



HAL
open science

Acceleration of Gossip Algorithms through the Euler-Poisson-Darboux Equation

Raphaël Berthier, Mufan Bill Li

► **To cite this version:**

Raphaël Berthier, Mufan Bill Li. Acceleration of Gossip Algorithms through the Euler-Poisson-Darboux Equation. 2022. hal-03582008

HAL Id: hal-03582008

<https://hal.science/hal-03582008v1>

Preprint submitted on 21 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Acceleration of Gossip Algorithms through the Euler–Poisson–Darboux Equation

Raphaël Berthier¹ and Mufan (Bill) Li²

¹EPFL*

²University of Toronto and Vector Institute

ABSTRACT. Gossip algorithms and their accelerated versions have been studied exclusively in discrete time on graphs. In this work, we take a different approach, and consider the scaling limit of gossip algorithms in both large graphs and large number of iterations. These limits lead to well-known partial differential equations (PDEs) with insightful properties. On lattices, we prove that the non-accelerated gossip algorithm of Boyd et al. [2006] converges to the heat equation, and the accelerated Jacobi polynomial iteration of Berthier et al. [2020] converges to the Euler–Poisson–Darboux (EPD) equation — a damped wave equation. Remarkably, with appropriate parameters, the fundamental solution of the EPD equation has the ideal gossip behaviour: a uniform density over an ellipsoid, whose radius increases at a rate proportional to t — the fastest possible rate for locally communicating gossip algorithms. This is in contrast with the heat equation where the density spreads on a typical scale of \sqrt{t} . Additionally, we provide simulations demonstrating that the gossip algorithms are accurately approximated by their limiting PDEs.

1. INTRODUCTION

In computer science, the large amount of data and the large size of computation networks motivate a growing interest in distributed algorithms, see [Assran et al., 2020] for instance. Among these, decentralized algorithms—where there is no master node aggregating information and distributing tasks—are appreciated for their flexibility, their robustness to node/links failures and their scalability. In this paper, we study the averaging problem, or gossip problem, a toy problem of decentralized computing where the network aims to compute the average of real values distributed along the nodes of the network, see [Boyd et al., 2006] for instance. The algorithms that tackle this task—called gossip algorithms—are used as primitives in more complex distributed algorithms, including distributed optimization or distributed reinforcement learning, see, e.g., [Dimakis et al., 2010, Assran et al., 2020, Szorenyi et al., 2013].

To be specific, we give each node v a real value $x_0(v)$, and gossip algorithms aim at computing the average of the values held by the nodes in a decentralized fashion. Boyd et al. [2006] proposed the following gossip algorithm, that we refer to as *simple gossip*: at each communication round n , each node v replaces its current value $x_n(v)$ by a weighted average of the value of its neighbors and its own current value. More specifically,

$$x_{n+1}(v) = \sum_{\eta:\eta\sim v} W_{v,\eta}x_n(\eta), \quad (1)$$

Here, we use $\eta \sim v$ to denote that η, v are neighbouring nodes, and $W_{v,\eta}$ to denote the weights in the averaging operation. As the number of iterations grows, the running value $x_n(v)$ of each node can be shown to converge to the global average of the initial values. However, the convergence is notoriously slow in networks with a finite-dimensional geometry, such as grids or random geometric graphs. Heuristically, simple gossip averages locally the values in the network, but fails at spreading information quickly at a larger distance. This is analogous to the heat diffusion in a continuous media that homogenizes quickly locally but only slowly on large scales: for this reason, the slow convergence of simple gossip is sometimes called a *diffusivity* problem. We make this analogy more precise later in this section.

To reach super-diffusive rates of convergence, several accelerations of the simple gossip algorithm were proposed. We mention a few of them that are representative or related to the approach of this paper. Dimakis et al. [2008] use the knowledge of the position of the nodes in space in order to give inertia to the information in a specific direction. Even et al. [2021] mimicked the proof of the accelerated coordinate gradient descent of Nesterov [2012] to obtain an accelerated gossip algorithm

*While RB is currently affiliated to EPFL, the research presented here was conducted mainly at Inria, Département d’informatique de l’ENS, PSL Research University, Paris, France.

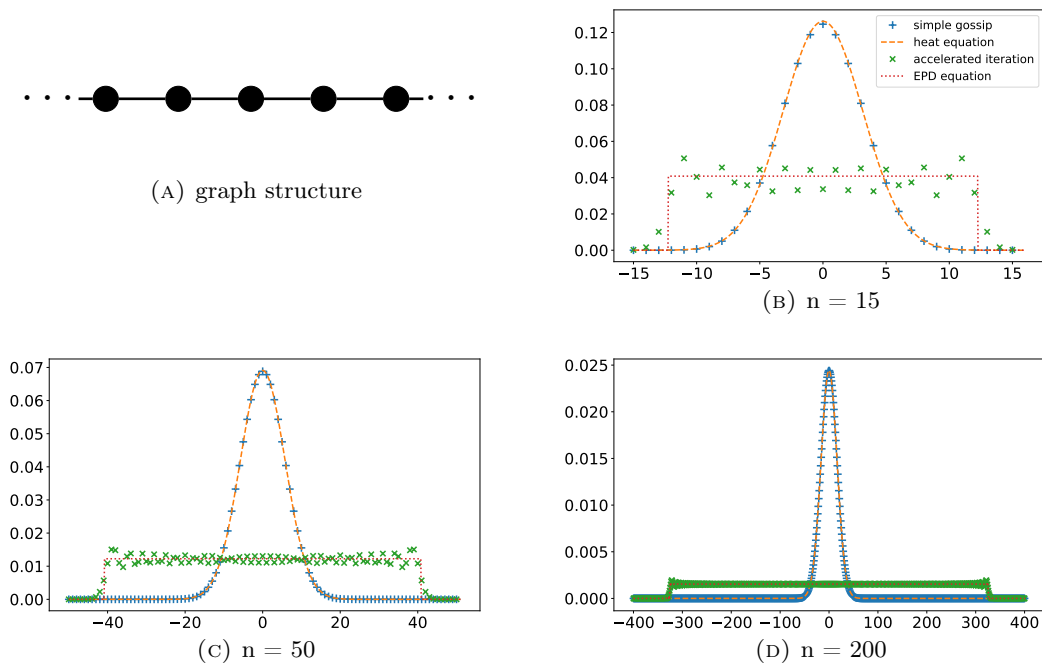


FIGURE 1. Comparison between gossip algorithms and their scaling limits: the simple gossip of [Boyd et al., 2006] and the heat equation, the accelerated Jacobi polynomial iteration of [Berthier et al., 2020] and the Euler–Poisson–Darboux equation. All iterations were run on the line graph \mathbb{Z} ($d = 1$) and initialized from $x_0 = \mathbf{1}_0$, the vector such that $\mathbf{1}_0(0) = 1$ and $\mathbf{1}_0(v) = 0$ otherwise. We show the results $x_n(v)$ as a function of $v \in \mathbb{Z}$ for different numbers of iterations $n = 15, 50, 200$. Note that as the number of iteration increases, the description through the scaling limits improves in accuracy. The accelerated Jacobi polynomial iteration diffuses faster; it has a different scaling than the simple gossip algorithm.

in the form of an iteration over several variables. Berthier et al. [2020] designed second-order iterations using an orthogonal polynomial point of view coupled with a Jacobi approximation of the spectral measure of the network graph: the resulting accelerated gossip algorithm is called the Jacobi polynomial iteration. Sardellitti et al. [2010] built on the analogy of simple gossip with diffusion processes; they proposed to add the discrete equivalent of an advection term in the iteration to accelerate the homogenization.

Contributions. We analyze and design gossip algorithms through associated partial differential equations (PDEs). We study gossip algorithms on regular lattices on \mathbb{Z}^d (for instance, grids), and draw a rigorous connection to their corresponding PDEs via scaling limits in time and space.

Our first contribution is to show that the scaling limit of the simple gossip algorithm on a lattice is the heat equation (see, e.g., [Evans, 1998])

$$\partial_t u = \frac{1}{2} \nabla_y \cdot (Q \nabla_y u), \quad u = u(t, y). \quad (2)$$

Here, ∇_y and $\nabla_y \cdot$ denote respectively the gradient and the divergence operator in the variable y . Q is a $d \times d$ matrix quantifying the potential anisotropy of the diffusion: it is a function of the local averaging operation, defined more precisely in Section 2. The fundamental solution of the heat equation (2) (the weak solution when initialized at the Dirac mass $u(0, \cdot) = \delta_0$ [Evans, 1998]) is a centered Gaussian density

$$u(t, y) = \frac{1}{(2\pi)^{d/2} t^{d/2} (\det Q)^{1/2}} \exp\left(-\frac{1}{2t} \langle y, Q^{-1} y \rangle\right). \quad (3)$$

The formula above shows the sub-optimality of the simple gossip method: the mass spreads on a typical scale $\|y\| \approx \sqrt{t}$, while we would like the scale to be $\|y\| \approx t$; indeed, the gossiped information

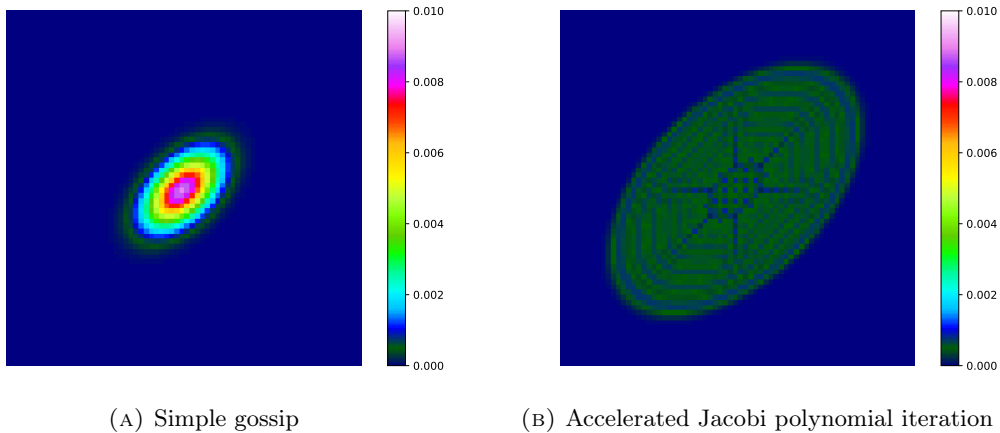


FIGURE 2. Comparison between simple gossip and the accelerated Jacobi polynomial iteration on the triangular lattice (see Equation (7) or Figure 3(B)). We initialize from $x_0 = \mathbf{1}_0$ and we show the iterates $x_{30} = (x_{30}(v))_{v \in \mathbb{Z}^2}$ using a color scale. The accelerated Jacobi polynomial iteration diffuses faster than simple gossip: the mass is distributed more evenly and on a larger ellipsoid.

can travel at most at distance $\Theta(t)$ in a time t (due to the speed limit of local communications), and we would like our gossip algorithms to match this optimal speed of diffusion. Equivalently, the solution decays to 0 at the rate $1/t^{d/2}$ in $\|\cdot\|_\infty$, while we would like the rate to be $1/t^d$.

