Automatica 49 (2013) 2496–2501

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

On the mean square error of randomized averaging algorithms☆

Paolo Frasca [a,1], Julien M. Hendrickx [b]

[a] Department of Mathematical Sciences, Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy

[b] ICTEAM Institute, Université catholique de Louvain, Avenue Georges Lemaitre 4, B-1348 Louvain-la-Neuve, Belgium

ARTICLE INFO

Article history:
Received 17 November 2011
Received in revised form
5 February 2013
Accepted 18 April 2013
Available online 28 May 2013

Keywords:
Consensus algorithm
Randomized algorithm
Networked estimation

ABSTRACT

This paper considers randomized discrete-time consensus systems that preserve the average "on average". As a main result, we provide an upper bound on the mean square deviation of the consensus value from the initial average. Then, we apply our result to systems in which few or weakly correlated interactions take place: these assumptions cover several algorithms proposed in the literature. For such systems we show that, when the network size grows, the deviation tends to zero, and that the speed of this decay is not slower than the inverse of the size. Our results are based on a new approach, which is unrelated to the convergence properties of the system.

1. Introduction

In modern control and signal processing applications, effective and easy-to-implement distributed algorithms for computing averages are an important tool. As a significant and motivational example, we consider the problem of estimating the expectation of a random variable of interest. By the law of large numbers, the sample average is an unbiased estimator, and its mean square error decreases as the inverse of the number of samples increases, provided that the random variables have finite second moment. In a distributed setting, the sample values are available at the nodes of a communication network, and the average needs to be approximated by running an iterative consensus system, which has the sample data as the initial condition. Clearly, we have to ensure that, along the iterations of the consensus system, no (or little) deviation from the correct average is introduced. However, a global property such as average preservation may be harder to satisfy when updates are performed asynchronously, unreliably, or following a random scheme. In the case of stochastic updates, a weaker requirement is the preservation of the expected average; such systems are known to converge to a consensus under mild conditions, but their consensus value is in general different from the average: it is actually a random variable whose expected value is the initial average. In this paper, we consider linear randomized asynchronous averaging algorithms, and we analyze the mean square deviation of the consensus value from the initial average. We want to ensure that this error is small, so that the averages are computed accurately. In particular, we aim to provide conditions under which the mean square error tends to zero when the number of samples, i.e., the number of nodes, grows. We will refer to this property as the *accuracy* of the algorithm.

The opportunity of using randomized algorithms to compute averages has already attracted significant interest, as testified by recent surveys and special issues (Dimakis, Kar, Moura, Rabbat, & Scaglione, 2010; Scaglione, Coates, Gastpar, Tsitsiklis, & Vetterli, 2011). Convergence theories for randomized linear averaging algorithms have been developed by several authors. A classic reference is (Cogburn, 1986), but more recently other conditions have been used in a few works, including (Fagnani & Zampieri, 2008b; Matei & Baras, 2011; Tahbaz-Salehi & Jadbabaie, 2008, 2010; Touri, 2012). As we will formally define later, a random linear averaging algorithm can be seen as a multiplication of the node-indexed state by a random update matrix. In principle, the variance of the consensus value can be exactly computed by the formula given as Equation 7 in Tahbaz-Salehi and Jadbabaie (2010), which involves the dominant eigenvectors of the first two moments of the update matrix. Unfortunately, little is known about these eigenvectors, and in particular explicit formulas are not available, so these results are difficult to apply. A few papers, on the other hand, have focused on specific examples of randomized algorithms, obtaining results which are interesting, although partial, from our perspective (Aysal, Yildiz, Sarwate, & Scaglione, 2009; Fagnani & Frasca, 2011a,b; Fagnani & Zampieri, 2008a). Typically, these results are obtained as a by-product of

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Hideaki Ishii under the direction of Editor Ian R. Petersen.

E-mail addresses: paolo.frasca@polito.it (P. Frasca), julien.hendrickx@uclouvain.be (J.M. Hendrickx).

[1] Tel.: +39 0907525; fax: +39 0907599.

0005-1098/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.automatica.2013.04.035

a convergence analysis, and they involve the eigenvalues of the update matrices, which are fairly well known for many families of communication graphs. We will come back to these results in Section 3 when discussing some example algorithms.

In this paper, we consider discrete-time consensus systems with random updates that preserve the expected average, and we provide new bounds on the mean square deviation of the current average from the initial average. We show that under certain conditions the expected increase of the deviation is bounded proportionally to the expected decrease of the disagreement. We then obtain bounds on the total deviation which are proportional to the initial disagreement and, unlike previous results, are actually independent of the convergence properties: indeed they hold at all times regardless of convergence. Compared to those already available in the literature, our bounds typically result in less conservative (and often more general) estimates of the deviation error and, remarkably, they are independent of the global properties, such as the connectivity or graph spectrum and eigensystem of the communication network. Instead, only local network properties, such as the degree, play a role in the examples. By contrast, we recall that results about convergence to consensus, and speed of convergence, depend on global network properties. Our estimates show that, under weak assumptions on the update law, the deviation tends to zero when the number of nodes grows. This is true for

(i) systems in which few updates take place simultaneously; and
(ii) systems in which the updates have small statistical dependence across the network.

Thanks to their generality and to their dependence on local network properties only, our results offer effective and easy-to-implement guidelines to the designer who needs to choose a network and an algorithm to solve an estimation problem.

