Performance of consensus algorithms in large-scale distributed estimation

Federica Garin and Sandro Zampieri
Dept. of Information Engineering (DEI), Università di Padova
Via Gradenigo 6/B, 35131 Padova, Italy
Email: garin@dei.unipd.it, zampi@dei.unipd.it

***Abstract*—When consensus algorithms are used in very large networks, spreading information across the whole graph requires a long time. Hence, traditional convergence analysis, studying the essential spectral radius of the transition matrix, predicts very poor performance. However, in estimation problems, it is clear that a growing number of measurements improves the quality of the estimate, and it is natural to expect such behaviour even though the best estimate is approximated using distributed algorithms. Then, it is important to define a suitable performance metric, depending on the actual estimation or control problem in which the consensus algorithm is used. This allows to study how performance scales when both computation time and number of agents grow to infinity, for different communication graphs and choices of the algorithm.**

I. Introduction

Average-consensus algorithms allow to compute an average in a distributed way. These algorithms have been extensively applied to the solution of problems of distributed estimation [13] and of sensor calibration for sensor networks [12], to load balancing for distributed computing systems [7], and to mobile multi-vehicles coordination [6].

In this paper, we focus on linear average-consensus. The analysis of such algorithms usually exploits results from Markov chains literature, and is focused on predicting the speed of convergence to the average, when computation time grows. There has been an extensive literature on this topic, with both analysis and optimization of convergence speed. However, we believe that when convergence to the average is not an objective per se, but is used to solve an estimation or control problem, it is important to consider different performance measures, more tightly related to the actual objective pursued. Literature along this research line is not very developed; some contributions have considered the effects of robustness to delays [14], noise on the communication links [16] and quantization [11].

In this paper, we focus on an estimation problem, for which the natural performance measure is mean quadratic error. We show that this cost is a function of all eigenvalues of the linear consensus update matrix, as opposed to classical convergence analysis which involved only the essential spectral radius. This remark calls for new results in spectral graph theory.

The study of a performance measure different from convergence speed is essential in large-scale networks, where not only computation time is large, but also the number of agents grows. In fact, the larger the network, the slower the convergence, but on the other side a larger number of measurements can give a better estimate. Our analysis of mean quadratic error allows to study the asymptotics with respect to both time and number of agents going to infinity.

In this paper, we consider families of consensus matrices associated with communication graphs having a structure, which allows to compute or at least to estimate the mean quadratic error. A first class is the one of Cayley graphs, which exhibit high symmetry and which have been largely studied in recent distributed estimation and control literature, see e.g. [1], [8] and [4]. Moreover, a result from [2] has allowed us to extend our results from Cayley graphs to grids; we believe this is an important preliminary work in order to tackle the more interesting problem of geometric graphs, which are related to grids, as suggested by results in [3], [15] and [5]. Finally, we have considered 'de Bruijn' graphs, introduced in consensus literature in [9] because of their well-known properties of fast information transfer, classically exploited in computer science.

II. 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[1].

The sensors' measurements form a vector $\boldsymbol{x}(0) \in \mathbb{R}^N$, with $\boldsymbol{x}(0)_k = \theta + n_k$, where the noises $n_1, \dots, n_N$ 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: $\boldsymbol{x}(t+1) = P\boldsymbol{x}(t)$ for some doubly-stochastic and primitive[2] $N \times N$ matrix P consistent with the communication graph G, i.e., such that $(u, v) \notin E$ implies $P_{uv} = 0$. It is well-known that, for $t \to \infty$, $\boldsymbol{x}(t) \to \frac{1}{N}\mathbf{1}^T\boldsymbol{x}(0)\mathbf{1}$, where $\mathbf{1}$ denotes a

[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 1; e.g., the presence of a self-loop implies aperiodicity.

[2] P is primitive if $\exists m$ such that $(P^m)_{uv} \neq 0 \,\forall u, v \in V$. Equivalently, the graph $\mathcal{G} = (V, \mathcal{E})$ with $\mathcal{E}$ defined by $(u, v) \in \mathcal{E} \iff P_{uv} \neq 0$ is strongly connected and aperiodic. Primitive stochastic matrices have dominant eigenvalue 1 with multiplicity 1.

column vector with all-1 entries, and that the speed of such convergence is given by the essential spectral radius of P, $\rho_{\mathrm{ess}}(P)$, i.e., the eigenvalue of P which has second largest modulus. For non-expander families of graphs, such as for example Cayley graphs on Abelian groups, when $N \to \infty$, $\rho_{\mathrm{ess}}(P) \to 1$. 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.

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

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

where T denotes transpose and where the error $\boldsymbol{e}(t)$ is defined as $\boldsymbol{e}(t) = \boldsymbol{x}(t) - \theta\mathbf{1}$. For our problem, it is easy to show that the cost $J_N(t)$ can be re-written as

$$J_N(t) = \tfrac{1}{N}\ \mathrm{trace}\left((P^t)^T P^t\right) \tag{1}$$

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

$$J_N(t) = \tfrac{1}{N}\sum_{i=0}^{N-1}|\lambda_i|^{2t} \tag{2}$$

where $\lambda_0, \dots, \lambda_{N-1}$ are the eigenvalues of P. In the next sections, we will study the asymptotic behaviour of $J_N(t)$ when both N and t grow to infinity, for some families of graphs.

