condition is non-trivial only if $p > d$. Here the random variable η and the constants involved depend only on p, d, the compact set K, and the Lipschitz bounds of the coefficients a and b, but are independent of the particular choice of the processes x_1, x_2. To estimate the supremum norm of the random field $y(u, \cdot)$ over K, we construct a δ-net $\{u_j\}_{j=1}^{N_\delta}$ of the compact set K, with $N_\delta \le C\delta^{-d}$. For each $u \in K$, there exists u_j in the net such that $\|u - u_j\| \le \delta$, which gives:

$$\|y(u,\cdot)\|_\infty \;\le\; \|y(u_j,\cdot)\|_\infty + \eta\,\|u-u_j\|^\gamma \;\le\; \max_{1\le j\le N_\delta} \|y(u_j,\cdot)\|_\infty + \eta\,\delta^\gamma.$$

Taking expectations and applying Minkowski's inequality for the supremum together with Jensen's inequality, we obtain

$$\mathbb{E}\sup_{u\in K}\|y(u,\cdot)\|_\infty \;\le\; \mathbb{E}[\eta\,\delta^\gamma] + \sum_{j=1}^{N_\delta}\mathbb{E}\,\|y(u_j,\cdot)\|_\infty \;\le\; C_1\,\delta^\gamma + C_2\,\delta^{-d}\,\gamma_2^2(\mu_0^1,\mu_0^2).$$

We now optimize the bound with respect to $\delta > 0$. Choosing

$$\delta \;=\; \left(\gamma_2^2(\mu_0^1,\mu_0^2)\right)^{\frac{1}{\gamma+d}},$$

we obtain the final estimate:

$$\mathbb{E}\sup_{u\in K}\|x_1(u,\cdot) - x_2(u,\cdot)\|_\infty \;\le\; C_3\,\gamma_2^2(\mu_0^1,\mu_0^2)^{\frac{\gamma}{\gamma+d}},$$

where the constant C_3 depends only on the domain K, the moment parameter p, and the Lipschitz constants of the coefficients a and b. This completes the proof. □

We now derive an estimate for the difference between the solution $x^N(u,t)$ of the stochastic differential equation with interaction corresponding to the initial measure μ_0^N, and the solution $x(u,t)$ associated with the original initial measure μ_0. To this end, we invoke Lemmas 1.8.1 and 1.8.2 from the proof of the Kolmogorov–Totoki theorem [4], which provide a quantitative version of Kolmogorov's continuity criterion for random fields. Under suitable moment and regularity assumptions, these lemmas yield an upper bound for the supremum norm of the difference between two stochastic processes. We now formulate the resulting estimate in a form adapted to our analytical framework.

Corollary 4.1. *Let $x^N(u,t)$, $u \in K$, $t \in [0;1]$ denote the solution of equation (1) with the initial measure μ_0^N, and let $x(u,t)$ denote the solution with the initial measure μ_0. Then there exist constants $C > 0$ and $0 < \sigma < 1$ such that*

$$\mathbb{E}\sup_{u\in K}\sup_{t\in[0,1]}\|x^N(u,t) - x(u,t)\| \;\le\; C\,\gamma_2^{2\sigma}(\mu_0^N,\mu_0).$$

Proof. This corollary is a direct application of Theorem 4.1 to the pair of solutions $x^N(u,\cdot)$ and $x(u,\cdot)$ of equation (1), corresponding to the initial measures μ_0^N and μ_0, respectively. According to Theorem 4.1, under the Lipschitz assumptions on a and b, we obtain

$$\mathbb{E}\sup_{u\in K}\sup_{t\in[0,1]}\|x^N(u,t) - x(u,t)\| \;\le\; C\,\gamma_2^{2\sigma}(\mu_0^N,\mu_0),$$

where $\sigma = \frac{\gamma}{\gamma+d}$, with $p > d$ and $\gamma \in \left(0, \frac{p-d}{p}\right)$. The exponent σ appears as a consequence of the Kolmogorov–Totoki continuity theorem (see [4], Lemmas 1.8.1–1.8.2), which yields Hölder-type regularity of the solution in both spatial and temporal variables and guarantees the existence of a jointly continuous modification. Finally, since the Wasserstein distance $\gamma_2(\mu_0^N,\mu_0)$ quantifies the discrepancy between the approximated and original distributions, the estimate above provides a quantitative rate for the convergence of $x^N(u,t)$ to $x(u,t)$, uniformly over compact sets, with respect to both space and time. This completes the proof. $\square$

Remark 4.1. Combining Theorem 4.1 with the estimate $\gamma_2^2(\mu_0^N,\mu_0) \le \frac{\tilde{C}_d}{N^2}$, we immediately obtain

$$\mathbb{E} \sup_{u\in K} \sup_{t\in[0,1]} \|x^N(u,t) - x(u,t)\| \;\le\; \frac{C_d}{N^{2\sigma}}.$$

Here $\sigma = \frac{\gamma}{\gamma+d}$, with $\gamma < \frac{p-d}{p}$. In particular, if p is large, then σ can be chosen arbitrarily close to $\frac{1}{d}$.

