MEAN SQUARE PERFORMANCE OF CONSENSUS-BASED DISTRIBUTED ESTIMATION OVER REGULAR GEOMETRIC GRAPHS*

FEDERICA GARIN† AND SANDRO ZAMPIERI‡

Abstract. Average-consensus algorithms allow to compute the average of some agents' data in a distributed way, and they are used as a basic building block in many algorithms for distributed estimation, load balancing, formation and distributed control.

Traditional analysis of such algorithms studies, for a given communication graph, the convergence rate (second largest eigenvalue of the transition matrix) and predicts that, for many graph families, performance degrades when the number of agents grows, because of the longer time required to spread information. However, in estimation problems, a growing number of sensor nodes improves the quality of the estimate. To understand whether such an improvement is possible also with distributed algorithms, it is important to specify a suitable performance metric, depending on the specific estimation problem in which the consensus algorithm is used, and to study how performance scales when both the number of iterations and the number of agents grow to infinity.

Here, we propose a simple example of a distributed estimation problem solved by average-consensus, and a performance index naturally arising in this context (mean square estimation error, MSE). To understand the performance limitations of sensor networks with limited-range communications, we consider graphs describing local interactions. We give analytic results for some families of such graphs whose symmetries allow the use of suitable mathematical tools. However, simulations indicate that a similar behavior occurs also for random geometric graphs. This suggests that the performance limitations of regular lattices are mainly due to the geometrically local interactions and not to the symmetries.

Key words. Multi-agent systems, consensus algorithm, distributed averaging, distributed estimation.

AMS subject classifications. 60G35, 62L12, 94C15, 05C50.

1. Introduction.

In recent years, the analysis of the coordination mechanisms of multi-agent systems is attracting a large attention in the engineering community. This is mainly due to the intrinsic robustness and to the degree of adaptation exhibited in nature by such systems, which makes their structure very attractive as an inspiring design paradigm for many engineering systems. This paradigm consists in the possibility of obtaining high performance levels through the cooperation of numerous simple and cheap local units.

The information dynamics which permit these systems to work properly is a challenging problem for the information engineering community; in fact, despite the variety of cooperating systems of different nature without a common underlying feature, it is clear that cooperation needs communication and so efficient cooperation has to be related to efficient information diffusion.

One of the simplest instances of coordinated task is averaging, i.e., computing the average of values initially separately known to the agents. One way to achieve this goal is given by the linear average-consensus algorithm [24, 28, 31, 27, 30, 9]; although

*The research leading to these results has received funding from the European Community's Seventh Framework Programme under agreement n. FP7-ICT-223866-FeedNetBack. Part of the results in this paper were presented at the 4^{th} ITA Workhop, San Diego, CA [10] and at the 10^{th} European Control Conference (ECC '09), Budapest, Hungary [21]

†NeCS, INRIA Grenoble Rhône-Alpes, 655 av. de l'Europe, Montbonnot, 38334 Saint-Ismier cedex, France, federica.garin@inrialpes.fr

‡DEI (Department of Information Engineering), Università di Padova, Via Gradenigo 6/b, 35131 Padova, Italy, zampi@dei.unipd.it

not being the most efficient method to compute the average in a distributed way [4], this technique is attracting a lot of attention mainly because of its simplicity, which makes it intrinsically robust to node or to communication failures [29, 7, 16].

This algorithm has been proposed in many contexts in which it is necessary to compute averages in a distributed way, namely in distributed estimation [26] and in sensor calibration for sensor networks [20, 8], in load balancing for distributed computing systems [12], and in mobile multi-vehicles coordination [11], as well as in distributed optimization and learning [25]. Linear average-consensus algorithm is a linear iterative algorithm in which a sequence of vectors is obtained by constructing the new vector from the previous one by multiplying it by a doubly-stochastic matrix. Through the theory of Markov chains it is possible to prove, under rather weak assumptions on the matrix, that the vector sequence converges to a vector with entries all equal to the average of the components of the initial vector. Traditionally, the index which is considered for determining the performance of a specific average-consensus algorithm is given by the exponential rate of convergence to the limit vector; this is given by the second largest eigenvalue of the matrix which is called the essential spectral radius. In the literature devoted to Markov chains and to the so-called 'spectral graph theory', this index has been deeply studied, and many bounds have been proposed in the different cases of matrices with symmetries [15], of randomly generated matrices [6], and of general matrices [15]. However, when this algorithm is used for specific applications requiring the distributed computation of averages, different performance indices become more natural instruments for comparing different choices of matrices. Literature along this research line is not very developed; some contributions have considered the effects of robustness to delays [28], noise on the communication links [30] and quantization [18], and a more recent work deals with a mobile vehicle coordination problem [1].

In the present paper, we consider a very simple example of a distributed estimation problem, solved by the average-consensus problem, and we propose as a natural performance metric the mean square estimation error (Section 2 describes the problem setting and the relevant cost function). Such a cost depends on all the eigenvalues of the consensus update matrix, and we show by a simple example that using it for performance evaluation can lead to significantly different results from looking at the essential spectral radius (see Section 3). Moreover, the study of performance indices different from convergence rate is essential in large-scale networks, namely in networks formed by a large number of cooperating agents. In fact, in this case a trade-off can be expected, since on the one hand, a larger number of sensors should give a better estimate, while on the other hand the more difficult communication between a larger number of agents will decrease this advantage. While this trade-off becomes quite clear when choosing correct performance indices, it is not highlighted by the essential spectral radius, which often simply underlines that the larger is the network, the slower is the convergence. Our analysis allows to correctly highlight this trade-off, as shown first in a simple example (Section 3), and then in our general results. Our main contribution (Sections 4 and 5) is the characterization of the asymptotic scaling of the estimation cost, both with the number of agents and with computation time, for families of lattice-like communication graphs, where the symmetries in the graph, and the associated algebraic structure, allow the use of analytical tools to get rigorous bounds. We consider grids over multi-dimensional tori, and over hyper-cubes. Such graphs have been largely studied in the recent literature devoted to distributed estimation and control, see e.g. [2], [13] and [9], because of the ease of analysis allowed by their structure, as well as because they are a prototypical example of geometrically local communication. Then, simulation results, presented in Section 6, show that connected realizations of random geometric graphs with a comparable number of nodes and of average number of neighbors exhibit a behavior very similar to the corresponding regular grids, thus suggesting that the behavior of the graphs that we analyzed is mainly due to the geometrically local interactions, and not to the symmetries and the regular structure.

Notation and preliminaries. Throughout this paper, we will use the following notational conventions. Vectors will be denoted with boldface letters, and matrices (equivalently, linear maps) with capital letters. Given a vector $\boldsymbol{v} \in \mathbb{R}^N$ and a matrix $M \in \mathbb{R}^{N\times N}$, we let $\boldsymbol{v}^T$ and M^T respectively denote the transpose of $\boldsymbol{v}$ and of M. We let $\Lambda(M)$ denote the set of eigenvalues of M, counted with their multiplicities. With the symbol $\mathbf{1}$ we denote the N-dimensional vector having all the entries equal to 1. We will denote by $|\boldsymbol{v}|$ the vector obtained by taking the modulus of each entry of a given vector $\boldsymbol{v}$, and we will write $\boldsymbol{w} \succ \boldsymbol{v}$ and $\boldsymbol{w} \succcurlyeq \boldsymbol{v}$ to denote that all entries satisfy $w_k > v_k$ and $w_k \geq v_k$ respectively.

Given any set A with finite cardinality $|A|$, $\mathbb{R}^A$ will denote the vector space isomorphic to $\mathbb{R}^{|A|}$, made of vectors where indices are elements of A instead of $\{1, 2, \dots, |A|\}$. Analogously, $\mathbb{R}^{A\times A}$ will denote the vector space of all linear maps from $\mathbb{R}^A$ to $\mathbb{R}^A$.

We use the convention that a summation over an empty set of indices is equal to zero, while a product over an empty set gives one. We also introduce the short-hand notation $[d] = \{1, 2, \dots, d\}$.

A directed graph G is a pair (V, E) where V is a set, called the set of vertices, and $E \subseteq V \times V$. A directed graph G is strongly connected if, for all $u, v \in V$, there exists a path connecting u to v. It is weakly connected if, disregarding edge orientations, for any for pair of vertices u, v, there exists an undirected path connecting u to v. The graph is aperiodic if the greatest common divisor of the lengths of all cycles is one. The presence of a self-loop implies aperiodicity.

A matrix $P \in \mathbb{R}^{V\times V}$ is said to be stochastic if it has non-negative entries and if $P\mathbf{1} = \mathbf{1}$. A stochastic matrix P is said to be doubly-stochastic if P^T is stochastic so that $P\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T P = \mathbf{1}^T$. Given a stochastic matrix P, the graph G_P associated with $P \in \mathbb{R}^{V\times V}$ is a directed graph $G_P = (V, E)$, with $(u, v) \in E$ if and only if $P_{uv} \neq 0$. A stochastic matrix P is primitive if there exists a positive integer m such that $(P^m)_{uv} \neq 0$ for all $u, v \in V$, or, equivalently, if the graph G_P associated with P is strongly connected and aperiodic. From the Perron-Frobenius theorem (see e.g. [3]) it follows that a primitive stochastic matrix P has an eigenvalue equal to 1, which has algebraic multiplicity one and is the dominant eigenvalue, i.e., all the other eigenvalues have absolute value smaller than one. The maximum absolute value of non-dominant eigenvalues is called the essential spectral radius of P and is denoted with the symbol $\rho_{\mathrm{ess}}(P)$. A matrix P is normal if it commutes with its transpose, namely $P^T P = P P^T$. Symmetric matrices are normal.

2. Problem formulation and performance measure. We consider the following simple problem of distributed estimation: N sensors measure the same real value θ plus i.i.d. noises. Clearly, the best estimate for θ is the average of such measurements, but sensors need to compute it in a distributed way. A directed graph $G = (V, E)$ describes the allowed communications: the vertices $v \in V$ are the sensors, and a pair (u, v) belongs to E if and only if u can communicate with v. We will assume that G is strongly connected and aperiodic

The sensors' measurements form a vector $\boldsymbol{x}(0) \in \mathbb{R}^V$, with $x_k(0) = \theta + w_k$, where the noises $w_1, \ldots, w_N$ are i.i.d. random variables with zero mean and finite variance (without loss of generality we will also assume variance is one).

Then we consider a linear average-consensus algorithm: $\boldsymbol{x}(t+1) = P\boldsymbol{x}(t)$ for some matrix $P \in \mathbb{R}^{V\times V}$ consistent with the communication graph G, i.e., such that the graph G_P associated with P is a subgraph of G. We assume that P is doubly-stochastic and primitive. Under these assumptions, P has dominant eigenvalue 1 with multiplicity 1 and

$$\forall k, \lim_{t\to\infty} x_k(t) = \tfrac{1}{N}\textstyle\sum_h x_h(0)\,.$$

The speed of convergence is given by the essential spectral radius of P, $\rho_{\mathrm{ess}}(P)$. For non-expander families of graphs, such as for example Cayley graphs on Abelian groups, when $N\to\infty$, $\rho_{\mathrm{ess}}(P)\to 1$ (see e.g. [9]). Clearly, this means that convergence to the average needs longer time as N grows, but this does not necessarily imply that larger N deteriorates performance in our specific application.

As our problem is estimating θ, a very natural performance measure is the mean quadratic error

$$J(P,t) := \tfrac{1}{N}\mathbb{E}\left[\boldsymbol{e}^T(t)\boldsymbol{e}(t)\right]\,,$$

where $\boldsymbol{e}(t) := \boldsymbol{x}(t) - \theta\mathbf{1}$, so that $J(P,t) = \frac{1}{N}\sum_k \mathbb{E}\left[(x_k(t)-\theta)^2\right]$. For our problem, we can easily show that the cost $J(P,t)$ can be re-written as

$$J(P,t) = \tfrac{1}{N}\ \mathrm{trace}\left((P^t)^T P^t\right). \tag{2.1}$$

Indeed, $J(P,t) = \frac{1}{N}\mathbb{E}\left[(P^t\boldsymbol{x}(0)-\theta\mathbf{1})^T(P^t\boldsymbol{x}(0)-\theta\mathbf{1})\right] = \frac{1}{N}\mathbb{E}\left[(P^t\boldsymbol{w})^T(P^t\boldsymbol{w})\right]$, where $\boldsymbol{w}$ is the measurement noise. Using linearity of expectation and of trace, plus the observation that for any scalar a we have $a = \mathrm{trace}\,a$, and the property $\mathrm{trace}(ABC) = \mathrm{trace}(CAB)$ where A, B, C are matrices of suitable size, we find the following equality:

$$\tfrac{1}{N}\mathbb{E}\left[(P^t\boldsymbol{w})^T(P^t\boldsymbol{w})\right] = \tfrac{1}{N}\mathbb{E}\left[\mathrm{trace}(\boldsymbol{w}^T(P^t)^TP^t\boldsymbol{w})\right] = \tfrac{1}{N}\,\mathrm{trace}\left((P^t)^TP^t\,\mathbb{E}(\boldsymbol{w}\boldsymbol{w}^T)\right),$$

which proves Eq. (2.1) because $\boldsymbol{w}$ has zero mean and identity covariance matrix.

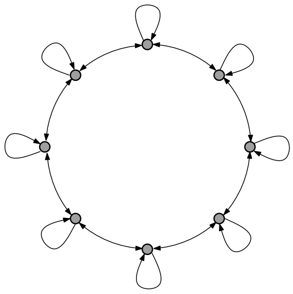
If P is normal, then Eq. (2.1) is equivalent to the following:

$$J(P,t) = \tfrac{1}{N}\sum_{\lambda\in\Lambda(P)}|\lambda|^{2t}\,. \tag{2.2}$$

3. Motivating examples.

3.1. Three examples motivating the study of the proposed cost. In this section we present three simple examples of families of graphs. Our first aim is to show that the cost $J(P,t)$ allows to study the trade-off between the two effects of a large number of nodes N, which improves performance after infinite time, but slows down computation. Moreover, we show that studying this cost can give performance results significantly different from traditional analysis of the speed of convergence via the essential spectral radius, although the two different performance measures are not completely unrelated.

