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, given by the essential spectral radius of the transition matrix (i.e. the second largest eigenvalue). For many graph families, such analysis predicts a performance which degrades when the number of agents grows, basically because spreading information across a larger graph requires a longer time. However, in estimation problems, a growing number of sensor nodes improves the quality of the estimate, and it is natural to ask whether such improvement is possible with distributed algorithms, in which a large number of nodes have contrasting effects. To answer this question, it is important to specify a suitable performance metric, depending on the actual estimation or control problem in which the consensus algorithm is used, and to study how performance scales when both number of iterations and number of agents grow to infinity, for different communication graphs.

Here we propose a simple example of distributed estimation problem solved by average-consensus algorithm. We propose a natural performance index (average quadratic estimation error) and we analyze its performance for graphs describing local interactions within a limited neighborhood, and with symmetries allowing for elegant mathematical tools. Despite their abstraction, such graphs capture the essential behavior of more general local interactions.

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 obtain this goal is given by the average-consensus algorithms [25, 21, 24, 28, 27, 8]; although not being the most efficient method to compute the average in a distributed way, this
This algorithm has been proposed in many contexts in which it is necessary to compute averages in a distributed way, namely in distributed estimation [23] and in sensor calibration for sensor networks [18, 7], in load balancing for distributed computing systems [11], and in mobile multi-vehicles coordination [10], as well as in distributed optimization and learning [22]. 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 estimates have been proposed in the different cases of matrices with symmetries [14], and for randomly generated matrices [5], and for general matrices [14]. 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 [25], noise on the communication links [27] and quantization [17], 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 (Section 4.2).

Our main contribution 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], [12] and [8], because of the ease of analysis allowed by their structure, as well as because they are a prototypical example of geographically local communication. Then, simulation results, presented in Section 4.3, 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 showing that this class of graphs captures the most significant aspects of geometrically local interaction.

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 $v \in \mathbb{R}^N$ and a matrix $P \in \mathbb{R}^{N \times N}$, we let $v^T$ and $P^T$ respectively denote the transpose of $v$ and of $P$. We let $\Lambda(M)$ denote the set of eigenvalues of $M$, counted with their multiplicities. With the symbol $1$ we denote the $N$-dimensional vector having all the entries equal to 1. We will denote by $|v|$ the vector obtained by taking the modulus of each entry of a given vector $v$, and we will write $w > v$ and $w \succeq v$ 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 $\{1, 2, \ldots, |A|\}$. 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 maximum or a product over an empty set gives one. We also introduce the short-hand notation $[d] = \{1, 2, \ldots, d\}$.

2. Problem formulation and performance measure. We consider the following simple problem of distributed estimation: $N$ sensors measure the same real value $\theta$ plus i.i.d. noises. Clearly, the best estimate for $\theta$ 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\(^1\).

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

Then we consider a linear average-consensus algorithm: $x(t + 1) = P x(t)$ 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\(^2\) with $P$ is a subgraph of $G$. We assume that $P$ is doubly-stochastic\(^3\) and primitive\(^4\). It is well-known that, under these assumptions, $P$ has dominant eigenvalue 1 with multiplicity 1 and

$$\forall i, \lim_{t \to \infty} x_i(t) = \frac{1}{N} \sum_j x_j(0).$$

The speed of such convergence is given by the essential spectral radius of $P$, $\rho_{\text{ess}}(P)$, i.e., the eigenvalue of $P$ which has second largest modulus. For non-expander families

\(^1\)G is strongly connected if, for all $u, v \in V$, there exists a path connecting $u$ to $v$. It is aperiodic if the greatest common divisor of the lengths of all cycles is one; e.g., the presence of a self-loop implies aperiodicity.

\(^2\)The graph $G_P$ associated with $P \in \mathbb{R}^{V \times V}$ is a graph $G_P = (V, E)$, with $(u, v) \in E \iff P_{uv} \neq 0$.

\(^3\)P is doubly-stochastic if $P_{uv} \geq 0$ for all $u, v$, $P 1 = 1$ and $1^T P = 1^T$.

\(^4\)P is primitive if $\exists m$ such that $(P^m)_{uv} \neq 0 \forall u, v \in V$. Equivalently, the graph $G_P$ associated with $P$ is strongly connected and aperiodic.
of graphs, such as for example Cayley graphs on Abelian groups, when \( N \to \infty \), \( \rho_{\text{ess}}(P) \to 1 \) (see e.g. [8]). 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 \( \theta \), a very natural performance measure is the mean quadratic error

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

where the error \( e(t) \) is defined as \( e(t) = x(t) - \theta 1 \), so that \( J(P, t) = \sum_j \mathbb{E} \left[ (x_j(t) - \theta)^2 \right] \).

For our problem, we can easily show that the cost \( J(P, t) \) can be re-written as

\[
J(P, t) = \frac{1}{N} \text{trace} \left( (P^t)^T P^t \right) \tag{2.1}
\]

Indeed,

\[
J(P, t) = \frac{1}{N} \mathbb{E} \left[ (P^t x(0) - \theta 1)^T (P^t x(0) - \theta 1) \right]
= \frac{1}{N} \mathbb{E} \left[ (P^t w)^T (P^t w) \right]
= \frac{1}{N} \mathbb{E} \left[ \text{trace}(w^T (P^t)^T P^t w) \right]
= \frac{1}{N} \text{trace} \left( (P^t)^T P^t \right) \mathbb{E}(ww^T).
\]

If \( P \) is normal, i.e. \( P^T P = PP^T \) (e.g. symmetric matrices are normal), then Eq. (2.1) is equivalent to

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

3. Two motivating examples. In this section we present two 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.

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), as depicted in Figure 3.1. For simplicity, we assume that each received message, as well as the agent’s own state, is weighted \( 1/3 \), so that the update matrix \( P_N \) is the following circulant symmetric matrix:

\[
P_N = \begin{bmatrix}
1/3 & 1/3 & 0 & \cdots & \cdots & 1/3 \\
1/3 & 1/3 & 1/3 & \cdots & 0 & 0 \\
0 & 1/3 & 1/3 & 1/3 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 1/3 & 1/3 & 1/3 \\
1/3 & \cdots & \cdots & 0 & 1/3 & 1/3
\end{bmatrix} \tag{3.1}
\]

As the eigenvalues of such a circulant matrix can be explicitly computed, it is easy to see that \( \rho_{\text{ess}}(P_N) \sim 1 - C/N^2 \), for some suitable constant \( C > 0 \). 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 $\sim (1 - C/N^2)$) 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 $t \to \infty$. The bounds derived in a more general setting presented in Section 4.2 (Coroll. 4.6) will show that indeed the asymptotic behavior of $J(P_N, t)$ in this example is given by $\max\left\{\frac{1}{N}, \frac{1}{\sqrt{t}}\right\}$. 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 $\lim_{N \to \infty} \rho_{ess}(P_N) = 1$. 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. 4.4). Nevertheless, a further look at Figure 3.2, and the results in Coroll. 4.3, give 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, as at time $t$ there is no way for a node to use information coming from other agents further than $t$ steps apart.