To this goal, we design an accelerated second-order gossip iteration: we choose the recursion coefficients so that the iteration converges to the Euler–Poisson–Darboux (EPD) equation [Euler, 1770, Poisson, 1823, Darboux, 1896]

$$\partial_{tt}u + \frac{d+1}{t}\partial_t u = \nabla_y \cdot (Q\nabla_y u). \quad (4)$$

See Appendix A and Bresters [1973] for an introduction to the EPD equation in a more general form. The EPD equation is a wave equation with a decaying damping term. Intuitively, the wave component gives inertia to the diffusion so that the resulting PDE mixes faster, while the damping term reduces oscillation. For the precise value of the damping coefficient $\frac{d+1}{t}$ and the initial conditions $u(0, \cdot) = \delta_0$, $\partial_t u(0, \cdot) = 0$, the fundamental solution has a remarkable formula

$$u(t, y) = \frac{\Gamma(d/2 + 1)}{\pi^{d/2}(\det Q)^{1/2}} \frac{1}{t^d} \mathbf{1}_{\{\langle y, Q^{-1}y \rangle \leq t^2\}}. \quad (5)$$

This method thus has an optimal scaling: the mass spreads on a typical scale $\|y\| \approx t$ and the solution decays to 0 at the rate $1/t^d$. Furthermore, the fundamental solution also has the perfect shape: the averaging is uniform on the ellipsoid $\{\langle y, Q^{-1}y \rangle \leq t^2\}$. We emphasize that this solution embodies the ideal gossip behaviour: uniformly spreading information at a sharp rate (again this is the fastest possible rate for locally communicating gossip algorithms).

Notably, the Jacobi polynomial iteration of Berthier et al. [2020] is one of the possible accelerations that we identify as scaling to the EPD equation (4). As a consequence, this paper can be seen as a more intuitive derivation of the Jacobi polynomial iteration, that was derived through algebraic methods on polynomials.

In Figures 1-2, we provide simulations in dimension $d = 1$ and $d = 2$. They show that the limiting PDEs are accurate in describing the behavior of gossip algorithms as the number of iterations grows, and that the accelerated methods achieve faster diffusion.

We intend this paper to be exploratory in the connection between gossip algorithms and PDEs, therefore our results are restricted to simplified settings. More precisely, our derivations and proofs assume synchronous communications—where all nodes communicate synchronously—as frequently done in the analysis of gossip algorithms, see for instance [Cao et al., 2006, Rebeschini and Tatikonda, 2017, Berthier et al., 2020]. Moreover, the network graph must be a lattice and the gossip operation must be translation-invariant. Indeed, this restriction is crucial to our proof technique which relies heavily on using the Fourier transform. However, a Fourier approach allows us to study the

fundamental solution, which cannot be easily handled by standard numerical analysis techniques [Genis, 1984, Celia and Gray, 1992]. At the same time, we believe that the insight from accelerating through the EPD equation can be used in much wider settings. For instance, the Jacobi polynomial iteration accelerates in many graphs with a finite-dimensional geometry, including regular grids, percolation bonds, or random geometric graphs [Berthier et al., 2020]. We expect the intuitions presented in this paper to have the same universality.

Structure of the paper. In Section 2, we set up our problem: we properly define the lattices, the gossip problem and the simple gossip algorithm. In Section 3, we give heuristic but intuitive derivations showing why simple gossip scales to the heat equation and how to build a method scaling to the EPD equation. Furthermore, we draw the connection with Jacobi polynomial iterations, and we discuss the open problem of extending the results to asynchronous settings or on more general graphs.

Section 4 provide some rigorous support to the above derivations. The simple gossip algorithm can be seen as the iteration of the law of a random walk on \mathbb{Z}^d ; the convergence to a Gaussian random variable is made rigorous by the central limit theorem and the local central limit theorem, see Section 4.1. In Section 4.2, we provide analog results for the convergence of the Jacobi polynomial iteration to the EPD equation (4): a weak limit theorem and a stronger result of local type. Finally, in Section 4.3, we apply the latter result to obtain an asymptotic equivalent² (not only a domination) of the convergence rate for the Jacobi polynomial iteration on \mathbb{Z}^d :

$$\sum_{v \in \mathbb{Z}^d} x_n(v)^2 \underset{n \rightarrow \infty}{\sim} \frac{1}{(\det Q)^{1/2} |B(0, 1)|} \frac{1}{n^d},$$

where $|B(0, 1)|$ is the volume of the unit ball in \mathbb{R}^d .

Notation. For $v \in \mathbb{Z}^d$, we denote $\mathbf{1}_v = (\mathbf{1}_v(w))_{w \in \mathbb{Z}^d}$ the vector with entry $\mathbf{1}_v(v) = 1$ and all other entries equal to 0. We denote e_1, \dots, e_d the canonical basis of \mathbb{R}^d . $[s]$ denotes the integer part of a real number s .

2. SETTING

Lattices and translation-invariant gossip operation. In this section, we introduce the notations and lattices for the gossip problem. In order to make the rescaling of the processes indexed by the vertices more natural, we consider here only graphs with vertex set \mathbb{Z}^d . Instead of defining the edge sets using graph notation, we will use a more convenient (and equivalent) definition through the local averaging operation. Let $\omega = (\omega(v))_{v \in \mathbb{Z}^d}$ be a vector of non-negative reals, representing a local averaging filter on \mathbb{Z}^d . Here we assume that ω has finite support and that $\sum_{v \in \mathbb{Z}^d} \omega(v) = 1$. For a vector $x = (x(v))_{v \in \mathbb{Z}^d}$, we define the local averaging operation by the discrete convolution on \mathbb{Z}^d

$$(\omega * x)(v) = \sum_{\eta \in \mathbb{Z}^d} \omega(v - \eta) x(\eta),$$

which represents a local weighted average of the neighboring values. For this operation to represent local communications on the graph, we would require all pairs $\{v, \eta\}$ such that $\omega(v - \eta) > 0$ to be connected in the graph. We thus take $\{\{v, \eta\} \in (\mathbb{Z}^d)^2 \mid \omega(v - \eta) > 0\}$ to be the edge set. The support of ω represents the communication range from any vertex in \mathbb{Z}^d . Typically, we can take

$$\omega = \frac{1}{2d} \sum_{i=1}^d (\mathbf{1}_{e_i} + \mathbf{1}_{-e_i}). \quad (6)$$

This corresponds to allowing only nearest neighbors communicate in \mathbb{Z}^d . In this case, the underlying graph is the standard lattice on \mathbb{Z}^d . The triangular lattice in dimension 2 can be obtained by taking

$$\omega = \frac{1}{6} (\mathbf{1}_{(1,0)} + \mathbf{1}_{(-1,0)} + \mathbf{1}_{(0,1)} + \mathbf{1}_{(0,-1)} + \mathbf{1}_{(1,1)} + \mathbf{1}_{(-1,-1)}). \quad (7)$$

See Figure 3 for drawings of these lattices.

²We use the notation $u_n \underset{n \rightarrow \infty}{\sim} v_n$ to mean that u_n/v_n converges to 1 as $n \rightarrow \infty$.

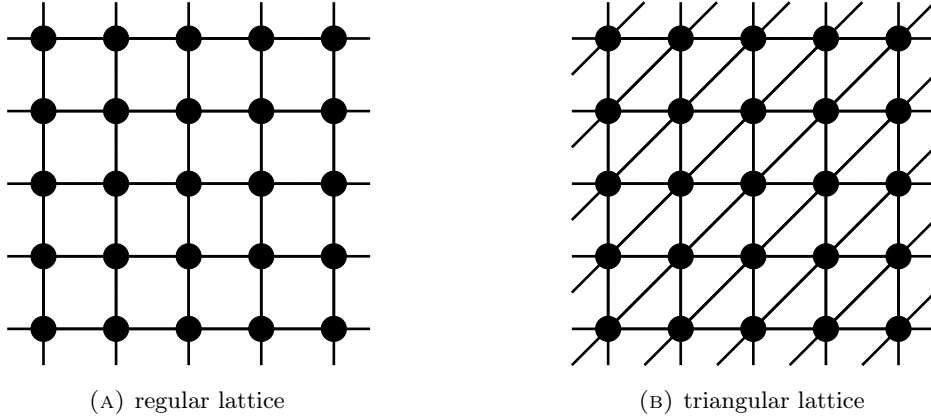


FIGURE 3. Different graph structures can be obtained from the vertex set \mathbb{Z}^2 by varying ω and thus the edge set. The regular lattice is obtained by choosing ω as in (6) (see left figure) and the triangular lattice is obtained by choosing ω as in (7) (see right figure).

The synchronous simple gossip algorithm iterates the local averaging operation:

$$x_{n+1}(v) = (\omega * x_n)(v) = \sum_{\eta \in \mathbb{Z}^d} \omega(v - \eta) x_n(\eta), \quad v \in \mathbb{Z}^d. \quad (8)$$

Note that this equation is (1) in the case where the averaging weights $W_{v-\eta} = \omega(v - \eta)$ are translation-invariant.

