Sparse Monte Carlo method for nonlocal diffusion problems

Dmitry Kaliuzhnyi-Verbovetskyi and Georgi S. Medvedev *

December 24, 2019

Abstract

A class of evolution equations with nonlocal diffusion is considered in this work. These are integro-differential equations arising as models of propagation phenomena in continuum media with nonlocal interactions including neural tissue, porous media flow, peridynamics, models with fractional diffusion, as well as continuum limits of interacting dynamical systems. The principal challenge of numerical integration of nonlocal systems stems from the lack of spatial regularity of the data and solutions intrinsic to nonlocal models. To overcome this problem we propose a semidiscrete numerical scheme based on the combination of sparse Monte Carlo and discontinuous Galerkin methods. Our method requires minimal assumptions on the regularity of the data. In particular, the kernel of the nonlocal diffusivity is assumed to be a square integrable function and may be singular or discontinuous. An important feature of our method is sparsity. Sparse sampling of points in the Monte Carlo approximation of the nonlocal term allows to use fewer discretization points without compromising the accuracy. For kernels with singularities, more points are selected automatically in the regions near the singularities.

We prove convergence of the numerical method and estimate the rate of convergence. There are two principal ingredients in the error of the numerical method related to the use of Monte Calro and Galerkin approximations respectively. We analyze both errors. Two representative examples of discontinuous kernels are presented. The first example features a kernel with a singularity, while the kernel in the second example experiences jump discontinuity. We show how the information about the singularity in the former case and the geometry of the discontinuity set in the latter translate into the rate of convergence of the numerical procedure. In addition, we illustrate the rate of convergence estimate with a numerical example of an initial value problem, for which an explicit analytic solution is available. Numerical results are consistent with analytical estimates.

1 Introduction

We propose a numerical method for the initial value problem (IVP) for a nonlinear heat equation with nonlocal diffusion

$$
\begin{aligned}
\partial_t u(t,x) &= f(u,x,t) + \int W(x,y) D\left(u(t,y) - u(t,x)\right) dy, \quad x \in Q \subset \mathbb{R}^d, && (1.1)\\
u(0,x) &= g(x). && (1.2)
\end{aligned}
$$

*Department of Mathematics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104; medvedev@drexel.edu

For analytical convenience, we take $Q = [0,1]^d$ as a spatial domain. Throughout this paper, when the domain of integration is not specified, it is assumed to be Q. Further, $g \in L^2(Q)$, $W \in L^2(Q^2)$, f is a bounded measurable function on $\mathbb{R} \times Q \times \mathbb{R}^+$, which is Lipschitz continuous in u, continuous in t, and integrable in x, and D is a Lipschitz continuous function on $\mathbb{R}$:

$$|D(u_1) - D(u_2)| \le L_D|u_1 - u_2| \quad \text{and} \quad |f(u_1, x, t) - f(u_2, x, t)| \le L_f|u_1 - u_2| \tag{1.3}$$

for all $(x, t) \in Q \times \mathbb{R}$. Throughout this paper we assume

$$\sup_{u \in \mathbb{R}} |D(u)| \le 1. \tag{1.4}$$

This assumption may be dropped if an apriori estimate on $\|u\|_{C(0,T;L^\infty(Q))}$ for $T > 0$ is available. Furthermore, the analysis below applies to models with the interaction function of a more general form $D(u_1, u_2)$ provided

$$|D(u_1, v_1) - D(u_2, v_2)| \le L_D \left(|u_1 - u_2| + |v_1 - v_2|\right) \quad \forall u_1, u_2, v_1, v_2 \in \mathbb{R}. \tag{1.5}$$

However, we keep $D(u_1, u_2) := D(u_2 - u_1)$ to emphasize the connection to diffusion problems.

Equation (1.1) is a nonlocal diffusion problem. It arises as a continuum limit of interacting particle systems [24, 17]. Equations of this form are used for modeling population dynamics [31, 7, 30, 2, 6], neural tissue [8], porous media flows [10, 11], and various other biological and physicochemical processes involving anomalous diffusion [32, 4]. The key distinction of the evolution equations with nonlocal diffusion from their classical counterparts is the lack of smoothening property. A priori the solution of (1.1), (1.2) is a square integrable function in x for all $t > 0$ [24] and it may not possess much more regularity beyond that, unless the initial data and kernel W are smooth [21, Theorem 3.3]. The lack of smoothness is a serious challenge for constructing numerical schemes for (1.1), (1.2) and for analyzing their convergence. All deterministic quadrature formulas require at least piecewise differentiability for a guaranteed convergence rate. The problem is even more challenging in high dimensional spatial domains. The main idea underlying our approach is to use the Monte Carlo approximation of the nonlocal term in (1.1). We take advantage of the essential feature of the Monte Carlo method: the independence of the convergence rate on the regularity of the integrand. The second key idea is sparsity, whose use is twofold: First, sparse sampling of points in the Monte Carlo method is used to minimize computation without compromising the accuracy. For W with jump discontinuity across Lipschitz hypersurfaces, the use of sparsity is computationally beneficial starting from $d = 2$. If the discontinuity set has nontrivial fractal dimension, the sparse Monte Carlo method performs better than its dense counterpart already for $d = 1$ (cf. Lemma 4.4). Furthermore, sparsity is the key for extending the Monte Carlo method for models with singular kernels (see § 4.1). Not only does it allow to apply the Monte Carlo method for unbounded functions, it also makes it automatically adaptive: more sample points are selected near the singularities. The combination of these ideas together with the discontinuous Galerkin method yields a numerical scheme for the IVP (1.1), (1.2) that performs well under minimal assumptions on the regularity of W and initial data.

This paper is based on our previous work on convergence of interacting particle systems on convergent graph sequences [21, 22, 18, 24]. Continuum limit is a powerful tool for studying various aspects of network dynamics including existence, stability, and bifurcations of spatiotemporal patterns [33, 23, 25, 26]. Very often the derivation of the continuum limit is based on heuristic considerations and its rigorous mathematical justification is a nontrivial problem. Recently, motivated by the theory of graph limits [20, 19, 5], we proved convergence to the continuum limit for a broad class of dynamical systems on graphs [21]. Importantly, our proof applies to models on random graphs [22] including sparse random graphs [18, 24]. These results prepared the ground for the numerical method proposed in this paper. There is an intimate relation between the problem of the continuum limit for interacting particle systems and numerical integration of nonlocal diffusion models. Given a continuum model (1.1), one can construct the corresponding particle system, approximating (1.1). This idea had been already mentioned in [21], but has never been detailed. Further, recent results for the continuum limit of coupled systems on sparse graphs indicate a strong potential of sparse discretization for numerical integration of nonlocal problems. It is the goal of this paper to present these ideas in detail.

In the next section, we present a discretization of (1.1), which can be viewed as an interacting dynamical system on a *sparse random* graph. The structure of the graph is determined by the kernel W, which defines the asymptotic connectivity of the graph sequence parametrized by the size of the graph. In the theory of graph limits, such functions are called graphons [19], the term we adopt for the reminder of this paper. In Section 3, we prove convergence of the semidiscrete (discrete in space and continuous in time) approximation of (1.1) and turn to estimating the rate of convergence in Sections 4 and 5. There are two main factors contributing to the error of approximation. The first is due to approximating the nonlocal term in (1.1) by a random sum (Monte Carlo method), while the second is due to approximating the kernel and the initial data by piecewise constant functions (discontinuous Galerkin method). The rate of convergence of the sparse Monte Carlo approximation follows from our previous results [24, Theorem 4.1]. Convergence of piecewise constant approximation in the L^2–norm follows from classical theorems of analysis (cf. the Lebesgue-Besicovitch Theorem [15] or L^2–Martingale Convergence Theorem [34]). However, neither of these theorems elucidates the rate of convergence. In fact, the example in § 4.2 shows that without additional hypotheses the algebraic convergence may be arbitrarily slow. To this end, we study what determines the rate of convergence of piecewise constant approximations for a square integrable function. For Hölder continuous functions the answer is simple (cf. Lemma 4.1). For discontinuous functions, on the other hand, the answer naturally depends on the type of discontinuity. In Section 4, we consider two examples elucidating this issue. The first example is based on a singular (unbounded) graphon. It shows how the information about the singularity translates into the rate of convergence estimate. Here, we also see how to use sparsity to optimize computation. The second example adapted from [21], on the other hand, presents a bounded graphon with jump discontinuity (§ 4.2). In this case, the convergence rate depends on the geometry of the set of the discontinuity (cf. Lemma 4.4). In the light of these examples, in Section 5, we perform convergence analysis under general assumptions on W. In Section 6, we illustrate rate of convergence estimates with a numerical example. Here, we choose a nonlinear problem, which has an explicit solution. This allows us to verify the rate of convergence of the L^2-error as the discretization step tends to zero. Special attention is paid to the dependence of the convergence rate on sparsity. Finally, in Section 7 we present a proof of a technical Lemma 3.5, which extends the corresponding result in [24] to models in multidimensional domains and affords a wider range of sparsification.

Numerical methods for nonlocal diffusion problems have been subject of intense research recently due to their increased use in modeling [28, 14, 13, 4, 3, 29]. Compared to the existing literature, the contribution of the present work is that our method applies to problems with nonlinear diffusivity as well as to problems with more general form of the interaction function (cf. (1.5)). The main focus of this paper is how to deal with models with low regularity of the data. We believe that the combination of the Monte Carlo and discontinuous Galerkin methods provides an effective tool for numerical integration of nonlocal problems under minimal regularity assumptions.

2 The model and its discretization

In this section, we formulate the technical assumptions on the kernel W and describe the numerical scheme for solving the IVP (1.1)-(1.2).

We assume that $W \in L^2(Q^2)$ is subject to the following assumptions:

$$\max\left\{\operatorname{ess\,sup}_{x\in Q}\int |W(x,y)|dy,\ \operatorname{ess\,sup}_{y\in Q}\int |W(x,y)|dx\right\} \le W_1, \tag{W-1}$$

Theorem 2.1. *Let* $W \in L^2(Q^2)$ *satisfy* (W-1). *Then for any* $g \in L^2(Q)$ *and* $T > 0$ *there is a unique solution of the IVP* (1.1), (1.2) $u \in C^1(0,T;L^2(Q))$.