Example 1: the circle. This first example is the simplest case of local communication, where N agents are disposed on a circle, and each agent communicates with its first neighbor on each side (left and right). For simplicity, we assume that each received message, as well as the agent's own state, is weighted 1/3. Figure 3.1 shows the circle graph and the corresponding circulant symmetric matrix P_N. As the eigen-



$$P_N = \begin{bmatrix} 1/3 & 1/3 & 0 & \cdots & \cdots & 1/3 \\ 1/3 & 1/3 & 1/3 & \cdots & \cdots & 0 \\ 0 & 1/3 & 1/3 & 1/3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1/3 & 1/3 & 1/3 \\ 1/3 & \cdots & \cdots & 0 & 1/3 & 1/3 \end{bmatrix}$$

Fig. 3.1: The circle graph and circulant matrix considered in Example 1.

values of such a circulant matrix can be explicitly computed [14], it is easy to see that $\rho_{\mathrm{ess}}(P_N) \sim 1-\pi^2/N^2$. This shows that, as N grows, the convergence of the algorithm tends to be very slow. Nonetheless we expect that, in case of distributed estimation, the presence of more sensors should instead improve performance. Figure 3.2 depicts $J(P_N, t)$ as a function of t, for various values of N. For any fixed N, we have evolutions which exponentially converge (with rate $\sim (1-\pi^2/N^2)$) to a constant value $1/N$. The different curves become lower as N grows, and their envelope, which corresponds to the limit for $N \to \infty$, converges to zero for $t \to \infty$. In Section 3.2 we will study the asymptotic behavior of $J(P_N, t)$ for circle graphs, and show that it scales as $\max\left\{\frac{1}{N}, \frac{1}{\sqrt{t}}\right\}$. In particular, $\lim_{N\to\infty} J(P_N, t)$ converges to zero as $1/\sqrt{t}$. This result shows that increasing N does not have the disadvantages predicted by observing that $\lim_{N\to\infty} \rho_{\mathrm{ess}}(P_N) = 1$. Also, it can be formally proven that, in this example, $J(P_N, t)$ is monotonic non-increasing w.r.t. N, for any fixed t (Coroll. 5.3). Nevertheless, a further look at Figure 3.2, together with the results in Coroll. 3.2, gives a caveat against the choice of too large values of N: when the number of iterations is not unlimited, there is a bound on the number of nodes being truly useful, after which there is no improvement in adding new nodes, since $J(P_N, t) = J(P_{2t+1}, t)$ for all $N \geq 2t+1$. This is very intuitive to understand, because at time t it is impossible for a node to use information coming from other agents located further than t steps apart.

Example 2: two (almost) disconnected cliques. We would like to show by two simple examples that the performance index $J(P, t)$ can behave very differently from the essential spectral radius. We start from an extreme case, an example of graph which is disconnected and thus the convergence to the average consensus does not occur. Nevertheless the estimation error can be quite small. Let N be an even number, and consider a graph consisting of two disconnected cliques, each with $N/2$ nodes; Fig. 3.3a depicts the case $N = 10$. Associate with each edge a coefficient $2/N$, so that P_N has the following form:

$$P_N = \left[\begin{array}{c|c} \frac{2}{N}\mathbf{1}\mathbf{1}^T & 0 \\ \hline 0 & \frac{2}{N}\mathbf{1}\mathbf{1}^T \end{array}\right].$$

The eigenvalues of P_N are easily computed: 1 with multiplicity 2 and 0 with multiplicity $N-2$. Therefore, the essential spectral radius is 1, which describes the fact that, the graph being disconnected, no convergence to the average consensus is possible. However, for all $t \geq 1$, $J(P_N, t) = \frac{2}{N}$, which is almost as good as the best possible error (the error variance in the case of centralized estimation is $\frac{1}{N}$), and, for large N and small t, is much better than the error obtained with the circle graph in

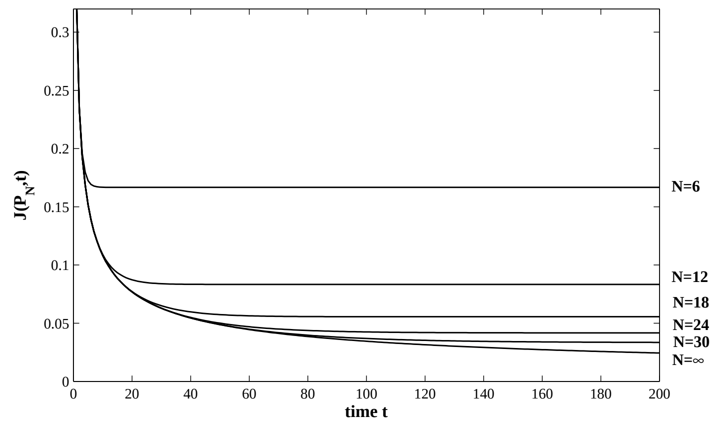
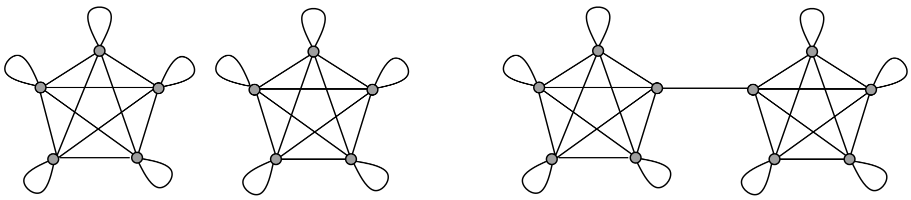


Fig. 3.2: The time evolution of $J(P_N, t)$ for the matrix P_N introduced in Example 1, for various values of N.



(a) Two disconnected cliques with $N/2$ nodes each.

(b) Modified connected graph.

Fig. 3.3: Communication graphs considered in Example 2.

Example 1.

The intuitive explanation is that, with the two cliques, the estimation error is very good for large N because, even if it is not possible to compute the average of all the measurements, it is possible to compute very quickly (in one iteration) the average of $N/2$ measurements. On the contrary, in the circle the average consensus can be reached asymptotically, as described by the essential spectral radius smaller than one, but the convergence is very slow for large N, and a reasonably good estimation error is achieved only after a long time.

The above example can be modified to obtain a slightly different matrix, whose associated graph is connected, so that the matrix is primitive. In this example, the average-consensus algorithm converges but slowly, while the estimation error is small. Indeed, consider the matrix

$$\tilde{P}_N = P_N + \left[\begin{array}{cc|cc} & & & \\ & -2/N & 2/N & \\ \hline & 2/N & -2/N & \\ & & & \end{array}\right].$$

The graph associated with $\tilde{P}_N$ is shown in Figure 3.3b. All eigenvalues of $\tilde{P}_N$ can be computed explicitly [5, Prop. 5.1]. There is one eigenvalue in 1 with multiplicity 1, one eigenvalue in 0 with multiplicity $N-3$ and two eigenvalues in $\frac{1}{2}-\frac{2}{N}\pm\frac{1}{2}\sqrt{1+\frac{8}{N}-\frac{16}{N^2}}$, each with multiplicity 1. Here the single edge connecting the two cliques is a bottleneck resulting in a quite slow convergence rate, since $\rho_{\mathrm{ess}}(\tilde{P}_N)\sim 1-\frac{8}{N^2}$. Nevertheless, the estimation error becomes good already from the first iteration, since $J(\tilde{P}_N,t)\leq\frac{3}{N}$ for all $t\geq 1$.

Example 3: regular expander graphs. Example 2 shows that the MSE cost $J(P,t)$ and the second largest eigenvalue ρ_{ess} are two performance measures that can be very different for some families of graphs. However, they are not completely unrelated. In fact, it is easy to see that $J(P,t)=\frac{1}{N}+\frac{1}{N}\sum_{\lambda\neq 1}|\lambda|^{2t}\leq\frac{1}{N}+\rho_{\mathrm{ess}}^{2t}$. Hence, if ρ_{ess} is small, then the cost $J(P,t)$ converges quickly to the asymptotic value $\frac{1}{N}$ when t grows. This means that graphs yielding good essential spectral radius also yield good MSE cost.

The study of the second largest eigenvalue of the adjacency matrix has been developed in a rich literature in spectral graph theory. In particular it has been shown that this index is strictly related to a property of the graph called expansion. Expander graphs are graphs with high expansion. For a survey on expander graphs see e.g. [23]. An example of a family of expander graphs is given by random regular graphs with N nodes and degree d. In fact, a theorem by Friedman [19] (see e.g. [23, Thm. 7.10]) ensures that, for any $\varepsilon>0$, asymptotically almost surely[1] random regular graphs with N nodes and fixed degree d have adjacency matrices A_N satisfying

$$\rho_{\mathrm{ess}}(\tfrac{1}{d}A_N)\leq 2\frac{\sqrt{d-1}}{d}+\varepsilon\,.$$

From this result, we can construct a family of matrices $P_N=\frac{1}{d+1}(I+A_N)$, so that, for all $\varepsilon>0$, asymptotically almost surely

$$\rho_{\mathrm{ess}}(P_N)\leq\frac{1+2\sqrt{d-1}}{d+1}+\varepsilon\,.$$

This shows that random regular graphs, with high probability, yield an essential spectral radius which does not converge to one as the number of nodes tends to infinity, although they have bounded degree. This is in contrast with the behavior of the circle graphs presented in Example 1, and of other graph families which yield an essential spectral radius converging to one, such as the grids which will be presented in Section 4 and the random geometric graphs which will be described in Section 6 and whose essential spectral radius is studied in [6].

3.2. The circle graph. The aim of the paper is to analyze the MSE performance for geometric graphs, namely for graphs which are generated by placing nodes in a metric space and by connecting nodes which are within a given distance. Unfortunately we have been unable to obtain results for general graphs of this kind. We will instead restrict to graphs possessing some symmetries, which enable an easier characterization of the spectral properties of the associated matrices P. In this subsection, we consider the simplest case, i.e., a circle, so as to show without cumbersome notation the main results and the main tools used in the proofs. Then, in Sections 4 and 5 we will extend these results to more general families of structured graphs. In the final section we will show through simulations that the performance

[1] i.e., with probability that tends to one when the number of vertices N tends to infinity.

for general geometric graphs presents similar behaviors to the ones which are proven mathematically in case of geometric graphs with symmetries.

In this subsection, we consider circular graphs similar to the one presented in Example 1, except that we allow connections not only to the two nearest neighbors, but also to the 2δ nearest ones, for some fixed positive integer δ. Also, we do not impose that all coefficients are the same. Thus, P_N is a circulant matrix, with first row given by $(p_0, p_1, \dots, p_\delta, 0, \dots, 0, p_{-\delta}, \dots, p_{-1})$. We introduce now two Laurent polynomials $p(z) := \sum_{k=-\delta}^{\delta} p_k z^k$ and $q(z) := p(z)p(z^{-1})$.

The circulant structure of P_N implies that P_N is normal, and that its eigenvalues $\lambda_0, \dots, \lambda_{N-1}$ are

$$\lambda_h = p(e^{-i\frac{2\pi}{N}h}), \quad h = 0, \dots, N-1\,,$$

where $i = \sqrt{-1}$. Notice that $|\lambda_h|^2 = q(e^{i\frac{2\pi}{N}h})$. Another consequence of the choice of a circulant matrix P_N is that the MSE is the same for all nodes, i.e., $\mathbb{E}(\boldsymbol{e}_h(t)^2) = \mathbb{E}(\boldsymbol{e}_k(t)^2) = J(P_N, t)$ for all vertices h, k. In particular, $J(P_N, t) = \mathbb{E}(\boldsymbol{e}_0(t)^2)$.

For a family of graphs with such a regular structure, it makes sense to consider also the infinite version of the graph and of the linear map P, with the same local neighborhoods and coefficients, i.e., the graph is an infinite line (nodes are labelled by integers in $\mathbb{Z}$) and the map P_∞ is linear banded-Toeplitz, with diagonal band $(p_{-\delta}, \dots, p_{-1}, p_0, p_1, \dots, p_\delta)$. Thanks to the remark that the MSE is the same for all nodes, it makes sense to define the cost to be the MSE of a reference node $J(P_\infty, t) = \mathbb{E}(\boldsymbol{e}_0(t)^2)$. We will show that $J(P_\infty, t)$ is the envelope of all the $J(P_N, t)$ as it was observed in Example 1.

We make the following assumptions on the coefficients $p_{-\delta}, \dots, p_{-1}, p_0, p_1, \dots, p_\delta$:

- $p_0 \neq 0$, i.e., the associated graphs have self-loops.
- $p_h \geq 0$ for all h, and $\sum_{k=-\delta}^{\delta} p_k = 1$. This ensures that P is doubly stochastic.
- $p_1 \neq 0$ or $p_{-1} \neq 0$. It is easy to see that this condition ensures that, for all N, the circle graph associated with P_N is strongly connected. This condition ensures also that the infinite line graph associated with P_∞ is weakly connected. It will turn out (Lemma 4.1) that it is also a necessary condition, and not only a sufficient one.

The first result that we obtain is the following characterization of the cost, involving only the coefficients of the polynomial $\big(q(z)\big)^t$. We will use the notation $p_h^{(t)}$ and $q_h^{(t)}$ denote the h-th coefficient of the polynomials $(p(z))^t$ and $(q(z))^t$, respectively.

Proposition 3.1. *With the above notation,*

- $J(P_\infty, t) = q_0^{(t)}$;
- $J(P_N, t) = \displaystyle\sum_{\substack{-2t\delta \leq h \leq 2t\delta \\ h = 0 \bmod N}} q_h^{(t)}$.

Proof. For the infinite line, notice that $J(P_\infty, t) = \mathbb{E}\big[\big((P_\infty^t \boldsymbol{w})_0\big)^2\big]$, where $\boldsymbol{w}$ is the noise in the initial measurements. Also notice that $(P_\infty^t \boldsymbol{w})_k = \sum_{h\in\mathbb{Z}} p_h^{(t)} w_{k-h}$. Therefore,

$$\mathbb{E}\big[\big((P_\infty^t \boldsymbol{w})_0\big)^2\big] = \mathbb{E}\big[\sum_{h,k\in\mathbb{Z}} p_h^{(t)} p_k^{(t)} w_{-h} w_{-k}\big] = \sum_{h\in\mathbb{Z}} \big(p_h^{(t)}\big)^2 = q_0^{(t)}\ .$$

For the circle, from Eq. (2.2), substituting $|\lambda_h|^2 = q(e^{i\frac{2\pi}{N}h})$ gives:

$$J(P_N, t) = \frac{1}{N} \sum_{h=0}^{N-1} \sum_{k=-2t\delta}^{2t\delta} q_k^{(t)} e^{i\frac{2\pi}{N}hk}\,,$$

which ends the proof, because

$$\frac{1}{N}\sum_{h=0}^{N-1} e^{i\frac{2\pi}{N}hk} = \begin{cases} 1 & \text{if } k = 0 \bmod N, \\ 0 & \text{otherwise.} \end{cases} \qquad \square$$

An immediate consequence of Prop. 3.1 is the following.

COROLLARY 3.2.

$$\lim_{N\to\infty} J(P_N, t) = J(P_\infty, t)$$

and, more precisely, for all $N > 2t\delta$, $J(P_N, t) = J(P_\infty, t)$.

We will discuss later (Coroll. 5.3) some sufficient conditions under which $J(P_N, t)$ is a decreasing function of N. The following result gives the asymptotic behavior of the cost whith respect to large t and N.