III. Simple examples: circle and line

We start by considering two simple examples, which clarify how the mean quadratic error can be decreasing for $N \to \infty$ even though $\rho_{\mathrm{ess}}(P) \to 1$. We discuss in this section the examples in full detail, as they allow to describe with lighter notation the same issues of the more general families of graph which are the topic of next section.

A. Circle and infinite line

As a communication graph, consider an undirected circle of N sensors, where each sensor is connected to its first δ neighbours to the right and to the left (for some positive integer $\delta < N$). Let P_N be a circulant matrix whose first row is $(p_0, \dots, p_\delta, 0, \dots, 0, p_{-\delta}, \dots, p_{-1})$. For future ease of notation, we define the Laurent polynomial $p(z) = \sum_{h=-\delta}^{\delta} p_h z^h$. We assume that $p_h \geq 0$ for all h and that $\sum_{h=-\delta}^{\delta} p_h = 1$. Moreover, we assume that $p_0 \neq 0$ and p_{-1} or p_1 (or both) are non-zero; this ensures primitivity of P_N for any N. We consider a family of such graphs and matrices, with growing N but fixed δ and same weights $(p_{-\delta}, \dots, p_0, \dots, p_\delta)$.

In this example, it makes sense to consider the infinite limit of the graph: an infinite line, so that $\boldsymbol{x}(t) \in \mathbb{R}^{\mathbb{Z}}$ (i.e., the set of bi-infinite sequences) and P is infinite banded-Toeplitz with diagonal band $(p_{-\delta}, \dots, p_0, \dots, p_\delta)$. Here, the average quadratic error $\mathbb{E}[(\boldsymbol{e}_u(t))^2]$ (where $\boldsymbol{e}_u(t) = \boldsymbol{x}_u(t) - \theta$) is the same for all $u \in \mathbb{Z}$, so that we can look, for example, at node 0 and define

$$J_\infty(t) = \mathbb{E}[(\boldsymbol{e}_0(t))^2]\,.$$

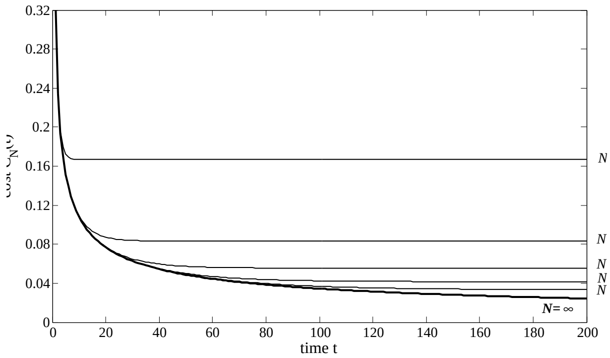

Fig. 1. Mean quadratic error (for different N's) for circle with weights $p_{-1} = p_0 = p_1 = \frac{1}{3}$.

Using the fact that $\boldsymbol{x}(0)_k = \theta + n_k$ (i.i.d. noise with zero mean and variance one), we can simplify the cost:

$$J_\infty(t) = \mathbb{E}\big[\big((P^t\boldsymbol{n})_0\big)^2\big]$$

and so

$$J_\infty(t) = \sum_{h\in\mathbb{Z}}\Big(\sum_{\substack{h_1,\dots,h_t\in\{-\delta,\dots,\delta\}\\ h_1+\dots+h_t=h}} p_{h_1}\cdot\ldots\cdot p_{h_t}\Big)^2$$

If we define the Laurent polynomial $q^{(t)}(z) = (p(z))^t\left(p(z^{-1})\right)^t = \sum_{h=-2\delta t}^{2\delta t} q_h^{(t)} z^h$, this expression can be re-written as simply

$$J_\infty(t) = q_0^{(t)} \tag{3}$$

The example of the circle graph is interesting because we can show that $\rho_{\mathrm{ess}}(P_N) \to 1$ for $N \to \infty$, but nevertheless $J_N(t) = J_\infty(t)$ for large N, and $J_\infty(t)$ decreases to zero as $\frac{1}{\sqrt{t}}$ when $t \to \infty$. More precisely, for both $N, t \to \infty$, $J_N(t)$ behaves like the maximum between $\frac{1}{N}$ and $\frac{1}{\sqrt{t}}$. As an example, Fig. 1 depicts the cost for various N's in the case where $p_{-1} = p_0 = p_1 = \frac{1}{3}$.

Let's see these facts in more detail, starting with the infinite line.