Proof. The proof is as in [18, Theorem 3.1] with minor adjustments. $\square$

Next, we note that the kernel in the nonlocal term may be assumed nonnegative. Indeed, by writing $W = W^+ - W^-$ as the difference of its positive and negative parts, one can rewrite (1.1) as

$$\partial_t u(t,x) = f(u,x,t) + \int W^+(x,y)D\left(u(t,y)-u(t,x)\right)dy - \int W^-(x,y)D\left(u(t,y)-u(t,x)\right)dy, \tag{2.1}$$

where the nonlocal terms splits into the difference of two terms with nonnegative kernels. Thus, without loss of generality, in the remainder of this paper we will assume

$$W \ge 0. \tag{2.2}$$

We approximate the IVP (1.1), (1.2) by the following system of ordinary differential equations

$$\dot{u}_{n,\bar{i}} = f_{n,\bar{i}}(u_{n,\bar{i}},t) + (\alpha_n n^d)^{-1}\sum_{\bar{j}\in[n]^d} a_{n,\bar{i}\bar{j}} D(u_{n,\bar{j}} - u_{n,\bar{i}}), \quad \bar{i}\in[n]^d, \tag{2.3}$$

$$u_{n,\bar{i}}(0) = g_{n,\bar{i}}, \tag{2.4}$$

where

$$Q_{n,\bar{i}} = \left[\frac{i_1-1}{n},\frac{i_1}{n}\right)\times\left[\frac{i_2-1}{n},\frac{i_2}{n}\right)\times\cdots\times\left[\frac{i_d-1}{n},\frac{i_d}{n}\right), \tag{2.5}$$

$$g_{n,\bar{i}} = n^d\int_{Q_{n,\bar{i}}} g(x)dx, \quad f_{n,\bar{i}}(u,t) = n^d\int_{Q_{n,\bar{i}}} f(u,x,t)dx, \quad \bar{i} := (i_1,i_2,\dots,i_d)\in[n]^d. \tag{2.6}$$

The semidiscrete system (2.3) can be viewed as a system of interacting particles on a random graph Γ_n with the node set $[n]^d$ and adjacency matrix $(a_{n,\bar{i}\bar{j}})$. The positive sequence

$$\alpha_n = n^{-d\gamma}, \quad 0\le\gamma<1, \tag{2.7}$$

is used to control the sparsity of Γ_n. The adjacency matrix $(a_{n,\bar{i}\bar{j}})$ is defined as follows. The case $W \in L^\infty(Q^2)$ is slightly different and so we treat it separately. Thus, there are two cases to consider.

(I) Suppose $W \in L^\infty(Q^2)$. Without loss of generality, we further assume that $0 \le W \le 1$. Then let

$$W_{n,\bar{i}\bar{j}} = n^{2d} \int_{Q_{n,\bar{i}} \times Q_{n,\bar{j}}} W(x,y)dxdy, \tag{2.8}$$

and

$$\mathbb{P}(a_{n,\bar{i}\bar{j}} = 1) = \alpha_n W_{n,\bar{i}\bar{j}}, \quad \bar{i}, \bar{j} \in [n]^d. \tag{2.9}$$

If $\alpha_n \equiv 1 (\gamma = 0)$ (cf. (2.7)), Γ_n is a dense W-random graph [20], otherwise Γ_n is sparse with the mean degree $O(n^{d(1-\gamma)}),\ 0 \le \gamma < 1$.

(II) Alternatively, if W is in $L^2(Q^2)$ but not in $L^\infty(Q^2)$ then let

$$\tilde{W}_n(x,y) := \alpha_n^{-1} \wedge W(x,y) \quad \text{and} \quad W_{n,\bar{i}\bar{j}} = n^{2d} \int_{Q_{n,\bar{i}} \times Q_{n,\bar{j}}} \tilde{W}(x,y)_n dxdy, \tag{2.10}$$

where α_n defined in (2.7) with $\gamma \in (0,1)$. Then

$$\mathbb{P}(a_{n,\bar{i}\bar{j}} = 1) = \alpha_n W_{n,\bar{i},\bar{j}}, \quad \bar{i}, \bar{j} \in [n]^d. \tag{2.11}$$

3 Convergence of the numerical method

In this section, we study convergence of the discrete scheme (2.3), (2.4). We first deal with the more general case of unbounded graphon W and then specialize the result for $W \in L^\infty(Q^2)$.

The following additional mild assumption on W is used to get a wider range of sparsity. Let nonnegative $W \in L^4(Q^2)$ satisfy

$$\max\left\{ \operatorname{ess\,sup}_{x \in Q} \int W^k(x,y)dy,\ \operatorname{ess\,sup}_{y \in Q} \int W^k(x,y)dx \right\} \le \bar{W}_k, \quad k \in [4]. \tag{W-1s}$$

Theorem 3.1. *Suppose nonnegative $W \in L^4(Q^2)$ is subject to* (W-1s)*, D, f, and g are as in* (1.1)*,* (1.2)*. Further, $\alpha_n = n^{-d\gamma}$ for some $\gamma \in (0,1)$. Then for arbitrary $0 < \delta < 1 - \gamma$, we have*

$$\begin{aligned} \sup_{t \in [0,T]} \|u(t,\cdot) - u_n(t,\cdot)\|_{L^2(Q)} \le C \Bigg(& \|g - g_n\|_{L^2(Q)} + \sup_{u \in \mathbb{R},\, t \in [0,T]} \|f(u,\cdot,t) - f_n(u,\cdot,t)\|_{L^2(Q)} \\ & + \|\tilde{W}_n - W\|_{L^2(Q^2)} + \|\tilde{W}_n - P_n \tilde{W}_n\|_{L^2(Q^2)} + n^{-d(1-\gamma-\delta)/2} \Bigg), \end{aligned} \tag{3.1}$$

where C is a positive constant independent of n, and $P_n \tilde{W}_n =: W_n$ stands for the L^2-projection of $\tilde{W}_n$ onto the finite–dimensional subspace $X_n = \operatorname{span}\{\mathbf{1}_{Q_{n,\bar{i}} \times Q_{n,\bar{j}}},\ (\bar{i},\bar{j}) \in [n]^{2d}\}$:

$$W_n(x,y) = \sum_{(\bar{i},\bar{j}) \in [n]^{d^2}} W_{n,\bar{i}\bar{j}} \mathbf{1}_{Q_{n,\bar{i}} \times Q_{n,\bar{j}}}(x,y).$$

and

$$f_n(u,x,t) = \sum_{\bar{i} \in [n]^d} f_{n,\bar{i}}(u,t) \mathbf{1}_{Q_{n,\bar{i}}}(x).$$

Estimate (3.1) *holds almost surely (a.s.) with respect to the random graph model.*

Remark 3.2. The theorem still holds without (W-1s), i.e., for square integrable W subject to (W-1). In this case, the last term on the right-hand side of (3.1) is replaced by $n^{-d(1/2-\gamma-\delta)}$, $\gamma \in (0, 1/2)$, and $\delta < 1/2-\gamma$.

The first two terms on the right–hand side of (3.1) correspond to the error of approximation of the initial data $g \in L^2(Q)$ and $f(u, x, t)$ by the step functions in x. Further, $\|\tilde{W}_n - W\|^2_{L^2(Q^2)}$ and $\|\tilde{W}_n - P_n\tilde{W}_n\|^2_{L^2(Q^2)}$ bound the error of approximation of W by a bounded step function W_n. Here, the first term $\|\tilde{W}_n - W\|^2_{L^2(Q^2)}$ is the error of truncating W and the second term $\|\tilde{W}_n - P_n\tilde{W}_n\|^2_{L^2(Q^2)}$ is the error of approximation of the truncated function $\tilde{W}_n$ by projecting it onto a finite–dimensional subspace. Finally, the last term on the right–hand side of (3.1) is the error of the approximation of the nonlocal term by the random sum in (2.3).

For bounded graphons W, Theorem 3.1 implies the following result.

Corollary 3.3. *Let $W \in L^\infty(Q^2)$. Then under the assumptions of Theorem 3.1 we have*

$$\begin{aligned}\sup_{t\in[0,T]} \|u(t,\cdot) - u_n(t,\cdot)\|_{L^2(Q)} \le C \Bigg(&\|g - g_n\|_{L^2(Q)} + \sup_{u\in\mathbb{R},\, t\in[0,T]} \|f(u,\cdot,t) - f_n(u,\cdot,t)\|_{L^2(Q)} \\ &+\|W_n - P_nW_n\|_{L^2(Q^2)} + n^{-d(1-\gamma-\delta)/2}\Bigg) \qquad a.s.,\end{aligned} \tag{3.2}$$

where C is a positive constant independent of n.

Remark 3.4. From (3.2) one can see how to use sparsity to optimize computation. Already for $d = 1$, if the largest of the two errors of approximation of g and W by step functions is $O(n^{-\kappa})$ with $1/2 < \kappa < 1$ (cf. §4.2) and the nonlinearity $f(\cdot)$ does not depend on x, then taking $\gamma = 1 - 2\kappa$ one can use sparse discretization without compromising the accuracy. This has obvious computational advantages over dense random and, moreover, deterministic spatial discretization schemes, e.g., Galerkin method. Sparse random discretization is even more efficient when $d > 1$.

The proof of Theorem 3.1 modulo a few minor details proceeds as the proof of convergence to the continuum limit in [22, 24]. First, the solution of the IVP (2.3), (2.4) is compared to that of the IVP for the averaged equation:

$$\dot{v}_{n,\bar{i}} = f_{n,\bar{i}}(v_{n,\bar{i}}, t) + n^{-d} \sum_{\bar{j}\in[n]^d} W_{n,\bar{i}\bar{j}} D(u_{n,\bar{j}} - u_{n,\bar{i}}), \quad \bar{i} \in [n]^d, \tag{3.3}$$

$$v_{n,\bar{i}}(0) = g_{n,\bar{i}}. \tag{3.4}$$

Then the solution of the averaged problem is compared to the solution of the IVP (1.1), (1.2). It is convenient to view the solution of the averaged problem as a function on $\mathbb{R}^+ \times Q$:

$$v_n(t, x) = \sum_{\bar{i}\in[n]^d} v_{n,\bar{i}}(t)\mathbf{1}_{Q_{n,\bar{i}}}(x). \tag{3.5}$$

Likewise, we interpret the solution of the discrete problem (2.3), (2.4) as a function on $\mathbb{R}^+ \times Q$:

$$u_n(t, x) = \sum_{\bar{i}\in[n]^d} u_{n,\bar{i}}(t)\mathbf{1}_{Q_{n,\bar{i}}}(x). \tag{3.6}$$

We recast the IVP (3.3), (3.4) as follows

$$\begin{aligned} \partial_t v_n(t,x) &= f_n(v_n(t,x),x,t) + \int W_n(x,y) D\left(v_n(t,y) - v_n(t,x)\right) dy, && (3.7)\\ v_n(0,x) &= g_n(x). && (3.8) \end{aligned}$$

The first step of the proof of convergence of the numerical scheme (2.3), (2.4) is accomplished in the following lemma.