THEOREM 3.3. *With the above notation and assumptions,*

- *for the infinite line, there exist positive constants* C, C', *depending* $p(z)$ *only, such that, for all* $t > 0$,
$$C\frac{1}{\sqrt{t}} \leq J(P_\infty, t) \leq C'\frac{1}{\sqrt{t}};$$
- *for the circle, there exist positive constants* C, C', *depending* $p(z)$ *only, such that, for all* $N > 2\delta$ *and* $t > 0$,

$$C \max\left\{\tfrac{1}{N}, \tfrac{1}{\sqrt{t}}\right\} \leq J(P_{\boldsymbol{n}}, t) \leq C' \max\left\{\tfrac{1}{N}, \tfrac{1}{\sqrt{t}}\right\}.$$

Proof. The proof is based on some simple remarks about the function $f : \mathbb{R} \to [0, +\infty)$ defined by

$$f(x) = |p(e^{ix})|^2 .$$

Notice that $f(x) = q(e^{ix}) = \sum_{\ell=-2\delta}^{\delta} q_\ell \cos(\ell x)$. Clearly f is a trigonometric polynomial, with $f(0) = 1$ and $0 \leq f(x) \leq 1$ for all x. The assumption that p_0 and that at least one between p_{-1} and p_1 are non-zero ensures that the same applies also to q_0, q_{-1} and q_1, and thus that $f(x) < 1$ for all $x \in (-2\pi, 2\pi)\setminus\{0\}$. The derivatives of f in zero are $f'(0) = 0$ and $f''(0) = -\sum_{\ell=-2\delta}^{\delta} \ell^2 q_\ell < 0$. Thus, we can choose α and β satisfying $0 < \alpha < -f''(0) < \beta$, and we can find a neighborhood of 0, say $(-a, a) \subseteq (-\pi, \pi)$, such that $e^{-\beta x^2} \leq f(x) \leq e^{-\alpha x^2}$ for any x in such neighborhood. Moreover, we can find a constant $c \in (0, 1)$ such that $f(x) < c$ for all $x \in [-\pi, \pi] \setminus (-a, a)$. We define the functions

$$f_{\mathrm{U}}(\boldsymbol{x}) = \begin{cases} e^{-\alpha\, x^2} & \text{for } \boldsymbol{x} \in (-a, a) \\ c & \text{otherwise}, \end{cases} \qquad f_{\mathrm{L}}(\boldsymbol{x}) = \begin{cases} e^{-\beta\, x^2} & \text{for } \boldsymbol{x} \in (-a, a) \\ 0 & \text{otherwise}. \end{cases}$$

so that we can write

$$f_{\mathrm{L}}(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq f_{\mathrm{U}}(\boldsymbol{x})\,, \forall x \in [-\pi, \pi]\,.$$

Now we can use such bounds on f to obtain bounds on the MSE cost. For the infinite line, Prop. 3.1 and the definition of the polynomial $q(z)$ give

$$J(P_\infty, t) = q_0^{(t)} = \sum_{h\in\mathbb{Z}} |p_h^{(t)}|^2\,.$$

By Parseval's identity applied to the function $(p(z))^t$, this expression can be re-written as follows

$$J(P_\infty, t) = \tfrac{1}{2\pi}\int_{-\pi}^{\pi} \left|(p(e^{ix}))^t\right|^2 \,\mathrm{d}x = \tfrac{1}{2\pi}\int_{-\pi}^{\pi} (f(x))^t \,\mathrm{d}x\,.$$

Using the upper bound $f(x) \leq f_{\mathrm{U}}(x)$ we get

$$J(P_\infty, t) \leq \tfrac{1}{2\pi}\int_{-\pi}^{\pi} (f_{\mathrm{U}}(x))^t \,\mathrm{d}x = \tfrac{1}{2\pi}\int_{-a}^{a} e^{-\alpha t x^2}\,\mathrm{d}x + \tfrac{1}{2\pi}(2\pi - 2a)c^t \leq \frac{1}{2\sqrt{\pi\,\alpha\,t}} + c^t\,,$$

where the last inequality comes from calculating $\int_{\mathbb{R}} e^{-\alpha\,t x^2} = \sqrt{\frac{\pi}{\alpha\,t}}$ (see Lemma A.2). Similarly, for the lower bound,

$$J(P_\infty, t) \geq \tfrac{1}{2\pi}\int_{[-\pi,\pi]} (f_{\mathrm{L}}(x))^t \,\mathrm{d}x = \tfrac{1}{2\pi}\int_{[-a,a]} e^{-\beta t x^2}\,\mathrm{d}x \geq \frac{1}{2\sqrt{\pi\,\alpha\,t}}\left(1 - e^{-\beta a^2 t}\right)\,,$$

where the last inequality comes from a well-known property of the the tail of the Gaussian distribution (see Lemma A.2 for more details). Finally notice that $1 - e^{-\beta a^2 t} \geq 1 - e^{-\beta a^2}$ for all $t \geq 1$.

Now we consider the circle. From Eq. (2.2),

$$J(P_N, t) = \frac{1}{N}\sum_{h=0}^{N} |\lambda_h|^2 = \frac{1}{N}\sum_{h=0}^{N}\left(f\left(\tfrac{2\pi}{N}h\right)\right)^t = \frac{1}{N}\sum_{h=-\lfloor (N-1)/2\rfloor}^{\lfloor N/2\rfloor}\left(f\left(\tfrac{2\pi}{N}h\right)\right)^t\,,$$

where the last equality comes from the fact that the trigonometric polynomial $f(x)$ has period 2π. Then, we can use the upper bound $f(x) \leq f_{\mathrm{U}}(x)$ and obtain

$$J(P_N, t) \leq \frac{1}{N}\sum_{h=-\lfloor (N-1)/2\rfloor}^{\lfloor N/2\rfloor}\left(f_{\mathrm{U}}\left(\tfrac{2\pi}{N}h\right)\right)^t \leq c^t + \frac{1}{N}\sum_{h=-\lfloor \frac{aN}{2\pi}\rfloor}^{\lfloor \frac{aN}{2\pi}\rfloor} e^{-\alpha\left(\frac{2\pi}{N}h\right)^2 t}\,.$$

The proof of the upper bound is concluded by noting that the latter sum is bounded by a suitable integral, as follows:

$$\frac{1}{N}\sum_{h=-\lfloor \frac{aN}{2\pi}\rfloor}^{\lfloor \frac{aN}{2\pi}\rfloor} e^{-\alpha\left(\frac{2\pi}{N}h\right)^2 t} = \frac{1}{N} + \frac{2}{N}\sum_{h=1}^{\lfloor \frac{aN}{2\pi}\rfloor} e^{-\alpha\left(\frac{2\pi}{N}h\right)^2 t} \leq \frac{1}{N} + \frac{1}{\pi}\int_0^{\frac{2\pi}{N}\lfloor \frac{aN}{2\pi}\rfloor} e^{-\alpha x^2 t}\,\mathrm{d}x \tag{3.1}$$

and finally

$$\frac{1}{\pi}\int_0^{\frac{2\pi}{N}\lfloor \frac{aN}{2\pi}\rfloor} e^{-\alpha x^2 t}\,\mathrm{d}x \leq \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha x^2 t}\,\mathrm{d}x = \frac{1}{2\sqrt{\pi\alpha t}}\,,$$

so that

$$J(P_N, t) \leq c^t + \frac{1}{N} + \frac{1}{2\sqrt{\pi\alpha t}} \leq \frac{1}{N} + \left(c + \frac{1}{2\sqrt{\pi\alpha}}\right)\frac{1}{\sqrt{t}} \leq 3\max\left\{\frac{1}{N}, \frac{1}{\sqrt{t}}\right\}\,.$$

The lower bound is obtained thanks to the following two simple remarks. First,

$$J(P_N, t) = \frac{1}{N}\sum_{h=0}^{N} |\lambda_h|^2 \geq \frac{1}{N}|\lambda_0|^2 = \frac{1}{N}\,.$$

Then, by Prop. 3.1,

$$J(P_N, t) = \sum_{\substack{-2t\delta \leq h \leq 2t\delta \\ h=0 \bmod N}} q_h^{(t)} \geq q_0^{(t)} = J(P_\infty, t)\,,$$

so that we can apply the result proved above for the infinite line and obtain:

$$J(P_N, t) \geq J(P_\infty, t) \geq \frac{C}{\sqrt{t}}\,. \qquad \square$$

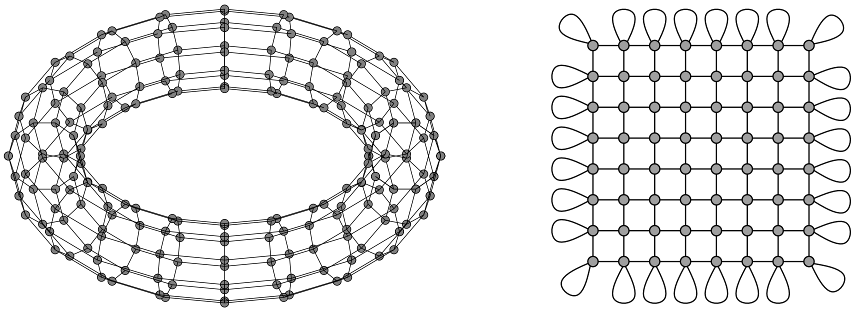
The asymptotic behavior predicted by Theorem 3.3 is well illustrated by Figure 6.1, already described in Example 1. In the next sections, we will show how the results presented here for the simple case of the circle can be generalized.

4. Regular geometric graphs. In this section we present the families of graphs and of the associated matrices that are the main object of our study. The idea is to consider graphs with enough structure so as to be able to obtain analytic results about the MSE cost, and which capture some interesting aspects of the geometrically local interactions. They are generalizations of the circle considered in Sect. 3.2. The circle is an example of one-dimensional local interaction, in the sense that nodes are aligned along some line, and can communicate only with a few neighbors on their right and on their left. We can consider a circle as a line except for the border conditions, which are periodic. The particular structure of the circle allows to associate with it some matrices which are circulant, and this allows to study the eigenvalues of such matrices. Our first generalization will be to consider the same kind of structure, both on graphs and on matrices, but in higher dimension. For example, dimension two corresponds to a grid on a torus. In Sect. 4.1 we will describe the notion of Cayley graph and of Cayley matrix on the Abelian group $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}$ and we will recall a result which generalizes to such graphs the well-known expression for the eigenvalues of circulant matrices.

Our second generalization will be to consider different border conditions, without imposing the unrealistic condition that the end-points of the line communicate closing the circle. For a particular choice of the coefficients corresponding to few nodes near the border, it is possible to obtain an explicit expression for the eigenvalues [5], and thus to study the MSE cost. Also this family of graphs and matrices can be constructed for any dimension, as we will describe in Sect. 4.3.

4.1. Grids on tori (Abelian Cayley graphs). This is an extension to general dimension d of the circle graph and circulant matrices, which can be seen as a particular case with $d = 1$. As an example, with dimension $d = 2$ the graph is a grid on a torus, as depicted in Figure 4.1a, and it is convenient to label vertices with double indices (h_1, h_2) in such a way that adding one to h_1 or to h_2 means considering the nearest vertex moving 'to the right' or 'up' respectively. More formally, the structure of such graphs is described by Cayley graphs over suitable groups, e.g., $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$ for the 2-dimensional case.

We recall the definition of Cayley graphs: given a group $(\Gamma, +)$ and a set $S \subseteq \Gamma$, the Cayley graph $\mathcal{G}(\Gamma, S)$ is a directed graph with vertex set Γ and edge set $E = \{(g, h) : h - g \in S\}$. We will consider finite graphs, with $|\Gamma| = N$, and matrices associated with such graphs, which respect the strong symmetries of the graph: we say that a matrix $P \in \mathbb{R}^{\Gamma \times \Gamma}$ (i.e. with entries labeled by indexes belonging to Γ) is Cayley if $P_{g,h} = P_{g+k,h+k}$ for all $g, h, k \in \Gamma$. This is equivalent to saying that there



(a) Grid on a torus (Cayley graph). (b) Grid on a cube.

Fig. 4.1: 2-dimensional grids.

exists a map $\pi : \Gamma \to \mathbb{R}$ such that $P_{h,k} = \pi(h-k)$; such a function is called the generator of the Cayley matrix P. Notice that a stochastic Cayley matrix is also doubly-stochastic. Also notice that Cayley matrices are normal.

In this paper, we limit our attention to Abelian groups, and we let $\Gamma_{n_1,\dots,n_d} := \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}$. We will use the notation $\boldsymbol{n} := (n_1, \dots, n_d)$, so that $\Gamma_{\boldsymbol{n}} = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}$, and we will write $N := |\Gamma_{\boldsymbol{n}}| = \prod_{j=1}^d n_j$.

When P is a Cayley matrix associated with $\Gamma_{\boldsymbol{n}}$, its eigenvalues have the following simple expression [3]: for any $\boldsymbol{h} = (h_1, \dots, h_d) \in \Gamma_{\boldsymbol{n}}$,

$$\lambda_{\boldsymbol{h}} = \sum_{\boldsymbol{k} \in \Gamma_{\boldsymbol{n}}} \pi(\boldsymbol{k}) e^{-i(\frac{2\pi}{n_1} h_1 k_1 + \cdots + \frac{2\pi}{n_d} h_d k_d)} .$$

Note that, with a slight abuse of notation, we write $e^{i\frac{2\pi}{n_r} h_r}$ with $h_r \in \mathbb{Z}_{n_r}$, meaning that we can substitute h_r with any integer which is equal to h_r mod n_r. In the sequel, we will need the specific choice of $h_r \in \{0, 1, \dots, n_r - 1\}$, which we will denote by $\boldsymbol{h} \in V_{\boldsymbol{n}}$, $V_{\boldsymbol{n}} := \{0, \dots, n_1 - 1\} \times \cdots \times \{0, \dots, n_d - 1\}$. When needed, we will actually identify the set of vertices of the graph with $V_{\boldsymbol{n}}$ rather than $\Gamma_{\boldsymbol{n}}$.