Notation and preliminaries

The set of real numbers is denoted by $\mathbb{R}$, and the set of nonnegative integers by $\mathbb{Z}_{\geq 0}$. In this work, we use the notion of a *(weighted directed) graph*, which we define as a pair $G = (I, A)$, where I is a finite set whose elements are called *nodes* and $A \in \mathbb{R}^{I\times I}$ is a matrix with nonnegative entries. Resorting to more standard graph-theoretic jargon, we may equivalently think of an implicit edge set $E = \{(i, j) \in I \times I : A_{ij} > 0\}$. For simplicity, we will sometimes assume that a graph may have no loops, that is, $A_{ii} = 0$ for every $i \in I$. Given a graph, that is, a nonnegative matrix A, we can define an associated Laplacian matrix $L(A) \in \mathbb{R}^{I\times I}$ by $[L(A)]_{ij} = -A_{ij}$ if $i \neq j$ and $[L(A)]_{ii} = \sum_{j:j\neq i} A_{ij}$. Observe that $L(A)$ is positive semidefinite and that $L(A)\mathbf{1} = 0$, provided that we denote by $\mathbf{1}$ the vector of suitable size whose components are all 1. Besides, to any matrix L satisfying $L\mathbf{1} = 0$ with non-positive off-diagonal elements, one can associate a corresponding weighted graph. Finally, the conjugate transpose of the matrix A is denoted by A^*, and inequalities $A \leq B$ between two matrices A and B denote the fact that $A - B$ is negative semidefinite.

2. Problem statement and main result

Given a set of nodes I of finite cardinality N, we consider the discrete-time random process $x(\cdot)$ taking values in $\mathbb{R}^I$ and defined as

$$x_i(t+1) = \sum_{j\in I} a_{ij}(t)x_j(t) \quad \text{for all } i \in I,\ t \in \mathbb{Z}_{\geq 0}, \tag{1}$$

where, for every $i, j \in I$, we assume $\{a_{ij}(t)\}_{t\in\mathbb{Z}_{\geq 0}}$ to be a sequence of independent and identically distributed (i.i.d.) random variables such that $a_{ij}(t) \geq 0$ and $\sum_{\ell\in I} a_{i\ell}(t) = 1$ for all $t \geq 0$. System (1) is run with the goal for the state of each node to provide a good estimate of the initial average $\frac{1}{N}\sum_{i\in I} x_i(0)$. Note that $x(0)$ is unknown but given, and that all our results will be valid for any $x(0) \in \mathbb{R}^I$. System (1) can be conveniently rewritten as

$$x_i(t+1) = x_i(t) + \sum_{j\in I} a_{ij}(t)(x_j(t) - x_i(t))$$

for all $i \in I,\ t \in \mathbb{Z}_{\geq 0}$,

or in matrix form as

$$x(t+1) = x(t) - L(t)x(t) \quad t \in \mathbb{Z}_{\geq 0}, \tag{2}$$

where the matrix $L(t)$ is defined so that $L_{ij}(t) = -a_{ij}(t)$ if $i \neq j$ and $L_{ii}(t) = \sum_{j:j\neq i} a_{ij}(t)$. Namely, $L(t)$ is the Laplacian matrix of a weighted graph $(I, A(t))$, where the entries of $A(t)$ are defined as $[A(t)]_{ij} = a_{ij}(t)$. The convergence of (2) has been addressed in the literature: rather than in convergence, in this paper we are interested in the quality of the convergence value, in terms of its distance from the initial average. For our convenience, we denote the average of $x(t)$ by

$$\bar{x}(t) = \frac{1}{N}\sum_{i\in I} x_i(t),$$

and we note that the average evolves according to $\bar{x}(t+1) = \bar{x}(t) - \mathbf{1}^*L(t)x(t)$. The expected evolution of $\bar{x}(t)$, conditional on the previous state, is written as $\mathbb{E}[\bar{x}(t+1)|x(t)]$. Since under our assumptions $L(t)$ is independent of $x(t)$, we immediately deduce that $\mathbb{E}[\bar{x}(t+1)|x(t)] = \bar{x}(t)$ if and only if $\mathbf{1}^*\mathbb{E}[L(t)] = 0$. In view of this fact, we restrict our attention to systems that preserve the expected average, that is, we will assume that $\mathbf{1}^*\mathbb{E}[L(t)] = 0$, implying that

$$\mathbb{E}[\bar{x}(t)] = \bar{x}(0) \quad \text{for all } t \geq 0.$$

Consequently, we are left with the problem of studying the variance of $\bar{x}(t)$, that is, $\mathbb{E}[(\bar{x}(t) - \bar{x}(0))^2]$. We will derive all our bounds from the following general result. For $y \in \mathbb{R}^I$, we denote $\bar{y} = \frac{1}{N}\sum_i y_i$ and $V(y) = \frac{1}{N}\sum_i \big(y_i - \bar{y}\big)^2$.

Theorem 1 (*Accuracy Condition*). *Let x be an evolution of system (2). If $\mathbf{1}^*\mathbb{E}[L(t)] = 0$ and there exists $\gamma > 0$ such that*

$$\mathbb{E}[L(s)^*\mathbf{1}\mathbf{1}^*L(s)] \leq \gamma\, \mathbb{E}[L(s) + L(s)^* - L(s)^*L(s)], \tag{3}$$

then, for every $t \geq 0$, it holds that

$$\mathbb{E}[(\bar{x}(t) - \bar{x}(0))^2] \leq \frac{\gamma}{N+\gamma}V(x(0)).$$

If, moreover, the system converges to consensus ($x(t) \to x_\infty\mathbf{1}$, for $x_\infty \in \mathbb{R}$), then $\mathbb{E}\left[(x_\infty - \bar{x}(0))^2\right] \leq \frac{\gamma}{N+\gamma}V(x(0))$.

Note that $\frac{\gamma}{\gamma+N}$ is increasing with γ: it is close to $\frac{\gamma}{N}$ for small values of γ, and close to 1 for large ones. The expected square error is thus always bounded by the initial disagreement when a valid γ can be found. And when a γ can be found which is independent of N, then the algorithm is *accurate*, according to the definition stated above. The bound of Theorem 1 may of course be conservative compared with the exact characterizations of the expected square error derived in Tahbaz-Salehi and Jadbabaie (2010), but it presents the main advantage of being easy to use. Indeed, we will see in the next section that general expressions of γ can be obtained for large classes of systems.