Lemma 3.5. *Let nonnegative* $W \in L^4(Q^2)$ *subject to* (W-1s)*, and* $\alpha_n = n^{-d\gamma}$*,* $\gamma \in (0,1)$ *(cf.* (4.24)*). Then for any* $T > 0$ *for solutions of* (2.3) *and* (3.3) *subject to the same initial conditions, we have*

$$\sup_{t\in[0,T]} \|u_n(t,\cdot) - v_n(t,\cdot)\|_{L^2(Q)} \le C n^{-d(1-\gamma-\delta)/2} \quad a.s., \tag{3.9}$$

where $0 < \delta < 1-\gamma$*, and positive constant* C *independent of* n*.*

The proof of the lemma is technical and is relegated to Section 7. The result still holds without for square integrable functions the additional assumption (W-1s) albeit for a narrower range of $\gamma \in (0, 0.5)$ (cf. [24, Theorem 4.1]).

Proof of Theorem 3.1. Denote the difference between the solutions of the original IVP (1.1), (1.2) and the averaged IVP (3.7), (3.8)

$$w_n(t,x) = u(t,x) - v_n(t,x). \tag{3.10}$$

By subtracting (3.3) from (1.1), multiplying the resultant equation by w_n, and integrating over Q, we obtain

$$\begin{aligned} \int \partial_t w_n(t,x) w_n(t,x) dx &= \int \left(f(u(t,x),x,t) - f(v_n(t,x),x,t)\right) w_n(t,x) dx \\ &+ \int \left(f(v_n(t,x),x,t) - f_n(v_n(t,x),x,t)\right) w_n(t,x) dx \\ &+ \int\int W(x,y) \left[D(u(t,y)-u(t,x)) - D(v_n(t,y)-v_n(t,x))\right] w_n(t,x) dy dx \\ &+ \int\int \left(W(x,y) - W_n(x,y)\right) D(v_n(t,y)-v_n(t,x)) w_n(t,x) dy dx. \end{aligned} \tag{3.11}$$

Using Lipschitz continuity of $f(u,x,t)$ in u and an elementary case of the Young's inequality, we obtain

$$\left| \int \left(f(u(t,x),x,t) - f(v_n(t,x),x,t)\right) w_n(t,x) dx \right| \le L_f \int w_n(t,x)^2 dx, \tag{3.12}$$

$$\begin{aligned} \left| \int \left(f(v_n(t,x),x,t) - f_n(v_n(t,x),x,t)\right) w_n(t,x) dx \right| &\le \frac{1}{2} \int \left(f(v_n(t,x),x,t) - f_n(v_n(t,x),x,t)\right)^2 dx \\ &+ \frac{1}{2}\|w_n(t,\cdot)\|^2, \end{aligned} \tag{3.13}$$

where $\|\cdot\|$ stands for the $L^2(Q)$-norm. Recall that D is bounded by 1 (cf. (1.4)). Using this bound and the Young's inequality, we obtain

$$\left|\int\int (W(x,y) - W_n(x,y))\, D(v_n(t,y) - v_n(t,x)) w_n(t,x) dydx\right| \le \frac{1}{2}\|W - W_n\|_{L^2(Q^2)} + \frac{1}{2}\|w_n\|^2. \tag{3.14}$$

Finally, using Lipschitz continuity of D and Young's inequality, we estimate

$$\begin{aligned}
&\left|\int\int W(x,y)\left[D(u(t,y) - u(t,x)) - D(v_n(t,y) - v_n(t,x))\right] w_n(t,x) dydx\right| \\
&\le L_D \int\int W(x,y) \left(|w_n(t,y)| + |w_n(t,x)|\right)|w_n(t,x)| dydx \\
&\le L_D \int\int W(x,y) \left(\frac{1}{2}|w_n(t,y)|^2 + \frac{3}{2}|w_n(t,x)|^2\right) dydx \\
&\le \frac{3L_D}{2}\int\int W(x,y)|w_n(t,x)|^2 dydx + \frac{L_D}{2}\int\int W(x,y)|w_n(t,y)|^2 dydx \\
&\le 2W_1 L_D \|w_n\|^2,
\end{aligned} \tag{3.15}$$

where we used Fubini theorem and (W-1) in the last line.

By combining (3.11)-(3.15), we arrive at

$$\frac{d}{dt}\|w_n(t,\cdot)\|^2 \le L\|w_n(t,\cdot)\|^2 + \sup_{u\in\mathbb{R},\, t\in[0,T]} \|f(u,\cdot,t) - f_n(u,\cdot,t)\|^2 + \|W_n - W\|^2, \tag{3.16}$$

where $L = 1 + 2L_f + L_D(3W_1 + W_2)$.

By Gronwall's inequality, we have

$$\begin{aligned}
\sup_{t\in[0,T]} \|w_n(t,\cdot)\| &\le e^{LT/2}\sqrt{\|w_n(0,\cdot)\|^2 + \sup_{u\in\mathbb{R},\, t\in[0,T]}\|f(u,\cdot t) - f_n(u,\cdot,t)\|^2 + \|W_n - W\|^2_{L^2(Q^2)}} \\
&\le e^{LT/2}\left(\|g - g_n\|_{L^2(Q)} + \sup_{u\in\mathbb{R},\, t\in[0,T]}\|f(u,\cdot,t) - f_n(u,\cdot,t)\| + \|W_n - W\|_{L^2(Q^2)}\right).
\end{aligned}$$

□

4 Two examples

The error of approximation of the nonlocal term by a random sum, the last term on the right–hand side of (3.1), is known explicitly. Next in importance is the error of approximation of the square integrable graphon W by the step function W_n. This error depends on the regularity of the graphon. In this section, we consider two representative examples of W: a singular graphon (§ 4.1) and a bounded graphon with jump discontinuities (§ 4.2). Motivated by these examples in the next section, we will analyze the rate of convergence estimates under general assumptions on graphon W.

We will begin with the following estimate for Hölder continuous functions. To this end, $\phi \in L^p(Q),\ p \geq 1$, and

$$\phi_n(x) = \sum_{\bar{i}\in[n]^d} \phi_{n,\bar{i}} \mathbf{1}_{Q_{n,\bar{i}}}(x), \quad \phi_{n,\bar{i}} = n^d \int_{Q_{n,\bar{i}}} \phi(x)dx.$$

Lemma 4.1. *Suppose $\phi \in L^p(Q),\ p \geq 1$, is a Hölder continuous function*

$$|\phi(x) - \phi(y)| \leq C|x-y|^\beta, \quad x, y \in Q,\ \beta \in (0,1]. \tag{4.17}$$

Then

$$\|\phi - \phi_n\|_{L^p(Q)} \leq Ch^\beta. \tag{4.18}$$

Here and below, $h := n^{-1}$.

Proof. Using Jensen's inequality and (4.18), we have

$$\begin{aligned} \|\phi - \phi_n\|^p_{L^p(Q)} &= \int_Q \left| \sum_{\bar{i}\in[n]^d} \left(\phi(x) - n^d \int_{Q_{n,\bar{i}}} \phi(y)dy \right) \mathbf{1}_{Q_{n,\bar{i}}}(x) \right|^p dx \\ &= \sum_{\bar{i}\in[n]^d} \int_{Q_{n,\bar{i}}} \left| n^d \int_{Q_{n,\bar{i}}} (\phi(x) - \phi(y))\, dy \right|^p dx \\ &\leq n^d \sum_{\bar{i}\in[n]^d} \int_{Q_{n,\bar{i}}} \int_{Q_{n,\bar{i}}} |\phi(x)-\phi(y)|^p\, dxdy \\ &\leq C^p h^{p\beta}. \end{aligned} \tag{4.19}$$

□

Remark 4.2. Below, we will freely apply Lemma 4.1 to functions on Q and on Q^2. The latter are clearly covered by the lemma by taking $Q := Q^2$.

4.1 A singular graphon

Consider the problem of approximation by step functions of the singular kernel graphon

$$W(x,y) = \frac{1}{|x-y|^\lambda}, \quad x, y \in Q = [0,1]^d, \tag{4.20}$$

where $0 < \lambda < d/2$.

Lemma 4.3. *For $\gamma \in (0, 1/2)$ and $0 < \lambda < d/2$ we have*

$$\|W - W_n\|_{L^2(Q^2)} \leq \max\left\{ O\left(h^{d\gamma(\frac{d}{2\lambda}-1)}\right), O\left(h^{1-d\gamma\left(1+\frac{1}{\lambda}\right)}\right)\right\}. \tag{4.21}$$

Proof. 1. Below, we will use the following change of variables $(x, y) = T(u, v)$ for (x, y) and (u, v) from $\mathbb{R}^{2d}$, defined by

$$u_i = x_i - y_i \quad \text{and} \quad v_i = x_i + y_i,\ i \in [d]. \tag{4.22}$$

2. Let $\alpha_n = n^{-d\gamma}$ and recall that $\tilde{W}_n = h^{-d\gamma} \wedge W$. Denote $\tilde{Q} = \left\{(x, y) \in Q^2 :\ |x-y|^{-\lambda} \geq h^{-d\gamma}\right\}$. Further,

$$\begin{aligned} \|W - \tilde{W}_n\|^2_{L^2(Q^2)} &= \int_{\tilde{Q}^2} \left(\frac{1}{|x-y|^\lambda} - n^{d\gamma}\right)^2 dxdy \\ &\leq C_1 \int_{\{|u| \leq h^{\frac{d\gamma}{\lambda}}\}} \left(\frac{1}{|u|^\lambda} - n^{d\gamma}\right)^2 du \\ &\leq C_2 \int_0^{h^{\frac{d\gamma}{\lambda}}} \left(\frac{1}{r^\lambda} - n^{d\gamma}\right)^2 r^{d-1} dr \\ &= O\left((d - 2\lambda)^{-1} h^{2d\gamma\left(\frac{d}{2\lambda} - 1\right)}\right), \end{aligned} \tag{4.23}$$

where we used (4.22) followed by the change to polar coordinates.