Proposition 1: There exist positive constants c_1, c_2 such that, for all $t \geq 1$:

$$\frac{c_1}{\sqrt{t}} \leq J_\infty(t) \leq \frac{c_2}{\sqrt{t}}$$

Proof: From Eq. (3), by Parseval's identity applied to the function $f(x) = \left(p(e^{jx})\right)^t$ (where $j = \sqrt{-1}$) we get:

$$J_\infty(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}|p(e^{jx})|^{2t}\,\mathrm{d}x\,.$$

Then we focus our attention on the function $g(x) = |p(e^{jx})|^2$, for which we find a lower bound g_L and an upper bound g_U. Notice that $g : \mathbb{R} \to [0, +\infty)$ is a trigonometric polynomial, with $g(0) = 1$ and with $g(x) < 1$ for all $x \in [-\pi, \pi] \setminus \{0\}$. The derivatives are $g'(0) = 0$ and $g''(0) = -\sum_{h=-2\delta}^{2\delta}\left(\sum_{h_1,h_2:h_1-h_2=h} p_{h_1}p_{h_2}\right)h^2 < 0$. Now let $0 < \alpha < -g''(0) < \beta$, so that there exists a neighborhood of 0, say $(-a, a)$, such that $e^{-\beta x^2} \leq g(x) \leq e^{-\alpha x^2}$ for all $x \in$

$(-a,a)$ and also $g(x) \leq e^{-\alpha a^2}$ for all $x \in [-\pi,-a] \cup [a,\pi]$. Now define functions

$$g_L(x) = \begin{cases} e^{-\beta x^2} & \text{for } x \in (-a,a) \\ 0 & \text{otherwise} \end{cases}$$

and

$$g_U(x) = \begin{cases} e^{-\alpha x^2} & \text{for } x \in (-a,a) \\ e^{-\alpha a^2} & \text{otherwise} \end{cases}$$

so that $g_L(x) \leq g(x) \leq g_U(x)$ in the interval $[-\pi,\pi]$. For the upper bound:

$$J_\infty(t) \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} (g_U(x))^t \,\mathrm{d}x \leq \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha t x^2}\,\mathrm{d}x + (e^{-\alpha a^2})^t$$

Then notice that $\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha t x^2}\,\mathrm{d}x = \frac{1}{2\sqrt{\pi\alpha}}\frac{1}{\sqrt{t}}$, which ends the proof of the upper bound. For the lower bound:

$$J_\infty(t) \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} (g_L(x))^t \,\mathrm{d}x = \frac{1}{2\pi}\int_{-a}^{a} e^{-\beta t x^2}\,\mathrm{d}x$$

and then use the well-known bound $\operatorname{erfc} x \leq e^{-x^2}$ for all $x \geq 0$, where $\operatorname{erfc} x = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-y^2}\,\mathrm{d}y$, obtaining:

$$\begin{aligned}\frac{1}{2\pi}\int_{-a}^{a} e^{-\beta t x^2}\,\mathrm{d}x &= \frac{1}{2\sqrt{\pi\beta}}\frac{1}{\sqrt{t}}(1-\operatorname{erfc}(\sqrt{\beta}a\sqrt{t})) \\ &\geq \frac{1}{2\sqrt{\pi\beta}}\frac{1}{\sqrt{t}}(1-e^{-\beta a^2 t})\end{aligned}$$

■

Now let's see the results for circular graphs. Let's recall that P_N is normal (so Eq. (2) applies), and its eigenvalues are, for $k = 0,\dots,N-1$, $\lambda_k = p(e^{j\frac{2\pi}{N}k})$. It is clear that $\lim_{N\to\infty} \rho_{\text{ess}}(P_N) = 1$. However the following facts are true for the mean quadratic error.

Proposition 2: $J_N(t)$ is non-increasing with respect to N and, for all $N > \delta t$, $J_N(t) = J_\infty(t)$ holds.

Proof: The cost is

$$\begin{aligned}J_N(t) &= \tfrac{1}{N}\sum_{h=0}^{N-1} q^{(t)}(e^{j\frac{2\pi}{N}h}) \\ &= \tfrac{1}{N}\sum_{k=-2t\delta}^{2t\delta} q_k^{(t)} \sum_{h=0}^{N-1} e^{j\frac{2\pi}{N}hk} \\ &= \sum_{\substack{-2t\delta\leq k\leq 2t\delta \\ k=0 \bmod N}} q_k^{(t)}\end{aligned}$$

This equality can also be interpreted as Parseval's identity for the Fourier transform over the group $\mathbb{Z}_N$ of the function $f(x) = \left(p(e^{jx})\right)^t$.

Notice that, if $N > 2t\delta$, then $J_N(t) = q_0^{(t)} = J_\infty(t)$. When N decreases, either the cost remains the same, or new positive terms are added in the summation, thus increasing the cost. ■

The proposition above states formally the very natural fact that at time t each node has received information only from $t\delta$ neighbours on his right and the same on his left, and for all $N > 2t\delta$ the node can't see any difference from an analogous node laying in the infinite line.

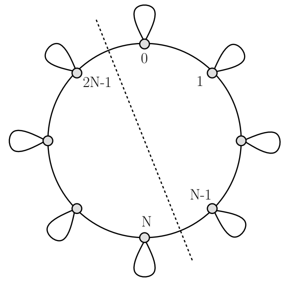

Fig. 2. Circle with $2N$ vertices and reflection axis corresponding to the map $l \mapsto 2N-1-l$, used in the construction of a line with N vertices.