Proof. We define $C(y) := N(\gamma + N)\bar{y}^2 + N\gamma V(y)$, for all $y \in \mathbb{R}^I$, a linear combination of the square average value and the disagreement,[2] with a ratio $\frac{\gamma}{\gamma+N}$ between the weights, with the

[2] The authors wish to thank Giacomo Como for suggesting formulating the proof in terms of this quantity $C(y)$.

intent of showing that the expectation of $C(x(t))$ is nonincreasing. We begin by developing a simpler expression for $C(y)$. Observe that $\bar{y} = \frac{1}{N}\mathbf{1}^*y$, and that $V(y) = \frac{1}{N}\sum_{i=1}^N (y_i - \bar{y})^2 = \frac{1}{N}(y - \frac{1}{N}\mathbf{1}\mathbf{1}^*y)^*(y - \frac{1}{N}\mathbf{1}\mathbf{1}^*y)$. Therefore,

$$\begin{aligned} C(y) &= y^*\left(\frac{N(N+\gamma)}{N^2}\mathbf{1}\mathbf{1}^* + \frac{N\gamma}{N} \right.\\ &\quad \left. \times \left(I - 2\frac{1}{N}\mathbf{1}\mathbf{1}^* + \frac{1}{N^2}\mathbf{1}\mathbf{1}^*\mathbf{1}\mathbf{1}^*\right)\right)y \\ &= y^*\left(\mathbf{1}\mathbf{1}^* + \gamma I\right)y. \end{aligned}$$

We now show that $\mathbb{E}(C(y - Ly)) \leq C(y)$ for any $y \in \mathbb{R}^I$, where the Laplacian L is a random variable having the same distribution as $L(t)$. We can write the difference as

$$\begin{aligned} C(y - Ly) - C(y) &= -y^*L^*(\mathbf{1}\mathbf{1}^* + \gamma I)y - y^*(\mathbf{1}\mathbf{1}^* + \gamma I)Ly \\ &\quad + y^*L^*(\mathbf{1}\mathbf{1}^* + \gamma I)Ly. \end{aligned}$$

Since it is assumed that $\mathbf{1}^*\mathbb{E}L = 0$, we then have

$$\begin{aligned} &\mathbb{E}[C(y - Ly) - C(y)] \\ &\quad = -y^*\mathbb{E}\left(-\gamma L^* - \gamma L + \gamma L^*L + L^*\mathbf{1}\mathbf{1}^*L\right)y \leq 0, \end{aligned} \tag{4}$$

where the last inequality follows from the assumption in (3). Eq. (4) implies that, if $x(t)$ follows the process (2), then $\mathbb{E}[C(x(t+1))|x(t)] \leq C(x(t))$. As a result, if $\bar{x}(0) = 0$, it holds that

$$N(\gamma + N)\mathbb{E}[(\bar{x}(t))^2] + N\gamma\,\mathbb{E}[V(x(t))] \leq N\gamma V(x(0)),$$

and thus $\mathbb{E}(\bar{x}(t))^2 \leq \frac{\gamma}{N+\gamma}V(x(0))$, since $V(x(t)) \geq 0$, which proves the result in that case. Otherwise, the result is obtained by applying the previous inequality to the translated system $x(t) - \bar{x}(0)\mathbf{1}$. □

3. Applications and examples

In this section, we consider classes of systems of type (2) for which we can apply Theorem 1, that is, for which we can find γ satisfying (3). Before presenting these example systems, we prove a general lemma which simplifies the search for γ: indeed, the proofs of our results will involve estimating $\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L)$ and $\mathbb{E}(L^*L)$ in terms of $\mathbb{E}(L + L^*)$, where we recall that an inequality between two matrices $A \leq B$ is intended as $A - B$ being negative semidefinite. Before the general lemma, we need the following preliminary result.

Lemma 2. *If the coefficients $c_1, \ldots, c_m$ are nonnegative, then it holds that*

$$\left(\sum_{i=1}^m c_i z_i\right)^2 \leq \sum_{i=1}^m c_i \sum_{i=1}^m c_i z_i^2.$$

Proof. Let $u, v \in \mathbb{R}^m$ be defined by $u_i = \sqrt{c_i}$ and $v_i = \sqrt{c_i}z_i$. Then, from the Cauchy–Schwartz inequality,

$$\begin{aligned} \left(\sum_{i=1}^m c_i z_i\right)^2 &= \left(u^*v\right)^2 \leq (||u||_2\,||v||_2)^2 \\ &= \left(\sum_{i=1}^m u_i^2\right)\left(\sum_{i=1}^m v_i^2\right) = \left(\sum_{i=1}^m c_i\right)\sum_{i=1}^m c_i z_i^2. \quad \square \end{aligned}$$

Lemma 3 (*Laplacian Bounds*)**.** *Let L be the Laplacian of a weighted directed graph with weight matrix A, define $a_{ii} := 1 - \sum_{j\neq i} a_{ij}$, and let $a^{\mathrm{d}}_{\min} > 0$ be such that $a_{ii} \geq a^{\mathrm{d}}_{\min}$ for all $i \in I$.*

(i) *If $\mathbf{1}^*L = 0$, then*

$$L^*L \leq (1 - a^{\mathrm{d}}_{\min})(L + L^*). \tag{5}$$

Now, let L be a random matrix such that the lower bound $a^{\mathrm{d}}_{\min}$ is valid almost surely.