Thus,

$$\|W - \tilde{W}_n\|_{L^2(Q^2)} = O\left(h^{d\gamma\left(\frac{d}{2\lambda} - 1\right)}\right), 0 < \lambda < d/2,\ \gamma > 0. \tag{4.24}$$

3. Next we turn to estimating $\left\|\tilde{W}_n - W_n\right\|_{L^2(Q^2)}$. Since the truncated function $\tilde{W}_n$ is Lipschitz continuous on Q^2, by Lemma 4.1,

$$\left\|\tilde{W}_n - W_n\right\|_{L^2(Q^2)} \leq L(\tilde{W}_n) h.$$

It remains to estimate the Lipschitz constant $L(\tilde{W}_n) \leq \operatorname{ess\,sup}_{Q^2} |\nabla \tilde{W}_n|$. On $Q^2 - \tilde{Q}$,

$$|\nabla \tilde{W}_n| = \lambda |x-y|^{-1-\lambda} |\nabla_{x,y} |x-y|| = \sqrt{2}\lambda |x-y|^{-1-\lambda}. \tag{4.25}$$

The gradient approaches its greatest value as $|x - y| \searrow h^{\frac{d\gamma}{\lambda}}$. Thus,

$$\operatorname{ess\,sup}_{Q^2} |\nabla \tilde{W}_n| = \sqrt{2}\lambda h^{-d\gamma\left(\frac{1}{\lambda} + 1\right)}$$

and

$$\left\|\tilde{W}_n - W_n\right\|_{L^2(Q^2)} = O\left(h^{1 - d\gamma\left(1 + \frac{1}{\lambda}\right)}\right). \tag{4.26}$$

4. The statement of the lemma follows (4.24) and (4.26) and the triangle inequality.

□

Next we choose γ to optimize the rate of convergence in (4.21). By setting the two exponents of h on the right–hand side of (4.21) equal, we see that the rate is optimal for

$$\gamma = \frac{2\lambda}{d(d+2)}.$$

With this choice of γ,
$$\|W - W_n\|_{L^2(Q^2)} = O(h^{\frac{2\lambda}{d+2}\left(\frac{d}{2\lambda}-1\right)}).$$

To optimize the rate of convergence of the numerical scheme (2.3), (2.4), one has to choose $\gamma \in (0,1)$ to maximize the smallest of the following three exponents

$$d\gamma\left(\frac{d}{2\lambda}-1\right), \quad 1-d\gamma\left(1+\frac{1}{\lambda}\right), \quad \frac{d}{2}\left(1-\gamma\right),$$

where the last exponent comes from the error of the Monte Carlo approximation (cf. (3.1)).

4.2 $\{0,1\}$-valued functions

The following example is adapted from [21]. It shows how jump discontinuities affect the rate of convergence of approximation by piecewise constant functions. The accuracy of approximation depends on the geometry of the hypersurface of discontinuity, more precisely, on its fractal dimension.

Let Q^+ be a closed subset of Q and consider

$$f(x) = \begin{cases} 1, & x \in Q^+ \\ 0, & \text{otherwise.} \end{cases} \tag{4.27}$$

Denote by ∂Q^+ the boundary of Q and recall the upper box-counting dimension of ∂Q^+

$$\beta := \overline{\lim}_{h\to 0} \frac{\log N_h(\partial Q^+)}{-\log h}, \tag{4.28}$$

where $N_h(\partial Q^+)$ stands for the number of $Q_{n,\bar{i}},\ \bar{i},\bar{j} \in [n]^d$, having nonempty intersection with ∂Q^+ (cf. [16]).

Lemma 4.4.
$$\|\phi - \phi_n\|_{L^p(Q)} \le C h^{\frac{d-\beta}{p}}, \tag{4.29}$$
for some positive C independent on n.

Proof. As in (4.19), we have

$$\|\phi - \phi_n\|^p_{L^p(Q^2)} \le h^{-d} \sum_{\bar{i}\in[n]^d} \int_{Q_{n,\bar{i}}} \int_{Q_{n,\bar{i}}} |f(x) - f(z)|^p \, dz dx. \tag{4.30}$$

Note that the only nonzero terms in the sum on the right–hand side of (4.30) are the integrals over $Q_{n,\bar{i}} \times Q_{n,\bar{j}}$'s having nonempty intersection with ∂Q^+. Thus,

$$\|\phi - \phi_n\|^p_{L^p(Q)} = h^d N_h(\partial Q^+) \le C h^{d-\beta}, \tag{4.31}$$

where we used (4.28). □

Remark 4.5. Note that as $\beta \to d - 0$ the rate of convergence in (4.29) can be made arbitrarily low.

5 The rate of convergence of the numerical method

5.1 Approximation by step functions

In this section, we address the rate of convergence of the Galerkin component of the numerical scheme (2.3), (2.4). Specifically, we study the error of the approximation of the graphon $W \in L^2(Q^2)$ and the initial data $g \in L^2(Q)$ by step functions.

We will need an L^p-modulus of continuity of function on a unit d-cube $Q = [0,1]^d$. In fact, we only need the L^2-modulus of continuity, but present the analysis in the more general L^p-setting, since this does not require any extra effort. For functions on the real line, the definition of the L^p-modulus of continuity can be found in [1, 12]. Here, we present a suitable adaptation of this definition for the problem at hand.

Definition 5.1. *For $\phi \in L^p(Q)$ we define the L^p-modulus of continuity*

$$\omega_p(\phi, \delta) = \sup_{|\xi|_\infty \le \delta} \|\phi(\cdot + \xi) - \phi(\cdot)\|_{L^p(Q_\xi)},\ \delta > 0, \tag{5.1}$$

where $|\xi|_\infty := \max_{i\in[d]} |\xi_i|$ and $Q_\xi = \{x \in Q :\ x + \xi \in Q\}$.

For $\alpha \in (0,1]$, we define a generalized Lipschitz space[1]

$$\mathrm{Lip}\,(\alpha, L^p(Q)) = \{\phi \in L^p(Q) :\ \exists C > 0 :\ \omega_p(\phi, \delta) \le C\delta^\alpha\}\,. \tag{5.2}$$

Clearly $\mathrm{Lip}\,(\alpha, L^p(Q))$ contains α-Hölder continuous functions. However, Lipschitz spaces are much larger than Hölder spaces. For instance, $\mathrm{Lip}\,(1/p, L^p(Q))$ contains discontinuous functions.

Below, we express the error of approximation of $\phi \in L^p(Q)$ by a step function through $\omega_p(\phi, h)$. The analysis works out a little cleaner for dyadic discretization of Q, which will be assumed for the remainder of this section. Thus, we approximate $\phi \in L^p(Q)$ by a piecewise constant function

$$\phi_{2^m}(x) = \sum_{\bar{i}\in[2^m]^d} \phi_{Q_{2^m,\bar{i}}} \mathbf{1}_{Q_{2^m,\bar{i}}}(x), \tag{5.3}$$

where $\phi_{Q_{2^m,\bar{i}}}$ stands for the mean value of ϕ on $Q_{2^m,\bar{i}}$

$$\phi_{Q_{2^m,\bar{i}}}(x) = 2^{md} \int_{Q_{2^m,\bar{i}}} \phi(x)dx,\ i \in [2^m].$$

Lemma 5.2. *For $\phi \in \mathrm{Lip}\,(\alpha, L^p(Q))$, we have*

$$\|\phi - \phi_{2^m}\|_{L^p(Q)} \le C2^{-\alpha m}, \tag{5.4}$$

where C is independent of m.

[1] Below, we will freely apply the definitions and various estimates established for functions on Q to functions on Q^2, for which they are trivially valid by setting $d := 2d$. In particular, the definitions of the modulus of continuity and the corresponding Lipschitz spaces obviously translate to functions on Q^2.

Proof. Fix $m \in \mathbb{N}$ and denote $h := 2^{-m}$. To simplify notation, throughout the proof we drop 2^m in the subscript of $x_{2^m,i}$ and $Q_{2^m,\bar{i}}$.

We write

$$\phi_{2^{m+1}}(x) = \sum_{\bar{i}\in[2^m]^d} \sum_{\bar{j}\in\{0,1\}^d} \left(\frac{2}{h}\right)^d \int_{Q'_{\bar{i}}} \phi\left(s + \bar{j}\frac{h}{2}\right) ds \mathbf{1}_{Q_{\bar{i}}^{\bar{j}}}(x), \tag{5.5}$$

where

$$s + \bar{j}\frac{h}{2} = \left(s_1 + j_1\frac{h}{2}, s_2 + j_2\frac{h}{2}, \dots, s_d + j_d\frac{h}{2}\right),$$

$$Q_{\bar{i}}^{\bar{j}} = \left[x_{i_1-1} + j_1\frac{h}{2}, x_{i_1-1} + (j_1+1)\frac{h}{2}\right) \times \cdots \times \left[x_{i_d-1} + j_d\frac{h}{2}, x_{i_d-1} + (j_d+1)\frac{h}{2}\right),$$

and

$$Q'_{\bar{i}} = \left[x_{i_1-1}, x_{i_1-1} + \frac{h}{2}\right) \times \cdots \times \left[x_{i_d-1}, x_{i_d-1} + \frac{h}{2}\right).$$

Rewrite (5.3)

$$\phi_{2^m}(x) = \sum_{\bar{i}\in[2^m]^d} \sum_{\bar{j}\in\{0,1\}^d} \left(\frac{1}{h}\right)^d \left\{ \sum_{\bar{k}\in\{0,1\}^d} \int_{Q'_{\bar{i}}} \phi\left(s + \bar{k}\frac{h}{2}\right) ds \right\} \mathbf{1}_{Q_{\bar{i}}^{\bar{j}}}(x). \tag{5.6}$$

By subtracting (5.5) from (5.6) we have

$$\phi_{2^m} - \phi_{2^{m+1}} = \sum_{\bar{i}\in[2^m]^d} \sum_{\bar{j}\in\{0,1\}^d} \frac{1}{h^d} \left\{ \sum_{\substack{\bar{k}\in\{0,1\}^d \\ \bar{k}\neq\bar{j}}} \int_{Q'_{\bar{i}}} \left[\phi\left(s + \bar{k}\frac{h}{2}\right) - \phi\left(s + \bar{j}\frac{h}{2}\right)\right] ds \right\} \mathbf{1}_{Q_{\bar{i}}^{\bar{j}}}.$$

Further,

$$|\phi_{2^m} - \phi_{2^{m+1}}|^p \leq \sum_{\bar{i}\in[2^m]^d} \sum_{\bar{j}\in\{0,1\}^d} \sum_{\substack{\bar{k}\in\{0,1\}^d \\ \bar{k}\neq\bar{j}}} \left| \int_{Q'_{\bar{i}}} \left[h^{-d}\phi\left(s + \bar{k}\frac{h}{2}\right) - \phi\left(s + \bar{j}\frac{h}{2}\right)\right] ds \right|^p \mathbf{1}_{Q_{\bar{i}}^{\bar{j}}}. \tag{5.7}$$

Integrating both sides of (5.7) over Q and using Jensen's inequality, we continue

$$\begin{aligned} \|\phi_{2^m} - \phi_{2^{m+1}}\|_{L^p(Q)}^p &\leq \sum_{\bar{j}\in\{0,1\}^d} \sum_{\substack{\bar{k}\in\{0,1\}^d \\ \bar{k}\neq\bar{j}}} \sum_{\bar{i}\in[2^m]^d} \int_{Q'_{\bar{i}}} \left| \phi\left(s + k\frac{h}{2}\right) - \phi\left(s + \bar{j}\frac{h}{2}\right) \right|^p ds \\ &\leq 2^d(2^d - 1)\omega_p^p(\phi, h). \end{aligned} \tag{5.8}$$

Thus,

$$\|\phi_{2^m} - \phi_{2^{m+1}}\|_{L^p(Q)} \leq \left[2^d(2^d-1)\right]^{1/p} \omega_p(\phi, 2^{-(m+1)}) =: C_{d,p}\omega_p(\phi, 2^{-(m+1)}).$$

Since $\phi \in \mathrm{Lip}\,(\alpha, L^p(Q))$, we have

$$\|\phi_{2^m} - \phi_{2^{m+1}}\|_{L^p(Q)} \leq C2^{-\alpha m}, \tag{5.9}$$

where C depends on $\phi, d,$ and p but not m.