Proposition 3: There exist positive constants c_1, c_2 such that, for all $t \geq 1$:

$$\max\left\{\frac{1}{N}, \frac{c_1}{\sqrt{t}}\right\} \leq J_N(t) \leq \frac{1}{N} + \frac{c_2}{\sqrt{t}}$$

Proof: Consider the same functions g, g_L and g_U defined for the proof of Prop. 1; note that both g_L and g_U are even functions, and they are decreasing on $[0,\pi]$. With this notation, $J_N(t) = \frac{1}{N}\sum_{h=0}^{N-1}(g(\frac{2\pi}{N}h))^t$. So, for the upper bound:

$$\begin{aligned}J_N(t) &\leq \tfrac{1}{N} + 2\tfrac{1}{N}\sum_{h=1}^{\lceil (N-1)/2\rceil} (g(\tfrac{2\pi}{N}h))^t \\ &\leq \tfrac{1}{N} + \tfrac{1}{\pi}\tfrac{2\pi}{N}\sum_{h=1}^{\lceil (N-1)/2\rceil} (g_U(\tfrac{2\pi}{N}h))^t \\ &\leq \tfrac{1}{N} + \tfrac{1}{\pi}2\int_0^\pi (g_U(x))^t\,\mathrm{d}x\end{aligned}$$

and then conclude, estimating $\int_{-\pi}^{\pi}(g_U(x))^t\,\mathrm{d}x$ as in the proof of Prop. 1. On the other side, we have two different lower bounds: the trivial $J_N(t) \geq \frac{1}{N}$, and also, by Propositions 2 and 1, $J_N(t) \geq J_\infty(t) \geq \frac{c_1}{\sqrt{t}}$. ■

The proposition above implies that if $N \ll \sqrt{t}$, then $J_N(t) \approx \frac{1}{N}$, while for $N \gg \sqrt{t}$, we have $J_N(t) \approx \frac{1}{\sqrt{t}}$.

B. Line

Thanks to a technique from [2], we can study also the mean quadratic error for a line of N nodes, each communicating with δ neighbours on the left and on the right, at least under some assumptions on the coefficients: we ask some symmetry and suitable modified weights at the borders of the line.

This allows to see the line as a projection of a circle with twice as many vertices, and so to obtain the eigenvalues of the consensus matrix by using [2, Prop. 3.2]. In fact, if you consider a circle with $2N$ vertices labeled consecutively as $0,\dots,2N-1$, and you take a symmetric consensus matrix $P \in \mathbb{R}^{2N\times 2N}$, circulant, with first row $(p_0,\dots,p_\delta,0,\dots,0,p_{-\delta},\dots,p_{-1})$ and with $p_{-l} = p_l$ for all l, then you have that the map $l \mapsto 2N-1-l$, (i.e. the reflection with respect to the line shown in Fig. 2) is a symmetry of the labeled graph. Thus, if we define $\overline{P} \in \mathbb{R}^{N\times N}$ as $\overline{P}_{l,m} = P_{l,m} + P_{l,2N-1-m}$ for all $l,m = 0,\dots,N-1$, then we can compute the eigenvalues of $\overline{P}$ using [2, Prop.

3.2]. Notice that $\overline{P}$ obtained in such a way is symmetric, and apart from the first and last δ rows and columns, is banded, with diagonal band $(p_{-\delta}, \ldots, p_0, \ldots, p_\delta)$.

We obtain that the eigenvalues are, for $k = 0, \ldots, N-1$, $\lambda_k = p(e^{j\frac{\pi}{N}k})$.

Now for the cost we can prove propositions very similar (both in the statement and in the proof) to those of the circle graph. With $J_\infty(t)$ we refer to the same infinite-line graph considered in Sect. III-A.

Proposition 4: $J_N(t)$ is non-increasing with respect to N, $J_N(t) > J_\infty(t)$ for all N and $\lim_{N\to\infty} J_N(t) = J_\infty(t)$.

Proof: The cost is

$$J_N(t) = \tfrac{1}{N}\sum_{h=0}^{N-1} q^{(t)}(e^{j\frac{\pi}{N}h}) = \tfrac{1}{N}\sum_{k=-2t\delta}^{2t\delta} q_k^{(t)} \sum_{h=0}^{N-1} e^{j\frac{\pi}{N}hk}$$

Now $\sum_{h=0}^{N-1} e^{j\frac{\pi}{N}hk} = N$ if k is multiple of $2N$; it is $\frac{1-e^{j\pi k}}{1-e^{j\frac{\pi}{N}k}}$ otherwise (so it's 0 for all even k). Notice that the assumption $p_k = p_{-k}$ for all k also implies $q_k^{(t)} = q_{-k}^{(t)}$ for all k and for all t, so that, for any odd k,

$$\begin{aligned} & q_k^{(t)} \sum_{h=0}^{N-1} e^{j\frac{\pi}{N}hk} + q_{-k}^{(t)} \sum_{h=0}^{N-1} e^{j\frac{\pi}{N}h(-k)} \\ & = q_k^{(t)} \left(\tfrac{1-e^{j\pi k}}{1-e^{j\frac{\pi}{N}k}} + \tfrac{1-e^{-j\pi k}}{1-e^{-j\frac{\pi}{N}k}} \right) \\ & = 2q_k^{(t)} = q_k^{(t)} + q_{-k}^{(t)} \end{aligned}$$