(ii) *If $\mathbf{1}^*\mathbb{E}(L) = 0$, then*

$$\mathbb{E}(L^*L) \leq (1 - a^{\mathrm{d}}_{\min})\mathbb{E}(L + L^*). \tag{6}$$

(iii) *If $\mathbf{1}^*\mathbb{E}(L) = 0$ and there exists $\beta > 0$ such that*

$$\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L) \leq \beta\,\mathbb{E}(L + L^*),$$

then $\mathbb{E}[L^*\mathbf{1}\mathbf{1}^*L] \leq \gamma\,\mathbb{E}[L + L^* - L^*L]$ holds for $\gamma = \frac{\beta}{a^{\mathrm{d}}_{\min}}$.

Proof. Let $y \in \mathbb{R}^I$ be arbitrary but fixed. To prove claim (i), we note that $(Ly)_i = \sum_j a_{ij}(y_i - y_j)$, and therefore $y^*L^*Ly = \sum_i \left(\sum_{j:j\neq i} a_{ij}(y_i - y_j)\right)^2$. For every i, since $1 - a^{\mathrm{d}}_{\min} \geq \sum_{j:j\neq i} a_{ij}$, Lemma 2 implies that

$$(Ly)_i^2 = \left(\sum_{j:j\neq i} a_{ij}(y_j - y_i)\right)^2 \leq (1 - a^{\mathrm{d}}_{\min})\sum_{j:j\neq i} a_{ij}\left(y_j - y_i\right)^2,$$

and by summing on i that

$$y^*L^*Ly \leq (1 - a^{\mathrm{d}}_{\min})\sum_i \sum_{j:j\neq i} a_{ij}(y_i - y_j)^2. \tag{7}$$

Statement (i) then follows by noting that $\sum_i \sum_{j:j\neq i} a_{ij}(y_j - y_i)^2 = y^*(L + L^*)y$ because $\mathbf{1}^*L = 0$. We now prove statement (ii). It follows from (7) that

$$\begin{aligned} y^*\mathbb{E}(L^*L)y &= \mathbb{E}(y^*L^*Ly) \\ &\leq \mathbb{E}\left[(1 - a^{\mathrm{d}}_{\min})\sum_i \sum_{j:j\neq i} a_{ij}(y_i - y_j)^2\right] \\ &= (1 - a^{\mathrm{d}}_{\min})\sum_i \sum_{j:j\neq i} \mathbb{E}(a_{ij})(y_i - y_j)^2. \end{aligned}$$

Since $\mathbb{E}(L)$ is a (deterministic) Laplacian and $1^*\mathbb{E}(L) = 0$, we can apply the same argument leading to (5) in order to argue that

$$y^*\mathbb{E}(L^*L)y \leq (1 - a^{\mathrm{d}}_{\min})y^*\mathbb{E}(L + L^*)y,$$

which implies (6). Finally, we prove claim (iii). It follows from (6) that $-(1 - a^{\mathrm{d}}_{\min})\mathbb{E}(L + L^*) \leq -\mathbb{E}(L^*L)$. Therefore, the existence of β implies that, for $\gamma = \frac{\beta}{a^{\mathrm{d}}_{\min}}$, it holds that

$$\begin{aligned} \mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L) &\leq \beta\,\mathbb{E}(L + L^*) \\ &\leq \gamma\,\mathbb{E}(L + L^*) - \gamma\,(1 - a^{\mathrm{d}}_{\min})\mathbb{E}(L + L^*) \\ &\leq \gamma\,\mathbb{E}\left(L + L^* - L^*L\right). \quad \square \end{aligned}$$

When we apply Lemma 3 to a system of type (2), the quantity $a^{\mathrm{d}}_{\min}$ is in fact a lower bound on the "self-confidence" $a_{ii}(t)$ of the nodes. For a constant β, the bound on the mean square error is thus inversely proportional to the minimal self-confidence. This remark is consistent with the intuition that, when $a_{ii}(t)$ is very small, the information held by some nodes may be almost entirely "forgotten" in one iteration, possibly resulting in large variations of the average.

3.1. Limited simultaneous updates

In this section, we show that a scalar γ which satisfies the condition in Theorem 1 can be found when the number of simultaneous updates, or at least their contribution, is small. The next result

has the following interpretation: the mean square deviation can be bounded proportionally to the ratio between the "strength" of the interactions in the system and the "self-confidence" of each node. Note that, from now on, when studying the evolution of system (2), we will for brevity avoid writing the dependence on time of the random variables a_{ij} and L, if this causes no confusion.

Theorem 4 (*Limited Updates*)**.** *Consider system (2), and let $a^{\text{all}}_{\max}$ and $a^{\text{d}}_{\min}$ be two positive constants such that almost surely $\sum_i \sum_{j:j\neq i} a_{ij} \leq a^{\text{all}}_{\max}$ and $a_{ii} \geq a^{\text{d}}_{\min}$ for all $i \in I$. If $\mathbf{1}^*\mathbb{E}(L) = 0$, then the condition of Theorem 1 holds for*

$$\gamma = \frac{a^{\text{all}}_{\max}}{a^{\text{d}}_{\min}}.$$

Proof. It follows from Lemma 2 that

$$y^*L^*\mathbf{1}\mathbf{1}^*Ly = \left(\sum_i \sum_{j:j\neq i} a_{ij}(y_j - y_i)\right)^2 \leq a^{\text{all}}_{\max} \sum_i \sum_{j:j\neq i} a_{ij}\left(y_j - y_i\right)^2.$$

Therefore,

$$\mathbb{E}\left(y^*L^*\mathbf{1}\mathbf{1}^*Ly\right) \leq a^{\text{all}}_{\max} \sum_i \sum_{j:j\neq i} \mathbb{E}(a_{ij})\left(y_j - y_i\right)^2 = a^{\text{all}}_{\max} y^T \mathbb{E}(L + L^*)y,$$

where we have used Lemma 3(i), so that $\mathbb{E}\left(L^*\mathbf{1}\mathbf{1}^*L\right) \leq a^{\text{all}}_{\max}\mathbb{E}[L + L^*]$. The result follows from Lemma 3 (iii). □

Theorem 4 can be applied to several particular cases involving small numbers of edges or small interactions; here, we discuss two of them, drawn from the literature.