Let $m \in \mathbb{N}$ be arbitrary but fixed. For any integer $M > m$ we have

$$\begin{aligned}
\|\phi_{2^M} - \phi_{2^m}\|_{L^p(Q)} &= \left\| \sum_{k=m}^{M-1} (\phi_{2^{k+1}} - \phi_{2^k}) \right\|_{L^p(Q)} \leq \sum_{k=m}^{\infty} \|\phi_{2^{k+1}} - \phi_{2^k}\|_{L^p(Q)} \\
&= \left\| \sum_{k=m}^{\infty} (\phi_{2^{k+1}} - \phi_{2^k}) \right\|_{L^p(Q)} \leq \sum_{k=m}^{\infty} \|\phi_{2^{k+1}} - \phi_{2^k}\|_{L^p(Q)} \\
&\leq \sum_{k=m}^{\infty} \omega_p(\phi, 2^{-(k+1)}) \leq 2^{-p+1} \sum_{k=m}^{\infty} C2^{\alpha(k+1)} \leq C2^{-\alpha m}.
\end{aligned} \tag{5.10}$$

By passing M to infinity in (5.10), we get (5.4). □

5.2 The rate of convergence

We now can combine Theorem 3.1 and Lemma 5.4 to estimate the convergence rate for (2.3), (2.4). For the model with a bounded graphon W (cf. **(I)**, Section 2) we have the following theorem.

Theorem 5.3. *Suppose that in addition to the assumptions of Theorem 3.1, for some $\alpha_i \in (0, 1],\ i \in [3]$, $g \in \mathrm{Lip}\left(\alpha_1, L^2(Q)\right)$, $W \in \mathrm{Lip}\left(\alpha_2, L^2(Q^2)\right) \cap L^\infty(Q^2)$, and $f(u, \cdot, t) \in \mathrm{Lip}\left(\alpha_3, L^2(Q)\right)$ uniformly for $(u, t) \in \mathbb{R} \times [0, T]$, i.e.,*

$$\omega_2(f(u, \cdot, t), \delta) \leq C\delta^{\alpha_3}$$

where $C > 0$ is independent of (u, t).

Then

$$\sup_{t\in[0,T]} \|u(t, \cdot) - u_n(t, \cdot)\| \leq Cn^{-\alpha}, \quad \alpha = \min\{\alpha_1, \alpha_2, \alpha_3, \frac{1}{2} - \gamma\}, \tag{5.11}$$

where C is independent of n.

If $W \in L^2(Q^2)$ has singularities then the convergence rate may also depend on the accuracy of approximation of W by the truncated function $\tilde{W}_n$. We do not estimate the truncation error for a general $W \in L^2(Q^2)$. For an example of how this error can be estimated for a given graphon in practice, we refer to the example in § 4.1.

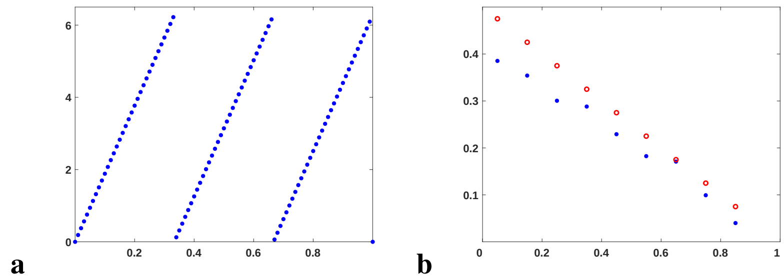

Figure 1: a) The 3-twisted state used to initialize the Kuramoto model (6.12). b) The numerically estimated exponent characterizing convergence of the numerical scheme (2.3), (2.4), α_γ, is plotted as a function of γ (see Section 6). The numerical estimates and the theoretical predictions are plotted using of the blue stars and red circles respectively.

6 Numerical example

In this section, we illustrate convergence analysis in the previous sections with a numerical example. To this end, we consider an IVP for the continuum Kuramoto model with nonlocal nearest–neighbor coupling [27]:

$$\begin{aligned} \partial_t u(t,x) &= \omega + \int_{[0,1]} K(y-x)\sin\left(u(t,x)-u(t,y)\right), &(6.12)\\ u(0,x) &= u^{(q)}(x), &(6.13)\end{aligned}$$

where $u(t,x)\in\mathbb{T},\ \mathbb{T}=\mathbb{R}/2\pi\mathbb{Z}$, stands for the phase of the oscillator at $x\in[0,1]$, $\omega\in\mathbb{R}$ is its intrinsic frequency. Function K, describing the connectivity of the network, is first defined on $[0,1/2)$ by

$$K(x)=\mathbf{1}_{\{y:\,|y|\le r\}}(x),\quad r\in(-1/2,1/2), \tag{6.14}$$

and then extended as a 1–periodic function on $\mathbb{R}$. The initial condition

$$u^{(q)}(x)=2\pi\left(qx \mod 1\right),\quad q\in\mathbb{Z}, \tag{6.15}$$

is called a q–twisted state (Figure 1a). For $\omega=0$, $u^{(q)}$ is a stationary solution of (6.12). Thus,

$$u(t,x)=(2\pi qx+\omega t) \mod 2\pi \tag{6.16}$$

solves the IVP (6.12), (6.13). We use the explicit solution (6.16) to compute the error of the numerical integration of (6.12), (6.13).

To estimate the rate of convergence of the numerical scheme (2.3), (2.4) we use the following values of parameters: $r=0.2$, $\omega=0.5$, and $q=3$. For these parameter values, the travelling wave solution (6.16) is unstable. We integrated (6.12) numerically for $t\in[0,1]$, using the fourth order Runge–Kutta method with the time step 10^{-2}. Note that the error of the Runge–Kutta method, i.e., of the discretization in time is significantly smaller than of that of the discretizing in space (c.f. (2.3), (2.4)). We integrated (6.12) numerically for different values of γ and for $n\in\{128,256\}$. For each pair (γ,n) we repeated the numerical

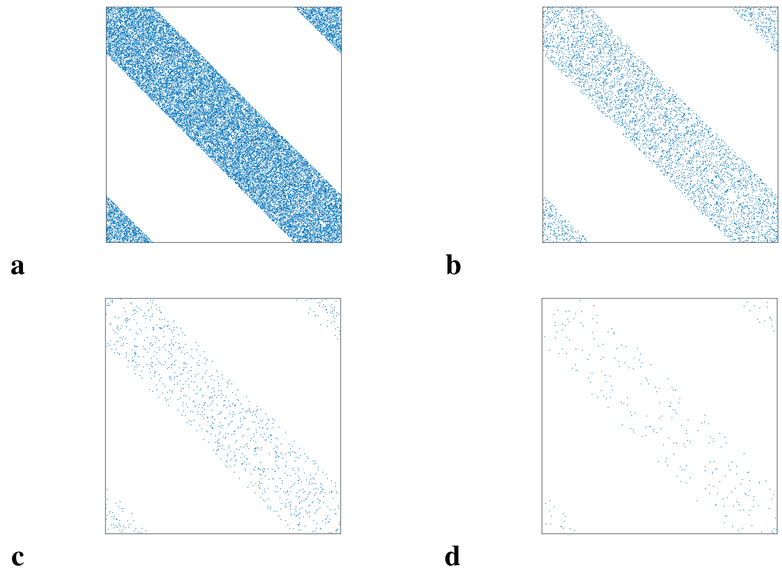

Figure 2: Pixel pictures of the adjacency matrices of sparse graphs generated with the following values of γ: a) 0.25, b) 0.5, c) 0.75, d) 0.95.

experiment 200 times and computed the mean value of the error of numerical integration (compared to the exact solution (6.16)). The mean errors $\bar{e}_{\gamma,128}$ and $\bar{e}_{\gamma,256}$ computed for $n = 128$ and $n = 256$ respectively are used to determine the convergence rate:

$$\alpha_\gamma = \frac{\ln\left(\bar{e}_{\gamma,128}/\bar{e}_{\gamma,256}\right)}{\ln 2}. \tag{6.17}$$

The results of this numerical experiment are shown in Figure 1b. Our main goal was to verity the dependence of the convergence rate on sparsity controlled by γ. The pixel pictures for the adjacency matrices of random graphs corresponding to the nonlocal nearest–neighbor coupling for $n = 512$ and different values of γ are shown in Figure 2. The plot in Figure 1b shows a clear linear relation between the exponent α_γ and γ. The numerical rates plotted by blue stars are slightly lower the theoretical rates $(1-\gamma)/2$ plotted in red. Overall numerical rates show a good fit with the analytical estimate.

7 Proof of Lemma 3.5

In this section, we prove Lemma 3.5. The proof follows the lines of the proof of Theorem 4.1 in [24], which covers $\gamma \in (0, 1/2)$ for $d = 1$. Extension to the multidimensional case $d > 1$ is straightforward. Lemmas 7.3 and 7.4 adapted from [9] allow to extend the range of γ to $(0, 1)$. The reader not interested in the extended range of γ may find a simpler proof in [24] easier to follow. For those interested in the full range of γ, below we present the following proof of Lemma 3.5.