This allows to re-write the cost as

$$J_N(t) = \sum_{\substack{-2t\delta\le k\le 2t\delta \\ k=0 \bmod 2N}} q_k^{(t)} + \tfrac{1}{N} \sum_{\substack{-2t\delta\le k\le 2t\delta \\ k \text{ odd}}} q_k^{(t)}$$

from which the statement of our proposition is proved. ■

Proposition 5: There exist positive constants c_1, c_2 such that, for all $t \ge 1$:

$$\max\left\{\frac{1}{N}, \frac{c_1}{\sqrt{t}}\right\} \le J_N(t) \le \frac{1}{N} + \frac{c_2}{\sqrt{t}}$$

Proof: Consider the same functions g, g_L and g_U defined for the proof of Propositions 1 and 3. With this notation, $J_N(t) = \frac{1}{N}\sum_{h=0}^{N-1}(g(\frac{\pi}{N}h))^t$. So, for the upper bound:

$$\begin{aligned} J_N(t) & \le \tfrac{1}{N} + \tfrac{1}{N}\sum_{h=1}^{N-1}(g(\tfrac{\pi}{N}h))^t \\ & \le \tfrac{1}{N} + \tfrac{1}{\pi}\tfrac{\pi}{N}\sum_{h=1}^{N-1}(g_U(\tfrac{\pi}{N}h))^t \\ & \le \tfrac{1}{N} + \tfrac{1}{\pi}\int_0^\pi (g_U(x))^t \, \mathrm{d}x \end{aligned}$$

and then conclude, estimating $\int_0^\pi (g_U(x))^t \, \mathrm{d}x$ as in previous proofs. On the other side, $J_N(t) \ge \frac{1}{N}$, and, by Propositions 4 and 1, $J_N(t) \ge J_\infty(t) \ge \frac{c_1}{\sqrt{t}}$. ■

IV. Cayley graphs and grids

The simple example of the circle can be generalized to more dimensions: for dimension 2 the graph is a grid on a torus, while for general dimension you get a Cayley graph on an Abelian group.

Now we'll give a formal definition of a linear consensus map P associated with a grid on a d-dimensional torus with N_i vertices in dimension i and connections to at most δ_i nearest neighbours in dimension i. The circle described and analyzed in Sect. III-A is the particular case when $d = 1$.

Consider the vertices of the grid to be the set $V = \{0, \ldots, N_1-1\}\times\cdots\times\{0, \ldots, N_d-1\}$ and fix positive integers $\delta_1, \ldots, \delta_d$ such that $\delta_i < N_i$ for all i. Consider now the group $G = \mathbb{Z}_{N_1} \times \cdots \times Z_{N_d}$ and define a linear map $P : \mathbb{R}^G \to \mathbb{R}^G$, as follows:

$$(P\boldsymbol{x})_{\boldsymbol{h}} = \sum_{k_1=-\delta_1}^{\delta_1} \cdots \sum_{k_d=-\delta_d}^{\delta_d} p_{\boldsymbol{k}} \boldsymbol{x}_{\boldsymbol{h}-\boldsymbol{k}}$$

for some coefficients $p_{\boldsymbol{k}} \in [0, 1]$ such that $\sum_{\boldsymbol{k}} p_{\boldsymbol{k}} = 1$. Notice that here indexes belong to G, so $\boldsymbol{h} - \boldsymbol{k}$ means $(h_1 - k_1 \bmod N_1, \ldots, h_d - k_d \bmod N_d)$. If now you identify indexes in G with indexes in V in the natural way, you get a map $P : \mathbb{R}^V \to \mathbb{R}^V$. For example, if $d = 1$, P is the circulant matrix corresponding to a circular graph, as we saw in Sect. III-A.

Throughout this section, we will assume that $p_{\boldsymbol{0}} \neq 0$ (i.e., the associated graph has self-loops), and that the set of vectors $\{\boldsymbol{k} : p_{\boldsymbol{k}} \neq 0\}$ generates $\mathbb{Z}^d$; these assumptions ensure primitivity of P. For later ease of notation, it's convenient to define the Laurent polynomial $p(z_1, \ldots, z_d) = \sum_{k_1=-\delta_1}^{\delta_1} \cdots \sum_{k_d=-\delta_d}^{\delta_d} p_{\boldsymbol{k}} z_1^{k_1} \ldots z_d^{k_d}$.