In our analysis we want to consider families of Cayley graphs, with a growing number of vertices, but with constant degree, and with the same algebraic structure and same values for the entries of P. More precisely, we fix d, while we let $n_1, \dots, n_d$ grow. In order to define the neighbors and weights, we fix a positive integer δ, we define the set $D_\delta = \{-\delta, -\delta{+}1, \dots, +\delta\}^d$ and we fix $|D_\delta|$ real numbers $p_{\boldsymbol{h}}$, $\boldsymbol{h} = (h_1, \dots, h_d) \in D_\delta$ such that $p_{\boldsymbol{h}} \geq 0 \, \forall \boldsymbol{h}$ and $\sum_{\boldsymbol{h} \in D_\delta} p_{\boldsymbol{h}} = 1$. Then, for any $\boldsymbol{n} \succ 2\delta\mathbf{1}$ (namely $n_j > 2\delta$ for all j) we construct the Cayley matrix $P_{\boldsymbol{n}} \in \mathbb{R}^{\Gamma_{\boldsymbol{n}} \times \Gamma_{\boldsymbol{n}}}$ with generator $\pi_{\boldsymbol{n}} : \Gamma_{\boldsymbol{n}} \to \mathbb{R}$ defined by $\pi_{\boldsymbol{n}}(\boldsymbol{g}) = p_{\boldsymbol{h}}$ if there is an $\boldsymbol{h} \in D_\delta$ such that, for all $\ell = 1, \dots, d$ $g_\ell = h_\ell$ mod n_ℓ, and $\pi_{\boldsymbol{n}}(\boldsymbol{g}) = 0$ otherwise. Note that, for any $\boldsymbol{n} \succ 2\delta\mathbf{1}$, $\pi_{\boldsymbol{n}}$ is well-defined. The matrix $P_{\boldsymbol{n}}$ defined in this way can be seen as a map $\mathbb{R}^{\Gamma_{\boldsymbol{n}}} \to \mathbb{R}^{\Gamma_{\boldsymbol{n}}}$ mapping $\boldsymbol{x} \in \mathbb{R}^{\Gamma_{\boldsymbol{n}}}$ to $P - \boldsymbol{n}\boldsymbol{x} \in \mathbb{R}^{\Gamma_{\boldsymbol{n}}}$ as follows: for all $\boldsymbol{h} \in \Gamma_{\boldsymbol{n}}$,

$$(P\boldsymbol{x})_{\boldsymbol{h}} = \sum_{\boldsymbol{k} \in D_\delta} p_{\boldsymbol{k}} x_{\boldsymbol{h}-\boldsymbol{k}} .$$

We introduce another useful notation, defining the Laurent polynomial

$$p(z_1, \dots, z_d) = \sum_{\boldsymbol{k} \in D_\delta} p_{\boldsymbol{k}} z_1^{k_1} \dots z_d^{k_d} .$$

We will refer to the above construction of a family of Cayley matrices for all $\boldsymbol{n} \succ 2\delta\mathbf{1}$ as the Cayley matrix family associated with $p(z_1, \ldots, z_d)$. With this notation, the eigenvalues of $P_{\boldsymbol{n}}$ are

$$\lambda_{\boldsymbol{h}} = p(e^{-i\frac{2\pi}{n_1}h_1}, \ldots, e^{-i\frac{2\pi}{n_d}h_d}), \quad \boldsymbol{h} \in \Gamma_{\boldsymbol{n}} .$$

An interesting remark is that, for Cayley matrices, thanks to the symmetries (the graph 'looks the same' from any vertex's perspective), the mean square error is the same for every node, so that $J(P_n, t) = \mathbb{E}\left[(e_{\boldsymbol{h}}(t))^2\right]$ for any node $\boldsymbol{h}$, and we can take for example node $\mathbf{0}$ as the reference node

$$J(P_n, t) = \mathbb{E}\left[(e_{\mathbf{0}}(t))^2\right] .$$

4.2. Infinite lattices. Similarly to the infinite line that was introduced for the 1-dimensional case, it makes sense to consider a lattice with the same local neighborhoods as the above-described grid on a torus, but with infinitely many nodes, namely the case when the group is $\Gamma = \mathbb{Z}^d$. We introduce the linear map $P_\infty : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}^{\mathbb{Z}^d}$ defined, for each $\boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}^d}$ and $\boldsymbol{h} \in \mathbb{Z}^d$, as

$$(P_\infty \boldsymbol{x})_{\boldsymbol{h}} := \sum_{\boldsymbol{k} \in D_\delta} p_{\boldsymbol{k}} x_{\boldsymbol{h}-\boldsymbol{k}} . \tag{4.1}$$

Notice that

$$(P_\infty^t \boldsymbol{w})_{\boldsymbol{k}} = \sum_{\boldsymbol{h} \in \mathbb{Z}^d} p_{\boldsymbol{h}}^{(t)} w_{\boldsymbol{k}-\boldsymbol{h}}$$

where $p_{\boldsymbol{h}}^{(t)}$ are the coefficients of the polynomial $p(z_1, \ldots, z_d)^t$.

If we consider the distributed estimation problem presented in Section 2 for an infinite lattice graph, and we solve it by updating the nodes' estimates with the linear map P described in Eq. (4.1), then the expectation of the quadratic error is the same for any vertex, and thus it makes sense to fix our attention on an arbitrarily chosen one, say vertex $\mathbf{0}$, and to define the MSE cost as follows:

$$J(P_\infty, t) = \mathbb{E}[(e_{\mathbf{0}}(t))^2] . \tag{4.2}$$

4.3. Grids on cubes. The families of Cayley graphs on the group $\Gamma_{\boldsymbol{n}} = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}$ presented in Sect. 4.1 can be seen as regular grids on multi-dimensional tori. An interesting result by Boyd et al. [5] on reversible Markov chains with symmetries allows to compute the eigenvalues and eigenvectors also of regular grids on a cube in $\mathbb{R}^d$ (for example, with $d = 1$ a line, with $d = 2$ a planar finite grid as in Fig. 4.1b). The graphs and the coefficients coincide with those of the regular grid on a torus except that they are suitably modified at the borders. This is particularly relevant because it allows to consider graphs which are still with a regular and idealized structure, but nevertheless are closer than the tori to represent realistic deployments of sensor networks in the Euclidean space.

Roughly speaking, the grids on cubes are obtained by considering an infinite lattice with suitable symmetries, by selecting a portion of the lattice, and then arranging the borders by 'folding back' the edges which are stranded because in the lattice they were connecting a selected vertex to a discarded vertex, then replacing the parallel edges resulting from the folding with a single edge labeled with the sum of the labels

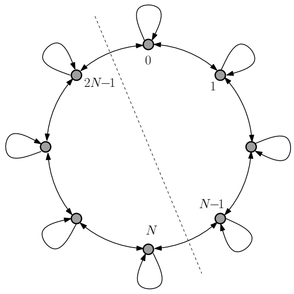


Fig. 4.2: Circle with $2N$ vertices and reflection axis corresponding to the map $\ell \mapsto 2N-1-\ell$, used in the construction of a line with N vertices.

of the parallel edges. However, the construction of the grid matrices can be more precisely defined by considering a Cayley graph with suitable symmetries, and by identifying into a single node the orbits of the action of such symmetries, as described below in detail following [5].

Let $P_{\boldsymbol{2n}}$ be a Cayley matrix on $\Gamma_{\boldsymbol{2n}} = \mathbb{Z}_{2n_1} \times \cdots \times \mathbb{Z}_{2n_d}$ associated with $p(z_1, \ldots, z_d)$, and assume that the coefficients $p_{\boldsymbol{h}}$ satisfy the following quadrantal symmetry

$$\forall \boldsymbol{h}, \quad p_{\boldsymbol{h}} = p_{|\boldsymbol{h}|}. \tag{4.3}$$

This assumption implies that P is symmetric and thus the associated Markov chain is reversible. Moreover, define for each $r = 1, \ldots, d$ the reflection σ_r on $\Gamma_{\boldsymbol{2n}}$ by letting $\sigma_r(\boldsymbol{h}) = \boldsymbol{k}$ with $k_\ell = h_\ell$ if $\ell \neq r$ and $k_r = 2n_r - 1 - h_r$. It is convenient here to identify $\Gamma_{\boldsymbol{2n}}$ with the set $V_{\boldsymbol{2n}}$, and consider $\sigma_r : V_{\boldsymbol{2n}} \to V_{\boldsymbol{2n}}$. For example, Fig. 4.2 shows the axis of reflection of σ_1 for the case $d = 1$, as in Example 1. In higher dimension, every σ_r simply keeps all coordinates invariant except for the r-th, where it is the reflection depicted for the one-dimensional case.

Notice that every σ_r is a symmetry of the labeled graph on the torus. Now denote by H the group generated by all reflections $\sigma_1, \ldots, \sigma_d$ and consider, for all $\boldsymbol{g} \in V_{\boldsymbol{n}} \subseteq V_{\boldsymbol{2n}}$, the orbit $O_{\boldsymbol{g}} = \{\eta(\boldsymbol{g}) : \eta \in H\} \subseteq V_{\boldsymbol{2n}}$. For example, if $d = 1$ there are N orbits, each containing two points: for $g = 0, \ldots, N-1$, the orbit O_g contains the point labeled with g and its reflection $\sigma_1(g) = 2N - 1 - g$. For higher dimension, there are N orbits, each containing 2^d points, for example with $d = 2$, for all $\boldsymbol{g} \in V_{\boldsymbol{n}} = \{0, \ldots, n_1 - 1\} \times \{0, \ldots, n_2 - 1\}$, the corresponding orbit $O_{\boldsymbol{g}}$ contains the four points $\boldsymbol{g}$, $\sigma_1(\boldsymbol{g}) = (2n_1 - 1 - g_1, g_2)$, $\sigma_2(\boldsymbol{g}) = (g_1, 2n_2 - 1 - g_2)$ and $\sigma_2 \circ \sigma_1(\boldsymbol{g}) = (2n_1 - 1 - g_1, 2n_2 - 1 - g_2)$.

Finally, define $\overline{P}_{\boldsymbol{n}} : \mathbb{R}^{V_{\boldsymbol{n}}} \to \mathbb{R}^{V_{\boldsymbol{n}}}$, for all $\boldsymbol{h}, \boldsymbol{k} \in V_{\boldsymbol{n}}$, by

$$(\overline{P}_{\boldsymbol{n}})_{\boldsymbol{h},\boldsymbol{k}} := \sum_{\boldsymbol{\ell} \in O_{\boldsymbol{k}}} (P_{\boldsymbol{2n}})_{\boldsymbol{h},\boldsymbol{\ell}} = \sum_{\eta \in H} (P_{\boldsymbol{2n}})_{\boldsymbol{h},\eta(\boldsymbol{k})} . \tag{4.4}$$

Notice that the entries of $\overline{P}_{\boldsymbol{n}}$ are actually equal to those of $P_{\boldsymbol{n}}$, except at the 'borders'. In fact, if the index $\boldsymbol{k} \in V_{\boldsymbol{n}}$ is such that $\delta < k_r < n_r - \delta$ is satisfied for all $r = 1, \ldots, d$, then for all $\boldsymbol{h} \in V_{\boldsymbol{n}}$ all the terms in the sum in Eq. (4.4) with $\boldsymbol{\ell} \neq \boldsymbol{k}$ are zero, so that $(\overline{P}_{\boldsymbol{n}})_{\boldsymbol{h},\boldsymbol{k}} = (P_{\boldsymbol{2n}})_{\boldsymbol{h},\boldsymbol{k}}$. Also notice that, for this choice of $\boldsymbol{k}$, $(P_{\boldsymbol{2n}})_{\boldsymbol{h},\boldsymbol{k}} = (P_{\boldsymbol{n}})_{\boldsymbol{h},\boldsymbol{k}}$. Moreover, an analogous equality $(\overline{P}_{\boldsymbol{n}})_{\boldsymbol{h},\boldsymbol{k}} = (P_{\boldsymbol{2n}})_{\boldsymbol{h},\boldsymbol{k}} = (P_{\boldsymbol{n}})_{\boldsymbol{h},\boldsymbol{k}}$ holds true whenever $\boldsymbol{h}$ and $\boldsymbol{k}$ are such that, for all $r = 1, \ldots, d$, at least one of the two following conditions is satisfied: $\delta < h_r < n_r - \delta$ or $\delta < k_r < n_r - \delta$.

For example, in the one-dimensional case, $\overline{P}_N$ shares with P_N the banded-diagonal central part of the matrix, and is different only in the initial and final part of the first δ and the last δ rows, for a total of at most $4\delta^2$ different entries. The circulant structure is substituted by modified rows, so that the corresponding graph is a line instead of a circle, and the weight of the edges removed in the construction of the line from the circle is suitably re-distributed along border edges of the line. As an illustrating example, consider the one-dimensional case with $\delta = 2$, where the matrices P_N and $\overline{P}_N$ are the following

$$P_N = \begin{bmatrix} p_0 & p_1 & p_2 & & & p_2 & p_1 \\ p_1 & p_0 & p_1 & p_2 & & & p_2 \\ p_2 & p_1 & p_0 & p_1 & p_2 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & p_2 & p_1 & p_0 & p_1 & p_2 \\ p_2 & & & p_2 & p_1 & p_0 & p_1 \\ p_2 & p_1 & & & p_2 & p_1 & p_0 \end{bmatrix}$$

and

$$\overline{P}_N = \begin{bmatrix} p_0{+}p_1 & p_1{+}p_2 & p_2 & & & & \\ p_1{+}p_2 & p_0 & p_1 & p_2 & & & \\ p_2 & p_1 & p_0 & p_1 & p_2 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & p_2 & p_1 & p_0 & p_1 & p_2 \\ & & & p_2 & p_1 & p_0 & p_1{+}p_2 \\ & & & & p_2 & p_1{+}p_2 & p_0{+}p_1 \end{bmatrix}.$$

We will refer to the above construction of a family of matrices $\overline{P}_{\boldsymbol{n}}$ for all $\boldsymbol{n} \succ 2\delta\mathbf{1}$ as the grid matrix family associated with $p(z_1, \ldots, z_d)$. Note that such a construction ensures that we can apply [5, Prop. 3.3], because both $P_{2\boldsymbol{n}}$ and $\overline{P}_{\boldsymbol{n}}$ are symmetric (and thus the corresponding Markov chain is reversible) and the latter is the lumped chain of the former, as defined in [5, Sect. 3]. Thus, the explicit expression for the eigenvalues of $\overline{P}_{\boldsymbol{n}}$ is the following

$$\bar{\lambda}_{\boldsymbol{h}} = p(e^{i\frac{\pi}{n_1}h_1}, \ldots, e^{i\frac{\pi}{n_d}h_d}), \qquad \boldsymbol{h} \in V_{\boldsymbol{n}}\,.$$

4.4. Assumptions ensuring primitivity. In Section 3.2, a simple assumption on the coefficients (that either p_{-1}, or p_1, or both were non-zero) was introduced, in order to ensure that, for any N, the associated circle graph was strongly connected, and that the infinite line graph was weakly connected. The generalization to the d-dimensional case is a requirement on the positions of the non-zero coefficients $p_{\boldsymbol{h}}$, as stated in the following Lemma.

LEMMA 4.1. *With the above notation, define* $S = \{\boldsymbol{h} \in D_\delta : p_{\boldsymbol{h}} \neq 0\}$. *The following conditions are equivalent:*

1. *the infinite Cayley graph associated with* $p(z)$ *and with the group* $\mathbb{Z}^d$ *is weakly connected;*
2. S *generates* $\mathbb{Z}^d$*;*
3. *for all* $\boldsymbol{n} \succ 2\delta\mathbf{1}$*, the Cayley graph associated with* $p(z)$ *and* $\Gamma_{\boldsymbol{n}}$ *is strongly connected;*
4. *for all* $\boldsymbol{n} \succ 2\delta\mathbf{1}$*,* S *generates* $\Gamma_{\boldsymbol{n}}$*.*

Proof. We start by proving 1. $\iff$ 2. By definition, the graph associated with $p(z_1, \ldots, z_d)$ is weakly connected if and only if for any pair of vertices $u, v \in \mathbb{Z}^d$ there exists a sequence of vertices $u = u_0, u_1, \ldots, u_{\ell-1}, u_\ell = v$ such that, for all i,

(u_i, u_{i+1}) or (u_{i+1}, u_i) is an edge of the directed Cayley graph. This means that $u_{i+1} - u_i \in S \cup (-S)$ for all i so that finally condition 1. turns out to be equivalent to the fact that for all $u, v \in \mathbb{Z}^d$, there exists $\ell \geq 1$ and $s_1, \ldots, s_\ell \in S \cup (-S)$ such that $v - u = s_1 + \cdots + s_\ell$. This is clearly equivalent to condition 2., which states that for all $g \in \mathbb{Z}^d$, there exists $\ell \geq 1$ and $s_1, \ldots, s_\ell \in S \cup (-S)$ such that $g = s_1 + \cdots + s_\ell$.