Theorem 7.1. *Let nonnegative $W \in L^4(Q^2)$ satisfy*

$$\max\left\{ \operatorname{ess\,sup}_{x\in Q} \int W^k(x,y)dy,\ \operatorname{ess\,sup}_{y\in Q} \int W^k(x,y)dx \right\} \le \bar{W}_k, \quad k \in [4] \tag{W-1s}$$

and

$$\liminf_{n\to\infty} \frac{\alpha_n n^d}{\ln n} > 0. \tag{7.1}$$

Then for solutions of (2.3) *and* (3.3) *subject to the same initial conditions and arbitrary* $0 < \epsilon < 1/2$, *we have*

$$\sup_{t\in[0,T]} \|u_n(t,\cdot) - v_n(t,\cdot)\|_{L^2(Q)} \le C(\alpha_n n^d)^{1/2-\epsilon} \quad a.s., \tag{7.2}$$

for arbitrary $T > 0$ and positive constant C independent of n. In particular, for $\alpha_n = n^{-d\gamma},\ \gamma \in (0,1)$, we have

$$\sup_{t\in[0,T]} \|u_n(t,\cdot) - v_n(t,\cdot)\|_{L^2(Q)} \le Cn^{-d(1-\gamma-\delta)/2} \quad a.s., \tag{7.3}$$

where $0 < \delta < 1 - \gamma$ can be taken arbitrarily small.

We precede the proof of Theorem 7.1 with several auxiliary estimates.

Lemma 7.2. *From* (W-1s) *it follows*

$$\max\left\{ \sup_{n\in\mathbb{N}} \max_{\bar{i}\in[n]^d} n^{-d} \sum_{\bar{j}\in[n]^d} W^k_{n,\bar{i}\bar{j}},\ \sup_{n\in\mathbb{N}} \max_{\bar{j}\in[n]^d} n^{-d} \sum_{\bar{i}\in[n]^d} W^k_{n,\bar{i}\bar{j}} \right\} \le \bar{W}_k, \quad k \in [4]. \tag{7.4}$$

Proof. We prove (7.4) assuming that nonnegative W is in $L^2(Q^2)$, but not in $L^\infty(Q^2)$. In this case, $W_{n,\bar{i}\bar{j}}$ are defined by (4.24). For arbitrary $\bar{k} \in [n]^d$ and $n \in \mathbb{N}$, we have

$$\begin{aligned}
\sum_{\jmath\in[n]^d} W^k_{n,\bar{\imath},\bar{\jmath}} &= \sum_{\bar{\jmath}\in[n]^d} \left(n^d \int_{Q_{n,\bar{\imath}}\times Q_{n,\bar{\jmath}}} \alpha_n^{-1} \wedge W(x,y)dxdy \right)^k \\
&\le \sum_{\bar{\jmath}\in[n]^d} n^d \int_{Q_{n,\bar{\imath}}\times Q_{n,\bar{\jmath}}} \left(\alpha_n^{-k} \wedge W(x,y)^k \right) dxdy \\
&\le n^d \sum_{\bar{\jmath}\in[n]^d} \int_{Q_{n,\bar{\imath}}\times Q_{n,\bar{\jmath}}} W(x,y)^k dxdy \\
&\le n^d \bar{W}_k,
\end{aligned} \tag{7.5}$$

where we used Jensen's inequality in the second line and (W-1s) in the last line. Thus,

$$\sup_{n\in\mathbb{N}} \max_{\bar{i}\in[n]^d} n^{-d} \sum_{\bar{j}\in[n]^d} W^k_{n,\bar{i}\bar{j}} \le \bar{W}_k, \quad k \in [k].$$

The bound for $\sup_{n\in\mathbb{N}} \max_{\bar{j}\in[n]^d} n^{-d} \sum_{\bar{i}\in[n]^d} W^k_{n,\bar{i}\bar{j}}$ is proved similarly. $\square$

Lemma 7.3. *For $K \geq 2\bar{W}_1$, we have*

$$\mathbb{P}\left(\max_{\bar{i}\in[n]^d}\sum_{\bar{j}\in[n]^d}\left|\frac{a_{n,\bar{i}\bar{j}}}{\alpha_n}-W_{n,\bar{i}\bar{j}}\right|\geq Kn^d\right)\leq n^d\exp\left\{\frac{\frac{-1}{2}\left(K-2\bar{W}_1\right)^2\alpha n^d}{\bar{W}_1+O(\alpha_n)+K}\right\}. \tag{7.6}$$

In particular, with probability 1 *there exists* $n_0\in\mathbb{N}$ *such that*

$$\max\left\{\max_{\bar{i}\in[n]^d}\sum_{\bar{j}\in[n]^d}\left|\frac{a_{n,\bar{i}\bar{j}}}{\alpha_n}-W_{n,\bar{i}\bar{j}}\right|,\max_{\bar{i}\in[n]^d}\sum_{\bar{j}\in[n]^d}\left|\frac{a_{n,\bar{j}\bar{i}}}{\alpha_n}-W_{n,\bar{j}\bar{i}}\right|\right\}\leq Kn^d \tag{7.7}$$

for all $n\geq n_0$.

For the next lemma, we will need the following notation

$$\begin{aligned} Z_{n,\bar{i}}(t) &= n^{-d}\sum_{\bar{j}\in[n]^d} b_{n,\bar{i}\bar{j}}(t)\eta_{n,\bar{i}\bar{j}}, && (7.8)\\ b_{n,\bar{i}\bar{j}}(t) &= D\left(v_{n,\bar{j}}(t)-v_{n,\bar{i}}(t)\right), && (7.9)\\ \eta_{n,\bar{i}\bar{j}} &= a_{n,\bar{i}\bar{j}}-\alpha_n W_{n,\bar{i}\bar{j}}, && (7.10)\end{aligned}$$

and $Z_n=(Z_{n,\bar{i}},\ \bar{i}\in[n]^d)$.

Lemma 7.4. *For arbitrary $\epsilon>0$, we have*

$$\alpha_n^{-2}\int_0^\infty e^{-Ls}\|Z_n(s)\|^2_{2,n^d}ds\leq C(\alpha_n n^d)^{1-\epsilon}, \tag{7.11}$$

where C is a positive constant independent of n and

$$\|Z_n(s)\|_{2,n^d}=\left(n^{-1}\sum_{\bar{j}\in[n]^d}Z_{n,\bar{j}}(s)^2\right)^{1/2}. \tag{7.12}$$

Proof of Theorem 7.1. Recall that $f(u,x,t)$ and D are Lipschitz continuous function in u with Lipschitz constants L_f and L_D respectively.

Further, $a_{n,ij}$, are Bernoulli random variables

$$\mathbb{P}(a_{n,\bar{i}\bar{j}}=1)=\alpha_n W_{n,\bar{i}\bar{j}}. \tag{7.13}$$

Denote $\psi_{n,\bar{i}}:=v_{n,\bar{i}}-u_{n,\bar{i}}$. By subtracting (2.3) from (3.3), multiplying the result by $n^{-d}\psi_{n,\bar{i}}$, and summing over $\bar{i} \in [n]^d$, we obtain

$$\frac{1}{2}\frac{d}{dt}\|\psi_n\|^2_{2,n^d} = \underbrace{N^{-1}\sum_{\bar{i}\in[n]^d}\left(f(v_{n,\bar{i}},t) - f(u_{n,\bar{i}},t)\right)\psi_{n,i}}_{I_1} + \underbrace{n^{-2d}\alpha_n^{-1}\sum_{\bar{i},\bar{j}\in[n]^d}\left(\alpha_n W_{n,\bar{i}\bar{j}} - a_{n,\bar{i}\bar{j}}\right) D(v_{n,\bar{j}} - v_{n,\bar{i}})\psi_{n,\bar{i}}}_{I_2}$$
$$+ \underbrace{N^{-2}\alpha_n^{-1}\sum_{\bar{i},\bar{j}\in[n]^d}^{n} a_{n,\bar{i}\bar{j}}\left[D(v_{n,\bar{j}} - v_{n,\bar{i}}) - D(u_{n,\bar{j}} - u_{n,\bar{i}})\right]\psi_{n,\bar{i}}}_{I_3} =: I_1 + I_2 + I_3, \tag{7.14}$$

where $\|\cdot\|^2_{2,n^d}$ is the discrete L^2-norm (cf. (7.12)).

Using Lipschitz continuity of f in u, we have

$$|I_1| \le L_f\|\psi_n\|^2_{2,n^d}. \tag{7.15}$$

Using Lipschitz continuity of D and the triangle inequality, we have

$$\begin{aligned}|I_3| &\le L_D n^{-2d}\alpha_n^{-1}\sum_{\bar{i},\bar{j}\in[n]^d} a_{n,\bar{i}\bar{j}}\left(|\psi_{n,\bar{i}}| + |\psi_{n,\bar{j}}|\right)|\psi_{n,\bar{i}}| \\ &\le L_D n^{-2d}\alpha_n^{-1}\left(\frac{3}{2}\sum_{\bar{i},\bar{j}\in[n]^d} a_{n,\bar{i},\bar{j}}\psi^2_{n,\bar{i}} + \frac{1}{2}\sum_{\bar{i},\bar{j}\in[n]^d} a_{n,\bar{i}\bar{j}}\psi^2_{n,\bar{j}}\right).\end{aligned} \tag{7.16}$$

Using Lemma 7.3 and (7.4), we obtain

$$\begin{aligned}\alpha_n n^{-2d}\sum_{\bar{i},\bar{j}\in[n]^d} a_{n,\bar{i}\bar{j}}\psi^2_{n,\bar{i}} &\le n^{-d}\sum_{\bar{i}\in[n]^d}\left\{n^{-d}\sum_{\bar{j}\in[n]^d}\left(\left|\frac{a_{n,\bar{i}\bar{j}}}{\alpha_n} - W_{n,\bar{i}\bar{j}}\right| + W_{n,\bar{i}\bar{j}}\right)\psi^2_{n,\bar{i}}\right\} \\ &\le n^{-d}\sum_{\bar{i}\in[n]^d}\left(K + \bar{W}_1\right)\psi^2_{n,\bar{i}} = (K + W_1)\|\psi_n\|^2_{2,n^d}.\end{aligned} \tag{7.17}$$

Similarly,

$$n^{-2d}\alpha_n^{-1}\sum_{\bar{i},\bar{j}\in[n]^d} a_{n,\bar{i}\bar{j}\in[n]^d}\psi^2_{n,\bar{j}} \le \left(K + \bar{W}_2\right)\|\psi_n\|^2_{2,n^d}. \tag{7.18}$$

By plugging (7.17) and (7.18) into (7.16), we have

$$|I_3| \le L_D\left(2K + \frac{3}{2}\bar{W}_1 + \frac{1}{2}\bar{W}_2\right)\|\psi\|^2_{n^d,2}. \tag{7.19}$$

It remains to bound I_2:

$$|I_2| = |n^{-d}\alpha_n^{-1}\sum_{\bar{i}\in[n]^d}^{n} Z_{n,\bar{i}}\psi_{n,\bar{i}}| \le 2^{-1}\alpha_n^{-2}\|Z_n\|^2_{2,n^d} + 2^{-1}\|\psi_n\|^2_{2,n^d}. \tag{7.20}$$

The combination of (7.14), (7.15), (7.19) and (7.20) yields

$$\frac{d}{dt}\|\psi_n(t)\|^2_{2,n} \le L\|\psi_n(t)\|^2_{2,n} + \frac{1}{\alpha_n^2}\|Z_n(t)\|^2_{2,n}, \tag{7.21}$$

where $L = L_f + L_D\left(2K + \frac{3}{2}\bar{W}_1 + \frac{1}{2}\bar{W}_2\right) + \frac{1}{2}$.