It is well-known that P is normal and has eigenvalues

$$\lambda_{\boldsymbol{h}} = p(e^{j\frac{2\pi}{N_1}h_1}, \ldots, e^{j\frac{2\pi}{N_d}h_d}), \qquad \boldsymbol{h} \in V$$

so that

$$J_N(t) = \sum_{\boldsymbol{h}\in V} \left| p(e^{j\frac{2\pi}{N_1}h_1}, \ldots, e^{j\frac{2\pi}{N_d}h_d}) \right|^{2t}$$

We can also define the infinite-limit of such graph and such map, considering an infinite grid in d-dimensions and defining $P : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}^{\mathbb{Z}^d}$ to be, for each $\boldsymbol{h} \in \mathbb{Z}^d$,

$$(P\boldsymbol{x})_{\boldsymbol{h}} = \sum_{k_1=-\delta_1}^{\delta_1} \cdots \sum_{k_d=-\delta_d}^{\delta_d} p_{\boldsymbol{k}} \boldsymbol{x}_{\boldsymbol{h}-\boldsymbol{k}} \,.$$

The mean quadratic error is the same for any vertex, so we consider for example vertex $\boldsymbol{0}$:

$$J_\infty(t) = \mathbb{E}[(\boldsymbol{e}_{\boldsymbol{0}}(t))^2] = \mathbb{E}\big[\big((P^t\boldsymbol{n})_{\boldsymbol{0}}\big)^2\big]$$

The same as in Sect. III, we can re-write this cost as $J_\infty(t) = q_{\boldsymbol{0}}^{(t)}$, where $q^{(t)}(z_1, \ldots, z_d) = p^t(z_1, \ldots, z_d)p^t(z_1^{-1}, \ldots, z_d^{-1})$.

Also the example of the line with N vertices can be generalized to any dimension d, obtaining a d-dimensional grid with $N_1 \cdot \ldots \cdot N_d$ vertices from a grid on a d-dimensional torus with $2N_1 \cdot \ldots \cdot 2N_d$ vertices, via the same projection technique from [2], as follows.

Let $V_{2N} = \{0,\ldots,2N_1-1\} \times \cdots \times \{0,\ldots,2N_d-1\}$ and $V_N = \{0,\ldots,N_1-1\} \times \cdots \times \{0,\ldots,N_d-1\}$. Take $P : \mathbb{R}^{V_{2N}} \to \mathbb{R}^{V_{2N}}$ a consensus map for a grid on a d-dimensional torus with $2N_i$ vertices in dimension i and connections to at most δ_i nearest neighbours in dimension i, associated with a Laurent polynomial $p(z_1,\ldots,z_d)$. Assume that coefficients $p_{\boldsymbol{h}}$ satisfy the following quadrantal symmetry: $p_{h_1,\ldots,h_d} = p_{k_1,\ldots,k_d}$ if $\forall i, h_i = \pm k_i$. This assumption implies that reflections σ_i, defined by $\sigma_i(\boldsymbol{h}) = \boldsymbol{k}$ with $k_l = h_l$ if $l \neq i$ and $k_i = 2N_i - 1 - i$, are symmetries of the labeled grid on the torus. Now denote by H the group generated by $\sigma_1,\ldots,\sigma_d$ and define, for all $\boldsymbol{g} \in V_N$, $O_{\boldsymbol{g}} = \{\eta(\boldsymbol{g}) : \eta \in H\} \subseteq V_{2N}$. Finally, define $\overline{P} : \mathbb{R}^{V_N} \to \mathbb{R}^{V_N}$ by $\overline{P}_{\boldsymbol{h},\boldsymbol{k}} = \sum_{\boldsymbol{l} \in O_{\boldsymbol{k}}} P_{\boldsymbol{h},\boldsymbol{l}}$, for all $\boldsymbol{h},\boldsymbol{k} \in V_N$. Notice that $\overline{P}$ is symmetric and that, apart from the borders, $\overline{P}$ associates to edges of the grid the same coefficients that P associates to edges of the grid on the torus.

Using [2, Prop. 3.2], we can find the eigenvalues of $\overline{P}$:

$$\lambda_{\boldsymbol{h}} = p(e^{j\frac{\pi}{N_1}h_1},\ldots,e^{j\frac{\pi}{N_d}h_d}), \qquad \boldsymbol{h} \in G$$

and thus obtain a cost

$$J_N(t) = \sum_{\boldsymbol{h} \in G} \left| p(e^{j\frac{\pi}{N_1}h_1},\ldots,e^{j\frac{\pi}{N_d}h_d}) \right|^{2t}.$$

Propositions 1, 2, 3, 4, and 5 generalize to this setting. We give here statements only; proofs use the same basic ingredients as the ones given in Section III, but are more technical, and are not given here for lack of space.

Proposition 6: For the infinite grid on $\mathbb{Z}^d$, there exist positive constants c_1, c_2 such that:

$$\frac{c_1}{(\sqrt{t})^d} \leq J_\infty(t) \leq \frac{c_2}{(\sqrt{t})^d}$$

□

Proposition 7: For the grid on a d-dimensional torus, with $N_1 \cdot \ldots \cdot N_d$ vertices, $J_N(t)$ is non-increasing in $N_1,\ldots,N_d$ and if $N_i > 2\delta_i t$ for all $i = 1,\ldots,d$, then $J_N(t) = J_\infty(t)$.