Example 1 (*Asynchronous Asymmetric Gossip Algorithm (AAGA)*)**.** Let a graph $G = (I, W)$ and $q \in (0, 1)$ be given, such that $\mathbf{1}^*W\mathbf{1} = 1$. For every $t \geq 0$, one edge (i, j) is sampled from a distribution such that the probability of selecting (i, j) is W_{ij}. Then,

$$x_i(t+1) = (1-q)\,x_i(t) + q\,x_j(t),$$

and $x_k(t+1) = x_k(t)$ for $k \neq i$.

Observe that, if $W\mathbf{1} = W^*\mathbf{1}$, then $\mathbf{1}^*\mathbb{E}[L(t)]$ holds for the AAGA, and we can apply Theorem 4 with $a^{\text{all}}_{\max} = 1 - a^{\text{d}}_{\min} = q$, since only one node is sending its state to another. This leads to $\gamma = \frac{q}{1-q}$, meaning that the expected deviation of the asymptotic value is not larger than $\frac{1}{N}\frac{q}{1-q+\frac{q}{N}}V(x(0))$. The AAGA system is also studied in Fagnani and Zampieri (2008a, Section 4): the authors prove, assuming that the components of $x(0)$ are i.i.d. random variables with variance σ^2, that the square deviation is not larger than $\frac{q-\frac{q}{N}}{1-q+\frac{q}{N}}\frac{1}{N}\sigma^2$. Taking into account that the expected value of $V(x(0))$ is $\left(1-\frac{1}{N}\right)\sigma^2$ in that case, we see that our bound allows retrieving their result.

The next example, which applies very naturally to wireless networks, has attracted significant attention (Aysal and Yildiz et al., 2009; Dimakis et al., 2010; Fagnani & Zampieri, 2008b).

Example 2 (*Broadcast Gossip Algorithm (BGA)*)**.** Let a graph $G = (I, W)$ and $q \in (0, 1)$ be given, such that $W \in \{0, 1\}^{I\times I}$. For every $t \geq 0$, one node j is sampled from a uniform distribution over I. Then, $x_i(t+1) = (1-q)\,x_i(t) + q\,x_j(t)$ if $W_{ij} > 0$, and $x_i(t+1) = x_i(t)$ otherwise. In other words, one randomly selected node broadcasts its value to all its neighbors, which update their values accordingly.

Previous results about the deviation of the BGA are dependent on the topology of the network. In Aysal and Yildiz et al. (2009, Proposition 3), it is proved that the expected square deviation is upper bounded by

$$V(x(0))\left(1 - \frac{\lambda_1}{\lambda_{N-1}}\frac{1}{1-\frac{1}{2}\frac{q}{N}\lambda_{N-1}}\right),$$

where λ_i is the ith-smallest non-zero eigenvalue of the Laplacian of the graph G. In Fagnani and Frasca (2011a, Proposition 3.3), the authors obtain the upper bound $2V(x(0))\frac{q}{1-q}\frac{d^2_{\max}}{N\lambda_1}$, where $d_{\max}$ is the maximum degree of the graph. None of these bounds suffices to show that the deviation goes to zero when N grows: for instance, $\frac{d^2_{\max}}{N\lambda_1} \geq \frac{N}{\pi^2}$ holds on a cycle graph. Accuracy is shown for cycles and some other sequences of graphs in Fagnani and Frasca (2011b), using Markov chain theory results from Fagnani and Delvenne (2010), but a general proof of accuracy is not available in the literature. Based on simulations, it was however conjectured in Fagnani and Frasca (2011a) that the mean square error of the BGA is proportional to the ratio between the degree and the number of nodes. This fact can actually be proved by applying Theorem 4, assuming that $W\mathbf{1} = W^*\mathbf{1}$. Indeed, when $W\mathbf{1} = W^*\mathbf{1}$, it holds that $\mathbb{E}[L(t)] = \frac{q}{N}L(W)$ (where we remind the reader that $L(W)$ is the Laplacian matrix corresponding to the weighted adjacency matrix W), and thus $\mathbf{1}^*\mathbb{E}[L(t)] = 0$. Observe moreover that $a^{\text{d}}_{\min} = 1 - q$ and $a^{\text{all}}_{\max} = q d^{\text{col}}_{\max}$, since one node may send its value to at most $d^{\text{col}}_{\max}$ neighbors. Theorem 4 then implies that $\frac{q}{1-q}d^{\text{col}}_{\max}$ is a valid value of γ, and a bound proportional to $\frac{d^{\text{col}}_{\max}}{N}$ then follows from Theorem 1. Finally, since every system admits a trivial $a^{\text{all}}_{\max} = N$, Theorem 4 also implies that a valid γ exists as soon as there is a $a^{\text{d}}_{\min} > 0$ for which $a_{ii} \geq a^{\text{d}}_{\min}$ holds for all i. It then follows from Theorem 1 that the expected square error is bounded by the initial disagreement in all these cases. On the other hand, the AAGA system with two nodes and $q = 1$, for which there is no such $a^{\text{d}}_{\min}$, is an example of a system for which no valid γ exists.

3.2. Uncorrelated updates

In this section we show that a small γ can still be found even if there are many simultaneous updates, provided that the correlation between the updates is sufficiently small. The next result considers three cases: (a) all update coefficients are uncorrelated, (b) nodes update their value according to any stochastic scheme, but their decisions of update are uncorrelated to that of the other nodes, (c) nodes transmit their values according to any stochastic scheme, but their decisions of transmission are uncorrelated to that of the other nodes.