Example 2: two complete sub-graphs. Let $N$ be an even number, and consider a graph consisting of two disconnected complete graphs, each on $N/2$ nodes; Fig. 3.3a depicts as an example the case $N = 10$. Associate with each edge a coefficient $2/N$, so that $P_N$ has the following form:

$$P_N = \begin{bmatrix} \frac{2}{N} & 11^T \\ 0 & \frac{2}{N} \end{bmatrix}.$$

We would like to compare performance of this $P_N$ with the circle presented as Example 1, by looking at the essential spectral radius, and then by looking at the estimation error $J_{estim}$, so as to show that the answer is actually quite different. The eigenvalues of $P_N$ are simply $1$ with multiplicity $2$ and the eigenvalue $0$ with multiplicity $N - 2$, so that $\rho_{ess}(P_N) = 1$, which is worse than the circle. 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, $\frac{1}{N}$), as opposed to the circle which, for large $N$, has a very slow convergence.

Behind computation of the eigenvalues, there is an intuitive explanation of what happens. The essential spectral radius $1$ describes the fact that the graph is disconnected, and thus no convergence is possible to the average of all initial values: simply
no information can transit from one group to another; nevertheless, the estimation error is very good for large $N$, because it is the average of $N/2$ measurements, and it is computed very fast, in one iteration, thanks to the complete graph which gives centralized computation within the group of $N/2$ agents. On the contrary, in the circle average consensus can be reached asymptotically, as described by the essential spectral radius smaller than one, but convergence is very slow for large $N$ ($\rho_{\text{ess}} \sim 1 - \frac{\pi^2}{N^2}$ for $N \to \infty$), and a reasonably good estimation error is achieved only after a long time.

The previous example can be modified to obtain a slightly different matrix whose associated graph is connected and which exhibits similar behavior as the matrix proposed in the previous example. Indeed, consider the matrix

$$
\tilde{P}_N = P_N + \begin{bmatrix}
-\frac{2/N}{2/N} & \frac{2/N}{2/N} \\
\frac{2/N}{2/N} & -\frac{2/N}{2/N}
\end{bmatrix}.
$$

The graph associated with $\tilde{P}_N$ is shown in Figure 3.3b. All eigenvalues of $\tilde{P}_N$ can be explicitly computed [4, Prop. 5.1]. There is one eigenvalue in 1 with multiplicity 1, one eigenvalue in 0 with multiplicity $N - 3$ and finally there are two eigenvalues in
The aim of the paper is to analyze this performance measures 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. All the proofs are postponed in the appendix.

4. Asymptotic behavior of costs for grids. This section contains the main results of this paper. First, we describe the families of graphs (and of associated matrices) that we consider, and then we state out results on the asymptotics of \( J(P_N, t) \) for such families. All the proofs are postponed in the appendix.


4.1.1. Grids on tori (Abelian Cayley graphs). First of all let’s recall the definition of Cayley graphs: given a group \((\Gamma, +)\) and a set \(S \subseteq \Gamma\), the Cayley graph \(G(\Gamma, S)\) is a directed graph with vertex set \(\Gamma\) 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 \(\Gamma\)) is Cayley if \(P_{g,h} = P_{g+k,h+k}\) for all \(g, h, k \in \Gamma\). This is equivalent to say that there exists a map \(\pi : \Gamma \rightarrow \mathbb{R}\) 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 \(\Gamma_{n_1, \ldots, n_d} := \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}\). We will use the notation \(n := (n_1, \ldots, n_d)\), so that \(\Gamma_n = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}\), and we will write \(N := |\Gamma_n| = \prod_{i=1}^d n_i\).

When \(P\) is a Cayley matrix associated with \(\Gamma_n\), its eigenvalues have the following simple expression [3] for any \(h = (h_1, \ldots, h_d) \in \Gamma_n\),

\[
\lambda_h = \sum_{k \in \Gamma_n} \pi(k) e^{-i \frac{2\pi}{n_1} h_1 + \cdots + i \frac{2\pi}{n_d} h_d}.
\]

As a simple example, when \(d = 1\), i.e. \(n = N\) and \(\Gamma_n = \mathbb{Z}_N\), you obtain that \(P\) is a circulant \(N \times N\) matrix [13] with first row equal to the vector \([\pi(0), \ldots, \pi(N-1)]^T\), and with eigenvalues \(\lambda_h = \sum_{k=0}^{N-1} \pi(k) e^{-i \frac{2\pi}{N} hk}\), for \(h = 0, \ldots, N - 1\). Note that, with a slight abuse of notation, we write \(e^{i \frac{2\pi}{N} h r}\) 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 \(h_r \in \{0, 1, \ldots, n_r - 1\}\), which we will denote by \(h \in V_{n_r}\), \(V_n := \{0, \ldots, n_1 - 1\} \times \cdots \times \{0, \ldots, n_d - 1\}\). When needed, we will actually identify the set of vertices of the graph with \(V_n\) rather than \(\Gamma_n\).

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 $n_1, \ldots, n_d$ grow. In order to define the neighbors and weights, we fix a positive integer $\delta$, we define the set $D_\delta = \{-\delta, -\delta + 1, \ldots, +\delta\}^d$ and we fix $|D_\delta|$ real numbers $p_h$, $h = (h_1, \ldots, h_d) \in D_\delta$ such that $p_h \geq 0 \forall h$ and $\sum_{h \in D_\delta} p_h = 1$. Then, for any $n > 2\delta 1$ (namely $n_j > 2\delta$ for all $j$) we construct the Cayley matrix $P_n \in \mathbb{R}^{\Gamma_n \times \Gamma_n}$ with generator $\pi_n : \Gamma_n \to \mathbb{R}$ defined by $\pi_n(g) = p_h$ if there is an $h \in D_\delta$ such that, for all $\ell = 1, \ldots, d$ $g_\ell = h_\ell \mod n_\ell$, and $\pi_n(g) = 0$ otherwise. Note that, for any $n > 2\delta 1$, $\pi_n$ is well-defined. The matrix $P$ defined in this way can be seen as a map $\mathbb{R}^{\Gamma_n} \to \mathbb{R}^{\Gamma_n}$.