Let $K \subset \mathbb{R}^d$ be compact and let $\{u_j^N\}_{j=1}^N$ be a regular grid in K. For each $u \in K$ denote by

$$\lfloor u \rfloor_N = \arg\min_{1\le j\le N} \|u - u_j^N\| \tag{13}$$

the closest grid point to u (representative of the cell, chosen to be its right corner). Define the grid projection of μ_0 by

$$\widetilde{\mu}_0^N \;=\; \mu_0 \circ \lfloor\cdot\rfloor_N^{-1}.$$

The continuous stochastic field $x(u,t)$ is approximated at $u \in K$ by

$$x_N\big(u_j^N,\, t,\, \widetilde{\mu}_s^N,\, \tfrac{1}{N}\big) \;\equiv\; x^N(u_j^N, t).$$

that is, by the value at the nearest grid node. Since the discretization scheme is defined only on the finite set $\{u_j^N\}$, each $u \in K$ is evaluated through its projection $\lfloor u \rfloor_N$.

Corollary 4.2. *Let $x(\cdot,t)$ be the solution of (1) with initial measure μ_0, and let $x^N(\cdot,t)$ be the solution with initial measure $\widetilde{\mu}_0^N = \mu_0 \circ \lfloor\cdot\rfloor_N^{-1}$. Let $x_n(\cdot,t)$ denote the Euler–Maruyama scheme with step $h = \frac{1}{n}$. Fix $p > d$ and any $\gamma \in \big(0, \frac{p-d}{p}\big)$, and set $\sigma = \frac{\gamma}{\gamma+d} \in (0,1)$.. Then there exists $C < \infty$ independent of n, N such that*

$$\mathbb{E} \sup_{u\in K} \sup_{t\in[0,1]} \big\|x(u,t) - x_n(\lfloor u\rfloor_N, t)\big\| \;\le\; C\Big(\gamma_2(\mu_0, \widetilde{\mu}_0^N)^{\sigma} + N^{-\gamma} + n^{-1/2}\Big).$$

If, in addition, $\widetilde{\mu}_0^N$ is the regular d-dimensional grid projection, so that $\gamma_2^2(\mu_0, \widetilde{\mu}_0^N) \le CN^{-2}$, then

$$\mathbb{E} \sup_{u\in K} \sup_{t\in[0,1]} \big\|x(u,t) - x_n(\lfloor u\rfloor_N, t)\big\| \;\le\; C\Big(N^{-2\sigma} + N^{-\gamma} + n^{-1/2}\Big).$$

Proof. We split the error into three terms:

$$\mathbb{E} \sup_{u\in K} \sup_{t\in[0,1]} \big\|x(u,t) - x_n(\lfloor u\rfloor_N, t)\big\| \le \mathbb{E}\sup_{u,t} \|x(u,t) - x^N(u,t)\| \tag{A}$$

$$+\,\mathbb{E} \sup_{u\in K} \sup_{t} \|x^N(u,t) - x^N(\lfloor u\rfloor_N, t)\| \tag{B}$$

$$+\,\mathbb{E} \sup_{u\in K} \sup_{t} \|x^N(\lfloor u\rfloor_N, t) - x_n(\lfloor u\rfloor_N, t)\|. \tag{C}$$

We will bound $\mathbb{E}\sup_{u\in K}\sup_{t\in[0,1]}$ of each term separately.