Theorem 5 (*Uncorrelated Updates*)**.** *Consider system (2), and let $a^{\text{ind}}_{\max}, a^{\text{r}}_{\max}, a^{\text{c}}_{\max}, a^{\text{d}}_{\min}$ be positive constants such that $a_{ij} \leq a^{\text{ind}}_{\max}$ (with $i \neq j$), $\sum_{j:j\neq i} a_{ij} \leq a^{\text{r}}_{\max}$, $\sum_{i:i\neq j} a_{ij} \leq a^{\text{c}}_{\max}$, and $a_{ii} \geq a^{\text{d}}_{\min}$, respectively, hold almost surely. Suppose that $\mathbf{1}^*\mathbb{E}(L) = 0$. The following implications about the value of γ in Theorem 1 hold true.*

(a) *Uncorrelated coefficients: If all the a_{ij} are uncorrelated, then $\gamma = \frac{a^{\text{ind}}_{\max}}{a^{\text{d}}_{\min}}$.*

(b) *Uncorrelated updates: If a_{ij} and a_{kl} are uncorrelated when $i \neq k$, then $\gamma = \frac{a^{\text{r}}_{\max}}{a^{\text{d}}_{\min}}$.*

(c) *Uncorrelated transmissions: If a_{ij} and a_{kl} are uncorrelated when $l \neq j$, then $\gamma = \frac{a^{\text{c}}_{\max}}{a^{\text{d}}_{\min}}$.*

Note that (b) implies that any scheme (preserving the expected average), where nodes update their values independently and have a minimal self-confidence, is accurate.

Proof. We begin by proving (b), bounding $\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L)$ proportionally to $\mathbb{E}(L+L^*)$ in order to apply Lemma 3. Since $\mathbf{1}^*\mathbb{E}L = 0$, observe that $\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L) = \mathbb{E}(L^*\mathbf{1}^*L) - \mathbb{E}(L^*)\mathbf{1}^*\mathbb{E}(L)$. Besides, $\mathbf{1}^*Ly = \sum_{i,j} a_{ij}(y_j - y_i)$ holds for $y \in \mathbb{R}^l$. Therefore, we have

$$y^*\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L)y = \sum_{i,j,k,l} \left(\mathbb{E}[a_{ij}a_{kl}] - \mathbb{E}a_{ij}\mathbb{E}a_{kl}\right)(y_j - y_i)(y_l - y_k). \quad (8)$$

According to assumption (b), if $i \neq k$, then a_{ij} and a_{kl} are uncorrelated, so $\mathbb{E}(a_{ij}a_{kl}) = \mathbb{E}a_{ij}\mathbb{E}a_{kl}$. We then have

$$\begin{aligned} y^*\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L)y &= \sum_{i,j,l} \mathbb{E}[a_{ij}a_{il}](y_j - y_i)(y_l - y_i) \\ &\quad - \sum_{i,j,l} \mathbb{E}a_{ij}\mathbb{E}a_{il}(y_j - y_i)(y_l - y_i) \\ &= \mathbb{E}\left[\sum_i \left(\sum_j a_{ij}(y_j - y_i)\right)^2\right] \\ &\quad - \sum_i \left(\sum_j \mathbb{E}a_{ij}(y_j - y_i)\right)^2. \end{aligned}$$

The second term in the last expression is clearly non-positive. Applying Lemma 2 for each i in the first term then leads to

$$\begin{aligned} y^*\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L)y &\leq \sum_i \mathbb{E}\left[\left(\sum_{j\neq i} a_{ij}\right)\left(\sum_j a_{ij}\left(y_j - y_i\right)^2\right)\right] \\ &\leq a^{\mathrm{r}}_{\max}\,\mathbb{E}\left[\sum_{i,j} a_{ij}\left(y_j - y_i\right)^2\right] \\ &= a^{\mathrm{r}}_{\max}\, y^*\mathbb{E}\left(L + L^*\right)y, \end{aligned} \quad (9)$$

where we have used the definition of $a^{\mathrm{r}}_{\max}$ and Lemma 3(i). Result (b) then follows from Lemma 3(iii). Part (c) of the result is obtained in a parallel way, using $\mathbb{E}(a_{ij}a_{kl}) = \mathbb{E}a_{ij}\mathbb{E}a_{kl}$ when $j \neq l$ instead of $i \neq k$ after Eq. (8), and $a^{\mathrm{c}}_{\max}$ instead of $a^{\mathrm{r}}_{\max}$ in Eq. (9). For part (a), one has $\mathbb{E}(a_{ij}a_{kl}) = \mathbb{E}a_{ij}\mathbb{E}a_{kl}$ unless $i = k$ and $j = l$. Therefore Eq. (8) becomes

$$\begin{aligned} y^*\mathbb{E}(L^*\mathbf{1}\mathbf{1}^*L)y &= \sum_{i,j} \mathbb{E}[a_{ij}^2](y_j - y_i)^2 - \sum_{i,j} (\mathbb{E}a_{ij})^2(y_j - y_i)^2 \\ &\leq a^{\mathrm{ind}}_{\max}\,\mathbb{E}\left[\sum_{i,j} a_{ij}^2(y_j - y_i)^2\right], \end{aligned}$$

which allows us to conclude, using again Lemma 3(i) and (iii). □

The following is a natural example of uncorrelated updates.

Example 3 (*Synchronous Asymmetric Gossip Algorithm (SAGA)*)**.** Let $q \in (0, 1)$ and a graph $G = (I, W)$ be given, such that $W\mathbf{1} = \mathbf{1}$. For every $t \geq 0$, and every $i \in I$, one edge (i, j_i) is sampled from a distribution such that the probability of selecting (i, j_i) is W_{i,j_i}. Then, for every $i \in I$, $x_i(t+1) = (1-q)\,x_i(t) + q\,x_{j_i}(t)$. In other words, every node chooses one neighbor, reads its value, and updates its own value accordingly.