The graph associated with such a map is an infinite $d$-dimensional lattice, with neighborhoods defined by the non-zero coefficients of $p(z_1, \ldots, z_d)$. For example, in the case of Example 1, we obtain an infinite line, where each node is connected to the first neighbors on its right and on its left, and with self-loops. Notice that

$$
(Px)_h = \sum_{k \in D_\delta} p_k x_{h-k}
$$

for all $h \in \Gamma_n$.

We introduce here another useful notation, defining the Laurent polynomial $p(z_1, \ldots, z_d) \in \mathbb{R}[z_1, z_1^{-1}, \ldots, z_d, z_d^{-1}]$ given by

$$
p(z_1, \ldots, z_d) = \sum_{k \in D_\delta} p_k z_1^{k_1} \cdots z_d^{k_d}.
$$

We will refer to the above construction of a family of Cayley matrices for all $n > 2\delta 1$ as the Cayley matrix family associated with $p(z_1, \ldots, z_d)$. With this notation, the eigenvalues of $P_n$ are $\lambda_h = p(e^{i2\pi k_1/n_1}, \ldots, e^{i2\pi k_d/n_d})$, $h \in \Gamma_n$.

We can also do a similar construction taking the limit case when $\Gamma = \mathbb{Z}^d$ by introducing the linear map $P_\infty : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}^{\mathbb{Z}^d}$ defined, for each $x \in \mathbb{R}^{\mathbb{Z}^d}$ and $h \in \mathbb{Z}^d$, as

$$
(P_\infty x)_h := \sum_{k \in D_\delta} p_k x_{h-k}.
$$

The graph associated with such a map is an infinite $d$-dimensional lattice, with neighborhoods defined by the non-zero coefficients of $p(z_1, \ldots, z_d)$. For example, in the case of Example 1, we obtain an infinite line, where each node is connected to the first neighbors on its right and on its left, and with self-loops. Notice that

$$
(P^t_\infty w)_k = \sum_{h \in \mathbb{Z}^d} p_h^{(t)} w_{h-k}
$$

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

Throughout this paper, we will assume that there are enough non-zero weights $p_0$ so as to ensure that the corresponding infinite graph is weakly connected\(^5\). This is enough to ensure that all finite graphs are strongly connected, as stated in the following lemma.

**Lemma 4.1.** With the above notation, define $S = \{h \in D_\delta : p_h \neq 0\}$. 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 $n > 2\delta 1$, the Cayley graph associated with $p(z)$ and $\Gamma_n$ is strongly connected;

---

\(^5\)A directed graph is weakly connected if, disregarding edge directions, for any pair of vertices $u, v$, there exists an undirected path connecting $u$ to $v$. 
for all $n > 2\delta 1$, $S$ generates $\Gamma_n$.

The proof is postponed in the Appendix. Notice that the second condition implies also that $S$ contains a basis of $\mathbb{R}^d$, and so it states that the connectivity requirement implies that the graph is somehow 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.

Finally, we assume that the graphs have self-loops, namely that $p_0 \neq 0$. This assumption, together with the connectivity assumption above, ensures that all matrices $P_n$ of the sequence we have constructed above are primitive; also recall that such matrices are doubly-stochastic and normal.

4.1.2. Grids on cubes. The families of Cayley graphs on the group $\Gamma_n = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_d}$ presented above can be seen as regular grids on multi-dimensional tori. An interesting result by Boyd et al. [4] on reversible Markov chains with symmetries allows to compute the eigenvalues and eigenvectors also of regular grids on a cube in $\mathbb{R}^d$, which are the graphs coinciding with 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.

More precisely, define the following family of matrices. Let $P_{2n}$ be a Cayley matrix on $\Gamma_{2n} = \mathbb{Z}_{2n_1} \times \cdots \times \mathbb{Z}_{2n_d}$ associated with $p(z_1, \ldots, z_d)$, and assume that the coefficients $p_h$ satisfy the following quadrantal symmetry

$$\forall h, \quad p_h = p(|h|).$$

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 $\sigma_r$ on $\Gamma_{2n}$ by letting $\sigma_r(h) = k$ with $k_\ell = h_\ell$ if $\ell \neq r$ and $k_r = 2n_r - 1 - h_r$. It is convenient here to identify $\Gamma_{2n}$ with the set $V_{2n}$, and consider $\sigma_r : V_{2n} \rightarrow V_{2n}$. For example, Fig. 4.1 shows the axis of reflection of $\sigma_1$ for the case $d = 1$, as in Example 1. In higher dimension, every $\sigma_r$ simply keeps all coordinates invariant except for the $r$-th, where it is the reflection depicted for the one-dimensional case.

Notice that every $\sigma_r$ 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

Fig. 4.1: 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.
$g \in V_n \subseteq V_{2n}$, the orbit $O_g = \{\eta(g) : \eta \in H\} \subseteq V_{2n}$. 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 $g \in V_n = \{0, \ldots, n_1 - 1\} \times \{0, \ldots, n_2 - 1\}$, the corresponding orbit $O_g$ contains the four points $g$, $\sigma_1(g) = (2n_1 - 1 - g_1, g_2)$, $\sigma_2(g) = (g_1, 2n_2 - 1 - g_2)$ and $\sigma_2 \circ \sigma_1(g) = (2n_1 - 1 - g_1, 2n_2 - 1 - g_2)$.

Finally, define $\mathcal{T}_n : \mathbb{R}^{V_n} \to \mathbb{R}^{V_n}$, for all $h, k \in V_n$, by

$$\mathcal{T}_n(h, k) := \sum_{\ell \in O_h} (P_{2n})_{h, \ell} = \sum_{\eta \in H} (P_{2n})_{h, \eta(k)}.$$  \hspace{1cm} (4.3)

**Remark 1.** The entries of $\mathcal{T}_n$ are actually equal to those of $P_n$, except at the ‘borders’. In fact, if you consider an index $k \in V_n$ such that $\delta < k_r < n_r - \delta$ is satisfied for all $r = 1, \ldots, d$, then for all $h \in V_n$ all terms in the sum in Eq. (4.3) with $\ell \neq k$ are zero, so that $(\mathcal{T}_n)_{h, k} = (P_{2n})_{h, k}$. Also notice that for such choice of $k$, $(P_{2n})_{h, k} = (P_n)_{h, k}$. Moreover, an analogous equality $(\mathcal{T}_n)_{h, k} = (P_{2n})_{h, k} = (P_n)_{h, k}$ holds true whenever $h$ and $k$ are such that, for all $r = 1, \ldots, d$, at least one of the two following conditions is satisfied: $\delta < k_r < n_r - \delta$ or $\delta < k_r < n_r - \delta$.