Throughout this paper, we assume that ω is centered, i.e.,

$$\sum_{v \in \mathbb{Z}^d} \omega(v) v = 0,$$

and that the covariance matrix

$$Q = \sum_{v \in \mathbb{Z}^d} \omega(v) v v^\top, \quad (9)$$

has full rank. Heuristically, this ensures that we do not have a drift term, and that we average in all directions. We note that our approach can still obtain similar results without these assumptions, however,

Goal. In the gossip problem, we give a real value $x_0(v)$ to each one of the nodes, and the goal is to compute the average of those values. In this paper, we consider graphs with an infinite number of nodes; the average of the infinite number of initial values is ill-defined. Thus the gossip problem is misspecified in this context. However, the gossip iterations like (1) make sense even on an infinite graph. In the following, we study the scaling limits of these iterations. This provides an intuition on the behavior on large but finite graphs.

More precisely, on \mathbb{Z}^d , we study the decay to 0 of the gossip algorithms when initialized from $x_0 = \mathbf{1}_0$, the vector with entry $\mathbf{1}_0(0) = 1$ and all other entries $\mathbf{1}_0(v)$, $v \in \mathbb{Z}^d \setminus \{0\}$, equal to 0. By analogy with PDEs, the solution we obtain with this initial condition is the *fundamental solution* of the gossip iterations; the solutions for other initializations x_0 can be obtained by convolution x_0 with the fundamental solution.

3. HEURISTIC DERIVATIONS AND DISCUSSIONS

Before we go into the rigorous convergence results, we would like to provide a heuristic derivation. While these are not proofs — they fail whenever the solution is not sufficiently smooth for the required Taylor expansions — we find these provide very strong intuition as to why the gossip algorithms converge to the corresponding PDEs.

3.1. Scaling limit of the simple gossip algorithm to the heat equation. Through a change of variables in (8), we rewrite the synchronous simple gossip algorithm iterates as

$$x_{n+1}(v) = \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) x_n(v - \eta), \quad v \in \mathbb{Z}^d. \quad (10)$$

Let $\Delta t, \Delta y > 0$ denote two scaling parameters. For $t \in (\Delta t)\mathbb{N} = \{(\Delta t)n, n \in \mathbb{N}\}$ and $y \in (\Delta y)\mathbb{Z}^d = \{(\Delta y)v, v \in \mathbb{Z}^d\}$, we define the scaled field

$$u(t, y) = x_{\frac{t}{\Delta t}} \left(\frac{y}{\Delta y} \right).$$

The iteration (10) can be reformulated in terms of u :

$$u(t + \Delta t, y) = \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) u(t, y - (\Delta y)\eta).$$

We now show that under a proper scaling for $\Delta t, \Delta y \rightarrow 0$, the above equation converges to a PDE in u . Recall that $\sum_{v \in \mathbb{Z}^d} \omega(v) = 1$, thus

$$u(t + \Delta t, y) - u(t, y) = \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) [u(t, y - (\Delta y)\eta) - u(t, y)].$$

Before we take $\Delta t, \Delta y \rightarrow 0$, we first Taylor expand the differences:

$$\begin{aligned} u(t + \Delta t, y) - u(t, y) &= (\Delta t) \partial_t u + o(\Delta t), \\ u(t, y - (\Delta y)\eta) - u(t, y) &= -(\Delta y) \langle \nabla_y u, \eta \rangle + \frac{(\Delta y)^2}{2} \langle \eta, (\nabla_y^2 u) \eta \rangle + o((\Delta y)^2), \end{aligned}$$

where all derivatives are taken in (t, y) . Note that we make a second-order expansion in space: this is due to the fact that the first-order terms cancel below. We obtain

$$\begin{aligned} &(\Delta t) \partial_t u + o(\Delta t) \\ &= -(\Delta y) \left\langle \nabla_y u, \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) \eta \right\rangle + \frac{(\Delta y)^2}{2} \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) \langle \eta, (\nabla_y^2 u) \eta \rangle + o((\Delta y)^2). \end{aligned}$$

As ω is centered, the first term of the right-hand side is zero. Moreover, we can use (9) to rewrite

$$\sum_{\eta \in \mathbb{Z}^d} \omega(\eta) \langle \eta, (\nabla_y^2 u) \eta \rangle = \text{Tr}(Q \nabla_y^2 u) = \nabla_y \cdot (Q \nabla_y u).$$

This gives us

$$(\Delta t) \partial_t u + o(\Delta t) = \frac{(\Delta y)^2}{2} \nabla_y \cdot (Q \nabla_y u) + o((\Delta y)^2).$$

Finally, we choose the scaling $\Delta t = (\Delta y)^2$ and by identifying the highest-order terms, we obtain the scaling to the heat equation in the limit as $\Delta t, \Delta y \rightarrow 0$

$$\partial_t u = \frac{1}{2} \nabla_y \cdot (Q \nabla_y u).$$

Here, Q quantifies the potential anisotropy of the diffusion. In the case of the standard grid (6), we have $Q = \frac{1}{d} \text{Id}$ and thus we obtain an isotropic heat equation $\partial_t u = \frac{1}{2d} \Delta_y u$, where Δ_y denotes the Laplacian in the variable y .

3.2. Second-order iteration scaling to the Euler–Poisson–Darboux equation. We now consider second-order iterations of the form

$$x_{n+1}(v) = a_n \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) x_n(v - \eta) + b_n x_n(v) - c_n x_{n-1}(v). \quad (11)$$

We impose $a_n + b_n - c_n = 1$ so that the sum of the coordinates of the vectors x_n remains constant. We show that, under specific asymptotics for a_n, b_n, c_n , the iteration (11) scales to the EPD equation. As in Section 3.1, we introduce scaling parameters $\Delta t, \Delta y > 0$ and the rescaled iterates

$$u(t, y) = x_{\frac{t}{\Delta t}} \left(\frac{y}{\Delta y} \right).$$

The iteration (11) can be reformulated in terms of u :

$$u(t + \Delta t, y) = a_n \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) u(t, y - (\Delta y)\eta) + b_n u(t, y) - c_n u(t - \Delta t, y).$$

Subtracting $u(t, y)$ and using $a_n + b_n - c_n = 1$, we obtain

$$\begin{aligned} u(t + \Delta t, y) - u(t, y) &= a_n \sum_{\eta \in \mathbb{Z}^d} \omega(\eta) [u(t, y - (\Delta y)\eta) - u(t, y)] - c_n [u(t - \Delta t, y) - u(t, y)]. \end{aligned}$$

We make the Taylor expansions of u , but this time a second-order expansion in t is necessary:

$$\begin{aligned} u(t + \Delta t, y) - u(t, y) &= (\Delta t) \partial_t u + \frac{(\Delta t)^2}{2} \partial_{tt} u + o(\Delta t), \\ u(t - \Delta t, y) - u(t, y) &= -(\Delta t) \partial_t u + \frac{(\Delta t)^2}{2} \partial_{tt} u + o(\Delta t), \\ u(t, y - (\Delta y)\eta) - u(t, y) &= -(\Delta y) \langle \nabla_y u, \eta \rangle + \frac{(\Delta y)^2}{2} \langle \eta, (\nabla_y^2 u) \eta \rangle + o((\Delta y)^2). \end{aligned}$$

We obtain

$$\frac{(\Delta t)^2}{2} (1 + c_n) \partial_{tt} u + (\Delta t) (1 - c_n) \partial_t u = a_n \frac{(\Delta y)^2}{2} \nabla_y \cdot (Q \nabla_y u).$$

To have the scaling to the Euler–Poisson–Darboux (EPD) equation, we take $\Delta t = \Delta y$, and

$$a_n \xrightarrow{n \rightarrow \infty} 2, \quad c_n = 1 - \frac{d+1}{n} + o\left(\frac{1}{n}\right). \quad (12)$$

Indeed, as $t = n\Delta t$, we have $1 - c_n \sim \frac{d+1}{t} \Delta t$ and thus

$$\frac{(\Delta t)^2}{2} (2 + o(1)) \partial_{tt} u + (\Delta t)^2 \left(\frac{d+1}{t} + o(1) \right) \partial_t u = (2 + o(1)) \frac{(\Delta y)^2}{2} \nabla_y \cdot (Q \nabla_y u),$$

thus by choosing the scaling $\Delta t = \Delta y$ and identifying highest-order terms, we obtain the EPD equation in the limit as $\Delta t, \Delta y \rightarrow 0$:

$$\partial_{tt} u + \frac{d+1}{t} \partial_t u = \nabla_y \cdot (Q \nabla_y u).$$

Note that there is the implicit condition $b_n \rightarrow 0$ implied by (12) as $a_n + b_n - c_n = 1$.

Different scalings. Note that in this section, the scaling is $\Delta t = \Delta y$ while for the simple gossip, the scaling is $\Delta t = (\Delta y)^2$. This is another illustration that the iteration of this section diffuses faster: to scale to a non-degenerate object, it needs go through a higher order rescaling in space.

3.3. Probabilistic interpretation. For the sake of mathematical curiosity, let us make an aside on the probabilistic interpretations of the heat equation and of the EPD equation. It is well-known that the heat equation represents the evolution of the probability density function of Brownian motion in \mathbb{R}^d [Le Gall, 2018]. While stochastic representation for wave equations remains largely an open problem [Dalang et al., 2008, Chatterjee, 2013], it is known that the solutions of (4) such that $\partial_t u(0, \cdot) = 0$ represent the evolution of a persistent random walk: $u(t, \cdot)$ is the probability density function of a random walker in \mathbb{R}^d , that moves according to a fixed unit speed, and, at a time-dependent Poisson rate $a(t) = (d+1)/(2t)$, resamples the direction of its speed uniformly over the unit sphere (see [Kac, 1974] for the seminal related work on the telegrapher’s equation, and for instance [Kaplan, 1964] or [Garra and Orsingher, 2014] for this precise statement). Thus, the EPD equation (4) is the density of a persistent random walk that gets more and more persistent over time. The rate $a = (d+1)/(2t)$ of the speed resampling is chosen so that the law of the random walker is uniform on the ball of radius t around the initial point. Note that as the random walker has unit speed, it can not be at a distance from the starting point larger than the elapsed time t .

This probabilistic point of view gives further credence to the high-level idea that acceleration is achieved by giving inertia to the gossiped information.