(A) By Theorem 4.1 with the pair of initial measures $(\mu_0, \widetilde{\mu}_0^N)$,

$$\mathbb{E} \sup_{u\in K} \sup_{t\in[0,1]} \|x(u,t) - x^N(u,t)\| \;\le\; C\left(\gamma_2^2(\mu_0, \widetilde{\mu}_0^N)\right)^{\sigma} \;=\; C\,\gamma_2(\mu_0, \widetilde{\mu}_0^N)^{2\sigma}.$$

(B) Spatial shift $u \mapsto \lfloor u \rfloor_N$. Lemma 2.2 is applicable to x^N as well, since x^N solves the same SDE with the modified initial measure $\widetilde{\mu}_0^N$. Therefore, for some $p > d$ and all $u, v \in K$ in $t \in [0,1]$, one has

$$\mathbb{E}\,\|x^N(u,t) - x^N(v,t)\|^p \;\le\; C\,\|u - v\|^p. \tag{14}$$

Setting $q = p-d > 0$, inequality (14) takes the form required by the Kolmogorov–Totoki continuity theorem (Kunita, [4, Theorems 1.8.1–1.8.2]):

$$\mathbb{E}\,\|Y(u)-Y(v)\|^p \;\le\; C\,\|u-v\|^{d+q}, \qquad u,v\in K.$$

Hence the theorem guarantees that the random field $u \mapsto x^N(u,\cdot)$ admits a modification which is Hölder continuous in u of every order

$$\gamma \;<\; \frac{q}{p} \;=\; \frac{p-d}{p}.$$

Consequently, there exists a modification and a nonnegative random variable ξ such that almost surely

$$\sup_{t\in[0,1]} \|x^N(u,t)-x^N(v,t)\| \;\le\; \xi\,\|u-v\|^{\gamma}, \qquad u,v\in K, \tag{15}$$

for every γ. Moreover, the Kolmogorov–Totoki theorem ensures that ξ has finite p-th moment:

$$\mathbb{E}\,[\xi^p] \;<\; \infty. \tag{16}$$

In particular, this rigorously justifies the statement that $\mathbb{E}[\xi^q] < \infty$ for some $q > 1$, since we may explicitly take $q = p > 1$. Finally, by (15) and the definition

$$\Delta_N \;=\; \sup_{u\in K}\|u-\lfloor u\rfloor_N\| \;\le\; C\,N^{-1},$$

we obtain

$$\mathbb{E}\,\sup_{u\in K}\,\sup_{t\in[0,1]} \|x^N(u,t)-x^N(\lfloor u\rfloor_N,t)\| \;\le\; C\,\Delta_N^{\gamma} \;\le\; C\,N^{-\gamma},$$

which completes the estimate of the spatial shift error.

(C) Euler–Maruyama time discretization at a fixed grid node. Fix the grid node

$$u^* \;=\; \lfloor u\rfloor_N$$

and consider the error

$$e(t) \;=\; x^N(u^*,t)-x_n(u^*,t), \qquad t\in[0,1].$$

Let $\kappa(t) = t_k$ be the left endpoint of the partition interval containing t, and define the measure discrepancy along the step

$$\Gamma(t) = \gamma_2(\mu_t^N,\mu_{\kappa(t)}^n), \qquad \mu_t^N = \widetilde{\mu}_0^N\circ(x^N(\cdot,t))^{-1}, \quad \mu_{t_k}^n = \mu_0\circ(x_n(\cdot,t_k))^{-1}.$$

Subtracting the x^N and x_n on $[0,t]$, using the Lipschitz property of a,b in state and in γ_2, and applying BDG to the stochastic integrals, we obtain the inequality (with a constant C depending only on K,p,d and the Lipschitz bounds, and independent of n,N and u^*):

$$\mathbb{E}\,\sup_{r\le t}\|e(r)\|^2 \;\le\; C\int_0^t \mathbb{E}\,\sup_{\tau\le s}\|e(\tau)\|^2\,ds \;+\; C\int_0^t \mathbb{E}\,\Gamma(s)^2\,ds \;+\; C\,h. \tag{17}$$

Here the last term Ch comes from the square of the increment of x^N over a time step: for each fixed u^*,

$$\mathbb{E}\,\sup_{s\in[t_k,t_{k+1}]} \|x^N(u^*,s)-x^N(u^*,t_k)\|^2 \;\le\; Ch. \tag{18}$$

To estimate the Γ–term, insert the intermediate measure

$$\widetilde{\mu}^n_{\kappa(s)} \;=\; \widetilde{\mu}_0^N\circ\big(x_n(\cdot,\kappa(s))\big)^{-1}.$$

and use the triangle inequality:

$$\Gamma(s) \;\le\; \underbrace{\gamma_2\big(\mu_s^N,\widetilde{\mu}^n_{\kappa(s)}\big)}_{T_1(s)} \;+\; \underbrace{\gamma_2\big(\widetilde{\mu}^n_{\kappa(s)},\mu^n_{\kappa(s)}\big)}_{T_2(s)}.$$