For example, in the one-dimensional case, $\mathcal{T}_n$ shares with $P_n$ the banded-diagonal central part of the matrix, and is different only in initial and final part of the first $\delta$ and the last $\delta$ 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 $\mathcal{T}_n$ are the following

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

and

$$\mathcal{T}_n = \begin{bmatrix}
p_0 + p_1 & p_1 + p_2 & p_2
p_1 + p_2 & p_0 & p_1 & p_2
p_2 & p_1 & p_0 & p_2
& \ddots & \ddots & \ddots & \ddots
p_2 & p_1 & p_0 & p_1 & p_2
p_2 & p_1 & p_0 & p_1 & p_2 + p_1
\end{bmatrix}$$

We will refer to the above construction of a family of matrices $\mathcal{T}_n$ for all $n \geq 2\delta$ as the grid matrix family associated with $p(z_1, \ldots, z_d)$. Note that such a construction ensures that we can apply [4, Prop. 3.3], because both $P_{2n}$ and $\mathcal{T}_n$ are symmetric (and thus the corresponding Markov chain is reversible) and the latter is the lumped
chain of the former, as defined in [4, Sect. 3]. Thus, the explicit expression for the
eigenvalues of $\mathcal{T}_n$ is the following

$$\tilde{\lambda}_h = p(e^{\frac{2\pi}{n}h_1}, \ldots, e^{\frac{2\pi}{n}h_d}), \quad h \in V_n.$$  

4.2. Main results. In this section, we give results for the cost $J(P, t)$ for the
above-described families of graphs. All proofs are postponed to the appendix.

A first 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 $J(P_n, t) = \mathbb{E}[(e_h(t))^2]$ for any node $h$, and we can
take for example node 0 as the reference node

$$J(P_n, t) = \mathbb{E}[(e_0(t))^2].$$

This remark suggests a way for defining a similar cost associated with the infinite Cay-
ley matrix $P_\infty$. In fact, if we consider the distributed estimation problem presented in
Section 2 for an infinite graph, and we solve it by updating the nodes’ estimates with
the linear map described in Eq. (4.1), then the expectation of the quadratic error
is the same for any vertex, and it makes sense to fix our attention on an arbitrarily
chosen one, say vertex 0, and to define

$$J(P_\infty, t) = \mathbb{E}[(e_0(t))^2]. \quad (4.4)$$

Given a Laurent polynomial $p(z_1, \ldots, z_d) = \sum_{h \in D_d} p_h z_1^{h_1} \cdots z_d^{h_d}$, define the poly-
nomial $q(z_1, \ldots, z_d) := p(z_1, \ldots, z_d)p(z_1^{-1}, \ldots, z_d^{-1})$; denote by $\{q_h^{(t)}\}_{h \in D_{2d}}$ the
coefficients of $p(z_1, \ldots, z_d)$’s. With this notation it is possible to characterize the cost
$J(P, t)$ in a way that involves the coefficients $q_h^{(t)}$ only, and that allows to draw some
interesting conclusions on the behavior of $J(P, t)$ when the number of nodes increases
(see in particular Corollary 4.3).

**Proposition 4.2.** With the above notation, given a polynomial $p(z_1, \ldots, z_d)$ and
$n = (n_1, \ldots, n_d) > 2d1$,

- if $P_\infty$ is the infinite map associated with $p(z_1, \ldots, z_d)$, then
  $$J(P_\infty, t) = q_0^{(t)}$$
- if $P_n$ is a Cayley matrix associated with $p(z_1, \ldots, z_d)$, then
  $$J(P_n, t) = \sum_{h \in F_n} q_h^{(t)},$$
  where $F_n := \{h : h_r = 0 \mod n_r \; \forall r\}$.
- if $p(z_1, \ldots, z_d)$ satisfies the quadrantal symmetries (4.2) and $\mathcal{T}_n$ is a grid
  matrix associated with $p(z_1, \ldots, z_d)$, then
  $$J(\mathcal{T}_n, t) = \sum_{K \subseteq [d]} \frac{1}{\prod_{r \in K} n_r} \sum_{h \in F_{K,n}} q_h^{(t)},$$
  where the first summation is over all subsets $K \subseteq [d] := \{1, \ldots, d\}$, including
  $K = \emptyset$ and $K = [d]$, and where
  $$F_{K,n} := \{(h_1, \ldots, h_d) : h_r \text{ is odd } \forall r \in K \text{ and } h_r = 0 \mod 2n_r \; \forall r \notin K\}.$$
COROLLARY 4.3. Under the assumptions of Proposition 4.2, if $P_n$ is a Cayley matrix, then in the limit when $n_r \to \infty$ for all $r = 1, \ldots, d$, $J(P_n, t) \to J(P_\infty, t)$ and moreover $J(P_n, t) = J(P_\infty, t)$ for all $n \geq 2\delta 1$. In the case of a grid matrix $\overline{P}_n$, only the limit result holds true.

COROLLARY 4.4. Under the assumptions of Proposition 4.2, if $p(z_1, \ldots, z_d)$ satisfies the quadrantal symmetries (4.2) and satisfies the following monotonicity assumption

$$|h| \geq |k| \Rightarrow p_h \leq p_k,$$

then, for the family of Cayley matrices $P_n$ associated with $p(z_1, \ldots, z_d)$, the cost $J(P_n, t)$ is monotonic non-increasing w.r.t. $n_1, \ldots, n_d$, namely

$$m \leq n \Rightarrow J(P_m, t) \geq J(P_n, t), \forall t.$$

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

Note that the assumptions on $p$ in Coroll. 4.4 are not necessary: for example, with $d = 1$, $p(z) = \frac{1}{5}z^{-2} + \frac{1}{5}z^{-1} + \frac{3}{5}z + \frac{1}{5}z^2$ violates both assumptions and nevertheless gives a monotonic cost $J(P_n, t)$. However, monotonicity of $J(P_n, t)$ is not always true, e.g., $p(z) = \frac{1}{5}z^{-2} + \frac{1}{5}z^{-1} + \frac{2}{5}z + \frac{1}{5}z^2$ gives a non-monotonic cost.

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.