3.4. Relation to the Jacobi polynomial iteration. For the convenience of the reader, we recall here the Jacobi polynomial iteration introduced by Berthier et al. [2020] to accelerate gossip algorithms:

$$x_1 = a_0\omega * x_0 + b_0x_0, \quad x_{n+1} = a_n\omega * x_n + b_nx_n - c_nx_{n-1}, \quad (13)$$

$$a_0 = \frac{d+4}{2(2+d)}, \quad b_0 = \frac{d}{2(2+d)}, \quad (14)$$

$$a_n = \frac{(2n+d/2+1)(2n+d/2+2)}{2(n+1+d/2)^2}, \quad b_n = \frac{d^2(2n+d/2+1)}{8(n+1+d/2)^2(2n+d/2)}, \quad (15)$$

$$c_n = \frac{n^2(2n+d/2+2)}{(n+1+d/2)^2(2n+d/2)}, \quad n \geq 1. \quad (16)$$

As explained in [Berthier et al., 2020], this iteration is associated to the Jacobi polynomials $P_n^{(\alpha,\beta)}$ with parameters $\alpha = d/2$ and $\beta = 0$. It is of the form (11) with coefficients satisfying (12). Thus the Jacobi polynomial iteration scales to the EPD equation. However, note the large difference between the approaches of [Berthier et al., 2020] and this paper: in [Berthier et al., 2020], the authors use the geometry of the graph to approximate the spectrum of the gossip problem and design a polynomial-based method adapted to this approximate spectrum; in this paper, we also use the geometry of the graph but to view gossip algorithms as PDEs when rescaled. It is remarkable that the two approaches lead to similar results.

The PDE perspective enriches our understanding of the Jacobi polynomial iteration. For instance, one can explore the effect of using the Jacobi polynomial $P_n^{(\alpha,\beta)}$ for a different value than $(\alpha, \beta) = (d/2, 0)$ used in the Jacobi polynomial iteration. The formula [Berthier et al., 2020, Equation (SM6.2)] gives the expression of the recurrence coefficients of the Jacobi polynomial iteration in this general case; it follows that

$$a_n \xrightarrow{n \rightarrow \infty} 2, \quad c_n = 1 - \frac{2\alpha + 1}{n} + o\left(\frac{1}{n}\right),$$

thus, repeating the computations of Section 3.2, the iteration converges to the more general EPD equation (17). Consider its fundamental solution (18). If $\alpha > d/2$, the mass concentrates at the center of the ball of radius t . On the contrary, if $\alpha < d/2$, the mass concentrates at the edge of the ball. Both effects are undesirable as uniform averaging is the optimal strategy. These effects are simulated in Figure 4.

3.5. Related numerical PDE methods and problems. Both the simple gossip algorithm and Jacobi polynomial iteration can be interpreted as numerical discretizations of their limiting PDEs, which has a rich literature studying both theory and implementation [Celia and Gray, 1992]. To our best knowledge, the only theoretical treatment of numerical discretization errors for the EPD equation is due to Genis [1984], where the author studied a large class of finite element methods (FEM). Here we note the local averaging operation used in gossip algorithms can be seen a special case of FEM, where averaging is approximating the Laplacian operator.

However, we emphasize two key differences. Firstly, FEM generally does not preserve total mass, which is a key property of gossip algorithms. In this sense, the Jacobi polynomial iteration is closer to a finite volume method (FVM) [LeVeque, 2002], which typically preserves a quantity of interest such as volume. Secondly, most numerical analysis approaches require some level of regularity for the initial condition, which cannot apply to the fundamental solution (with Dirac delta initial condition). In the case of Genis [1984], all of the error bounds were in terms of an equivalent fractional Sobolev norm of the initial condition, which is unbounded for the Dirac delta. Instead, we were able to avoid regularity issues by studying the discretization error in the Fourier domain, see Section 4 and Appendix B.

There is a body of empirical work on numerical methods for EPD and related equations (for example Glowinski and Quaini [2013]), however these generally cannot be transformed to a locally communicating algorithm, which is important to build practical gossip algorithms. We also mention a Hamilton–Jacobi equation of the type $\partial_t u = \|\nabla u\|$, which has a similar fundamental solution of the form $\mathbf{1}_{\{\|y\| \leq t\}}$. It is well known this equation can be solved by the fast sweeping method [Zhao, 2005] and the fast marching method [Sethian, 1996]. However, also due to the lack of normalization

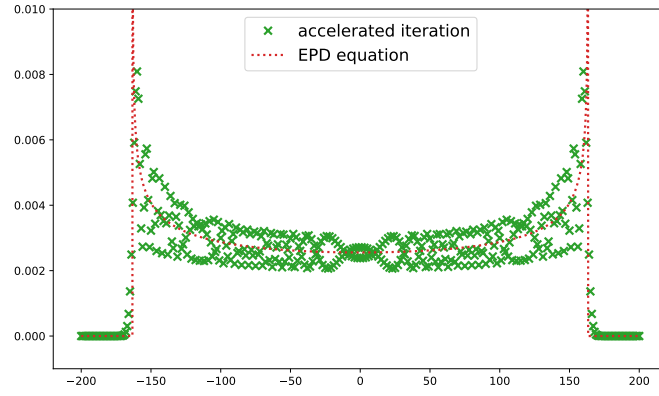
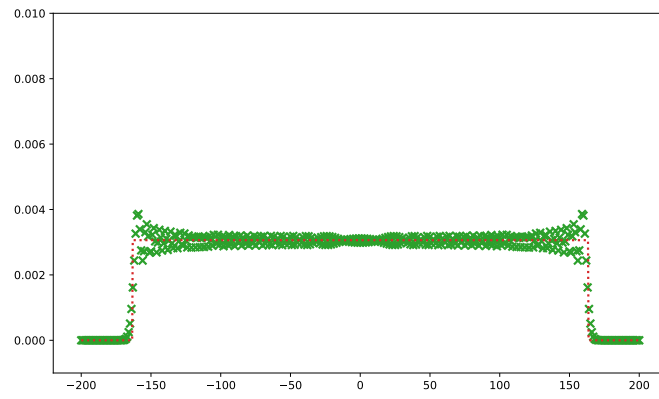
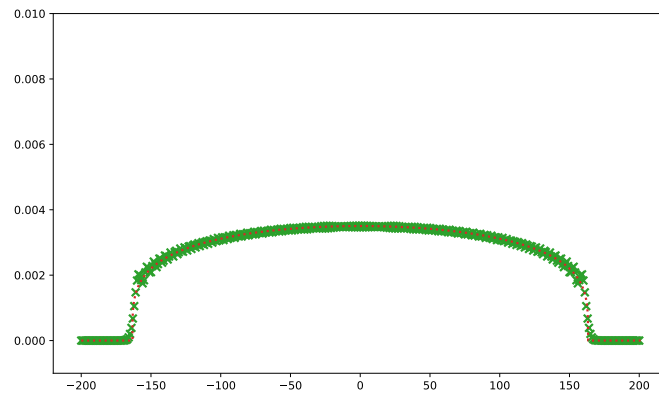
(A) $\alpha = \frac{1}{4} < \frac{1}{2} = \frac{d}{2}$ (B) $\alpha = \frac{1}{2} = \frac{d}{2}$ (C) $\alpha = \frac{3}{4} > \frac{1}{2} = \frac{d}{2}$

FIGURE 4. Same simulation as in Figure 1(C), but we now study the effect of varying the parameter α of the Jacobi polynomial iteration. Varying α also changes the fundamental solution (18) of the EPD equation (17).

(to preserve mass), and the fact these algorithms are not locally communicating, they are not good candidates for gossip algorithms.

3.6. Open problems: other geometries, stochastic case. An important limitation of this work is that we only study synchronous gossip on a regular lattice. It is natural to ask what could happen in an asynchronous setting, or when the graph is microscopically perturbed (percolation graph, random geometric graph, etc).

For the simple gossip, or equivalently, for the random walk or for heat diffusion, answering this question is the subject of the field of homogenization, see, e.g., [Armstrong et al., 2019, Armstrong and Dario, 2018, Biskup, 2011]. The heuristic is that on a large scale and for long diffusion times, microscopic fluctuations of the connectivity (in space and in time) are homogenized: the process scales to an homogeneous diffusion with some constant effective diffusion matrix Q .

Our work raises the following question: is there homogenization for the EPD equation? Berthier et al. [2020] prove that there is some robustness of the Jacobi polynomial iteration to microscopic details of the graphs, as the rates are the same on all graphs of spectral dimension d . However, we do not know if the process scales to the same limit on those graphs.

4. RIGOROUS CONVERGENCE RESULTS

In this section, we provide rigorous justification to the heuristic derivations of Section 3. In Section 4.1, we start with the convergence of simple gossip to the heat equation. This case is simple as it is equivalent to the central limit theorem: we obtain a weak convergence result. A stronger convergence result, of local type, is deduced from the local central limit theorem.

Section 4.1 illustrates that two types of convergence are possible: weak and local. In Section 4.2, we prove analog results for the convergence of the Jacobi polynomial iteration to the EPD equation. We restrict ourselves to the Jacobi polynomial iteration—and not to any method satisfying (12)—for technical reasons: we use fine asymptotic properties of the Jacobi polynomials. However, we end this section with a remark on why we conjecture the same scaling for all iterations satisfying (12).

In Section 4.3, we apply the local convergence result to obtain convergence rates of the Jacobi polynomial iteration. These rates are sharp up to constants.

4.1. Simple gossip and the heat equation. Consider the simple gossip iteration

$$x_0 = \mathbf{1}_0, \quad x_{n+1} = \omega * x_n.$$

The iteration x_n can be interpreted as the probability density function of a random walk on \mathbb{Z}^d , initialized from 0, with increments of law ω . As ω is centered, the random walk is unbiased; the matrix Q is the covariance of the increments. The asymptotic law x_n is described by the central limit theorems: here, we interpret them with our notations. Let $u(t, y)$ denote the fundamental solution (3) of the heat equation (2). We denote δ_y the Dirac mass at $y \in \mathbb{R}^d$.

Theorem 1 (Central Limit Theorem, see, e.g., [Billingsley, 2008]). *We have the following weak convergence in the space of positive measures: for any $t \geq 0$,*

$$\sum_{v \in \mathbb{Z}^d} x_{\lfloor t/\varepsilon^2 \rfloor}(v) \delta_{\varepsilon v} \xrightarrow{\varepsilon \rightarrow 0} u(t, y) dy.$$

A stronger local result holds assuming that ω is aperiodic, i.e., that the random walk with increments ω is an aperiodic Markov chain on \mathbb{Z}^d [Billingsley, 2008, Section 8]. For instance, the vector ω of Equation (7), corresponding to the triangular lattice, is aperiodic, while the vector ω of Equation (6), corresponding to the regular grid, is not.