We omit the proof that 3. $\iff$ 4., because it follows exactly the same lines.

We will conclude by proving that 2. $\iff$ 4. For ease of notation, let $S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_r\}$. Now notice that 4. can be equivalently re-stated as follows. For all $\boldsymbol{n} \succ 2\delta\mathbf{1}$, $\exists X \in \mathbb{Z}^{r \times d}$ and $\exists Y \in \mathbb{Z}^{d \times d}$ such that $I = AX + MY$, where I is the $d \times d$ identity matrix, A is a $d \times r$ matrix whose columns are $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_r$, and M is a diagonal matrix with diagonal elements $n_1, \ldots, n_d$. On the other hand, 2. is equivalent to the fact that $I = AZ$ for some $Z \in \mathbb{Z}^{r \times d}$. This shows that 2. implies 4. To see that also the converse holds true, first notice that 4. implies that A is a full row-rank matrix. Indeed, if $\exists \boldsymbol{z} \in \mathbb{Z}^d$ such that $\boldsymbol{z}^T A = 0$, then $\boldsymbol{z}^T = \boldsymbol{z}^T MY$. Taking in particular $M = bI$ with $b > 2\delta$, we have that $\boldsymbol{z}^T = b\boldsymbol{z}^T Y$, which implies that the entries of $\boldsymbol{z}$ are multiple of b. Since this holds for all $b > 2\delta$, this implies that $\boldsymbol{z} = \mathbf{0}$. Now from the fact that A is a full row-rank matrix it follows that there exists $\bar{X} \in \mathbb{Z}^{r \times d}$ such that $A\bar{X} = aI$, where we can choose $a > 2\delta$. Now observe that 4. implies that there exist $X \in \mathbb{Z}^{r \times d}$ and $Y \in \mathbb{Z}^{d \times d}$ such that $I = AX + aY$ and so $I = AX + aY = AX + A\bar{X}Y = A(X + \bar{X}Y)$ which is equivalent to 2. □

Notice that the second condition in Lemma 4.1 (S generates $\mathbb{Z}^d$) implies that S contains a basis of $\mathbb{R}^d$, and so it states that the connectivity requirement implies that the graph is truly d-dimensional, and not with a lower dimension. Moreover it provides a very easy way to check whether the assumption is satisfied; for example, if S contains all vectors of the canonical basis of $\mathbb{R}^d$ the condition is surely satisfied.

Similarly to the case of the circle, we also assume that the graphs have self-loops, namely that $p_{\mathbf{0}} \neq 0$. This assumption, together with the connectivity assumption above, ensures that the Cayley matrices $P_{\boldsymbol{n}}$ are primitive; also recall that such matrices are doubly-stochastic and normal.

For the grid matrices, notice that the construction ensures that the associated graph is strongly connected when the initial Cayley matrix had this property. Thus, under the connectivity assumption above and with $p_{\mathbf{0}} \neq 0$, the grid matrices are primitive. Moreover, grid matrices are doubly-stochastic and symmetric.

5. Main results. In this section, we give our results on the asymptotics of $J(P_N, t)$ for the graph families introduced in Sect. 4. They are generalizations of the results presented for the circle in Sect. 3.2, and the proofs use the same ideas, although they require a few technicalities.

5.1. Behavior for increasing N. In this section, we give a characterization of the cost which generalizes Prop. 3.1, and from which it is possible to derive corollaries describing the behavior of the MSE cost when the number of nodes increases.

Given a Laurent polynomial $p(z_1, \ldots, z_d) = \sum_{\boldsymbol{h} \in D_\delta} p_{\boldsymbol{h}} z_1^{h_1} \ldots z_d^{h_d}$, we define the polynomial $q(z_1, \ldots, z_d) := p(z_1, \ldots, z_d) p(z_1^{-1}, \ldots, z_d^{-1})$ and we denote by $\{p_{\boldsymbol{h}}^{(t)}\}_{\boldsymbol{h} \in D_{t\delta}}$ and $\{q_{\boldsymbol{h}}^{(t)}\}_{\boldsymbol{h} \in D_{2t\delta}}$ the coefficients of $(p(z_1, \ldots, z_d))^t$ and $(q(z_1, \ldots, z_d))^t$, respectively. With this notation it is possible to characterize the cost $J(P, t)$ in a way that involves the coefficients $q_{\boldsymbol{h}}^{(t)}$ only.

PROPOSITION 5.1. *With the above notation, given a polynomial $p(z_1, \ldots, z_d)$ and $\boldsymbol{n} = (n_1, \ldots, n_d) \succ 2\delta\mathbf{1}$,*

- *if P_∞ is the infinite map associated with $p(z_1,\dots,z_d)$, then*
$$J(P_\infty,t)=q^{(t)}_{\mathbf{0}}$$
- *if $P_{\boldsymbol{n}}$ is a Cayley matrix associated with $p(z_1,\dots,z_d)$, then*
$$J(P_{\boldsymbol{n}},t)=\sum_{\boldsymbol{h}\in\mathcal{F}_{\boldsymbol{n}}} q^{(t)}_{\boldsymbol{h}},$$
where $\mathcal{F}_{\boldsymbol{n}}:=\{\boldsymbol{h}:\ h_r=0 \bmod n_r\ \forall r\}$.
- *if $p(z_1,\dots,z_d)$ satisfies the quadrantal symmetries (4.3) and $\overline{P}_{\boldsymbol{n}}$ is a grid matrix associated with $p(z_1,\dots,z_d)$, then*
$$J(\overline{P}_{\boldsymbol{n}},t)=\sum_{K\subseteq[d]}\frac{1}{\prod_{r\in K}n_r}\sum_{\boldsymbol{h}\in\mathcal{F}_{K,\boldsymbol{n}}} q^{(t)}_{\boldsymbol{h}},$$
where the first summation is over all subsets $K\subseteq[d]:=\{1,\dots,d\}$, including $K=\emptyset$ and $K=[d]$, and where
$$\mathcal{F}_{K,\boldsymbol{n}}:=\{(h_1,\dots,h_d): h_r \textit{ is odd } \forall r\in K \textit{ and } h_r=0 \bmod 2n_r\ \forall r\notin K\}.$$

Proof. We consider separately the three families of graphs.
Infinite lattice. From Eq. (4.2) it follows that
$$J(P_\infty,t)=\mathbb{E}\big[\big((P_\infty^t\boldsymbol{w})_{\mathbf{0}}\big)^2\big].$$
Now notice that $(P_\infty^t\boldsymbol{w})_{\boldsymbol{k}}=\sum_{\boldsymbol{h}\in\mathbb{Z}^d}p^{(t)}_{\boldsymbol{h}}w_{\boldsymbol{k}-\boldsymbol{h}}$. Therefore,
$$\mathbb{E}\big[\big((P_\infty^t\boldsymbol{w})_{\mathbf{0}}\big)^2\big]=\mathbb{E}\big[\sum_{\boldsymbol{h},\boldsymbol{k}\in\mathbb{Z}^d}p^{(t)}_{\boldsymbol{h}}p^{(t)}_{\boldsymbol{k}}w_{-\boldsymbol{h}}w_{-\boldsymbol{k}}\big]=\sum_{\boldsymbol{h}\in\mathbb{Z}^d}\big(p^{(t)}_{\boldsymbol{h}}\big)^2=q^{(t)}_{\mathbf{0}}.$$

Cayley matrix. In this case,
$$\begin{aligned}J(P_{\boldsymbol{n}},t)&=\frac{1}{N}\sum_{h_1=0}^{n_1-1}\cdots\sum_{h_d=0}^{n_d-1}\sum_{\boldsymbol{k}\in D_{2t\delta}}q^{(t)}_{\boldsymbol{k}}e^{i(\frac{2\pi}{n_1}h_1k_1+\cdots+\frac{2\pi}{n_d}h_dk_d)}\\&=\sum_{\boldsymbol{k}\in D_{2t\delta}}q^{(t)}_{\boldsymbol{k}}\prod_{r=1}^{d}\frac{1}{n_r}\sum_{h_r=0}^{n_r-1}e^{i\frac{2\pi}{n_r}h_rk_r},\end{aligned}$$
which ends the proof, because
$$\frac{1}{n_r}\sum_{h_r=0}^{n_r-1}e^{i\frac{2\pi}{n_r}h_rk_r}=\begin{cases}1 & \text{if } k_r=0 \bmod n_r,\\ 0 & \text{otherwise.}\end{cases}$$

Grid matrix. For the grid matrix,
$$J(\overline{P}_{\boldsymbol{n}},t)=\frac{1}{N}\sum_{\boldsymbol{k}\in D_{2t\delta}}q^{(t)}_{\boldsymbol{k}}\sum_{h_1=0}^{n_1-1}\cdots\sum_{h_d=0}^{n_d-1}e^{i(\frac{\pi}{n_1}h_1k_1+\cdots+\frac{\pi}{n_d}h_dk_d)}.\tag{5.1}$$
Now we use the assumption of quadrantal symmetries: denoting by $\odot$ the entrywise product of two vectors, we have $q_{\boldsymbol{k}}=q_{\boldsymbol{\omega}\odot\boldsymbol{k}}$ for all $\boldsymbol{\omega}\in\{-1,1\}^d$, and so
$$q_{\boldsymbol{k}}=\frac{1}{2^d}\sum_{\boldsymbol{\omega}\in\{-1,1\}^d}q^{(t)}_{\boldsymbol{\omega}\odot\boldsymbol{k}}.$$

By plugging this into Eq. (5.1), we get

$$J(\overline{P}_{\boldsymbol{n}}, t) = \sum_{\boldsymbol{k}\in D_{2t\delta}} \sum_{\boldsymbol{\omega}\in\{-1,1\}^d} q^{(t)}_{\boldsymbol{\omega}\odot\boldsymbol{k}} \prod_{r=1}^{d} \frac{1}{2n_r} \sum_{h_r=0}^{n_r-1} e^{i\frac{\pi}{n_r}h_r k_r},$$

from which, by exchanging the order of summations and letting $\boldsymbol{k}' := \boldsymbol{\omega}\odot\boldsymbol{k}$, we obtain

$$\begin{aligned} J(\overline{P}_{\boldsymbol{n}}, t) &= \sum_{\boldsymbol{\omega}\in\{-1,1\}^d} \sum_{\boldsymbol{k}'\in D_{2t\delta}} q^{(t)}_{\boldsymbol{k}'} \prod_{r=1}^{d} \frac{1}{2n_r} \sum_{h_r=0}^{n_r-1} e^{i\frac{\pi}{n_r}h_r k'_r \omega_r} \\ &= \sum_{\boldsymbol{k}'\in D_{2t\delta}} q^{(t)}_{\boldsymbol{k}'} \prod_{r=1}^{d} \frac{1}{2n_r} \sum_{h_r=0}^{n_r-1} \sum_{\omega_r\in\{-1,1\}} e^{i\frac{\pi}{n_r}h_r k'_r \omega_r}. \end{aligned}$$

The proof ends by computing

$$\sum_{h_r=0}^{n_r-1} \left(e^{i\frac{\pi}{n_r}h_r k'_r} + e^{-i\frac{\pi}{n_r}h_r k'_r} \right) = \begin{cases} 2n_r & \text{if } k'_r = 0 \bmod 2n_r, \\ 2 & \text{if } k'_r \text{ is odd}, \\ 0 & \text{otherwise}. \end{cases} \qquad \square$$

The following corollary (generalizing Coroll. 3.2) shows that the behavior of large Cayley or grid matrices with a large number of nodes, at a fixed iteration time t tends to the cost of the corresponding infinite lattice.

COROLLARY 5.2. *Under the assumptions of Proposition 5.1, if $P_{\boldsymbol{n}}$ is a Cayley matrix, then in the limit when $n_r \to \infty$ for all $r = 1, \dots, d$, $J(P_{\boldsymbol{n}}, t) \to J(P_\infty, t)$ and moreover $J(P_{\boldsymbol{n}}, t) = J(P_\infty, t)$ for all $\boldsymbol{n} \succ 2t\delta\mathbf{1}$. In the case of a grid matrix $\overline{P}_{\boldsymbol{n}}$, only the limit result holds true.*

Proof. For the Cayley matrix $P_{\boldsymbol{n}}$, in the expression for $J(P_{\boldsymbol{n}}, t)$ given in Prop. 5.1, notice that the only coefficients $q^{(t)}_{\boldsymbol{h}}$ in the summation which can be non-zero are those where $\boldsymbol{h} \in D_{2t\delta}$, namely $-2t\delta \le h_r \le 2t\delta$ for all $r = 1, \dots, d$. If $\boldsymbol{n} \succ 2\delta\mathbf{1}$, then $D_{2t\delta} \cap \mathcal{F}_{\boldsymbol{n}} = \{\mathbf{0}\}$, so that $J(P_{\boldsymbol{n}}, t) = q^{(t)}_{\mathbf{0}} = J(P_\infty, t)$.

For the grid matrix $P_{\boldsymbol{n}}$, the expression for $J(\overline{P}_{\boldsymbol{n}}, t)$ given in Prop. 5.1 can be re-written as

$$J(\overline{P}_{\boldsymbol{n}}, t) = \sum_{\boldsymbol{h}\in\mathcal{F}_{\emptyset,\boldsymbol{n}}} q^{(t)}_{\boldsymbol{h}} + \sum_{\substack{K\subseteq[d]\\K\neq\emptyset}} \frac{1}{\prod_{r\in K} n_r} \sum_{\boldsymbol{h}\in\mathcal{F}_{K,\boldsymbol{n}}} q^{(t)}_{\boldsymbol{h}}.$$

For the first summation, we can conclude analogously to the Cayley case that if $\boldsymbol{n} \succ \delta\mathbf{1}$ then the only non-zero term is $q^{(t)}_{\mathbf{0}} = J(P_\infty, t)$. Finally, all summations corresponding to a non-empty K go to zero when $n_r \to \infty$ for all r. $\square$

Under some assumptions, Proposition 5.1 also implies that the MSE cost is a monotone function of the number of nodes (for fixed t).