THEOREM 4.5. Given a polynomial $p(z) = \sum_{h \in D_d} p_h z^{h_1} \cdots z^{h_d}$ 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 $P_\infty$ is the infinite map associated with $p(z_1, \ldots, z_d)$, then there exist positive constants $c, c'$, depending on $d$ and $p(z_1, \ldots, z_d)$ 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 $P_n$ is a Cayley matrix associated with $p(z_1, \ldots, z_d)$, then there exist positive constants $C, C'$, depending on $d$ and $p(z_1, \ldots, z_d)$ only, such that, for all $n \geq 2\delta 1$ and $t > 0$

$$C \max_{0 \leq k \leq d} \frac{\prod_{1 \leq r \leq k} n_r}{N} \frac{1}{(\sqrt{t})^k} \leq J(P_n, t) \leq C' \max_{0 \leq k \leq d} \frac{\prod_{1 \leq r \leq k} n_r}{N} \frac{1}{(\sqrt{t})^k};$$

- if $p(z_1, \ldots, z_d)$ satisfies the quadrantal symmetries (4.2) and $\overline{P}_n$ is a grid matrix associated with $p(z_1, \ldots, z_d)$, then there exist constants $\overline{C}, \overline{C}'$, depending on $d$ and $p(z)$ only, such that, for all $n \geq 2\delta 1$ and $t > 0$

$$\overline{C} \max_{0 \leq k \leq d} \frac{\prod_{1 \leq r \leq k} n_r}{N} \frac{1}{(\sqrt{t})^k} \leq J(\overline{P}_n, t) \leq \overline{C}' \max_{0 \leq k \leq d} \frac{\prod_{1 \leq r \leq k} n_r}{N} \frac{1}{(\sqrt{t})^k}.$$
Corollary 4.6. If $n_r = n$ for all $r = 1, \ldots, d$, under the assumptions of Theorem 4.5, if $P_n$ is a Cayley matrix associated with $p(z_1, \ldots, z_d)$, then there exist $C, C' > 0$ (depending on $d$ and and $p(z_1, \ldots, z_d)$ only) such that

$$C \max \left\{ \frac{1}{N}, \frac{1}{(\sqrt{N})^d} \right\} \leq J(P_n, t) \leq C' \max \left\{ \frac{1}{N}, \frac{1}{(\sqrt{N})^d} \right\}$$

An analogous bound holds true for the grid matrix $\overline{P}_n$ associated with $p(z_1, \ldots, z_d)$.

4.3. 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 geographical 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 ([20], see also the recent book [16]): 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). To make a fair comparison with the families of graphs considered in previous sections, we consider random geometric graphs of increasing number of nodes $N$, with a threshold $r$ chosen 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 [28].

Figure 4.2 provides examples of numerical results for the 2-dimensional case, where the behavior of the Cayley graph predicts a cost scaling as max $\left\{ \frac{1}{N}, \frac{1}{N} \right\}$ (Corollary 4.6). 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.

5. 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 4.3) 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.

Acknowledgments. The authors are grateful to dr. 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.

REFERENCES

CONSENSUS PERFORMANCE IN LARGE-SCALE ESTIMATION


Appendix. Proofs.

A.1. Abelian Cayley graphs. Proof of Lemma 4.1: 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 \(u = u_0, u_1, \ldots, u_{d+1}, u_d = v\) 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 \(u_{i+1} = u_i + S \cup (-S)\) 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 \(d \geq 1\) and \(s_1, \ldots, s_d \in S \cup (-S)\) such that \(v = u + s_1 + \cdots + s_d\). 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 \( S = \{s_1, \ldots, s_r\} \). Now notice that 4. can be equivalently re-stated as follows. For all \( n > 2\delta \), \( \exists X \in \mathbb{Z}^{r \times d} \) 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 \( s_1, \ldots, s_r \), \( M \) 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 \( \exists z \in \mathbb{Z}^d \) such that \( z^T A = 0 \), then \( z^T = z^T Y \). Taking in particular \( M = bI \) with \( b > 2\delta \), we have that \( z^T = bz^T Y \), which implies that the entries of \( z \) are multiple of \( b \). Since this holds for all \( b > 2\delta \), this implies that \( z = 0 \). Now from the fact that \( A \) is a full row-rank matrix it follows that there exists \( X \in \mathbb{Z}^{r \times d} \) such that \( AX = aI \), 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 \( I = AX + aY = AX + AXY = A(X + XY) \) which is equivalent to 2.

### A.2. Proof of Prop. 4.2 and of its corollaries.

**Proof of Prop. 4.2 – infinite Cayley matrix.** From Eq. (4.4) it follows that

\[
J(P_\infty, t) = \mathbb{E}[(P_\infty^t w)_0^2].
\]

Now notice that

\[
(P_\infty^t w)_k = \sum_{h \in \mathbb{Z}^d} p_h^{(t)} w_{k-h}
\]

where \( p_h^{(t)} \) are the coefficients of the polynomial \( p(z_1, \ldots, z_d)^t \). Therefore

\[
\mathbb{E}[(P_\infty^t w)_0^2] = \mathbb{E} \left[ \sum_{h, k \in \mathbb{Z}^d} p_h^{(t)} p_k^{(t)} w_{k-h} \right] = \sum_{h \in \mathbb{Z}^d} (p_h^{(t)})^2 = q_h^{(t)}.
\]

**Proof of Prop. 4.2 – finite Cayley matrix.** In this case,

\[
J(P_n, t) = \frac{1}{N} \sum_{h_1=0}^{n_1-1} \cdots \sum_{h_d=0}^{n_d-1} \sum_{k \in \mathbb{Z}^d} q_h^{(t)} e^{i \left( \frac{2\pi}{n_r} h_1 k_1 + \cdots + \frac{2\pi}{n_r} h_d k_d \right)}
\]

\[
= \sum_{k \in \mathbb{Z}^d} q_k^{(t)} \prod_{r=1}^{d} \frac{1}{n_r} \sum_{h_r=0}^{n_r-1} e^{i \frac{2\pi}{n_r} h_r k_r}
\]

which ends the proof, because

\[
\frac{1}{n_r} \sum_{h_r=0}^{n_r-1} e^{i \frac{2\pi}{n_r} h_r k_r} = \begin{cases} 
1 & \text{if } k_r = 0 \mod n_r, \\
0 & \text{otherwise}.
\end{cases}
\]

**Proof of Prop. 4.2 – grid matrix.** For the grid matrix,

\[
J(\overline{P}_n, t) = \frac{1}{N} \sum_{k \in \mathbb{Z}^d} \sum_{h_1=0}^{n_1-1} \cdots \sum_{h_d=0}^{n_d-1} e^{i \left( \frac{2\pi}{n_r} h_1 k_1 + \cdots + \frac{2\pi}{n_r} h_d k_d \right)}.
\]