Theorem 2 (Local Central Limit Theorem, [Gnedenko, 1948]). *Assume that ω is aperiodic. Then*

$$\sup_{v \in \mathbb{Z}^d} |x_n(v) - u(n, v)| = o\left(\frac{1}{n^{d/2}}\right) \quad \text{as } n \rightarrow \infty.$$

A pedagogical introduction to the local central limit theorem is provided by Curien [2020]. The beauty of the local central limit theorem is that no rescaling is required: we simply discretize the heat equation in time and space.

4.2. The Jacobi polynomial iteration and the Euler–Poisson–Darboux equation. We now give analogs of Theorems 1 and 2 for the convergence of the Jacobi polynomial iteration to the EPD equation. Let x_n denote the iterates of the Jacobi polynomial iteration (13)–(16) initialized from $x_0 = \mathbf{1}_0$ and $u(t, y)$ the fundamental solution (5) of the EPD equation (4).

Assumptions. In this section, we assume that ω is symmetric ($\omega(-v) = \omega(v)$) and aperiodic. While the aperiodicity assumption is clearly necessary for Theorem 4 to hold, we do not know if these assumptions are necessary otherwise.

Theorem 3 (Weak Convergence). *We have the following weak convergence in the space of signed measures: for all $t > 0$,*

$$\sum_{v \in \mathbb{Z}^d} x_{\lfloor t/\varepsilon \rfloor}(v) \delta_{\varepsilon v} \xrightarrow{\varepsilon \rightarrow 0} u(t, y) dy.$$

Theorem 4 (Local Convergence). *Let $\psi(x) := \prod_{i=1}^d \frac{\sin(\pi x_i)}{\pi x_i}$. Then we have that*

$$\sum_{v \in \mathbb{Z}^d} (x_n(v) - (u(n, \cdot) * \psi)(v))^2 = o\left(\frac{1}{n^d}\right), \quad \text{as } n \rightarrow \infty.$$

The two theorems are proved in Appendix B.

Remark 1. *We believe that Theorem 4 should hold without the convolution with ψ , namely, $\sum_{v \in \mathbb{Z}^d} (x_n(v) - (u(n, v))(v))^2 = o\left(\frac{1}{n^d}\right)$, especially in light of Theorem 3. However, there are some technical challenges with the proof that we are unable to resolve.*

Remark 2. *The statements of Theorems 3 and 4 and their proofs can be easily adapted to study the Jacobi polynomial iterations for other parameters $(\alpha, \beta) \neq (d/2, 0)$, as long as $\alpha > d/2 - 1/2$ and $\beta \leq \alpha$. In this case, the limiting PDE depends on α . We have the convergence to the fundamental solution (18) of the general EPD equation (17).*

Remark 3 (Extension beyond the Jacobi polynomial iteration). *Our theorems are stated for the Jacobi polynomial iteration only for a technical reason: the proofs are based on well-known asymptotic properties of the Jacobi polynomials, stated in Proposition 3. We conjecture that all other sequences of polynomials with recursion coefficients satisfying (12) also satisfy the same properties: this would prove the scaling to the EPD equation for all second-order gossip algorithms satisfying (12).*

This conjecture is supported by Aptekarev [1993]: he shows that a sequence of orthogonal polynomial must satisfy the Mehler–Heine asymptotics (Proposition 3.(1)) provided that the recurrence coefficients of the polynomials satisfy some conditions that resemble (12). Interestingly, he explains that the asymptotics of the recurrence coefficients are related to the shape of the orthogonality measure of the associated orthogonal polynomials near 1: this links the approaches of [Berthier et al., 2020] and this paper.

4.3. Application: sharp rates of the Jacobi polynomial iteration on \mathbb{Z}^d . In this section, we apply Theorem 4 to obtain sharp rates for the Jacobi polynomial iteration.

Corollary 1. *Assume that ω is symmetric and aperiodic. Let x_n be the iterates of the Jacobi polynomial iteration (13)–(16), initialized at $x_0 = \mathbf{1}_0$. Then we have the asymptotic equivalence*

$$\sum_{v \in \mathbb{Z}^d} x_n(v)^2 \underset{n \rightarrow \infty}{\sim} \frac{1}{(\det Q)^{1/2} |B(0, 1)|} \frac{1}{n^d},$$

where $|B(0, 1)| = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$ is the volume of the Euclidean unit ball in dimension d .

We compare with Berthier et al. [2020, Theorem SM7.1]. Here our theorem applies only to regular lattices, while the previous result applies to all graphs of spectral dimension d ; but we obtain an asymptotic equivalent, while the previous result gave only the exponent in n . In Figure 5, we compare the two asymptotic equivalent quantities in the case of the Jacobi polynomial iteration on the triangular lattice. Note that similarly, one could obtain sharp rates for simple gossip from the local central limit Theorem 2.

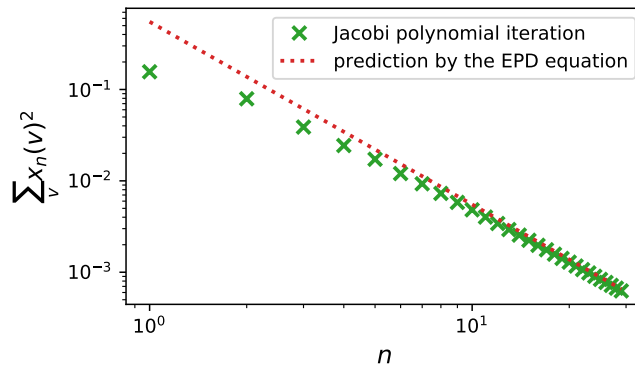


FIGURE 5. Comparison between the empirical error $\sum_{v \in \mathbb{Z}^d} x_n(v)^2$ and the asymptotic rate predicted by Corollary 1. Here, x_n are the iterates of the Jacobi polynomial iteration on the triangular lattice (7). Note that the corollary predicts sharply not only the scaling in n (the asymptotic slope in the logarithmic plot) but also the leading constant (the intercept of the asymptotic line).

ACKNOWLEDGEMENTS

We would like to thank the Saint-Flour Probability Summer School for the opportunity to meet and initiate this project. We would also like to thank Patric Bonnier, Christina Christara, Dejan Slepčev for insightful discussions. ML is supported by the Ontario Graduate Scholarship and the Vector Institute.

REFERENCES

- A. I. Aptekarev. Asymptotics of orthogonal polynomials in a neighborhood of the endpoints of the interval of orthogonality. *Sbornik: Mathematics*, 76(1):35, 1993.
- S. Armstrong and P. Dario. Elliptic regularity and quantitative homogenization on percolation clusters. *Communications on Pure and Applied Mathematics*, 71(9):1717–1849, 2018.
- S. Armstrong, T. Kuusi, and J.-C. Mourrat. *Quantitative stochastic homogenization and large-scale regularity*, volume 352. Springer, 2019.
- M. Assran, A. Aytekin, H. R. Feyzmahdavian, M. Johansson, and M. G. Rabbat. Advances in asynchronous parallel and distributed optimization. *Proceedings of the IEEE*, 108(11):2013–2031, 2020.
- L. Baez-Duarte. Central limit theorem for complex measures. *Journal of Theoretical Probability*, 6(1):33–56, 1993.
- R. Berthier, F. Bach, and P. Gaillard. Accelerated gossip in networks of given dimension using Jacobi polynomial iterations. *SIAM Journal on Mathematics of Data Science*, 2(1):24–47, 2020.
- P. Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- M. Biskup. Recent progress on the random conductance model. *Probability Surveys*, 8, 2011.
- S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.
- D. Bresters. On the equation of Euler–Poisson–Darboux. *SIAM Journal on Mathematical Analysis*, 4(1):31–41, 1973.
- M. Cao, D. A. Spielman, and E. M. Yeh. Accelerated gossip algorithms for distributed computation. In *44th Annual Allerton Conference on Communication, Control, and Computation*, pages 952–959, 2006.
- M. A. Celia and W. G. Gray. *Numerical methods for differential equations: fundamental concepts for scientific and engineering applications*. Pearson College Division, 1992.
- S. Chatterjee. Stochastic solutions of the wave equation. *arXiv preprint arXiv:1306.2382*, 2013.
- N. Curien. Random walks and graphs, 2020. lecture notes, available at <https://www.imo.universite-paris-saclay.fr/~curien/enseignement.html>.

- R. Dalang, C. Mueller, and R. Tribe. A feynman-kac-type formula for the deterministic and stochastic wave equations and other pde's. *Transactions of the American Mathematical Society*, 360(9):4681–4703, 2008.
- G. Darboux. *Leçons sur la théorie générale des surfaces*. 1896.
- A. Dimakis, A. Sarwate, and M. Wainwright. Geographic gossip: efficient averaging for sensor networks. *IEEE Transactions on Signal Processing*, 56(3):1205–1216, 2008.
- A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847–1864, 2010.
- L. Euler. *Institutiones calculi integralis*, vol iii, Petropoli. 1770.
- L. Evans. Partial differential equations. *Graduate studies in mathematics*, 19(2), 1998.
- M. Even, R. Berthier, F. Bach, N. Flammarion, P. Gaillard, H. Hendrikx, L. Massoulié, and A. Taylor. A continuized view on Nesterov acceleration for stochastic gradient descent and randomized gossip. *arXiv preprint arXiv:2106.07644*, 2021.
- R. Garra and E. Orsingher. Random flights related to the Euler-Poisson-Darboux equation. *arXiv preprint arXiv:1411.0648*, 2014.
- A. M. Genis. On finite element methods for the Euler–Poisson–Darboux equation. *SIAM journal on numerical analysis*, 21(6):1080–1106, 1984.
- R. Glowinski and A. Quaini. When Euler-Poisson-Darboux meets Painlevé and Bratu: On the numerical solution of nonlinear wave equations. *Methods and Applications of Analysis*, 20(4):405–424, 2013.
- B. V. Gnedenko. On a local limit theorem of the theory of probability. *Uspekhi Matematicheskikh Nauk*, 3(3):187–194, 1948.
- M. Kac. A stochastic model related to the telegrapher's equation. *The Rocky Mountain Journal of Mathematics*, 4(3):497–509, 1974.
- S. Kaplan. Differential equations in which the Poisson process plays a role. *Bulletin of the American Mathematical Society*, 70(2):264–268, 1964.
- J.-F. Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Springer, 2018.
- R. J. LeVeque. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.
- S. Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.
- Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.
- S. D. Poisson. *Mémoire sur l'intégration des équations linéaires aux différences partielles*. J. de l'Ecole Polytechnique, 1823.
- P. Rebeschini and S. Tatikonda. Accelerated consensus via min-sum splitting. In *Advances on Neural Information Processing Systems*, 2017.
- S. Sardellitti, M. Giona, and S. Barbarossa. Fast distributed average consensus algorithms based on advection-diffusion processes. *IEEE Transactions on Signal Processing*, 58(2):826–842, 2010.
- J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595, 1996.
- G. Szegő. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939.
- B. Szorenyi, R. Busa-Fekete, I. Hegedus, R. Ormándi, M. Jelasity, and B. Kégl. Gossip-based distributed stochastic bandit algorithms. In *International Conference on Machine Learning*, pages 19–27, 2013.
- H. Zhao. A fast sweeping method for eikonal equations. *Mathematics of computation*, 74(250):603–627, 2005.