Previous results on the SAGA are only able to guarantee accuracy on certain sequences of graphs: in Fagnani and Zampieri (2008a, Section 5), the authors derive an upper bound on the deviation of the limit value, which for symmetric W and large N is asymptotically equivalent to $\frac{q}{1-q}\frac{1}{2N}\frac{1}{1-\mathrm{esr}(W)}V(x(0))$, where $\mathrm{esr}(W)$ is the second-largest absolute value of the eigenvalues of W. This result fails to prove accuracy for some sequences of graphs: for instance, on a cycle graph with positive W_{ij} equal to $1/2$, we have $\frac{1}{2N}\frac{1}{1-\mathrm{esr}(W)} = \frac{1}{2N}\frac{1}{\cos\left(\frac{2\pi}{N}\right)}V(x(0)) \geq \frac{q}{1-q}\frac{N}{4\pi^2}V(x(0))$. Our approach allows proving asymptotic accuracy independently of the topology of the networks, provided that W is such that $\mathbf{1}^*W = \mathbf{1}^*$. Observe indeed that $\mathbb{E}[L(t)] = qL(W)$, and thus that $\mathbf{1}^*\mathbb{E}[L(t)] = 0$. Moreover, since every node receives information from exactly one neighbor, it holds that $a^{\mathrm{r}}_{\max} = q$ and $a^{\mathrm{d}}_{\min} = 1-q$. Since the choices of neighbors are independent, we can apply Theorem 5(b) to show that $\frac{q}{1-q}$ is a valid value of γ, so that the expected square deviation is bounded by $\frac{q}{1-q+\frac{q}{N}}V(x(0))$, as in the case of the AAGA.

3.3. Simultaneous correlated updates

We have seen that accurate systems are obtained when there are few simultaneous updates or when the updates are uncorrelated. When these two conditions are not met, one can have systems whose expected square deviation is large with respect to $V(x(0))$, or does not decrease when N grows. However, one should not conclude that every system with unbounded and not strictly uncorrelated updates must not be accurate. In particular, small mean square errors can still occur for systems in which the updates follow a probability law involving some partial correlations. An example is the following algorithm, which generalizes the BGA and has been proposed in Aysal, Sarwate, and Dimakis (2009).

Example 4 (*Probabilistic Broadcast Gossip Algorithm (PBGA)*)**.** Let $q \in (0, 1)$ and $G = (I, W)$. At each time step, one node j, sampled from a uniform distribution over I, broadcasts its current value. Every node i receives the value with a probability $W_{ij} \in [0, 1]$. When node i receives the value from j, it updates its value to $x_i(t+1) = x_i(t) + q(x_j(t) - x_i(t))$. Otherwise, $x_k(t+1) = x_k(t)$.

Proposition 6 (*PBGA is Accurate*)**.** *Assume that $W = W^*$. Then, Theorem 1 holds with $\gamma = (W_{\max} + 1)\frac{q}{1-q}$, where $W_{\max} = \max_{i\in I}\sum_{j\in I} W_{ij}$.*

Proof. From Aysal and Sarwate et al. (2009, Lemma 2), we can quickly derive that, for every $t \geq 0$,

$$\mathbb{E}[L(t)] = \frac{q}{N}L(W)$$

$$\mathbb{E}[L(t)^*L(t)] = 2\frac{q^2}{N}L(W)$$

$$\mathbb{E}[L(t)^*\mathbf{1}\mathbf{1}^*L(t)] = \frac{q^2}{N}L(W)^2 + 2\frac{q^2}{N}L(W) - 2\frac{q^2}{N}L(W\cdot W),$$

where $W \cdot W$ denotes an entrywise product. The assumption on W implies that $\mathbf{1}^*\mathbb{E}[L(t)] = 0$, and in order to apply Theorem 1 we have to find γ which satisfies the inequality

$$\begin{aligned} &\frac{q^2}{N}L(W)^2 + 2\frac{q^2}{N}L(W) - 2\frac{q^2}{N}L(W\cdot W) \\ &\quad \leq \gamma\left(2\frac{q}{N}L(W) - 2\frac{q^2}{N}L(W)\right), \end{aligned}$$

that is,

$$L(W)^2 - 2L(W\cdot W) \leq 2\left(\gamma\frac{1-q}{q} - 1\right)L(W).$$

Since any Laplacian – and in particular $L(W \cdot W)$ – is positive semidefinite, a sufficient condition for the previous inequality to hold is

$$L(W)^2 \leq 2\left(\gamma\frac{1-q}{q} - 1\right)L(W).$$

Gershgorin's disk lemma implies that the spectral radius of $L(W)$ is not larger than $2W_{\max}$, and the result follows. □

4. Conclusion and perspectives

We have developed a new way of evaluating the mean square error of decentralized consensus protocols that preserve the expected average. Our results ensure that, under mild conditions, distributed averaging can be performed via asymmetric and asynchronous algorithms, with a loss in the quality of the estimate which vanishes when the number of samples (and nodes) is increased. This fact strongly supports the application of these algorithms to large networks. Our analysis complements the results about the speed of convergence, which has been thoroughly studied in the literature and was not reconsidered in this paper. Regarding design issues, we indeed note that optimizing an algorithm for accuracy may entail a slower convergence rate: for instance, it is intuitive that, in the AAGA, larger values of q imply faster convergence but poorer accuracy. Thanks to our results, the speed/accuracy trade-off can be more precisely studied in a wide range of examples.

Unlike certain previous approaches, which relied on the convergence speed of these systems, our results are based on the fact that the increase of the error can be bounded proportionally to the decrease of the disagreement. As such, they are independent of the speed at which the system converges, and therefore of the spectral properties of the network, which determines this speed. Notably, our bounds only involve local quantities such as the degree of the nodes or the weight that they give to their neighbors' values, as opposed to global ones such as the eigenvalues of the network Laplacian. As local quantities are much easier to control in distributed systems, our results are of immediate application in design.