Using the Gronwall's inequality and Lemma 7.4, we have

$$\begin{aligned}\sup_{t\in[0,T]}\|\psi_n(t)\|^2_{2,n^d} &\le \alpha_n^{-2}e^{LT}\int_0^\infty e^{-Ls}\|Z_n(s)\|^2_{2,n^d}ds \\ &\le \alpha_n^{-2}e^{LT}(n^d\alpha_n)^{-1+\epsilon}.\end{aligned} \tag{7.22}$$

□

Proof of Lemma 7.3. Let

$$\xi_{n,\bar{i}\bar{j}} = \left|\frac{a_{n,\bar{i}\bar{j}}}{\alpha_n} - W_{n,\bar{i}\bar{j}}\right| - 2W_{n,\bar{i}\bar{j}}\left(1-\alpha_n W_{n,\bar{i}\bar{j}}\right),\ \bar{i},\bar{j}\in[n]^d. \tag{7.23}$$

Note that for fixed $\bar{i}\in[n]^d$, $\{\xi_{n,\bar{i}\bar{j}},\ \bar{j}\in[n]^d\}$ are mean zero independent RVs. Further, using the definition of $\xi_{n,\bar{i}\bar{j}}$, it is straightforward to bound

$$|\xi_{n,\bar{i}\bar{j}}| \le \alpha_n^{-1} + 2W_{n,\bar{i}\bar{j}} \le 3\alpha_n^{-1} =: M, \tag{7.24}$$

$$\mathbb{E}\,\xi^2_{n,\bar{j}} \le 2\alpha_n^{-1}W_{n,\bar{i}\bar{j}} + 2W^2_{n,\bar{i}\bar{j}} + 4\alpha_n W^2_{n,\bar{i}\bar{j}} + 4\alpha_n W^3_{n,\bar{i}\bar{j}}. \tag{7.25}$$

From (7.25), we have

$$\begin{aligned}\mathbb{E}\left(\sum_{\bar{j}\in[n]^d}\xi^2_{n,\bar{i}\bar{j}}\right) &\le \alpha_n^{-1}\sum_{\bar{j}\in[n]^d}\left(2W_{n,\bar{i}\bar{j}} + \alpha_n 2W^2_{n,\bar{i}\bar{j}} + 4\alpha_n^2 W^2_{n,\bar{i}\bar{j}} + 4\alpha_n^2 W^3_{n,\bar{i}\bar{j}}\right), \\ &\le \alpha_n^{-1}n^d W_1 + O\left(\alpha_n\right).\end{aligned} \tag{7.26}$$

Using Bernstein's inequality and the union bound, we have

$$\begin{aligned}\mathbb{P}\left(\max_{\bar{i}\in[n]^d}\sum_{\bar{j}\in[n]^d}\xi_{n,\bar{i}\bar{j}} \ge \left(K-2\bar{W}_1\right)n^d\right) &\le n^d\exp\left\{\frac{\frac{-1}{2}\left(K-2\bar{W}_1\right)^2 n^{2d}}{\sum_{\bar{j}\in}\mathbb{E}\,\xi^2_{n,\bar{i}\bar{j}} + (1/3)M\left(K-2\bar{W}_1\right)n^d}\right\} \\ &\le n^d\exp\left\{\frac{\frac{-1}{2}\left(K-2\bar{W}_1\right)^2 n^{2d}}{\alpha_n^{-1}n^d\left(\bar{W}_1 + O(\alpha_n)\right) + \alpha_n^{-1}\left(K-2\bar{W}_1\right)n^d}\right\} \\ &\le N\exp\left\{\frac{\frac{-1}{2}\left(K-2\bar{W}_1\right)^2\alpha_n n^d}{\bar{W}_1 + O(\alpha_n) + K}\right\}.\end{aligned} \tag{7.27}$$

Finally, the combination of (7.23) and (7.27) yields

$$\begin{aligned}\mathbb{P}\left(\max_{\ni\in[n]^d}\sum_{\bar{j}\in[n]^d}\left|\frac{a_{n,\bar{i}\bar{j}}}{\alpha_n}-W_{n,\bar{i}\bar{j}}\right|\geq Kn^d\right)&\leq\mathbb{P}\left(\max_{\ni\in[n]^d}\sum_{\bar{j}\in[n]^d}\xi_{n,\bar{i}\bar{j}}\geq\left(K-\frac{2}{n^d}\sum_{\bar{j}\in[n]^d}W_{n,\bar{i}\bar{j}}\right)n^d\right)\\&\leq\mathbb{P}\left(\max_{\ni\in[n]^d}\sum_{\bar{j}\in[n]^d}\xi_{n,\bar{i}\bar{j}}\geq\left(K-2\bar{W}_1\right)n^d\right)\\&\leq n^d\exp\left\{\frac{\frac{-1}{2}\left(K-2\bar{W}_1\right)^2\alpha n^d}{\bar{W}_1+O(\alpha_n)+K}\right\}.\end{aligned}$$

This proves (7.13). By Borel-Cantelli Lemma, (7.7) follows. □

Proof of Lemma 7.4. Recall (7.8)-(7.10) and rewrite

$$\int_0^\infty e^{-Ls}\|Z_n(s)\|^2_{2,n^d}ds=n^{-3d}\sum_{\bar{i},\bar{k},\bar{l}\in[n]^d}c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}},\tag{7.28}$$

where

$$c_{n,\bar{i}\bar{k}\bar{l}}=\int_0^\infty e^{-Ls}b_{n,\bar{i}\bar{k}}(s)b_{n\bar{i}\bar{l}}(s)ds\quad\text{and}\quad|c_{n,\bar{i}\bar{k}\bar{l}}|\leq L^{-1}=:\bar{c}.\tag{7.29}$$

By (7.1), one can choose a sequence $\delta_n\searrow 0$ such that

$$n^d\delta_n\gg\alpha_n^{-1}.\tag{7.30}$$

Specifically, let

$$\delta_n:=\frac{1}{\sqrt{\ln n}}.\tag{7.31}$$

and define events

$$\Omega_n=\left\{(n^d\alpha_n)^{-2}\sum_{\bar{i},\bar{k},\bar{l}\in[n]^d}^{n}c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}}>\delta_n n^d\right\},\tag{7.32}$$

$$A_{n,\bar{i}}=\left\{\sum_{\bar{j}\in[n]^d}\left|\frac{a_{n,\bar{i}\bar{j}}}{\alpha_n}-W_{n,\bar{i}\bar{j}}\right|>Kn^d\right\},\ \text{and } A_n=\bigcup_{\bar{i}\in[n]^d}A_{n,\bar{i}}.\tag{7.33}$$

Clearly,

$$\mathbb{P}(\Omega_n)\leq\mathbb{P}(\Omega_n\cap A_n^c)+\mathbb{P}(A_n).\tag{7.34}$$

We want to show that $\mathbb{P}\left(\Omega_n\text{ infinitely often}\right)=0$. By Borel-Cantelli Lemma, it is sufficient to show that

$$\sum_{n\geq 1}\mathbb{P}(\Omega_n)<\infty.$$

From Lemma 7.3, we know that $\sum_{n\geq1}\mathbb{P}(A_n)<\infty$ for $K>2\bar{W}_1$. In the remainder of the proof, we show that $\sum_{n\geq1}\mathbb{P}(\Omega_n\cap A_n^c)$ is convergent.

Applying the exponential Markov inequality to $\mathbb{P}(\Omega_n \,|A_n^c\,)$, from $\mathbb{P}(\Omega_n \cap A_n^c) \le \mathbb{P}(\Omega_n \,|A_n^c\,)$ and (7.32), we have

$$\mathbb{P}(\Omega_n \cap A_n^c) \le \exp\left\{ -n^d\delta_n + \ln \mathbb{E}\left[\mathbf{1}_{A_n^c} \exp\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{i},\bar{k},\bar{l}\in[n]^d}^{n} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right\}\right]\right\}. \tag{7.35}$$

Using the independence of $\eta_{n,\bar{i}\bar{k}\bar{l}}$ in $\bar{i}\in[n]^d$, we have

$$\mathbb{E}\left[\mathbf{1}_{A_n^c} \exp\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{i},\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right\}\right] = \prod_{\bar{i}\in[n]^d} \mathbb{E}\left[\mathbf{1}_{A_n^c} \exp\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}}. \right\}\right] \tag{7.36}$$

Using

$$e^x \le 1 + |x|e^{|x|}, \quad x\in\mathbb{R},$$

and the Cauchy-Schwartz inequality, we bound the right–hand side of (7.36) as follows

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{A_n^c} \exp\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right\}\right] \\
&\le 1 + \mathbb{E}\left[\mathbf{1}_{A_n^c} \left| (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right| \exp\left\{ \left| (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right| \right\}\right] \\
&\le 1 + \left(\mathbb{E}\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right\}^2 \right)^{1/2} \\
&\times \left(\mathbb{E}\left\{ \mathbf{1}_{A_n^c} \exp\left\{ 2(n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right\}\right\}\right)^{1/2}.
\end{aligned} \tag{7.37}$$

From (7.10), (7.29), and under A_n^c (cf. (7.33)), we have

$$\mathbf{1}_{A_n^c} 2(n^d\alpha_n)^{-2} \left| \sum_{\bar{k},\bar{l}\in[n]^d}^{n} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right| \le 2K\bar{c}. \tag{7.38}$$

Further,

$$\begin{aligned}
&\mathbb{E}\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d} c_{n,\bar{i}\bar{k}\bar{l}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}} \right\}^2 \le (n^d\alpha_n)^{-4} \sum_{\bar{j},\bar{k},\bar{l},\bar{p}\in[n]^d} \mathbb{E}\left(\eta_{n,\bar{i}\bar{j}}\eta_{n,\bar{i}\bar{k}}\eta_{n,\bar{i}\bar{l}}\eta_{n,\bar{i}\bar{p}}\right) c_{n,\bar{i}\bar{j}\bar{p}}c_{n,\bar{i}\bar{k}\bar{l}} \\
&\le \frac{(\bar{c})^2}{(n^d\alpha_n)^4} \left\{ \sum_{\bar{j}\in[n]^d} \mathbb{E}\,\eta_{n,\bar{i}\bar{j}}^4 + 6\left(\sum_{\bar{j}\in[n]^d} \mathbb{E}\,\eta_{\bar{i}\bar{j}}^2 \right)^2 \right\}.
\end{aligned} \tag{7.39}$$