COROLLARY 5.3. *Under the assumptions of Proposition 5.1, if $p(z_1, \dots, z_d)$ satisfies the quadrantal symmetries (4.3) and satisfies the following monotonicity assumption*

$$|\boldsymbol{h}| \succcurlyeq |\boldsymbol{k}| \;\Rightarrow\; p_{\boldsymbol{h}} \le p_{\boldsymbol{k}},$$

then, for the family of Cayley matrices $P_{\boldsymbol{n}}$ associated with $p(z_1, \dots, z_d)$, the cost $J(P_{\boldsymbol{n}}, t)$ is monotonic non-increasing w.r.t. $n_1, \dots, n_d$, namely

$$\boldsymbol{m} \preccurlyeq \boldsymbol{n} \;\Rightarrow\; J(P_{\boldsymbol{m}}, t) \ge J(P_{\boldsymbol{n}}, t),\ \forall t.$$

The same property holds true for the family of grid matrices $\overline{P}_{\boldsymbol{n}}$ associated with $p(z_1,\dots,z_d)$.

Proof. This property is a consequence of Proposition 5.1 and of Lemma A.1, whose proof is postponed to the Appendix, which ensures that, if $p(z_1,\dots,z_d)$ satisfies the assumptions of Coroll. 5.3, then also $\left(q(z_1,\dots,z_d)\right)^t$ satisfies them.

We want to prove that, under the assumptions of Coroll. 5.3, $\boldsymbol{n} \succcurlyeq \boldsymbol{m}$ implies $J(P_{\boldsymbol{n}},t) \le J(P_{\boldsymbol{m}},t)$ and $J(\overline{P}_{\boldsymbol{n}},t) \le J(\overline{P}_{\boldsymbol{m}},t)$. For ease of notation, without loss of generality, we will consider the case where $\boldsymbol{n} = (n_1, \boldsymbol{n'})$ and $\boldsymbol{m} = (m_1, \boldsymbol{n'})$, with $n_1 \ge m_1$. The key point we will exploit is that, by the assumptions and by Lemma A.1, for all $h_1, k_1 \in \mathbb{Z}$ and for all $\boldsymbol{h'} \in \mathbb{Z}^{d-1}$, $|h_1| \ge |k_1|$ implies that $q^{(t)}_{h_1,\boldsymbol{h'}} \le q^{(t)}_{k_1,\boldsymbol{h'}}$.

For the Cayley case, by Prop. 5.1,

$$J(P_{\boldsymbol{n}},t) = \sum_{\boldsymbol{h}\in\mathcal{F}_{\boldsymbol{n}}} q^{(t)}_{\boldsymbol{h}},$$

where $\mathcal{F}_{\boldsymbol{n}} := \{\boldsymbol{h} :\ h_r = 0 \bmod n_r\ \forall r\}$. We only need to re-write this summation highlighting the first component n_1 of $\boldsymbol{n}$, and to compare it with the analogous expression for $J(P_{\boldsymbol{m}},t)$, as follows

$$J(P_{\boldsymbol{n}},t) = \sum_{\ell\in\mathbb{Z}} \sum_{\boldsymbol{h'}\in\mathcal{F}_{\boldsymbol{n'}}} q^{(t)}_{\ell n_1,\boldsymbol{h'}} \le \sum_{\ell\in\mathbb{Z}} \sum_{\boldsymbol{h'}\in\mathcal{F}_{\boldsymbol{n'}}} q^{(t)}_{\ell m_1,\boldsymbol{h'}} = J(P_{\boldsymbol{m}},t)\,.$$

For the grid case, by Prop. 5.1,

$$J(\overline{P}_{\boldsymbol{n}},t) = \sum_{K\subseteq[d]} \frac{1}{\prod_{r\in K} n_r} \sum_{\boldsymbol{h}\in\mathcal{F}_{K,\boldsymbol{n}}} q^{(t)}_{\boldsymbol{h}}\,,$$

where $\mathcal{F}_{K,\boldsymbol{n}} := \{(h_1,\dots,h_d) : h_r \text{ is odd } \forall r \in K \text{ and } h_r = 0 \bmod 2n_r\ \forall r \notin K\}$. Now we can consider separately each term corresponding to a set $K \subseteq [d]$, and compare it with the corresponding term in the analogous expression for $J(\overline{P}_{\boldsymbol{m}},t)$. If $1 \notin K$, then

$$\frac{1}{\prod_{r\in K} n_r} = \frac{1}{\prod_{r\in K} m_r}$$

and, defining $K' := \{r-1 : r \in K\}$,

$$\sum_{\boldsymbol{h}\in\mathcal{F}_{K,\boldsymbol{n}}} q^{(t)}_{\boldsymbol{h}} = \sum_{\ell\in\mathbb{Z}} \sum_{\boldsymbol{h'}\in\mathcal{F}_{K',\boldsymbol{n'}}} q^{(t)}_{\ell n_1,\boldsymbol{h'}} \le \sum_{\ell\in\mathbb{Z}} \sum_{\boldsymbol{h'}\in\mathcal{F}_{K',\boldsymbol{n'}}} q^{(t)}_{\ell m_1,\boldsymbol{h'}} = \sum_{\boldsymbol{h}\in\mathcal{F}_{K,\boldsymbol{m}}} q^{(t)}_{\boldsymbol{h}}\,.$$

If $1 \in K$, then

$$\frac{1}{\prod_{r\in K} n_r} \le \frac{1}{\prod_{r\in K} m_r}$$

and $\mathcal{F}_{K,\boldsymbol{n}} = \mathcal{F}_{K,\boldsymbol{m}}$, which ends the proof. ∎

Note that the assumptions on the polynomial p in Coroll. 5.3 are not necessary: for example, with $d=1$, $p(z) = \frac{1}{3}z^{-2} + \frac{1}{18}z^{-1} + \frac{1}{9} + \frac{1}{6}z + \frac{1}{3}z^2$ violates both assumptions and nevertheless gives a monotonic cost $J(P_{\boldsymbol{n}},t)$. However, monotonicity of $J(P_{\boldsymbol{n}},t)$ is not always true, for example, $p(z) = \frac{2}{5}z^{-2} + \frac{1}{25}z^{-1} + \frac{3}{10} + \frac{1}{4}z + \frac{1}{100}z^2$ gives a non-monotonic cost.

5.2. Behavior for increasing N and t. A further step in the analysis is to understand the exact scaling of $J(P_N, t)$ when both t and N grow to infinity. Such an asymptotic behavior is given in the following theorem, which generalizes Theorem 3.3.

THEOREM 5.4. *Given a polynomial $p(\boldsymbol{z}) = \sum_{\boldsymbol{h}\in D_\delta} p_{\boldsymbol{h}} z_1^{h_1} \dots z_d^{h_d}$ such that the corresponding infinite graph is weakly connected and has self-loops and given $n_1 \geq n_2 \geq \cdots \geq n_d > 2\delta$, then*

- *if P_∞ is the infinite map associated with $p(z_1, \dots, z_d)$, then there exist positive constants c, c', depending on d and $p(z_1, \dots, z_d)$ only, such that, for all $t > 0$*
$$c\frac{1}{(\sqrt{t})^d} \leq J(P_\infty, t) \leq c'\frac{1}{(\sqrt{t})^d};$$
- *if $P_{\boldsymbol{n}}$ is a Cayley matrix associated with $p(z_1, \dots, z_d)$, then there exist positive constants C, C', depending on d and $p(z_1, \dots, z_d)$ only, such that, for all $\boldsymbol{n} \succ 2\delta\mathbf{1}$ and $t > 0$*
$$C \max_{0\leq k\leq d} \frac{\prod_{1\leq r\leq k} n_r}{N}\frac{1}{(\sqrt{t})^k} \leq J(P_{\boldsymbol{n}}, t) \leq C' \max_{0\leq k\leq d} \frac{\prod_{1\leq r\leq k} n_r}{N}\frac{1}{(\sqrt{t})^k};$$
- *if $p(z_1, \dots, z_d)$ satisfies the quadrantal symmetries (4.3) and $\overline{P}_{\boldsymbol{n}}$ is a grid matrix associated with $p(z_1, \dots, z_d)$, then there exist constants $\overline{C}, \overline{C}'$, depending on d and $p(z)$ only, such that, for all $\boldsymbol{n} \succ 2\delta\mathbf{1}$ and $t > 0$*
$$\overline{C} \max_{0\leq k\leq d} \frac{\prod_{1\leq r\leq k} n_r}{N}\frac{1}{(\sqrt{t})^k} \leq J(\overline{P}_{\boldsymbol{n}}, t) \leq \overline{C}' \max_{0\leq k\leq d} \frac{\prod_{1\leq r\leq k} n_r}{N}\frac{1}{(\sqrt{t})^k}.$$

The rest of this section is devoted to the proof of this result, but before going through the proof it is interesting to see that, in the case when $n_1 = \cdots = n_d$, the statement can be re-written in the following simpler way.

COROLLARY 5.5. *If $n_r = n$ for all $r = 1, \dots, d$, under the assumptions of Theorem 5.4, if $P_{\boldsymbol{n}}$ is a Cayley matrix associated with $p(z_1, \dots, z_d)$, then there exist $C, C' > 0$ (depending on d and and $p(z_1, \dots, z_d)$ only) such that*
$$C \max\left\{\tfrac{1}{N}, \tfrac{1}{(\sqrt{t})^d}\right\} \leq J(P_{\boldsymbol{n}}, t) \leq C' \max\left\{\tfrac{1}{N}, \tfrac{1}{(\sqrt{t})^d}\right\}$$
An analogous bound holds true for the grid matrix $\overline{P}_{\boldsymbol{n}}$ associated with $p(z_1, \dots, z_d)$.

The proof of Theorem 5.4 follows the lines of the proof of Theorem 3.3, and is based on the study of the function $f : \mathbb{R}^d \to [0, +\infty)$ defined by
$$f(\boldsymbol{x}) = |p(e^{ix_1}, \dots, e^{ix_d})|^2.$$
Notice that $f(\boldsymbol{x}) = q(e^{ix_1}, \dots, e^{ix_d}) = \sum_{\boldsymbol{\ell}\in D_{2\delta}} q_{\boldsymbol{\ell}} \cos\big(\ell_1 x_1 + \cdots + \ell_d x_d\big)$, where $q(z_1, \dots, z_d) = p(z_1, \dots, z_d)p(z_1^{-1}, \dots, z_d^{-1})$.

Clearly f is a trigonometric polynomial, with $f(\mathbf{0}) = 1$ and $0 \leq f(\boldsymbol{x}) \leq 1$ for all $\boldsymbol{x}$. Under the assumptions of Theorem 5.4 we can also guarantee that the maximum in $\boldsymbol{x} = \mathbf{0}$ is unique in the region $(-2\pi, 2\pi)^d$. To this aim, notice that, by Lemma 4.1, the assumption that the infinite Cayley graph on $\mathbb{Z}^d$ associated with $p(z_1, \dots, z_d)$ is weakly connected and has self-loops is equivalent to the assumption that the support set of $p(z_1, \dots, z_d)$, defined as $S(p) := \{\boldsymbol{k} \in D_\delta : p_{\boldsymbol{k}} \neq 0\}$, contains the origin and generates $\mathbb{Z}^d$. Under such assumption, we can prove the following properties of the maximum of f in the origin.

LEMMA 5.6. *With the above notation and assumptions,*

$$f(\boldsymbol{x}) < 1 \text{ for all } \boldsymbol{x} \in (-2\pi, 2\pi)^d \setminus \{\mathbf{0}\}$$

and moreover the Hessian matrix of f in $\boldsymbol{x} = \mathbf{0}$ is negative definite.

Proof. Let $S(q)$ be the support of $q(z_1, \ldots, z_d)$. Assume that $S(q) = \{\boldsymbol{\ell}^{(1)}, \ldots, \boldsymbol{\ell}^{(s)}\}$ and let $L \in \mathbb{Z}^{d\times s}$ be the matrix whose columns are $\boldsymbol{\ell}^{(1)}, \ldots, \boldsymbol{\ell}^{(s)}$. Since $S(q) \supseteq S(p)$, the assumptions ensure that $\boldsymbol{\ell}^{(1)}, \ldots, \boldsymbol{\ell}^{(s)}$ generate $\mathbb{Z}^d$ and so there exists $Y \in \mathbb{Z}^{s\times d}$ such that $LY = I$. Assume now that $\boldsymbol{x} \in (-2\pi, 2\pi)^d$ is such that $f(\boldsymbol{x}) = 1$. It follows that, for all $i = 1, \ldots, s$, $(\boldsymbol{\ell}^{(i)})^T \boldsymbol{x} = 2\pi b_i$ where $b_i \in \mathbb{Z}$. This implies that $L^T \boldsymbol{x} = 2\pi \boldsymbol{b}$ where $\boldsymbol{b} \in \mathbb{Z}^s$. Consequently $\boldsymbol{x} = Y^T L^T \boldsymbol{x} = 2\pi Y^T \boldsymbol{b}$ which, recalling that $\boldsymbol{x} \in (-2\pi, 2\pi)^d$, implies that $\boldsymbol{x} = 0$.

For the second claim, denote by H the Hessian matrix of f in $\mathbf{0}$, which is given by $H_{r,s} := \frac{\partial^2 f}{\partial x_r \partial x_s} = -\sum_{\boldsymbol{h}\in S(q)} q_{\boldsymbol{h}} \boldsymbol{h}_r \boldsymbol{h}_s$. Our aim is to prove that $-H$ is positive definite. First observe that $H = -\sum_{\boldsymbol{h}\in S(q)} q_{\boldsymbol{h}} \boldsymbol{h}\boldsymbol{h}^T = -LDL^T$ where L is the matrix defined in the first part of the proof and where D is a $s \times s$ diagonal matrix with diagonal entries equal to $q_{\boldsymbol{\ell}^{(1)}}, \ldots, q_{\boldsymbol{\ell}^{(s)}}$. Since D is positive definite, and since L has full rank, LDL^T is positive definite. ∎

The following bound on $f(\boldsymbol{x})$ is an immediate consequence of Lemma 5.6.

LEMMA 5.7. *Under the assumptions of Lemma 5.6, there exists $a \in (0, \pi)$, $\alpha, \beta > 0$, and $c \in (0, 1)$, depending only on $p(\boldsymbol{z})$ and d, such that, for all $\boldsymbol{x} \in [-\pi, \pi]^d$,*

$$f_{\mathrm{L}}(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq f_{\mathrm{U}}(\boldsymbol{x}),$$

where the functions f_{U} and f_{L} are defined as

$$f_{\mathrm{U}}(\boldsymbol{x}) = \begin{cases} e^{-\alpha\, \boldsymbol{x}^T \boldsymbol{x}} & \text{for } \boldsymbol{x} \in (-a, a)^d \\ c & \text{otherwise}, \end{cases} \qquad f_{\mathrm{L}}(\boldsymbol{x}) = \begin{cases} e^{-\beta\, \boldsymbol{x}^T \boldsymbol{x}} & \text{for } \boldsymbol{x} \in (-a, a)^d \\ 0 & \text{otherwise}. \end{cases}$$

Now we can use the bounds on f to find bounds for the MSE cost. We will consider separately the three cases.

