Brief paper

Distributed averaging on digital erasure networks☆

Ruggero Carli [a], Giacomo Como [b], Paolo Frasca [c,*], Federica Garin [d]

[a] Department of Information Engineering, University of Padova, Padova, Italy
[b] Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[c] Dipartimento di Matematica, Politecnico di Torino, Torino, Italy
[d] INRIA Rhône-Alpes, Grenoble, France

ARTICLE INFO

ABSTRACT

Iterative distributed algorithms are studied for computing arithmetic averages over networks of agents connected through memoryless broadcast erasure channels. These algorithms do not require the agents to have any knowledge about the global network structure or size. Almost sure convergence to state agreement is proved, and the communication and computational complexities of the algorithms are analyzed. Both the number of transmissions and the number of computations performed by each agent of the network are shown to grow not faster than poly-logarithmically in the desired precision. The impact of the graph topology on the algorithms' performance is analyzed as well. Moreover, it is shown how, in the presence of noiseless communication feedback, one can modify the algorithms, significantly improving their performance versus complexity trade-off.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Many scenarios of current applicative interest can be modeled as large networks of identical anonymous agents, which have access to some partial information, and aim at computing an application-specific function of the global information. The main requirements are that the network be reconfigurable and scalable, and the computation be completely distributed, i.e., each agent can only communicate with a restricted group of neighbors while processing the available information. A special instance, which has been the object of recent extensive work, is the average consensus problem, in which a large number of agents aim at computing the arithmetic average of some initial scalar measurements. While most of the literature on consensus algorithms has modeled

communication constraints in the average consensus algorithm by a communication graph in which a link between two nodes is assumed to support the noise-free transmission of a real value, there is a clear demand for more realistic communication models. In fact, some recent work has addressed the cases of quantized communication (Aysal, Coates, & Rabbat, 2008; Carli, Bullo, & Zampieri, 2010; Frasca, Carli, Fagnani, & Zampieri, 2009; Kar & Moura, 2010; Li, Fu, Xie, & Zhang, in press), or transmission affected by additive noise (Huang & Manton, 2009; Kar & Moura, 2009; Rajagopal & Wainwright, 2008). However, to the best of our knowledge, there has been no contribution yet toward the design of consensus algorithms on networks in which the communication links are modeled as digital noisy channels. The latter models of communication are particularly significant as, in practice, bandwidth limitations imply that the channels have finite capacity. For such digital noisy networks, information-theoretic bounds on the performance of distributed computation algorithms have been established in Ayaso, Shah, and Dahleh (in press) and Como and Dahleh (2009). Related problems of distributed computation have been considered, for instance, in Giridhar and Kumar (2006), Hendrickx, Olshevsky, and Tsitsiklis (2009) and Ying, Srikant, and Dullerud (2007).

In this paper, we study iterative distributed averaging algorithms for networks whose nodes can communicate through memoryless erasure broadcast channels. In order to compare the performance of different algorithms, we define suitable

* Corresponding author. Tel.: +39 0115647525; fax: +39 0115647599.
E-mail addresses: carlirug@engineering.ucsb.edu (R. Carli), giacomo@mit.edu (G. Como), paolo.frasca@polito.it, paolo.frasca@gmail.com (P. Frasca), federica.garin@inrialpes.fr (F. Garin).

complexity measures, which account for the number of channel transmissions (*communication complexity*), and, respectively, of in-node computations (*computational complexity*) required to achieve a desired precision. These performance measures are particularly relevant, as they allow for directly estimating the energy consumption of such distributed computation systems, as well as their time complexity. Related measures to evaluate distributed algorithms have been proposed in various settings; see, for instance, (Denantes, Benezit, Thiran, & Vetterli, 2008; Kushilevitz & Nisan, 1997; Martínez, Bullo, Cortés, & Frazzoli, 2007). The algorithms proposed in this paper combine the classical iterative linear consensus algorithm with coding schemes for the reliable transmission of real numbers on noisy channels, recently proposed in Como, Fagnani, and Zampieri (2010). They involve a sequence of transmission phases, of increasing duration, in which the agents attempt to broadcast their state, i.e., their current estimate of the global average, to their neighbors, alternated to averaging steps, in which the agents' states are updated. These algorithms are fully distributed, and they do not require the agents to have any global knowledge of the network structure or size. Our main result – stated as Theorem 4 – shows that such algorithms drive the agents to state agreement – or consensus – which can be made arbitrarily close to the true average. The number of channel transmissions and the number of in-node computations are shown to grow at most poly-logarithmically in the desired precision. We also show how communication feedback, when available, allows one to modify the algorithms, achieving asymptotic average consensus (i.e., state agreement on the average of the initial observations), and reducing the computational and communication complexities—see Theorem 5.