APPENDIX A. THE EULER–POISSON–DARBOUX (EPD) EQUATION

The EPD equation is the partial differential equation

$$\partial_{tt}u + \frac{2\alpha + 1}{t}\partial_t u = \nabla_y \cdot (Q\nabla_y u). \quad (17)$$

Posing a rigorous framework for solving this equation is subtle because there is a diverging coefficient $\frac{2\alpha+1}{t}$ as $t \rightarrow 0$. Moreover, we see below that fundamental solutions are irregular; they are defined in a weak sense. Thankfully, we do not have to bother with these technical details as our rigorous results only require to know the expression of the fundamental solution of the EPD equation (Proposition 1) and its Fourier transform (Proposition 2). These expressions are given by Bresters [1973] in the case $Q = \text{Id}$; here, we easily extend the expressions for a general matrix Q .

Proposition 1. *The fundamental solution of the EPD equation, i.e., the solution initialized from $u(0, \cdot) = \delta_0$, $\partial_t u(0, \cdot) = 0$, is*

$$u(t, y) = \frac{\Gamma(\alpha + 1)}{\pi^{d/2}\Gamma(\alpha + 1 - d/2)(\det Q)^{1/2} t^{2\alpha}} (t^2 - \langle y, Q^{-1}y \rangle)_+^{\alpha-d/2}, \quad (18)$$

where $(\cdot)_+$ denotes the positive part of a real number.

The case $\alpha = d/2$ is particularly important to us; in this case we recover (5) from (18).

Proof of Proposition 1. In the case $Q = \text{Id}$, the solution is given by Bresters [1973, Equation (3.4)]:

$$u(t, y) = \frac{\Gamma(\alpha + 1)}{\pi^{d/2}\Gamma(\alpha + 1 - d/2) t^{2\alpha}} (t^2 - \|y\|^2)_+^{\alpha-d/2}.$$

In the general case, consider $v(t, y) = u(t, Q^{1/2}y)(\det Q)^{1/2}$. Computations give that $v(t, y)$ is the fundamental solution of the EPD equation (17) with $Q = \text{Id}$, thus

$$u(t, Q^{1/2}y)(\det Q)^{1/2} = v(t, y) = \frac{\Gamma(\alpha + 1)}{\pi^{d/2}\Gamma(\alpha + 1 - d/2) t^{2\alpha}} (t^2 - \|y\|^2)_+^{\alpha-d/2}.$$

This gives the desired formula. □

Proposition 2. *The Fourier transform in space of the fundamental solution (18) is*

$$\widehat{u}(t, \xi) = \int_{\mathbb{R}^d} dy e^{i\langle \xi, y \rangle} u(t, y) = 2^\alpha \Gamma(\alpha + 1) \langle \xi, Q\xi \rangle^{-\alpha/2} t^{-\alpha} J_\alpha \left(t \langle \xi, Q\xi \rangle^{1/2} \right),$$

where J_α denotes the Bessel function of the first kind of order α [Szegő, 1939, Section 1.71].

Proof. In the case $Q = \text{Id}$, the result is given by Bresters [1973, Equation (3.1)]:

$$\widehat{u}(t, \xi) = 2^\alpha \Gamma(\alpha + 1) \|\xi\|^{-\alpha} t^{-\alpha} J_\alpha(t\|\xi\|).$$

In the general case, $v(t, y) = u(t, Q^{1/2}y)(\det Q)^{1/2}$ is a solution of the EPD equation (17) with $Q = \text{Id}$. Moreover,

$$\begin{aligned} \widehat{v}(t, \xi) &= \int_{\mathbb{R}^d} dy e^{i\langle \xi, y \rangle} v(t, y) \\ &= (\det Q)^{1/2} \int_{\mathbb{R}^d} dy e^{i\langle \xi, y \rangle} u(t, Q^{1/2}y). \end{aligned}$$

In the last integral, we change the variable to $x = Q^{1/2}y$. Then $dx = \det(Q^{1/2})dy = (\det Q)^{1/2}dy$.

$$\begin{aligned} \widehat{v}(t, \xi) &= \int_{\mathbb{R}^d} dx e^{i\langle \xi, Q^{-1/2}x \rangle} u(t, x) \\ &= \widehat{u}(t, Q^{-1/2}\xi). \end{aligned}$$

Thus

$$\widehat{u}(t, \xi) = \widehat{v}(t, Q^{1/2}\xi) = 2^\alpha \Gamma(\alpha + 1) \langle \xi, Q\xi \rangle^{-\alpha/2} t^{-\alpha} J_\alpha \left(t \langle \xi, Q\xi \rangle^{1/2} \right).$$

□

APPENDIX B. PROOF OF THEOREMS 3 AND 4

In the following, we use the notation of Berthier et al. [2020]: $P_n^{(\alpha,\beta)}$ denotes the classical Jacobi polynomials, orthogonal for the Jacobi measure $d\sigma(\lambda) = (1-\lambda)^\alpha(1+\lambda)^\beta$ [Szegő, 1939]. We also denote $\pi_n^{(\alpha,\beta)} = P_n^{(\alpha,\beta)}/P_n^{(\alpha,\beta)}(1)$ the Jacobi polynomial rescaled so that $\pi_n^{(\alpha,\beta)}(1) = 1$.

Consider the algebra $(\ell^2(\mathbb{Z}^d), +, *)$ where $+$ denotes the classical addition over functions and $*$ denotes the convolution. As $\omega \in \ell^2(\mathbb{Z}^d)$, this algebra gives a meaning to the evaluation $P(\omega) \in \ell^2(\mathbb{Z}^d)$ of a polynomial P at ω . By its definition from [Berthier et al., 2020], the Jacobi polynomial iteration (13)-(16) can be written more compactly as

$$x_n = \pi_n^{(d/2,0)}(\omega).$$

Fourier analysis plays a central role in the proofs below. If $x \in \ell^2(\mathbb{Z}^d)$, the Fourier transform $\widehat{x} \in L^2([-\pi, \pi]^d)$ is defined as $\widehat{x}(\xi) = \sum_{v \in \mathbb{Z}^d} e^{i\langle \xi, v \rangle} x(v)$. Consider the algebra $(L^2([-\pi, \pi]^d), +, \cdot)$ where $+$ denotes the classical addition over functions and \cdot denotes the pointwise multiplication of functions. As $\widehat{\omega} \in L^2([-\pi, \pi]^d)$, this algebra gives a meaning to the evaluation $P(\widehat{\omega}) \in L^2([-\pi, \pi]^d)$ of a polynomial P at $\widehat{\omega}$.

The Fourier transform of a sum is the sum of the Fourier transforms, and the Fourier transform of a convolution is the pointwise product of the Fourier transforms, thus if P is a polynomial, $\widehat{P(\omega)} = P(\widehat{\omega})$. In particular, in the following, we analyze the Jacobi polynomial iteration by using the relation

$$\widehat{x}_n = \pi_n^{(\widehat{d/2,0})}(\omega) = \pi_n^{(d/2,0)}(\widehat{\omega}).$$

The proofs below use the following well-known results on Jacobi polynomials.

Proposition 3. (1) *(Mehler–Heine asymptotic) The Jacobi polynomials satisfy the following asymptotic at the edge of the orthogonality measure*

$$\lim_{n \rightarrow \infty} \pi_n^{(d/2,0)} \left(1 - \frac{z^2}{2n^2} \right) = 2^{d/2} \Gamma \left(\frac{d}{2} + 1 \right) z^{-d/2} J_{d/2}(z),$$

where $J_{d/2}$ denotes the Bessel function of the first kind of order $d/2$ [Szegő, 1939, Section 1.71]. The convergence is uniform for z in compact sets.

(2) *On the whole support of the orthogonality measure, we have the following bounds: there exists constants $C_1, C_2 > 0$ such that for all $n \geq 0$,*

$$\left| \pi_n^{(d/2,0)}(\lambda) \right| \leq \begin{cases} C_1 (\arccos |\lambda|)^{-d/2-1/2} n^{-d/2-1/2} & \text{if } |\lambda| \leq 1 - \frac{1}{n^2}, \\ C_2 & \text{otherwise.} \end{cases}$$

Proof. (1) Szegő [1939, Theorem 8.1.1] gives the Mehler–Heine asymptotic for the classical Jacobi polynomials $P_n^{(d/2,0)}$:

$$\lim_{n \rightarrow \infty} n^{-d/2} P_n^{(d/2,0)} \left(1 - \frac{z^2}{2n^2} \right) = 2^{d/2} z^{-d/2} J_{d/2}(z),$$

with uniform convergence for z in compact sets. As

$$\pi_n^{(d/2,0)} = \frac{P_n^{(d/2,0)}}{P_n^{(d/2,0)}(1)} = \frac{P_n^{(d/2,0)}}{\binom{n+d/2}{n}}, \quad \binom{n+d/2}{n} \sim \frac{n^{d/2}}{\Gamma(\frac{d}{2}+1)},$$

we obtain the desired formula.