Infinite lattice. From Eq. (4.1), we have $J(P_\infty, t) = \sum_{\boldsymbol{h}\in\mathbb{Z}^d} |p_{\boldsymbol{h}}^{(t)}|^2$. By Parseval's identity, this expression can be re-written as

$$J(P_\infty, t) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} |p^t(e^{ix_1}, \ldots, e^{ix_d})|^2 \,\mathrm{d}\boldsymbol{x} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} (f(x_1, \ldots, x_d))^t \,\mathrm{d}\boldsymbol{x}.$$

Then, an upper and a lower bound which conclude the proof can be obtained by using Lemma 5.7 and then applying well-known properties of the tail of a Gaussian distribution, see Lemma A.2 in the Appendix for more detail.

Cayley matrix. In this case,

$$J(P_{\boldsymbol{n}}, t) = \frac{1}{N} \sum_{\boldsymbol{h}\in V_{\boldsymbol{n}}} |\lambda_{\boldsymbol{h}}|^{2t} = \frac{1}{N} \sum_{\boldsymbol{h}\in V_{\boldsymbol{n}}} \left[f\left(\tfrac{2\pi}{n_1} h_1, \ldots, \tfrac{2\pi}{n_d} h_d \right) \right]^t.$$

Define

$$V'_{\boldsymbol{n}} := \left\{ -\left\lfloor \tfrac{n_1-1}{2} \right\rfloor, \ldots, 0, \ldots, +\left\lfloor \tfrac{n_1}{2} \right\rfloor \right\} \times \cdots \times \left\{ -\left\lfloor \tfrac{n_d-1}{2} \right\rfloor, \ldots, 0, \ldots, +\left\lfloor \tfrac{n_d}{2} \right\rfloor \right\}.$$

Clearly, $f(\boldsymbol{x})$ has period 2π in each of its variables, and so

$$J(P_{\boldsymbol{n}}, t) = \frac{1}{N} \sum_{\boldsymbol{h}\in V'_{\boldsymbol{n}}} \left[f\left(\tfrac{2\pi}{n_1} h_1, \ldots, \tfrac{2\pi}{n_d} h_d \right) \right]^t.$$

From Lemma 5.7 it follows

$$J(P_{\boldsymbol{n}},t) \leq \frac{1}{N} \sum_{\boldsymbol{h}\in V'_{\boldsymbol{n}}} \left[f_{\mathrm{U}} \left(\tfrac{2\pi}{n_1} h_1, \ldots, \tfrac{2\pi}{n_d} h_d \right) \right]^t ,$$

where f_{U} is defined in Lemma 5.7. Now consider the following set (recall that $a < \pi$):

$$V''_{\boldsymbol{n}} := \left\{ \boldsymbol{h} \in V'_{\boldsymbol{n}} \ : \ -a \leq \tfrac{2\pi}{n_r} h_r \leq a \ \forall r \right\} = \left\{ \boldsymbol{h} \in \mathbb{Z}^d \ : \ -\left\lfloor \tfrac{a n_r}{2\pi} \right\rfloor \leq h_r \leq \left\lfloor \tfrac{a n_r}{2\pi} \right\rfloor \ \forall r \right\} .$$

Then, using the definition of f_{U}, we get

$$J(P_{\boldsymbol{n}},t) \leq \frac{1}{N} \sum_{\boldsymbol{h}\in V''_{\boldsymbol{n}}} e^{-\alpha \left[\left(\frac{2\pi}{n_1} h_1 \right)^2 + \cdots + \left(\frac{2\pi}{n_d} h_d \right)^2 \right] t} + c^t.$$

We conclude the proof by approximating such a Riemann sum with an integral, similarly to the technique used in the proof of Theorem 3.3 (see Eq. (3.1)), as explained in detail in Lemma A.3 in the Appendix.

For the lower bound, the proof is very similar. Indeed,

$$\begin{aligned} J(P_{\boldsymbol{n}},t) &= \frac{1}{N} \sum_{\boldsymbol{h}\in V'_{\boldsymbol{n}}} \left[f \left(\tfrac{2\pi}{n_1} h_1, \ldots, \tfrac{2\pi}{n_d} h_d \right) \right]^t \\ &\geq \frac{1}{N} \sum_{\boldsymbol{h}\in V'_{\boldsymbol{n}}} \left[f_{\mathrm{L}} \left(\tfrac{2\pi}{n_1} h_1, \ldots, \tfrac{2\pi}{n_d} h_d \right) \right]^t \\ &= \frac{1}{N} \sum_{\boldsymbol{h}\in V''_{\boldsymbol{n}}} e^{-\beta \left[\left(\frac{2\pi}{n_1} h_1 \right)^2 + \cdots + \left(\frac{2\pi}{n_d} h_d \right)^2 \right] t}. \end{aligned}$$

Finally, the conclusion is obtained by using Lemma A.3 (see the Appendix).

Grid matrices. The proof is very similar to the one for Cayley matrices. In this case,

$$J(\overline{P}_{\boldsymbol{n}},t) = \frac{1}{N} \sum_{\boldsymbol{h}\in V_{\boldsymbol{n}}} \left[f \left(\tfrac{\pi}{n_1} h_1, \ldots, \tfrac{\pi}{n_d} h_d \right) \right]^t ,$$

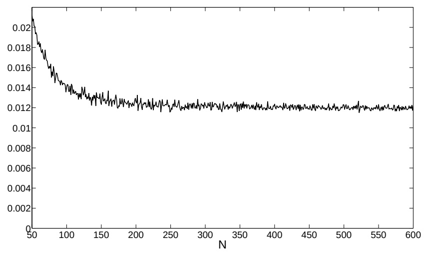
and so

$$\begin{aligned} J(\overline{P}_{\boldsymbol{n}},t) &\leq \frac{1}{N} \sum_{\boldsymbol{h}\in V_{\boldsymbol{n}} \cap V''_{2\boldsymbol{n}}} e^{-\alpha \left[\left(\frac{\pi}{n_1} h_1 \right)^2 + \cdots + \left(\frac{\pi}{n_d} h_d \right)^2 \right] t} + c^t \\ &\leq \frac{1}{N} \sum_{\boldsymbol{h}\in V''_{2\boldsymbol{n}}} e^{-\alpha \left[\left(\frac{\pi}{n_1} h_1 \right)^2 + \cdots + \left(\frac{\pi}{n_d} h_d \right)^2 \right] t} + c^t \end{aligned}$$

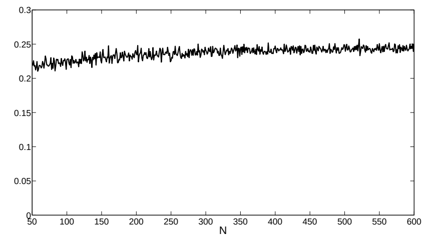
and

$$\begin{aligned} J(\overline{P}_{\boldsymbol{n}},t) &\geq \frac{1}{N} \sum_{\boldsymbol{h}\in V_{\boldsymbol{n}} \cap V''_{2\boldsymbol{n}}} e^{-\beta \left[\left(\frac{\pi}{n_1} h_1 \right)^2 + \cdots + \left(\frac{\pi}{n_d} h_d \right)^2 \right] t} \\ &\geq \frac{1}{N} \frac{1}{2^d} \sum_{\boldsymbol{h}\in V''_{2\boldsymbol{n}}} e^{-\beta \left[\left(\frac{\pi}{n_1} h_1 \right)^2 + \cdots + \left(\frac{\pi}{n_d} h_d \right)^2 \right] t}. \end{aligned}$$

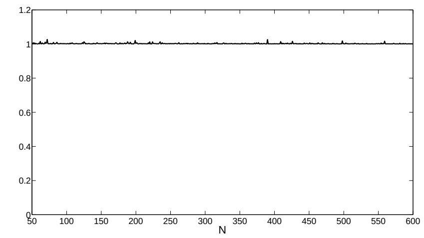
Then again the conclusion comes from Lemma A.3 (see the Appendix).



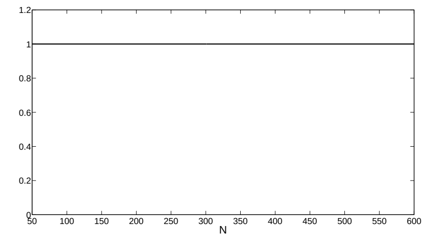
(a) $J(P_N, 20)$



(b) $\sqrt{N}\, J(P_N, \sqrt{N})$



(c) $N\, J(P_N, N)$



(d) $N\, J(P_N, N^{3/2})$

Fig. 6.1: Plots of $J(P_N, t_N)$ for random geometric graphs (averaged), as a function of N, for various computation times (differently growing w.r.t. N), and re-scaled by the predicted factor of decay.

6. More general geometric graphs. Our main results concern a class of highly-structured, regular graphs, for which it was possible to derive precise bounds. However, we believe that such results can also provide guidelines for the design of more realistic sensor networks, because deploying agents in a portion of the 2-dimensional or 3-dimensional space, in a roughly uniform way, and with the constraint of local communication (meaning connection only with geometric neighbors, i.e. within some given distance range) results in graphs resembling to portions of lattices, with some additional irregularities. Our conjecture is supported by some simulation results. We consider random geometric graphs, as in the Gilbert model for wireless communication networks ([22], see also the recent book [17]): nodes are placed on a d-dimensional unit cube uniformly at random and then pairs of nodes are connected by an edge if and only if the two are within a given distance r (the resulting graph is undirected). Similarly to the results we have obtained for lattice-like graphs, we are interested in the asymptotic behavior when the number of nodes N increases, while the number of neighbors roughly remains constant. Thus, we choose the distance threshold r in such a way that the average degree is kept constant. Moreover, we only consider connected realizations, discarding disconnected graphs. Then we choose a classical way of associating a consensus matrix P to an undirected graph, the so-called *Metropolis weights* rule [31]. Figure 6.1 provides examples of numerical results for the 2-dimensional case, where the behavior of the Cayley graph predicts a cost scaling as $\max\left\{\frac{1}{N}, \frac{1}{t}\right\}$ (Corollary 5.5). We plot $J(P_N, t)$ as a function of N, for various choices of growth of t w.r.t. N (respectively, constant $t = 20$, $t = \sqrt{N}$, $t = N$, $t = N^{3/2}$), and then we already pre-multiply $J(P_N, t)$ by the predicted scaling factor, so that a perfectly flat line represents the asymptotic predicted behavior. In average, for large N the prediction turns out to be quite accurate.

7. Conclusions. In this paper the behavior of an estimation performance index is analyzed. More precisely, it is studied how this index varies with the number of nodes and the number of iterations. In this way it is possible to determine the minimum number of iterations which allow to exploit the estimation power of a sensor network. The limitation of these results is given by the fact that they apply only to regular grids. However simulation results (presented in Section 6) show that connected realizations of random geometric graphs with a comparable number of nodes and of average number of neighbors exhibit a behavior very similar to the corresponding Cayley graphs. This suggests that for those graphs the performance index behaves similarly as for regular grids.

A mathematical proof of this fact seems not to be trivial since studying the properties of graphs which are 'small perturbations' of known graphs is not a trivial task. First, classical literature on small perturbation of matrices does not apply, as here 'small' is meant as a significant modification of a little number of entries compared to the size, not as a infinitesimal variation of each entry. Secondly, suitable assumptions should be made on the perturbation so as to rule out those strongly affecting performance, e.g., disconnecting the graph. The goal of rigorously characterizing the behavior of large classes of 'grid-like' graphs is left as an open and interesting research area.

Appendix. In this appendix we present the proofs of some technical lemmas that were used in Section 5.

The following Lemma is the main tool in the proof of Coroll. 5.3. It ensures that if $p(z_1, \dots, z_d)$ satisfies the assumptions of Coroll. 5.3, then also $q^{(t)}(z_1, \dots, z_d)$ satisfies them (notice that $q^{(t)}(z_1, \dots, z_d) = (p(z_1, \dots, z_d))^{2t}$ due to the quadrantal symmetries).

LEMMA A.1. *Given two sequences of non-negative numbers* $\{a_{\boldsymbol{h}}\}, \{b_{\boldsymbol{h}}\} \in \mathbb{R}^{\mathbb{Z}^d}$ *satisfying the assumptions of Coroll. 5.3, then also their convolution* $\{c_{\boldsymbol{h}}\}$ *defined by* $c_{\boldsymbol{h}} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} a_{\boldsymbol{k}} b_{\boldsymbol{h}-\boldsymbol{k}}$ *satisfies the same assumptions.*

Proof. Non-negativity and quadrantal symmetries are immediate. What is left to prove is that, for all $i = 1, \dots, d$, if $k_j = h_j \geq 0$ for all $j \neq i$ and $k_i = h_i + 1 \geq 1$, then $c_{\boldsymbol{k}} \leq c_{\boldsymbol{h}}$. For ease of notation, we give the proof with $i = 1$, and we write indexes as $\boldsymbol{h} = (h, \boldsymbol{h}')$, with $h \in \mathbb{Z}$, $\boldsymbol{h}' \in \mathbb{Z}^{d-1}$, so that what we want to prove is that, for all $\boldsymbol{h} = (h, \boldsymbol{h}') \succcurlyeq \boldsymbol{0}$, $c_{h,\boldsymbol{h}'} - c_{h+1,\boldsymbol{h}'} \geq 0$. We start by noting that

$$
\begin{aligned}
c_{h,\boldsymbol{h}'} - c_{h+1,\boldsymbol{h}'} &= \sum_{(k,\boldsymbol{k}')\in\mathbb{Z}^d} (a_{k,\boldsymbol{k}'} - a_{k+1,\boldsymbol{k}'}) b_{h-k,\boldsymbol{h}'-\boldsymbol{k}'} \\
&= \sum_{\boldsymbol{k}'\in\mathbb{Z}^{d-1}} \left[\sum_{k\geq 0} (a_{k,\boldsymbol{k}'} - a_{k+1,\boldsymbol{k}'}) b_{h-k,\boldsymbol{h}'-\boldsymbol{k}'} + \sum_{k\geq 1} (a_{k,\boldsymbol{k}'} - a_{k-1,\boldsymbol{k}'}) b_{h+k,\boldsymbol{h}'-\boldsymbol{k}'} \right],
\end{aligned}
$$

where the quadrantal symmetry has been used for terms with $k \leq -1$. This can also been re-written by changing the index in the last summation (so as to start from 0), getting

$$
c_{h,\boldsymbol{h}'} - c_{h+1,\boldsymbol{h}'} = \sum_{\boldsymbol{k}'\in\mathbb{Z}^{d-1}} \sum_{k\geq 0} (a_{k,\boldsymbol{k}'} - a_{k+1,\boldsymbol{k}'}) (b_{h-k,\boldsymbol{h}'-\boldsymbol{k}'} - b_{h+k+1,\boldsymbol{h}'-\boldsymbol{k}'}) \,.
$$

Notice that $a_{k,\boldsymbol{k}'} - a_{k+1,\boldsymbol{k}'} \geq 0$ by assumption, and also $b_{h-k,\boldsymbol{h}'-\boldsymbol{k}'} - b_{h+k+1,\boldsymbol{h}'-\boldsymbol{k}'} \geq 0$, because either $0 \leq h-k \leq h+k+1$, or $h-k \leq 0$ and in the latter case $0 \leq k-h \leq k+h+1$ and $b_{h-k,\boldsymbol{h}'-\boldsymbol{k}'} = b_{k-h,\boldsymbol{h}'-\boldsymbol{k}'} \geq b_{h+k+1,\boldsymbol{h}'-\boldsymbol{k}'}$. ∎

The following two lemmas are used in the proof of Theorem 5.4. The first one concerns the area under a Gaussian distribution in a neighborhood of the origin, while the second one is a discrete version where we consider Riemann sums, generalizing the technique used in the proof of Theorem 3.3 (see Eq. (3.1)).