For the grid on a d-dimensional cube, with $N_1 \cdot \ldots \cdot N_d$ vertices, $J_N(t)$ is non-increasing in $N_1,\ldots,N_d$, and, if $N_1,\ldots,N_d \to \infty$, then $J_N(t) \to J_\infty(t)$. □

Proposition 8: Both for the grid on a d-dimensional torus and for the grid on a d-dimensional cube, with $N_1 \cdot \ldots \cdot N_d$ vertices, there exist positive constants c_1, c_2 such that for all $t \geq 1$:

$$J_N(t) \geq \max_{I \subseteq \{1,\ldots,d\}} \left\{ \frac{1}{\prod_{i \notin I} N_i} \frac{c_1}{(\sqrt{t})^{|I|}} \right\}$$

and

$$J_N(t) \leq \sum_{I \subseteq \{1,\ldots,d\}} \left[\frac{1}{\prod_{i \notin I} N_i} \frac{c_2}{(\sqrt{t})^{|I|}} \right]$$

□

This last proposition means that, for $N_1,\ldots,N_d,t \to \infty$,

$$J_N \approx \max_{I \subseteq \{1,\ldots,d\}} \left\{ \frac{1}{\prod_{i \notin I} N_i} \frac{1}{(\sqrt{t})^{|I|}} \right\}$$

The result is cleaner in the case when $N_1 = \cdots = N_d = n$, where:

$$J_N \approx \max\left\{ \frac{1}{n^d}, \frac{1}{(\sqrt{t})^d} \right\}$$

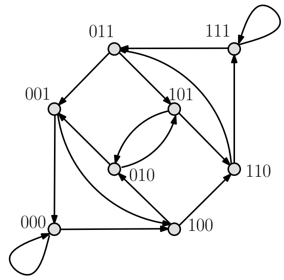

Fig. 3. 'De Bruijn' graph, $k = 2$, $n = 3$

V. 'De Bruijn' graphs

We consider now a different example, a family of graphs known for good expansion properties, introduced in the consensus literature in [9]. They are regular directed graphs, with in-degree and out-degree independent from the number of vertices N and nevertheless they have diameter only logarithmic in N.

A 'de Bruijn' graph is a directed graph (V, E), where V is the set of all n-tuples of elements of a fixed alphabet $\mathcal{A}$ of cardinality k, say $\mathcal{A} = \{0,\ldots,k-1\}$. Clearly the cardinality of V is $N = k^n$. An edge is drawn from a word $u = (u_1,\ldots,u_n)$ to a word $v = (v_1,\ldots,v_n)$ if and only if $v = (a, u_1,\ldots,u_{n-1})$ for some $a \in \mathcal{A}$. Fig. 3 shows as an example the binary 'de Bruijn' graph with $k = 2$ and $n = 3$. Notice the presence of k self-loops (on words with $u_1 = u_2 = \cdots = u_n$), which ensures that the graph is aperiodic. Also notice that both the in-degree and the out-degree of each node is k, and that the diameter of the graph is $n = \log_k N$. We will consider k as a fixed parameter and we will let n grow, in such a way that $N \to \infty$, but the degree is kept constant and the diameter is logarithmic in N.

We associate to the graph a consensus matrix with uniform weight $1/k$ on each edge. Such matrix P is not normal, so Eq. (2) does not hold, but we can explicitly compute the power P^t and the cost $J_N(t)$ by using Eq. (1). In fact, if we order vertices by reading their labels as integers written in base-k, then we can write P as

$$P = \tfrac{1}{k}\mathbf{1}_k \otimes I_{k^{n-1}} \otimes \mathbf{1}_k^T$$

where $\otimes$ denotes Kronecker product of matrices, $\mathbf{1}_l$ is a length-l all-ones vector, and I_l is a $l \times l$ identity matrix. This expression, together with properties of the Kronecker product, gives the following result.

Proposition 9: There is convergence in finite time n to average consensus:

$$P^n \boldsymbol{x}(0) = \frac{1}{N} \sum_{k=1}^{N} \boldsymbol{x}_k(0)$$

and moreover:

$$J_N(t) = \begin{cases} \frac{1}{k^t} & \text{if } t \leq n \\ \frac{1}{N} & \text{if } t \geq n. \end{cases}$$

Proof: You can re-write P as $P = \frac{1}{k}\mathbf{1}_k \otimes I_k^{\otimes(n-1)} \otimes \mathbf{1}_k^T$, where $A^{\otimes a} = A \otimes \cdots \otimes A$ (a times). From this, a proof by induction, using the the property that $(A\otimes B)(C\otimes D) = (AC)\otimes(BD)$ whenever the sizes of matrices A, B, C, D allow to write the right-hand side expression, allows to obtain:

$$P^t = \frac{1}{k^t}\mathbf{1}_k^{\otimes t} \otimes I_k^{\otimes(n-t)} \otimes (\mathbf{1}_k^T)^{\otimes t}.$$

Then, it immediately follows that

$$P^t = \begin{cases} \frac{1}{k^t}\mathbf{1}_{k^t} \otimes I_{k^{n-t}} \otimes \mathbf{1}_{k^t}^T & \text{if } t \leq n \\ \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T & \text{if } t \geq n. \end{cases}$$

Hence, $P^n\boldsymbol{x}(0) = \frac{1}{N}\sum_{k=1}^N \boldsymbol{x}(0)_k$.