Using (7.10), we estimate sum of the fourth moments of $\eta_{n,\bar{i}\bar{j}}$

$$\begin{aligned}\sum_{\bar{j}\in[n]^d} \mathbb{E}\,\eta_{\bar{i}\bar{j}}^4 &= \sum_{\bar{j}\in[n]^d} \left\{\alpha_n W_{n,\bar{i}\bar{j}} \left(1-\alpha_n W_{n,\bar{i}\bar{j}}\right)^4 + \alpha_n^4 W_{n,\bar{i}\bar{j}}^4 \left(1-\alpha_n W_{n,\bar{i}\bar{j}}\right)\right\} \\ &\le n^d \alpha_n \left(N^{-1} \sum_{\bar{j}\in[n]^d} W_{n,\bar{i}\bar{j}} + \alpha_n^3 n^{-d} \sum_{\bar{j}\in[n]^d} W_{n,\bar{i}\bar{j}}^4 \right) \\ &\le n^d \alpha_n \left(\bar{W}_1 + \alpha_n^3 \bar{W}_4\right) = O(\alpha_n n^d),\end{aligned} \tag{7.40}$$

where we also use (7.4). Similarly,

$$\begin{aligned}\sum_{\bar{j}\in[n]^d} \mathbb{E}\,\eta_{\bar{i}\bar{j}}^2 &= \sum_{\bar{j}\in[n]^d} \left\{\alpha_n W_{n,\bar{i}\bar{j}} \left(1-\alpha_n W_{n,\bar{i}\bar{j}}\right)^2 + \alpha_n^2 W_{n,\bar{i}\bar{j}}^2 \left(1-\alpha_n W_{n,\bar{i}\bar{j}}\right)\right\} \\ &\le n^d \alpha_n \left(N^{-1} \sum_{\bar{j}\in[n]^d} W_{n,\bar{i}\bar{j}} + \alpha_n n^{-d} \sum_{\bar{j}\in[n]^d} W_{n,\bar{i}\bar{j}}^2 \right) \\ &\le n^d \alpha_n \left(\bar{W}_1 + \alpha_n \bar{W}_2\right) = O(\alpha_n N).\end{aligned} \tag{7.41}$$

By combining (7.39)-(7.41), we obtain

$$\mathbb{E}\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d}^{n} c_{n,\bar{i}\bar{k}\bar{l}} \eta_{n,\bar{i}\bar{k}} \eta_{n,\bar{i}\bar{l}} \right\}^2 = O\left((n^d\alpha_n)^{-2} \right). \tag{7.42}$$

By plugging (7.38) and (7.42) into (7.37), we obtain

$$\mathbb{E}\left[\mathbf{1}_{A_n^c} \exp\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{k},\bar{l}\in[n]^d}^{n} c_{n,\bar{i}\bar{k}\bar{l}} \eta_{n,\bar{i}\bar{k}} \eta_{n,\bar{i}\bar{l}} \right\} \right] \le 1 + \frac{C_1}{n^d \alpha_n} e^{C_2}. \tag{7.43}$$

Using this bound on the right–hand side of (7.36), we further obtain

$$\mathbb{E}\left[\mathbf{1}_{A_n^c} \exp\left\{ (n^d\alpha_n)^{-2} \sum_{\bar{i},\bar{k},\bar{l}\in[n]^d}^{n} c_{n,\bar{i}\bar{k}\bar{l}} \eta_{n,\bar{i}\bar{k}} \eta_{n,\bar{i}\bar{l}} \right\} \right] \le e^{C_3 \alpha_n^{-1}}. \tag{7.44}$$

Using (7.44), from (7.35) we obtain

$$\mathbb{P}(\Omega_n \cap A_n^c) \le \exp\left\{ -n^d \delta_{n^d} + C_3 \alpha_n^{-1} \right\} \to 0, \quad n \to \infty. \tag{7.45}$$

Furthermore, using (7.31) it is straightforward to check that

$$\sum_{n=1}^{\infty} \mathbb{P}(\Omega_n \cap A_n^c) < \infty.$$

The statement of the lemma then follows from (7.32)-(7.34) via Borel-Cantelli Lemma. □

Acknowledgements. This work was supported in part by the NSF grant DMS 1715161.

References

[1] N. I. Akhieser, *Theory of Approximation*, Dover Publications, Inc., New York, 1992, Translated from the Russian and with a preface by Charles J. Hyman, Reprint of the 1956 English translation.

[2] Fuensanta Andreu-Vaillo, José M. Mazón, Julio D. Rossi, and J. Julián Toledo-Melero, *Nonlocal Diffusion Problems*, Mathematical Surveys and Monographs, vol. 165, American Mathematical Society, Providence, RI; Real Sociedad Matemática Española, Madrid, 2010.

[3] Lehel Banjai, Jens M. Melenk, Ricardo H. Nochetto, Enrique Otárola, Abner J. Salgado, and Christoph Schwab, *Tensor FEM for spectral fractional diffusion*, Found. Comput. Math. **19** (2019), no. 4, 901–962.

[4] Andrea Bonito, Juan Pablo Borthagaray, Ricardo H. Nochetto, Enrique Otárola, and Abner J. Salgado, *Numerical methods for fractional diffusion*, Comput. Vis. Sci. **19** (2018), no. 5-6, 19–46.

[5] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao, *An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions*, Trans. Amer. Math. Soc. **372** (2019), no. 5, 3019–3062.

[6] Nikolaos Bournaveas and Vincent Calvez, *The one-dimensional Keller-Segel model with fractional diffusion of cells*, Nonlinearity **23** (2010), no. 4, 923–935.

[7] C. Carrillo and P. Fife, *Spatial effects in discrete generation population models*, J. Math. Biol. **50** (2005), no. 2, 161–188.

[8] Stephen Coombes, Peter beim Graben, and Roland Potthast, *Tutorial on neural field theory*, Neural fields, Springer, Heidelberg, 2014, pp. 1–43.

[9] Fabio Coppini, Helge Dietert, and Giambattista Giacomin, *A Law of Large Numbers and Large Deviations for interacting diffusions on Erdos-Renyi graphs*, arXiv e-prints (2018), arXiv:1807.10921.

[10] Arturo de Pablo, Fernando Quirós, and Ana Rodríguez, *Nonlocal filtration equations with rough kernels*, Nonlinear Anal. **137** (2016), 402–425.

[11] Félix del Teso, Jø rgen Endal, and Espen R. Jakobsen, *Uniqueness and properties of distributional solutions of nonlocal equations of porous medium type*, Adv. Math. **305** (2017), 78–143.

[12] Ronald A. DeVore and George G. Lorentz, *Constructive approximation*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 303, Springer-Verlag, Berlin, 1993.

[13] Qiang Du, *An invitation to nonlocal modeling, analysis, and computation*, Proc. Int. Cong. of Math. (2018), Rio de Janeiro, vol. 3, pp. 3523-3552.

[14] Qiang Du, Lili Ju, and Jianfang Lu, *A discontinuous Galerkin method for one-dimensional time-dependent nonlocal diffusion problems*, Math. Comp. **88** (2019), no. 315, 123–147.

[15] Lawrence C. Evans and Ronald F. Gariepy, *Measure theory and fine properties of functions*, revised ed., Textbooks in Mathematics, CRC Press, Boca Raton, FL, 2015.

[16] Kenneth Falconer, *Fractal geometry*, third ed., John Wiley & Sons, Ltd., Chichester, 2014, Mathematical foundations and applications.

[17] François Golse, *On the dynamics of large particle systems in the mean field limit*, Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity, Lect. Notes Appl. Math. Mech., vol. 3, Springer, [Cham], 2016, pp. 1–144.

[18] Dmitry Kaliuzhnyi-Verbovetskyi and Georgi S. Medvedev, *The semilinear heat equation on sparse random graphs*, SIAM J. Math. Anal. **49** (2017), no. 2, 1333–1355.

[19] L. Lovász, *Large networks and graph limits*, AMS, Providence, RI, 2012.

[20] L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957.

[21] Georgi S. Medvedev, *The nonlinear heat equation on dense graphs and graph limits*, SIAM J. Math. Anal. **46** (2014), no. 4, 2743–2766.

[22] ———, *The nonlinear heat equation on W-random graphs*, Arch. Ration. Mech. Anal. **212** (2014), no. 3, 781–803.

[23] ———, *Small-world networks of Kuramoto oscillators*, Phys. D **266** (2014), 13–22.

[24] ———, *The continuum limit of the Kuramoto model on sparse random graphs*, Communications in Mathematical Sciences **17** (2019), no. 4, 883–898.

[25] Georgi S. Medvedev and Xuezhi Tang, *Stability of twisted states in the Kuramoto model on Cayley and random graphs*, J. Nonlinear Sci. **25** (2015), no. 6, 1169–1208.

[26] ———, *The Kuramoto model on power law graphs: Synchronization and contrast states*, Journal of Nonlinear Science (2018).

[27] Georgi S. Medvedev and J. Douglas Wright, *Stability of twisted states in the continuum Kuramoto model*, SIAM J. Appl. Dyn. Syst. **16** (2017), no. 1, 188–203.

[28] Ricardo H. Nochetto, Enrique Otárola, and Abner J. Salgado, *A PDE approach to space-time fractional parabolic problems*, SIAM J. Numer. Anal. **54** (2016), no. 2, 848–873.

[29] ———, *A PDE approach to space-time fractional parabolic problems*, SIAM J. Numer. Anal. **54** (2016), no. 2, 848–873.

[30] Russell W. Schwab and Luis Silvestre, *Regularity for parabolic integro-differential equations with very irregular kernels*, Anal. PDE **9** (2016), no. 3, 727–772.

[31] Wenxian Shen and Xiaoxia Xie, *Approximations of random dispersal operators/equations by nonlocal dispersal operators/equations*, J. Differential Equations **259** (2015), no. 12, 7375–7405.

[32] Juan Luis Vázquez, *The mathematical theories of diffusion: nonlinear and fractional diffusion*, Nonlocal and nonlinear diffusions and interactions: new methods and directions, Lecture Notes in Math., vol. 2186, Springer, Cham, 2017, pp. 205–278.

[33] D.A. Wiley, S.H. Strogatz, and M. Girvan, *The size of the sync basin*, Chaos **16** (2006), no. 1, 015103, 8.

[34] David Williams, *Probability with Martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.