(2) For $\lambda \geq 0$, this is only a reformulation of [Szegő, 1939, Theorem 7.32.2]. For $\lambda < 0$, we use the symmetry of the Jacobi polynomials $P_n^{(d/2,0)}(\lambda) = (-1)^n P_n^{(0,d/2)}(\lambda)$ [Szegő, 1939, Equation (4.1.3)] and use again [Szegő, 1939, Theorem 7.32.2]. □

B.1. Proof of Weak Convergence.

Proof of Theorem 3. Denote

$$\mu_{t,\varepsilon} = \sum_{v \in \mathbb{Z}^d} x_{[t/\varepsilon]}(v) \delta_{\varepsilon v}, \quad \mu_t = u(t, y) dy$$

The proof is based on [Baez-Duarte, 1993, Theorem 2.1], a variant of Lévy’s theorem for signed measures: in order to prove the weak convergence $\mu_{t,\varepsilon} \rightarrow \mu_t$ as $\varepsilon \rightarrow 0$, it is sufficient to check that

the family of measures $\mu_{t,\varepsilon}$, $\varepsilon > 0$ is tight, bounded in total variation, that we have the pointwise convergence of the Fourier transform $\widehat{\mu}_{t,\varepsilon}(\xi) = \int_{\mathbb{R}^d} \mu_{t,\varepsilon}(dy) e^{i(\xi,y)} \rightarrow \widehat{\mu}_t(\xi) = \int_{\mathbb{R}^d} \mu_t(dy) e^{i(\xi,y)}$ almost everywhere. These three conditions are checked below.

Tightness of $\mu_{t,\varepsilon}$, $\varepsilon > 0$. ω has a finite support, thus there exists $R > 0$ such that the support of ω is included in $B(0, R)$. Then for all $n \geq 0$, the support of $w^{*n} = w * \dots * w$ (with n terms) is included in $B(0, nR)$. The vector x_n is a linear combination of the w^{*l} for $l \leq n$, thus is also included in $B(0, nR)$. Finally, when rescaling by ε , the support of $\mu_{t,\varepsilon} = \sum_{v \in \mathbb{Z}^d} x_{\lfloor t/\varepsilon \rfloor}(v) \delta_{\varepsilon v}$ is included in $B(0, \varepsilon \lfloor t/\varepsilon \rfloor R) \subset B(0, tR)$. The latter set is independent of ε , thus the family of measures $\mu_{t,\varepsilon}$, $\varepsilon > 0$ is tight.

Boundedness of $\mu_{t,\varepsilon}$, $\varepsilon > 0$. Note that $\mu_{t,\varepsilon}(\mathbb{R}^d) = 1$, but as $\mu_{t,\varepsilon}$ is a signed measure, we need to show that the total mass $\|\mu_{t,\varepsilon}\| = |\mu_{t,\varepsilon}|(\mathbb{R}^d)$ of the total variation $|\mu_{t,\varepsilon}|$ is bounded independently of ε . By Hölder's inequality,

$$\|\mu_{t,\varepsilon}\| = \|x_{\lfloor t/\varepsilon \rfloor}\|_{l^1(\mathbb{Z}^d)} \leq |\text{Supp } x_{\lfloor t/\varepsilon \rfloor}|^{1/2} \|x_{\lfloor t/\varepsilon \rfloor}\|_{l^2(\mathbb{Z}^d)}^{1/2}, \quad (19)$$

where $|\text{Supp } x_{\lfloor t/\varepsilon \rfloor}|$ denotes the cardinal of the support of $x_{\lfloor t/\varepsilon \rfloor}$. As this support is included in $B(0, \lfloor t/\varepsilon \rfloor R)$, its cardinal can be bounded by the number of integer points in $B(0, \lfloor t/\varepsilon \rfloor R)$. This is dominated by ε^{-d} as $\varepsilon \rightarrow 0$. Thus

$$|\text{Supp } x_{\lfloor t/\varepsilon \rfloor}| = O(\varepsilon^{-d}).$$

We now bound the second term in (19), namely the norm $\|x_{\lfloor t/\varepsilon \rfloor}\|_{l^2(\mathbb{Z}^d)}$. By Plancherel identity,

$$\begin{aligned} \|x_n\|_{\ell^2(\mathbb{Z}^d)}^2 &= \left\| \pi_n^{(d/2,0)}(\omega) \right\|_{\ell^2(\mathbb{Z}^d)}^2 = \frac{1}{(2\pi)^d} \left\| \widehat{\pi_n^{(d/2,0)}(\omega)} \right\|_{L^2([-\pi, \pi]^d)}^2 \\ &= \frac{1}{(2\pi)^d} \left\| \pi_n^{(d/2,0)}(\widehat{\omega}) \right\|_{L^2([-\pi, \pi]^d)}^2 = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} d\xi \left| \pi_n^{(d/2,0)}(\widehat{\omega}(\xi)) \right|^2. \end{aligned}$$

Here, as ω is symmetric, $\widehat{\omega}(\xi)$ is real. We can use the bounds of Proposition 3.(2). We need to estimate $\widehat{\omega}(\xi)$. We use the following lemma.

Lemma 1. *As ω is aperiodic, there exists $\lambda > 0$ such that*

$$|\widehat{\omega}(\xi)| \leq 1 - \lambda \|\xi\|^2, \quad \xi \in [-\pi, \pi]^d.$$

This lemma is simple and proved by Curien [2020, Section 7.1]. We now return to our estimate of $\left| \pi_n^{(d/2,0)}(\widehat{\omega}(\xi)) \right|^2$.

- If $\|\xi\| \geq \frac{1}{\sqrt{\lambda n}}$, we have $|\widehat{\omega}(\xi)| \leq 1 - \lambda \|\xi\|^2 \leq 1 - \frac{1}{n^2}$. Thus by Proposition 3.(2),

$$\begin{aligned} \left| \pi_n^{(d/2,0)}(\widehat{\omega}(\xi)) \right| &\leq C_1 (\arccos |\widehat{\omega}(\xi)|)^{-d/2-1/2} n^{-d/2-1/2} \\ &\leq C_1 (\arccos(1 - \lambda \|\xi\|^2))^{-d/2-1/2} n^{-d/2-1/2}. \end{aligned}$$

- If $\|\xi\| < \frac{1}{\sqrt{\lambda n}}$, we can only say $\left| \pi_n^{(d/2,0)}(\widehat{\omega}(\xi)) \right| \leq C_2$.

Thus

$$\begin{aligned} \|x_n\|_{\ell^2(\mathbb{Z}^d)}^2 &= \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} d\xi \left| \pi_n^{(d/2,0)}(\widehat{\omega}(\xi)) \right|^2 \\ &\leq C_3 n^{-d-1} \int_{\{\|\xi\| \geq 1/(\sqrt{\lambda n})\}} d\xi (\arccos(1 - \lambda \|\xi\|^2))^{-d-1} + C_4 \int_{\{\|\xi\| < 1/(\sqrt{\lambda n})\}} d\xi \end{aligned}$$

where we use the notation C_i to denote constants independent of n . We use a spherical change of variables in the first integral:

$$\|x_n\|_{\ell^2(\mathbb{Z}^d)}^2 \leq C_5 n^{-d-1} \int_{1/(\sqrt{\lambda n})}^{\sqrt{d}\pi} dr r^{d-1} (\arccos(1 - \lambda r^2))^{-d-1} + C_6 n^{-d}. \quad (20)$$

As $r \rightarrow 0$, $\arccos(1 - \lambda r^2) \sim \sqrt{2\lambda} r$ and thus

$$r^{d-1} (\arccos(1 - \lambda r^2))^{-d-1} \sim \sqrt{2\lambda}^{-d-1/2} r^{-2}.$$

Thus $r^{d-1} (\arccos(1 - \lambda r^2))^{-d-1}$ is not integrable at 0. We thus have, as $n \rightarrow \infty$,

$$\int_{1/(\sqrt{\lambda n})}^{\sqrt{d}\pi} dr r^{d-1} (\arccos(1 - \lambda r^2))^{-d-1} \sim C_7 \int_{1/(\sqrt{\lambda n})}^{\sqrt{d}\pi} dr r^{-2} \sim C_8 n.$$

Putting back in (20), we obtain $\|x_n\|_{\ell^2(\mathbb{Z}^d)}^2 = O(n^{-d})$. Finally, getting back to (19), we obtain as $\varepsilon \rightarrow 0$

$$\|\mu_{t,\varepsilon}\| \leq |\text{Supp } x_{\lfloor t/\varepsilon \rfloor}|^{1/2} \|x_{\lfloor t/\varepsilon \rfloor}\|_{\ell^2(\mathbb{Z}^d)}^{1/2} = O(\varepsilon^{-d/2}) O\left(\left\lfloor \frac{t}{\varepsilon} \right\rfloor^{-d/2}\right) = O(1).$$

This shows that the family of measures $\mu_{t,\varepsilon}$, $\varepsilon > 0$ is bounded in total variation.

Pointwise convergence of the Fourier transform.

$$\widehat{\mu}_{t,\varepsilon}(\xi) = \int_{\mathbb{R}^d} d\mu_{t,\varepsilon}(y) e^{i\langle \xi, y \rangle} = \sum_{v \in \mathbb{Z}^d} x_{\lfloor t/\varepsilon \rfloor}(v) e^{i\langle \xi, \varepsilon v \rangle} = \widehat{x}_{\lfloor t/\varepsilon \rfloor}(\varepsilon \xi) = \pi_{\lfloor t/\varepsilon \rfloor}^{(d/2,0)}(\widehat{\omega}(\varepsilon \xi)).$$

As $\varepsilon \rightarrow 0$,

$$\widehat{\omega}(\varepsilon \xi) = 1 - \frac{\varepsilon^2}{2} \langle \xi, Q \xi \rangle + o(\varepsilon^2).$$

We now apply Proposition 3.(1):

$$\begin{aligned} \widehat{\mu}_{t,\varepsilon}(\xi) &= \pi_{\lfloor t/\varepsilon \rfloor}^{(d/2,0)} \left(1 - \frac{\varepsilon^2}{2} (\langle \xi, Q \xi \rangle + o(1)) \right) = \pi_{\lfloor t/\varepsilon \rfloor}^{(d/2,0)} \left(1 - \frac{t^2 \langle \xi, Q \xi \rangle + o(1)}{2 \lfloor t/\varepsilon \rfloor^2} \right) \\ &\xrightarrow{\varepsilon \rightarrow 0} 2^{d/2} \Gamma\left(\frac{d}{2} + 1\right) \langle \xi, Q \xi \rangle^{-d/4} t^{-d/2} J_{d/2}\left(t \langle \xi, Q \xi \rangle^{1/2}\right) = \widehat{u}(t, \xi) = \widehat{\mu}_t(\xi), \end{aligned}$$

where we used the formula for $\widehat{u}(t, \xi)$ from Proposition 2. This finishes the proof of the weak convergence. \square

B.2. Proof of Local Convergence with Sinc Filter.