Lemma A.2. *Given $\gamma > 0$,*

$$\sqrt{\frac{\pi}{\gamma t}}\,(1 - e^{-\gamma t a^2}) \leq \int_{-a}^{a} e^{-\gamma x^2 t}\,\mathrm{d}x \leq \sqrt{\frac{\pi}{\gamma t}}\,.$$

Proof. The proof exploits well-known properties of the Gaussian distribution. For the upper bound, simply

$$\int_{-a}^{a} e^{-\gamma x^2 t}\,\mathrm{d}x \leq \int_{\mathbb{R}} e^{-\gamma x^2 t}\,\mathrm{d}x = \sqrt{\frac{\pi}{\gamma t}}\,.$$

The lower bound exploits the well-known property of the complementary error function $\operatorname{erfc}(\zeta) := \frac{2}{\sqrt{\pi}} \int_{\zeta}^{+\infty} e^{-\xi^2}\,\mathrm{d}\xi$, which satisfies $\operatorname{erfc}(\zeta) < e^{-\zeta^2}$ for all $\zeta > 0$, so that

$$\int_{-a}^{a} e^{-\gamma x^2 t}\,\mathrm{d}x = \int_{\mathbb{R}} e^{-\gamma x^2 t}\,\mathrm{d}x - 2\frac{1}{\sqrt{\gamma t}}\operatorname{erfc}(\sqrt{\gamma t}\,a) > \frac{\sqrt{\pi}}{\sqrt{\gamma t}}(1 - e^{-\gamma t a^2})\,. \qquad ∎$$

Lemma A.3. *Assume that $n_1 \geq n_2 \geq \cdots \geq n_d$ and take any constants $c \in (0, \frac{1}{2})$ and $\gamma > 0$. Define*

$$A_{\boldsymbol{n},c,\gamma}(t) = \tfrac{1}{N} \sum_{\boldsymbol{h}:\forall r, -\lfloor cn_r \rfloor \leq h_r \leq \lfloor cn_r \rfloor} e^{-\gamma\left[\left(\frac{2\pi}{n_1}h_1\right)^2 + \cdots + \left(\frac{2\pi}{n_d}h_d\right)^2\right]t}\,.$$

Then there exist $c', c'' > 0$ (depending on c, γ and d only) such that, for all $t \geq 1$,

$$\frac{c'}{N} \max_{\ell=0,\ldots,d} \left\{ \frac{1}{t^{\ell/2}} \prod_{r=1}^{\ell} n_r \right\} \leq A_{\boldsymbol{n},c,\gamma}(t) \leq \frac{c''}{N} \sum_{\ell=0,\ldots,d} \frac{1}{t^{\ell/2}} \prod_{r=1}^{\ell} n_r\,.$$

Proof. We start from the upper bound. Let $\mathcal{I} := \{(h_1, \ldots, h_d) \in \mathbb{Z}^d \ :\ -\lfloor cn_r \rfloor \leq h_r \leq \lfloor cn_r \rfloor\ \forall r\}$. Moreover, for any set $K \subseteq [d]$, define

$$\mathcal{I}_K := \{(h_1, \ldots, h_d) \in \mathcal{I} \ :\ h_i \neq 0\ \forall i \in K \text{ and } h_i = 0\ \forall i \notin K\}$$

and notice that they form a partition of $\mathcal{I}$ as K varies over all the possible subsets of $[d]$ (including $K = \emptyset$ and $K = [d]$). Then

$$A_{\boldsymbol{n},c,\gamma}(t) = \frac{1}{N} \sum_{K \subseteq [d]} \sum_{\boldsymbol{h} \in \mathcal{I}_K} \prod_{r=1}^{d} e^{-\gamma(\frac{2\pi}{n_r}h_r)^2 t}\,.$$

Now, we want to estimate each term of the sum. Fix $K \subseteq [d]$. Except in the trivial case $K = \emptyset$, with no loss of generality we can assume that $K = \{1, \dots, s\}$. Then

$$
\begin{aligned}
\sum_{\boldsymbol{h} \in \mathcal{I}_K} \prod_{r=1}^{d} e^{-\gamma(\frac{2\pi}{n_r} h_r)^2 t} &= \sum_{\substack{-\lfloor cn_1 \rfloor \leq h_1 \leq \lfloor cn_1 \rfloor \\ h_1 \neq 0}} \cdots \sum_{\substack{-\lfloor cn_s \rfloor \leq h_s \leq \lfloor cn_s \rfloor \\ h_s \neq 0}} e^{-\gamma(\frac{2\pi}{n_1} h_1)^2 t} \cdots e^{-\gamma(\frac{2\pi}{n_s} h_s)^2 t} \\
&= \left(\sum_{\substack{\lfloor cn_1 \rfloor \leq h_1 \leq \lfloor cn_1 \rfloor \\ h_1 \neq 0}} e^{-\gamma(\frac{2\pi}{n_1} h_1)^2 t} \right) \cdots \left(\sum_{\substack{\lfloor cn_s \rfloor \leq h_s \leq \lfloor cn_s \rfloor \\ h_s \neq 0}} e^{-\gamma(\frac{2\pi}{n_s} h_s)^2 t} \right) \\
&= 2^s \prod_{r=1}^{s} \left(\sum_{1 \leq h_r \leq \lfloor cn_r \rfloor} e^{-\gamma(\frac{2\pi}{n_r} h_r)^2 t} \right).
\end{aligned}
$$

Then, using the following upper bound

$$
\frac{2}{n_r} \sum_{1 \leq h_r \leq \lfloor cn_r \rfloor} e^{-\gamma(\frac{2\pi}{n_r} h_r)^2 t} \leq \frac{1}{\pi} \int_0^{\frac{2\pi}{n_r} \lfloor cn_r \rfloor} e^{-\gamma x^2 t} \, \mathrm{d}x \leq \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\gamma x^2 t} \, \mathrm{d}x = \frac{1}{2} \left(\frac{1}{\pi \gamma t} \right)^{\frac{1}{2}}
$$

we obtain

$$
A_{\boldsymbol{n},c,\gamma}(t) \leq \sum_{K \subseteq [d]} \frac{1}{\prod_{r \notin K} n_r} \left(\frac{1}{4\pi \gamma t} \right)^{\frac{|K|}{2}}.
$$

For the lower bound, we use subsets of indexes quite similar to the above-defined $\mathcal{I}_K$, but in this case we do not look for a partition of $\mathcal{I}$. Rather, we define, for any $K \subseteq [d]$,

$$
\mathcal{J}_K := \{(h_1, \dots, h_d) \in \mathcal{I} \ : h_i = 0 \ \forall i \notin K\}
$$

without the additional request that $h_i \neq 0 \ \forall i \in K$. Then we make a different lower bound for any K, by discarding the terms with $\boldsymbol{h} \notin \mathcal{J}_K$ in the summation which defines $A_{\boldsymbol{n},c,\gamma}(t)$, namely we use the fact that, for all $K \subseteq [d]$ we have that

$$
A_{\boldsymbol{n},c,\gamma}(t) \geq \sum_{\boldsymbol{h} \in \mathcal{J}_K} \prod_{r=1}^{d} e^{-\gamma(\frac{2\pi}{n_r} h_r)^2 t}.
$$

The choice $K = \emptyset$ simply gives

$$
A_{\boldsymbol{n},c,\gamma}(t) \geq \frac{1}{N}.
$$

The choice $K = \{1, \dots, s\}$ gives

$$
\begin{aligned}
A_{\boldsymbol{n},c,\gamma}(t) &\geq \frac{1}{N} \sum_{-\lfloor cn_1 \rfloor \leq h_1 \leq \lfloor cn_1 \rfloor} \cdots \sum_{-\lfloor cn_s \rfloor \leq h_s \leq \lfloor cn_s \rfloor} e^{-\gamma(\frac{2\pi}{n_1} h_1)^2 t} \cdots e^{-\gamma(\frac{2\pi}{n_s} h_s)^2 t} \\
&\geq \left(\prod_{r=1}^{s} \frac{1}{n_r} \sum_{0 \leq h_r \leq \lfloor cn_r \rfloor} e^{-\gamma(\frac{2\pi}{n_r} h_r)^2 t} \right) \left(\prod_{r=s+1}^{d} \frac{1}{n_r} \right) \\
&\geq \left(\frac{1}{2\pi} \int_0^{2\pi c} e^{-\gamma t x^2} \, \mathrm{d}x \right)^s \frac{\prod_{r=1}^{s} n_r}{N}.
\end{aligned}
$$

Then, we end by using Lemma A.2, which gives

$$\int_0^{2\pi c} e^{-\gamma t x^2}\,\mathrm{d}x > \frac{1}{2}\sqrt{\frac{\pi}{\gamma t}}(1-e^{-\gamma t(2\pi c)^2}) \geq \frac{1}{2}\sqrt{\frac{\pi}{\gamma t}}(1-e^{-\gamma(2\pi c)^2})$$

when $t \geq 1$. $\square$

Acknowledgments. The authors are grateful to Ruggero Carli for the simulation results on random geometric graphs, as well as for his contribution in their preliminary work on this topic. They also thank Enrico Lovisari for many useful discussions and Paolo Frasca for helpful comments.

REFERENCES

[1] B. Bamieh, M. Jovanovic, P. Mitra, and S. Patterson, *Effect of topological dimension on rigidity of vehicle formations: Fundamental limitations of local feedback*, in Proc. 47th IEEE Conf. Decision and Control, 2008.

[2] B. Bamieh, F. Paganini, and M. A. Dahleh, *Distributed control of spatially invariant systems*, IEEE Transactions on Automatic Control, 47 (2002), pp. 1091–1107.

[3] E. Behrends, *Introduction to Markov Chains (with Special Emphasis on Rapid Mixing)*, Vieweg Verlag, 1999.

[4] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall, 1989.

[5] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, *Symmetry analysis of reversible Markov chains*, Internet Mathematics, 2 (2005), pp. 31–71.

[6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory/ACM Trans. Netw., 14 (2006), pp. 2508–2530.

[7] M. Cao, A. S. Morse, and B. D. O. Anderson, *Reaching a consensus in a dynamically changing environment: A graphical approach*, SIAM Journal on Control and Optimization, 47 (2008), pp. 575–600.

[8] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, *A PI consensus controller for networked clocks synchronization*, in IFAC World Congress, Seoul, Corea, July 2008.

[9] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in the average consensus problem*, Automatica, 44 (3) (2008), pp. 671–684.

[10] R. Carli, F. Garin, and S. Zampieri, *Quadratic indices for the analysis of consensus algorithms*, in Proc. of the 4th Information Theory and Applications Workshop, La Jolla, CA, USA, Feb. 2009, pp. 96–104.

[11] J. Cortes, S. Martinez, and F. Bullo, *Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions*, IEEE Transactions on Automatic Control, 51 (2006), pp. 1289–1298.

[12] G. Cybenko, *Dynamic load balancing for distributed memory multiprocessors*, J. Parallel Distrib. Comput., 7 (1989), pp. 279–301.

[13] R. D'Andrea and G. E. Dullerud, *Distributed control design for spatially interconnected systems*, IEEE Transactions on Automatic Control, 48 (2003), pp. 1478–1495.

[14] P. J. Davis, *Circulant matrices*, A Wiley-Interscience Publication, Pure and Applied Mathematics, John Wiley & Sons, New York-Chichester-Brisbane, 1979.

[15] P. Diaconis and D. Stroock, *Geometric bounds for eigenvalues of Markov chains*, Annals of Applied Probability, 1 (1991), pp. 36–61.

[16] F. Fagnani and S. Zampieri, *Average consensus with packet drop communication*, SIAM Journal on Control and Optimization, 48 (2009), pp. 102–133. Special issue on "Control and Optimization in Cooperative Networks".

[17] M. Franceschetti and R. Meester, *Random networks for communication*, Cambridge University Press, Cambridge, 2007.

[18] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, *Average consensus on networks with quantized communication*, International Journal of Robust and Nonlinear Control, 19 (2009), pp. 1787–1816.

[19] J. Friedman, *A proof of Alon's second eigenvalue conjecture and related problems*, Memoirs of the American Mathematical Society, 195 (2008), pp. 1–100.

[20] S. Ganeriwal, R. Kumar, and M. B. Srivastava, *Timing-sync protocol for sensor networks*, in SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems, New York, NY, USA, 2003, ACM, pp. 138–149.

[21] F. Garin and S. Zampieri, *Performance of consensus algorithms in large-scale distributed estimation*, in Proceedings of the European Control Conference 2009, 23-26 Aug. 2009, pp. 755–760.

[22] E. N. Gilbert, *Random plane networks*, Journal of SIAM, 9 (1961), pp. 533–543.

[23] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bulletin of the American Mathematical Society, 43 (2006), pp. 439–561.

[24] L. Moreau, *Stability of multi-agent systems with time-dependent communication links*, IEEE Transactions on Automatic Control, 50 (2005), pp. 169–182.

[25] A. Nedić and A. Ozdaglar, *Distributed subgradient methods for multi-agent optimization*, IEEE Transactions on Automatic Control, 54 (2009), pp. 48–61.

[26] R. Olfati-Saber, *Distributed Kalman filter with embedded consensus filters*, 44th IEEE Conference on Decision and Control and 2005 European Control Conference (CDC-ECC '05), (2005), pp. 8179–8184.

[27] R. Olfati-Saber, J. A. Fax, and R. M. Murray, *Consensus and cooperation in networked multi-agent systems*, Proceedings of IEEE, 95 (2007), pp. 215–233.

[28] R. Olfati-Saber and R. M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Transactions on Automatic Control, 49 (2004), pp. 1520–1533.

[29] J. N. Tsitsiklis, *Problems in Decentralized Decision Making and Computation*, PhD thesis, MIT, Nov. 1984. Technical Report LIDS-TH-1424, Laboratory for Information and Decision Systems.

[30] L. Xiao, S. Boyd, and S.-J. Kim, *Distributed average consensus with least-mean-square deviation*, Journal of Parallel and Distributed Computing, 67 (2007), pp. 33–46.

[31] L. Xiao, S. Boyd, and S. Lall, *A scheme for asynchronous distributed sensor fusion based on average consensus*, in International Conference on Information Processing in Sensor Networks (IPSN'05), Los Angeles, CA, Apr. 2005, pp. 63–70.