(A.1)
By plugging this into Eq. (A.1), we get

$$q_k = \frac{1}{2^d} \sum_{w \in \{-1, 1\}^d} q_{w \otimes k}^{(t)}.$$ 

By plugging this into Eq. (A.1), we get

$$J(F_n, t) = \sum_{k \in D_{2^d}} \sum_{w \in \{-1, 1\}^d} q_{w \otimes k}^{(t)} \prod_{r=1}^d \frac{1}{2n_r} \sum_{h_r=0}^{n_r-1} e^{j\pi h_r k_r},$$

from which, by exchanging the order of summations and letting $k^t := w \otimes k$, we obtain

$$J(F_n, t) = \sum_{w \in \{-1, 1\}^d} \sum_{k' \in D_{2^d}} q_k^{(t)} \prod_{r=1}^d \frac{1}{2n_r} \sum_{h_r=0}^{n_r-1} e^{j\pi h_r k'_r w_r},$$

or for the grid matrix $P_n$, the expression for $J(F_n, t)$ given in Prop. 4.2 can be re-written as

$$J(F_n, t) = \sum_{h \in \mathcal{F}_{\delta, n}} q_h^{(t)} + \sum_{k \in D_{2^d}} \prod_{r=1}^d \frac{1}{2n_r} \sum_{h \in \mathcal{F}_{k, n}} q_h^{(t)}.$$

Now we use the assumption of quadrantal symmetries: denoting by $w \otimes w$ the entrywise product, we have $q_k = q_{w \otimes k}$ for all $w \in \{-1, 1\}^d$, and so

$$J(F_n, t) = \prod_{r=1}^d \frac{1}{2n_r} \sum_{h_r=0}^{n_r-1} e^{j\pi h_r k_r}.$$ 

Proof of Corollary 4.3. For the Cayley matrix $P_n$, in the expression for $J(P_n, t)$ given in Prop. 4.2, notice that the only coefficients $q_h^{(t)}$ in the summation which can be non-zero are those where $h \in D_{2^d}$, namely $-2\delta h_r \leq h_r \leq 2\delta$ for all $r = 1, \ldots, d$. If $n > 2\delta 1$, then $D_{2^d} \cap \mathcal{F}_n = \{0\}$, so that $J(P_n, t) = q_{0}^{(t)} = J(P_{\infty}, t)$.

For the grid matrix $P_n$, the expression for $J(F_n, t)$ given in Prop. 4.2 can be re-written as

$$J(F_n, t) = \sum_{h \in \mathcal{F}_{\delta, n}} q_h^{(t)} + \sum_{k \in D_{2^d}} \prod_{r=1}^d \frac{1}{2n_r} \sum_{h \in \mathcal{F}_{k, n}} q_h^{(t)}.$$ 

The proof ends by computing

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

Proof of Corollary 4.4. The second Corollary is less immediate. The main tool is the following Lemma, which ensures that if $p(z_1, \ldots, z_d)$ satisfies the assumptions of Coroll. 4.4, then also $q^{(t)}(z_1, \ldots, z_d)$ satisfies them (notice that $q^{(t)}(z_1, \ldots, z_d) = (p(z_1, \ldots, z_d))^{2t}$ due to the quadrant symmetries).

**Lemma A.1.** Given two sequences of non-negative numbers $\{a_h\}, \{b_h\} \in \mathbb{R}_{\geq 0}^d$ satisfying the assumptions of Coroll. 4.4, then also their convolution $\{c_h\}$ defined by

$$c_h = \sum_{k \in \mathbb{Z}^d} a_k b_{h-k}$$

satisfies the same assumptions.

**Proof.** Non-negativity and quadrant symmetries are immediate. What is left to prove is that, for all $i = 1, \ldots, d$, if $k_j = h_j \geq 0$ for all $j \neq i$ and $k_i = h_i + 1 \geq 1$, then
where \( J \) lighting the first component

\[ n + k \]

Notice that \( F \) been re-written by changing the index in the last summation (so as to start from 0), getting

\[ c_{h, h'} - c_{h+1, h'} = \sum_{(k, k') \in \mathbb{Z}^d} (a_{k, k'} - a_{k+1, k'}) b_{h-k, h' - k'} \]

\[ = \sum_{k' \in \mathbb{Z}^{d-1}} \left[ \sum_{k \geq 0} (a_{k, k'} - a_{k+1, k'}) b_{h-k, h' - k'} + \sum_{k \geq 1} (a_{k, k'} - a_{k-1, k'}) b_{h+k, h' - k'} \right] \]

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

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

Notice that \( a_{k, k'} - a_{k+1, k'} \geq 0 \) by assumption, and also \( b_{h-k, h' - k'} - b_{h+k, h' - k'} \geq 0 \), 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 \( b_{h-k, h' - k'} = b_{h+k, h' - k'} \geq b_{h+k, h' - k'} \).

We want to prove that, under the assumptions of Coroll. 4.4, \( n \gg m \) implies \( J(P_n, t) \leq J(P_m, t) \) and \( J(P_n, t) \leq J(P_m, t) \). For ease of notation, without loss of generality, we will consider the case where \( n = (n_1, n') \) and \( m = (m_1, n') \), with \( n_1 \geq m_1 \). The key point we will exploit is that, by the assumptions and by Lemma A.1, for all \( h, k_1 \in \mathbb{Z} \) and for all \( h' \in \mathbb{Z}^{d-1}, |h_1| \geq |k_1| \) implies that \( q_{h_1, h'}^{(t)} \leq q_{k_1, h'}^{(t)} \).

For the Cayley case, by Prop. 4.2,

\[ J(P_n, t) = \sum_{h \in \mathcal{F}_n} q_h^{(t)}, \]

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

\[ J(P_n, t) = \sum_{\ell \in \mathbb{Z}} \sum_{h' \in \mathcal{F}_{n'}} q_{\ell n_1, h'}^{(t)} \leq \sum_{\ell \in \mathbb{Z}} \sum_{h' \in \mathcal{F}_{n'}} q_{\ell n_1, h'}^{(t)} = J(P_m, t). \]

For the grid case, by Prop. 4.2,

\[ J(P_n, t) = \sum_{K \subseteq [d]} \frac{1}{\prod_{r \in K} n_r} \sum_{h \in \mathcal{F}_K} q_h^{(t)}, \]

where \( \mathcal{F}_K := \{(h_1, \ldots, h_d) : h_r \text{ is odd } \forall r \in K \text{ and } h_r = 0 \text{ mod } 2n_r \ \forall r \notin K \} \). 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 \( J(P_m, t) \). If \( 1 \notin K \), then

\[ \prod_{r \in K} n_r = \prod_{r \in K} m_r \]

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

\[ \sum_{h \in \mathcal{F}_K} q_h^{(t)} = \sum_{\ell \in \mathbb{Z}} \sum_{h' \in \mathcal{F}_{K', h'}} q_{\ell n_1, h'}^{(t)} \leq \sum_{\ell \in \mathbb{Z}} \sum_{h' \in \mathcal{F}_{K', h'}} q_{\ell n_1, h'}^{(t)} = \sum_{h \in \mathcal{F}_{K, m}} q_h^{(t)}. \]
If $1 \in K$, then
\[
\frac{1}{\prod_{r \in K} m_r} \leq \frac{1}{\prod_{r \in K} m_r}
\]
and $\mathcal{F}_{K,n} = \mathcal{F}_{K,m}$, which ends the proof.