Finally, you compute

$$(P^T)^t P^t = \begin{cases} \frac{1}{k^t}\left(\mathbf{1}_{k^t}\mathbf{1}_{k^t}^T\right) \otimes I_{k^{n-t}} & \text{if } t \leq n \\ \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^T & \text{if } t \geq n. \end{cases}$$

and conclude by taking the trace. ■

If we compare communication on a 'de Bruijn' graph to the one using a Cayley graph with bounded size of the neighborhood, we see that for any fixed N the convergence to the average is obtained in finite time $\log_k N$ for the former, and is asymptotical for $t \to \infty$ in the latter (with exponential speed, but slower as N grows). The limit for $N \to \infty$ of the mean quadratic error is $\frac{1}{k^t}$ for the 'de Bruijn' graph, as opposed to a cost that asymptotically decreases only as $\frac{1}{\sqrt{t}}$ for the circle. The different result is justified by the long-range connections allowed in the 'de Bruijn' graph, which improve performance, but are expensive or unrealistic in some communication scenarios. Comparing these two examples suggests a further investigation of the influence of graph properties on the mean quadratic error.

VI. Conclusion and open problems

We have considered linear average-consensus algorithms applied to solving a simple distributed estimation problem. We have underlined how, in this context, a very natural performance metric—mean quadratic error—depends on the whole spectrum of the consensus matrix, and not just on the essential spectral radius. A rigorous asymptotic analysis of such cost for some families of communication graphs confirms intuition: a growing number of sensors can improve performance, despite the slower convergence to average consensus.

We leave as an open problem to extend our analysis to more general families of graphs. In particular, we believe it is interesting to study (random) geometric graphs, which are known to be realistic models for many wireless communication and sensor networks (see e.g. the recent book [10]). Our results on the grid graphs can be considered as a preliminary work towards this aim, as we believe that behaviour of random geometric graphs is very related to the one of grids. In fact, the analysis in [3] of the mixing time of a random walk on a random geometric graph is done using bounds involving the essential spectral radius of a matrix associated with the grid. Moreover, in [15] there is a proof of concentration of the asymptotic distribution of the spectrum of a simple random walk on a random geometric graph towards the one on a grid. These results encourage the search for bounds on the mean quadratic error when communication follows a random geometric graph involving the cost of a corresponding grid, for which we have already established the asymptotic behaviour. The conjecture of a strong similarity of random geometric graphs to grids is also supported by simulations, as it is shown in [5].

VII. Acknowledgements

The first author is grateful to E. Lovisari for the example using the 'de Bruijn' graph.

The research leading to these results has received funding from the European Community's Seventh Framework Programme under agreement n. FP7-ICT-223866-FeedNetBack.

References

[1] B. Bamieh, F. Paganini, and M. A. Dahleh. Distributed control of spatially invariant systems. *IEEE Trans. Automatic Control*, 47(7):1091–1107, July 2002.

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

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

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

[5] 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*, pages 96–104, La Jolla, CA, USA, Feb. 2009.

[6] J. Cortes, S. Martinez, and F. Bullo. Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions. *IEEE Trans. Automatic Control*, 51(8):1289–1298, 2006.

[7] G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. *J. Parallel Distrib. Comput.*, 7(2):279–301, 1989.

[8] R. D'Andrea and G. E. Dullerud. Distributed control design for spatially interconnected systems. *IEEE Trans. Automatic Control*, 48(9):1478–1495, Sept. 2003.

[9] J.-C. Delvenne, R. Carli, and S. Zampieri. Optimal strategies in the average consensus problem. In *Proc. of the 46th IEEE Conf. on Decision and Control*, pages 2498–2503, New Orleans, LA, USA, Dec. 2007.

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

[11] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri. Average consensus on networks with quantized communication. *Intern. Journ. on Non-linear and Robust Control*, to appear, 2009.

[12] S. Ganeriwal, R. Kumar, and M. B. Srivastava. Timing-sync protocol for sensor networks. In *Proc. of the 1st international conference on Embedded networked sensor systems (SenSys '03)*, pages 138–149, 2003.

[13] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. In *Proceedings of 44th IEEE Conf. on Decision and Control and 2005 European Control Conference*, pages 8179–8184, Dec. 2005.

[14] 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(9):1520–1533, Sept. 2004.

[15] S. Rai. The spectrum of a random geometric graph is concentrated. *J. Theoret. Probab.*, 20(2):119–132, 2007.

[16] L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *J. Parallel Distrib. Comput.*, 67(1):33–46, 2007.