Proof of Theorem 4. This proof is similar to the one of Theorem 3. By Plancherel's formula,

$$\sum_{v \in \mathbb{Z}^d} (x_n(v) - (u(n, \cdot) * \psi)(v))^2 = \|x_n - u(n, \cdot) * \psi\|_{\ell^2(\mathbb{Z}^d)}^2 = \frac{1}{(2\pi)^d} \left\| \widehat{x}_n - \widehat{u(n, \cdot) * \psi} \right\|_{L^2([- \pi, \pi]^d)}^2 \quad (21)$$

In this last expression, we take the Fourier transform of $u(n, \cdot) * \psi$ as a function of $v \in \mathbb{Z}^d$. However, to decompose the computation, let us first compute the Fourier transform of $y \in \mathbb{R}^d \mapsto (u(n, \cdot) * \psi)(y)$. The Fourier product of this convolution is the product of the Fourier transforms, and ψ is chosen specifically so that its Fourier transform is $\widehat{\psi}(\xi) = \mathbf{1}_{\{\xi \in [-\pi, \pi]^d\}}$. As a consequence, the Fourier transform of $y \in \mathbb{R}^d \mapsto (u(n, \cdot) * \psi)(y)$ is $\xi \in \mathbb{R}^d \mapsto \widehat{u}(t, \xi) \mathbf{1}_{\{\xi \in [-\pi, \pi]^d\}}$.

We now discretize this function and seek the Fourier transform of $v \in \mathbb{Z}^d \mapsto (u(n, \cdot) * \psi)(v)$. The Fourier transform of the discretization is the periodization of the Fourier transform [Mallat, 1999, Theorem 3.1], thus the Fourier transform of $v \in \mathbb{Z}^d \mapsto (u(n, \cdot) * \psi)(v)$ is $\xi \in [-\pi, \pi]^d \mapsto \widehat{u}(n, \xi)$.

We obtain

$$\left\| \widehat{x}_n - \widehat{u(n, \cdot) * \psi} \right\|_{L^2([- \pi, \pi]^d)}^2 = \int_{[- \pi, \pi]^d} d\xi \left(\pi_n^{(d/2,0)}(\widehat{\omega}(\xi)) - \widehat{u}(n, \xi) \right)^2.$$

We make the change of variables $\zeta = n\xi$:

$$\left\| \widehat{x}_n - \widehat{u(n, \cdot) * \psi} \right\|_{L^2([- \pi, \pi]^d)}^2 = n^{-d} \int_{\mathbb{R}^d} d\zeta \left(\pi_n^{(d/2,0)} \left(\widehat{\omega} \left(\frac{\zeta}{n} \right) \right) - \widehat{u} \left(n, \frac{\zeta}{n} \right) \right)^2 \mathbf{1}_{\{\zeta \in [-n\pi, n\pi]^d\}}. \quad (22)$$

Fix $\zeta \in \mathbb{R}^d$. Using the Mehler–Heine asymptotic (Proposition 3.(1)), we prove that

$$\pi_n^{(d/2,0)} \left(\widehat{\omega} \left(\frac{\zeta}{n} \right) \right) - \widehat{u} \left(n, \frac{\zeta}{n} \right) \xrightarrow{n \rightarrow \infty} 0.$$

We do not repeat the computations because they are similar to the pointwise convergence in the proof of Theorem 3. This proves that the integrand of (22) converges pointwise to 0. We want to apply the dominated convergence theorem to conclude, and thus seek a domination of

$$\left(\pi_n^{(d/2,0)} \left(\widehat{\omega} \left(\frac{\zeta}{n} \right) \right) - \widehat{u} \left(n, \frac{\zeta}{n} \right) \right)^2 \mathbf{1}_{\{\zeta \in [-n\pi, n\pi]^d\}} \quad (23)$$

$$\leq 2\pi_n^{(d/2,0)} \left(\widehat{\omega} \left(\frac{\zeta}{n} \right) \right)^2 \mathbf{1}_{\{\zeta \in [-n\pi, n\pi]^d\}} + 2\widehat{u} \left(n, \frac{\zeta}{n} \right)^2. \quad (24)$$

By scale invariance of the EPD equation (or, more simply, from Proposition 2), $\widehat{u} \left(n, \frac{\zeta}{n} \right) = \widehat{u}(1, \zeta)$. Further, by Plancherel's theorem,

$$\int_{\mathbb{R}^d} d\zeta \widehat{u}(1, \zeta)^2 = (2\pi)^d \int_{\mathbb{R}^d} dy u(1, y)^2 < \infty,$$

thus the second term of (24) is independent of n and integrable. We now need to find a domination for the first term. Here, the reasoning is similar to the boundedness of $\mu_{t,\varepsilon}$ in the proof of Theorem 3.

- If $\|\zeta\| \geq \frac{1}{\sqrt{\lambda}}$, by Lemma 1, $\left| \widehat{\omega} \left(\frac{\zeta}{n} \right) \right| \leq 1 - \frac{1}{n^2}$, thus by Proposition 3.(2),

$$\begin{aligned} \pi_n^{(d/2,0)} \left(\widehat{\omega} \left(\frac{\zeta}{n} \right) \right)^2 &\leq C_1^2 \left(\arccos \left| \widehat{\omega} \left(\frac{\zeta}{n} \right) \right| \right)^{-d-1} n^{-d-1} \\ &\leq C_1^2 \left(\arccos \left(1 - \lambda \frac{\|\zeta\|^2}{n^2} \right) \right)^{-d-1} n^{-d-1}. \end{aligned}$$

There exists $C_9 > 0$ such that $\arccos(1 - z) \geq C_9 \sqrt{z}$. Thus

$$\pi_n^{(d/2,0)} \left(\widehat{\omega} \left(\frac{\zeta}{n} \right) \right)^2 \leq C_{10} \|\zeta\|^{-d-1}.$$

- If $\|\zeta\| < \frac{1}{\sqrt{\lambda}}$, then

$$\pi_n^{(d/2,0)} \left(\widehat{\omega} \left(\frac{\zeta}{n} \right) \right)^2 \leq C_2^2.$$

We thus define the domination

$$g(\zeta) = \begin{cases} C_{10} \|\zeta\|^{-d-1} & \text{if } \|\zeta\| \geq \frac{1}{\sqrt{\lambda}}, \\ C_2^2 & \text{if } \|\zeta\| < \frac{1}{\sqrt{\lambda}}. \end{cases}$$

This domination is integrable on \mathbb{R}^d ; this concludes the theorem. \square

APPENDIX C. PROOF OF COROLLARY 1

Note that $\sum_{v \in \mathbb{Z}^d} x_n(v)^2 = \|x_n\|_{l^2(\mathbb{Z}^d)}^2$ and by Theorem 4,

$$\left| \|x_n\|_{l^2(\mathbb{Z}^d)} - \|u(n, \cdot) * \psi\|_{l^2(\mathbb{Z}^d)} \right| \leq \|x_n - u(n, \cdot) * \psi\|_{l^2(\mathbb{Z}^d)} = o(n^{-d/2}).$$

It is thus sufficient to prove that

$$\|u(n, \cdot) * \psi\|_{l^2(\mathbb{Z}^d)}^2 \sim \frac{1}{(\det Q)^{1/2} |B(0, 1)|} \frac{1}{n^d}.$$

In the proof of Theorem 4, we explain that the Fourier transform of $v \in \mathbb{Z}^d \mapsto (u(n, \cdot) * \psi)(v)$ is $\xi \in [-\pi, \pi]^d \mapsto \widehat{u}(n, \xi)$. Thus by Plancherel's theorem,

$$\|u(n, \cdot) * \psi\|_{l^2(\mathbb{Z}^d)}^2 = \frac{1}{(2\pi)^d} \|\widehat{u}(n, \cdot)\|_{L^2([-\pi, \pi]^d)}^2 = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} d\xi \widehat{u}(n, \xi)^2.$$

We make the change of variables $\zeta = \xi/n$. Note that by scale invariance of the EPD equation (or, more simply, from Proposition 2), $\widehat{u} \left(n, \frac{\zeta}{n} \right) = \widehat{u}(1, \zeta)$. Thus

$$\|u(n, \cdot) * \psi\|_{l^2(\mathbb{Z}^d)}^2 = \frac{1}{(2\pi)^d n^d} \int_{[-n\pi, n\pi]^d} d\zeta \widehat{u}(1, \zeta)^2 = \frac{1}{(2\pi)^d n^d} \left(\int_{\mathbb{R}^d} d\zeta \widehat{u}(1, \zeta)^2 + o(1) \right). \quad (25)$$

We use again Plancherel's theorem and then (5):

$$\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} d\zeta \widehat{u}(1, \zeta)^2 = \int_{\mathbb{R}^d} dy u(1, y)^2 = \left(\frac{\Gamma(d/2 + 1)}{\pi^{d/2} (\det Q)^{1/2}} \right)^2 |\{y | \langle y, Q^{-1}y \rangle \leq 1\}|.$$

As the volume of the d -dimensional unit ball is $|B(0, 1)| = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$, the volume of the ellipsoid $|\{y | \langle y, Q^{-1}y \rangle \leq 1\}|$ is $\frac{\pi^{d/2} (\det Q)^{1/2}}{\Gamma(d/2+1)}$, thus

$$\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} d\zeta \widehat{u}(1, \zeta)^2 = \frac{\Gamma(d/2 + 1)}{\pi^{d/2} (\det Q)^{1/2}} = \frac{1}{(\det Q)^{1/2} |B(0, 1)|}.$$

Substituting in (25), this concludes the proof.