**A.3. Preliminary bounds.** Our bounds are based on a simple but effective technique for getting a bound for the eigenvalues. In fact, the essential object in the expressions of the eigenvalues is the function $f: \mathbb{R}^d \to [0, +\infty)$ defined by
\[
f(x) = |p(e^{ix_1}, \ldots, e^{ix_d})|^2.
\]
Notice that $f(x) = q(e^{ix_1}, \ldots, e^{ix_d}) = \sum_{\ell \in D_\pi} q_\ell \cos (\ell_1 x_1 + \cdots + \ell_d x_d)$, where $q(z_1, \ldots, z_d) = p(z_1, \ldots, z_d)p(z_1^{-1}, \ldots, z_d^{-1})$.

Clearly $f$ is a trigonometric polynomial, with $f(0) = 1$ and $0 \leq f(x) \leq 1$ for all $x$. Under the assumptions of Theorem 4.5 we can also guarantee that the maximum in $x = 0$ is unique in the region $(-2\pi, 2\pi)^d$. To this aim, in the sequel we will assume that the support set of $p(z_1, \ldots, z_d)$, defined as $S(p) := \{ k \in D_\delta : p_k \neq 0 \}$, contains the origin and generates $\mathbb{Z}^d$. Note that, by Lemma 4.1, this is equivalent to assuming that the infinite Cayley graph on $\mathbb{Z}^d$ associated with $p(z_1, \ldots, z_d)$ is weakly connected and has self-loops. Under such assumptions, we can prove the following properties the maximum of $f$ in the origin.

**Lemma A.2.** With the above notations and assumptions,
\[
f(x) < 1 \text{ for all } x \in (-2\pi, 2\pi)^d \setminus \{0\}
\]
and moreover the Hessian matrix of $f$ in $x = 0$ is negative definite.

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

For the second claim, denote by $H$ the Hessian matrix of $f$ in 0, which is given by $H_{rs} := \frac{\partial^2 f}{\partial x_r \partial x_s} = -\sum_{h \in S(q)} q_h h_r h_s$. Our aim is to prove that $-H$ is positive definite. First observe that $H = -\sum_{h \in S(q)} q_h hh^T = -LDL^T$ 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 $q_{\ell^{(1)}}^2, \ldots, q_{\ell^{(s)}}^2$. Since, by definition $D$ is positive definite, and since $L$ has full rank, then $LDL^T$ has to be positive definite. \(\square\)

From Lemma A.2, it immediately follows the following bound on $f(x)$, which will be useful in the proof of Theorem 4.5.

**Lemma A.3.** Under the assumptions of Lemma A.2, there exists $a \in (0, \pi)$, $\alpha, \beta > 0$, $c \in (0, 1)$, depending only on $p(z)$ and $d$, such that, for all $x \in [-\pi, \pi]^d$,
\[
f_L(x) \leq f(x) \leq f_U(x),
\]
where the functions $f_U$ and $f_L$ are defined as
\[
f_U(x) = \begin{cases} 
    e^{-\alpha x^T x} & \text{for } x \in (-a, a)^d \\
    c & \text{otherwise,}
\end{cases}
\]
and
\[
f_L(x) = \begin{cases} 
    e^{-\beta x^T x} & \text{for } x \in (-a, a)^d \\
    0 & \text{otherwise.}
\end{cases}
\]
We conclude with the following lemma which provides useful lower and upper bounds.

**Lemma A.4.** Given $\gamma > 0$,

$$\sqrt{\frac{\pi}{\gamma t}} \left( 1 - e^{-\gamma a^2} \right) \leq \int_{[-a, a]} e^{-\gamma x^2 t} dx \leq \sqrt{\frac{\pi}{\gamma t}}$$

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

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

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

$$\int_{[-a, a]} e^{-\gamma x^2 t} dx = \int_{\mathbb{R}} e^{-\gamma x^2 t} dx - 2 \frac{1}{\sqrt{\gamma t}} \text{erfc}(\sqrt{\gamma t} a) > \sqrt{\frac{\pi}{\gamma t}} (1 - e^{-\gamma a^2}) .$$

\[ \square \]

**A.4. Proof of Thm. 4.5.** Let’s start with the infinite lattice. From Eq. (4.1),

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

where $p_h^{(t)}$ denotes the coefficients of the polynomial $(p(z_1, \ldots, z_d))^t$.

By Parseval’s identity, this expression can be re-written as

$$J(\infty, t) = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} |p(x_1, \ldots, x_d)|^2 dx = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} (f(x_1, \ldots, x_d))^t dx .$$

Now it is easy to conclude, using the bounds from Lemma A.3 and then applying the Lemma A.4 to each component.

The proof for the Cayley matrix and for the grid is slightly more involved, and will use the explicit expressions for the eigenvalues, Lemma A.3 and the following lemma.

**Lemma A.5.** 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

$$A_{n,c,\gamma}(t) = \frac{1}{N} \sum_{h : |h| \leq |cn_r|} e^{-\gamma \left( \left( \frac{2\pi h_1}{n_1} \right)^2 + \cdots + \left( \frac{2\pi h_d}{n_d} \right)^2 \right)^t} .$$

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

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

**Proof.**

We start from the upper bound. Let $\mathcal{I} := \{(h_1, \ldots, h_d) \in \mathbb{Z}^d : -|cn_r| \leq h_r \leq |cn_r| \ \forall r \}$. 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 \( I \) as \( K \) varies over all the possible subsets of \([d]\) (including \( K = \emptyset \) and \( K = [d] \)). Then

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

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

\[
\sum_{h \in \mathcal{I}_K} \prod_{r=1}^d e^{-\gamma (\frac{2\pi}{nh_r})^2 t} = \sum_{[cn_1] \leq h_1 \leq [cn_1]} \cdots \sum_{[cn_s] \leq h_s \leq [cn_s]} e^{-\gamma (\frac{2\pi}{nh_1})^2 t} \cdots e^{-\gamma (\frac{2\pi}{nh_s})^2 t}
\]

\[
= \left( \sum_{[cn_1] \leq h_1 \leq [cn_1]} e^{-\gamma (\frac{2\pi}{nh_1})^2 t} \right) \cdots \left( \sum_{[cn_s] \leq h_s \leq [cn_s]} e^{-\gamma (\frac{2\pi}{nh_s})^2 t} \right)
\]

\[
= 2^s \prod_{r=1}^s \left( \sum_{1 \leq h_r \leq [cn_r]} e^{-\gamma (\frac{2\pi}{nh_r})^2 t} \right)
\]