For $T_1(s)$ we apply the bound:$\widetilde{\mu}_0^N$:

$$T_1(s) \le \sup_{w\in K} \|x^N(w,s) - x_n(w,\kappa(s))\|$$

$$\le \underbrace{\sup_{w\in K} \|x^N(w,s) - x^N(w,\kappa(s))\|}_{\text{increments of } x^N}$$

$$+ \underbrace{\sup_{w\in K} \|e(w,\kappa(s))\|}_{\text{previous-step EM error}} .$$

Squaring and integrating over $s \in [0,1]$, the first part yields Ch by (18) (uniformly in w after applying Kolmogorov–Totoki in the spatial variable using Lemma 2 moments); the second gives $\int_0^1 \mathbb{E}\sup_{u\le s}\|e(u)\|^2 ds$. For $T_2(s)$ we use stability (Theorem 4.1) at time $\kappa(s)$ with the pair of initial measures $\widetilde{\mu}_0^N$ and μ_0, which gives

$$\sup_{s\in[0,1]} \mathbb{E}\, T_2(s)^2 \le C\,\gamma_2\big(\widetilde{\mu}_0^N,\mu_0\big)^{2\sigma}.$$

Combining the two parts,

$$\int_0^1 \mathbb{E}\,\Gamma(s)^2\,ds \le C\,h + C\,\gamma_2\big(\widetilde{\mu}_0^N,\mu_0\big)^{2\sigma} + C\int_0^1 \mathbb{E}\sup_{u\le s}\|e(u)\|^2\,ds. \tag{19}$$

Define the error:

$$F(t) = \mathbb{E}\sup_{w\in K}\sup_{r\le t}\|x^N(w,r) - x_n(w,r)\|^2.$$

By Lemma 2 and Kolmogorov–Totoki the field admits a jointly continuous modification, so $F(t) < \infty$. Using (17) and (19) yields

$$F(t) \le C\int_0^t F(s)\,ds + C\big(h + \gamma_2(\widetilde{\mu}_0^N,\mu_0)^{2\sigma}\big).$$

Gronwall's lemma gives

$$F(1) \le C\big(h + \gamma_2(\widetilde{\mu}_0^N,\mu_0)^{2\sigma}\big).$$

Taking square roots,

$$\mathbb{E}\sup_{w\in K}\sup_{t\in[0,1]}\|x^N(w,t) - x_n(w,t)\| \le C\Big(h^{1/2} + \gamma_2(\widetilde{\mu}_0^N,\mu_0)^{\sigma}\Big).$$

□

References

1. D. A. Dawson, *Measure-valued markov processes*, Lecture Notes in Mathematics, vol. 1541, Springer, 1993, Lecture Notes from the 1991 Saint-Flour Probability Summer School. https://doi.org/10.1007/bfb0084190
2. A. A. Dorogovtsev, *Stochastic flows with interaction and measure-valued processes*, International Journal of Mathematics and Mathematical Sciences **2003** (2003), no. 34, 2145–2162. https://doi.org/10.1155/S0161171203301073
3. Benjamin Gess, Rahul S. Gvalani, and Vadim Konarovskyi, *Conservative spdes as fluctuating mean field limits of stochastic gradient descent*, Probability Theory and Related Fields **192** (2024), 1447–1515. https://doi.org/10.1007/s00440-024-01353-6
4. Hiroshi Kunita, *Stochastic flows and jump-diffusions*, Probability Theory and Stochastic Modelling, vol. 92, Springer, 2019. https://doi.org/10.1007/978-981-13-3801-4
5. D. F. Kuznetsov, *Numerical modeling of stochastic differential equations and stochastic integrals*, Nauka, St. Petersburg, 1999 (Russian).
6. R. S. Liptser and A. N. Shiryaev, *Statistics of random processes: General theory*, Springer, 1977.

7. A.-S. Sznitman, *Topics in propagation of chaos*, École d'Été de Probabilités de Saint-Flour 1989, Lecture Notes in Mathematics, vol. 1464, Springer, 1991, Lecture Notes from the 1989 Saint-Flour Probability Summer School, pp. 165–251. https://doi.org/10.1007/bfb0085169
8. Cédric Villani, *Optimal transport: Old and new*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, 2009. https://doi.org/10.1007/978-3-540-71050-9

Institute of Mathematics of NAS of Ukraine, Tereschenkivska st. 3, 01024 Kyiv, Ukraine.
E-mail address: katekustareva@gmail.com