Our method has been applied to several known protocols: although we have sometimes been very conservative when deriving our bounds, we have obtained bounds that either match or improve upon those available in the literature. In addition, results from algorithmic simulations are closely matched by our bounds, which appear to accurately capture the qualitative dependence on the network size. Note that we have limited the number of applications of our results presented here, in the interest of concision and simplicity: some additional applications can be found in Frasca and Hendrickx (in press).

Overall, two classes of systems were proved to be *accurate*: those with sufficiently few or small simultaneous updates, and those with sufficiently uncorrelated simultaneous updates. These two apparently unrelated situations in reality present strong similarities, because the updates taking place at different times are assumed to be uncorrelated. This suggests that the real parameter, which determines the mean square error, is the level of correlation between the updates taking place across the history of the system. Further work could be devoted to formalize and quantify this intuition on the importance of the correlations between the updates. Finally, we note that – to the best of our knowledge – the distribution of the final values for processes which do not preserve the expected average has not been studied yet.

Acknowledgments

The work of P. Frasca was partly supported by the Italian Ministry of University and Research under grant PRIN-20087W5P2K. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The research is also supported by the Concerted Research Action (ARC) "Large Graphs and Networks" of the French Community of Belgium.

References

Aysal, T. C., Sarwate, A. D., & Dimakis, A. G. (2009). Reaching consensus in wireless networks with probabilistic broadcast. In *Allerton conf. on communications, control and computing* (pp. 732–739). Monticello, IL. September.

Aysal, T. C., Yildiz, M. E., Sarwate, A. D., & Scaglione, A. (2009). Broadcast gossip algorithms for consensus. *IEEE Transactions on Signal Processing*, *57*(7), 2748–2761.

Cogburn, R. (1986). On products of random stochastic matrices. In *Contemporary mathematics: vol. 50. Random matrices and their applications* (pp. 199–213). American Mathematical Society.

Dimakis, A. G., Kar, S., Moura, J. M. F., Rabbat, M. G., & Scaglione, A. (2010). Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, *98*(11), 1847–1864.

Fagnani, F., & Delvenne, J.-C. (2010). Democracy in Markov chains and its preservation under local perturbations. In *IEEE conf. on decision and control* (pp. 6620–6625). Atlanta, GA. December.

Fagnani, F., & Frasca, P. (2011a). The asymptotical error of broadcast gossip averaging algorithms. In *IFAC world congress* (pp. 10027–10031). Milan, Italy. August.

Fagnani, F., & Frasca, P. (2011b). Broadcast gossip averaging: interference and unbiasedness in large Abelian Cayley networks. *IEEE Journal of Selected Topics in Signal Processing*, *5*(4), 866–875.

Fagnani, F., & Zampieri, S. (2008a). Asymmetric randomized gossip algorithms for consensus. In *IFAC world congress* (pp. 9052–9056).

Fagnani, F., & Zampieri, S. (2008b). Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas in Communications*, *26*(4), 634–649.

Frasca, P., & Hendrickx, J. M. (2013). Large network consensus is robust to packet losses and interferences. In *European control conference* (in press).

Matei, I., & Baras, J. S. (2011). Convergence results for the agreement problem on Markovian random topologies. In *IFAC world congress* (pp. 8860–8865). Milan, Italy. August.

Scaglione, A., Coates, M., Gastpar, M., Tsitsiklis, J. N., & Vetterli, M. (2011). Special issue on gossiping algorithms design and applications. *IEEE Journal of Selected Topics in Signal Processing*, *5*(4), 645–648.

Tahbaz-Salehi, A., & Jadbabaie, A. (2008). A necessary and sufficient condition for consensus over random networks. *IEEE Transactions on Automatic Control*, *53*(3), 791–795.

Tahbaz-Salehi, A., & Jadbabaie, A. (2010). Consensus over ergodic stationary graph processes. *IEEE Transactions on Automatic Control*, *55*(1), 225–230.

Touri, B. (2012). *Springer Theses, Product of random stochastic matrices and distributed averaging*. Springer.

Paolo Frasca received his Ph.D. degree from Politecnico di Torino, Italy, in 2009. Since 2010, he has been with the Department of Mathematical Sciences (DISMA) of the Politecnico di Torino as a research fellow. He has held research and visiting positions at the IAC-CNR, Rome, and at the University of California, Santa Barbara, USA. Starting June 1st, 2013, he will be an assistant professor at the University of Twente, Netherlands. His research includes systems theory and applications of probability and graph theory to distributed and networked systems: specific interests are the modeling and control of robotic, sensor, and social networks. On these topics, Dr. Frasca has (co)authored more than 30 papers in journals, books, and conference proceedings, and has given several invited talks at international institutions, both in Europe and overseas.

Julien M. Hendrickx received an engineering degree in applied mathematics and a Ph.D. in mathematical engineering from the Université catholique de Louvain, Belgium, in 2004 and 2008, respectively.

He was a visiting researcher at the University of Illinois at Urbana Champaign in 2003–2004, at the National ICT Australia in 2005 and 2006, and at the Massachusetts Institute of Technology in 2006 and 2008. He was a postdoctoral fellow at the Laboratory for Information and Decision Systems of the Massachusetts Institute of Technology in 2009 and 2010, holding postdoctoral fellowships of the F.R.S.-FNRS (Fund for Scientific Research) and of Belgian American Education Foundation. Since September 2010, he has been an assistant professor (chargé de cours) at the Université catholique de Louvain, in the Ecole Polytechnique de Louvain.

Doctor Hendrickx is the recipient of the 2008 EECI award for the best Ph.D. Thesis in Europe in the field of Embedded and Networked Control, and of the Alcatel-Lucent-Bell 2009 award for a Ph.D. thesis on original new concepts or applications in the domain of information or communication technologies.