Then, using the following upper bound

\[
\frac{2}{\sqrt{n}} \sum_{1 \leq h_r \leq [cn_r]} e^{-\gamma (\frac{2\pi}{nh_r})^2 t} \leq \frac{1}{\pi} \int_0^{\frac{2\pi}{nh_r}} e^{-\gamma x^2 t} dx \leq \frac{1}{2\pi} \int_\mathbb{R} e^{-\gamma x^2 t} dx = \frac{1}{2} \left( \frac{1}{\pi \gamma t} \right)^{\frac{1}{2}}
\]

we obtain

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

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 \( I \). Rather, we define, for any \( K \subseteq [d] \),

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

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 \( h \notin \mathcal{J}_K \) in the summation which defines \( A_{n,c,\gamma}(t) \), namely we use the fact that, for all \( K \subseteq [d] \) we have that

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

The choice \( K = \emptyset \) simply gives

\[
A_{n,c,\gamma}(t) \geq \frac{1}{N}.
\]
The choice $K = \{1, \ldots, s\}$ gives

\[
A_{n,c,\gamma}(t) \geq \frac{1}{N} \sum_{|cn_1| \leq h_1 \leq |cn_s|} \cdots \sum_{|cn_s| \leq h_s \leq |cn_s|} e^{-\gamma(\frac{2\pi}{n_r}h_r)^2t} \geq \left( \prod_{r=1}^{s} \frac{1}{n_r} \sum_{0 \leq h_r \leq |cn_r|} e^{-\gamma(\frac{2\pi}{n_r}h_r)^2t} \right) \left( \prod_{r=s+1}^{d} \frac{1}{n_r} \right) \geq \left( \frac{1}{2\pi} \int_{0}^{2\pi c} e^{-\gamma t x^2} \, dx \right)^s \prod_{r=1}^{s} \frac{n_r}{N}.
\]

Then, we end by using Lemma A.4, which gives

\[
\int_{0}^{2\pi c} e^{-\gamma t x^2} \, dx > \frac{1}{2} \sqrt{\frac{\pi}{\gamma t}} (1 - e^{-\gamma t(2\pi c)^2}) \geq \frac{1}{2} \sqrt{\frac{\pi}{\gamma t}} (1 - e^{-\gamma(2\pi c)^2})
\]

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

Now we have all the tools for the proof of the theorem. Let’s first consider the torus.

\[
J(P_n,t) = \frac{1}{N} \sum_{h \in V_n} |\lambda_h|^{2t} = \frac{1}{N} \sum_{h \in V_n} \left[ f(e^{i\frac{2\pi}{n_1}h_1}, \ldots, e^{i\frac{2\pi}{n_d}h_d}) \right]^t
\]

Define

\[
V_n' := \left\{ -\left[ \frac{an_r}{2\pi} \right], \ldots, 0, \ldots, +\left[ \frac{an_r}{2\pi} \right] \right\} \times \cdots \times \left\{ -\left[ \frac{an_d}{2\pi} \right], \ldots, 0, \ldots, +\left[ \frac{an_d}{2\pi} \right] \right\}.
\]

Clearly, $f(x)$ has period $2\pi$ in each of its variables, and so

\[
J(P_n,t) = \frac{1}{N} \sum_{h \in V_n'} \left[ f(e^{i\frac{2\pi}{n_1}h_1}, \ldots, e^{i\frac{2\pi}{n_d}h_d}) \right]^t
\]

From Lemma A.3 it follows

\[
J(P_n,t) \leq \frac{1}{N} \sum_{h \in V_n''} \left[ f_U(e^{i\frac{2\pi}{n_1}h_1}, \ldots, e^{i\frac{2\pi}{n_d}h_d}) \right]^t,
\]

where $f_U$ is defined in Lemma A.3. Now consider the set

\[
V_n''' := \left\{ h \in V_n' : -a \leq \frac{2\pi}{n_r} h_r \leq a \quad \forall r \right\} = \left\{ h \in V_n' : -\left[ \frac{an_r}{2\pi} \right] \leq h_r \leq \left[ \frac{an_r}{2\pi} \right] \quad \forall r \right\}.
\]

Then, using the definition of $f_U$, we get

\[
J(P_n,t) \leq \frac{1}{N} \sum_{h \in V_n'''} e^{-a \left( \left( \frac{2\pi}{n_1}h_1 \right)^2 + \cdots + \left( \frac{2\pi}{n_d}h_d \right)^2 \right)} + c^t
\]

Now you can conclude using Lemma A.5.
For the lower bound, the proof is very similar. Indeed,
\[
J(P_n, t) = \frac{1}{N} \sum_{h \in V_n} \left[ f(e^{i \frac{2\pi}{n} h_1}, \ldots, e^{i \frac{2\pi}{n} h_d}) \right]^t
\geq \frac{1}{N} \sum_{h \in V_n} \left[ f_L(e^{i \frac{2\pi}{n} h_1}, \ldots, e^{i \frac{2\pi}{n} h_d}) \right]^t
= \frac{1}{N} \sum_{h \in V_n} e^{-\beta \left( \frac{2\pi}{n} h_1 \right)^2 + \cdots + \left( \frac{2\pi}{n} h_d \right)^2} t
\]
Finally, you conclude using Lemma A.5.

When you consider grids instead of toruses, the proof is very similar. In this case,
\[
J(P_n, t) = \frac{1}{N} \sum_{h \in V_n} \left[ f(e^{i \frac{\pi}{n} h_1}, \ldots, e^{i \frac{\pi}{n} h_d}) \right]^t
\]
and so
\[
J(P_n, t) \leq \frac{1}{N} \sum_{h \in V_n \cap V_{2n}} e^{-\alpha \left( \frac{\pi}{n} h_1 \right)^2 + \cdots + \left( \frac{\pi}{n} h_d \right)^2} t + e^t \leq \frac{1}{N} \sum_{h \in V_n \cap V_{2n}} e^{-\alpha \left( \frac{\pi}{n} h_1 \right)^2 + \cdots + \left( \frac{\pi}{n} h_d \right)^2} t + e^t
\]
and
\[
J(P_n, t) \geq \frac{1}{N} \sum_{h \in V_n \cap V_{2n}} e^{-\beta \left( \frac{\pi}{n} h_1 \right)^2 + \cdots + \left( \frac{\pi}{n} h_d \right)^2} t \geq \frac{1}{N} \sum_{h \in V_n \cap V_{2n}} e^{-\beta \left( \frac{\pi}{n} h_1 \right)^2 + \cdots + \left( \frac{\pi}{n} h_d \right)^2} t
\]
Then again you conclude using Lemma A.5.