The remainder of this paper is organized as follows. In Section 2, we formally state the problem and introduce the relevant performance measures. In Section 3, we revise some results on the transmission of continuous information through digital noisy channels. In Section 4, we present our algorithms and present the main convergence and complexity results. In Section 5, we discuss how to efficiently modify our algorithms in the presence of communication feedback. Section 6 contains some concluding remarks, and all proofs are collected in the Appendix.

Before proceeding, let us establish some notation to be used throughout the paper. We denote by $\mathbb{N}$, $\mathbb{Z}^+$, and $\mathbb{R}$, respectively, the sets of naturals, nonnegative integers, and real numbers. The set of the smallest t naturals is denoted by $[t] := \{1, 2, \ldots, t\}$. The transposes of a vector $\mathbf{v} \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{n \times n}$, are denoted by $\mathbf{v}^*$ and M^*, respectively. Given two matrices M, M', we denote by $M \odot M'$ their entrywise product. With the symbol $\mathbf{1}$ we denote the n-dimensional vector all of whose entries equal 1. A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the pair of a finite vertex set $\mathcal{V}$ and of a set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ of directed edges. For a vertex $v \in \mathcal{V}$, we denote by $\mathcal{N}_v^+ := \{w \in \mathcal{V} : (v, w) \in \mathcal{E}\}$, and $\mathcal{N}_v^- := \{w \in \mathcal{V} : (w, v) \in \mathcal{E}\}$, respectively, the sets of its out-neighbors and in-neighbors. Given a matrix $M \in \mathbb{R}^{n \times n}$, we define the induced graph $\mathcal{G}_M$ by taking $\mathcal{V} = \{1, \ldots, n\}$ and putting an edge (j, i) in $\mathcal{E}$ if $i \neq j$ and $M_{ij} > 0$; M is adapted to a graph $\mathcal{G}$ if $\mathcal{G}_M$ is a subgraph of $\mathcal{G}$.

2. Problem setting

In this section, we present a formal statement of the problem, and introduce the main performance measures. We consider a finite set of agents $\mathcal{V}$ of cardinality n, and assume that each agent $v \in \mathcal{V}$ has access to some partial information consisting in the observation of a scalar value θ_v. The full vector of observations is denoted by $\boldsymbol{\theta} = (\theta_v)_{v \in \mathcal{V}}$. We consider the case when all θ_v's take values in the same bounded interval $\Theta \subseteq \mathbb{R}$. Such an interval may represent the common measurement range of the agents, possibly dictated by technological constraints, and assumed to be known a priori to all the agents. For ease of exposition, we shall assume

throughout that Θ coincides with the unitary interval [0, 1].[1] For the network, the goal is to compute the average of such values,

$$y := f(\boldsymbol{\theta}) = n^{-1} \sum_{v \in \mathcal{V}} \theta_v,$$

through repeated exchanges of information among the agents and without a centralized computing system. Communication among the agents takes place as follows. At each time instant $t = 1, 2, \ldots$, every agent v broadcasts a binary signal $a_v(t) \in \{0, 1\}$ to its out-neighborhood $\mathcal{N}_v^+$. Every agent $w \in \mathcal{N}_v^+$ receives a possibly erased version $b_{v \to w}(t) \in \{0, 1, ?\}$ of $a_v(t)$. Here, the symbol ? represents a lost binary signal. We denote by $b_v(t) = (b_{w \to v}(t))_{w \in \mathcal{N}_v^-}$ and $b_v'(t) = (b_{v \to w}(t))_{w \in \mathcal{N}_v^+}$ the vector of signals received by agent v at time t and, respectively, the vector of signals received from agent v by its out-neighbors. At time t, each agent $v \in \mathcal{V}$ makes an estimate $\hat{y}_v(t)$ of y. The compact notation $\mathbf{a}(t) = (a_v(t))_{v \in \mathcal{V}}$, $\mathbf{b}(t) = (b_v(t))_{v \in \mathcal{V}}$, and $\hat{\mathbf{y}}(t) = (\hat{y}_v(t))_{v \in \mathcal{V}}$ is used for the full vectors of transmitted signals, received signals, and estimates at time t, respectively.

We assume the communication network to be memoryless, i.e., that $\mathbf{b}(t)$ is conditionally independent from the initial observations $\boldsymbol{\theta}$ and the previous transmissions $\{\mathbf{a}(s), \mathbf{b}(s) : 1 \leq s < t\}$, given the currently broadcasted signals $\mathbf{a}(t)$. Further, we assume that, given $\mathbf{a}(t)$, for every $v \in \mathcal{V}$ and $w \in \mathcal{N}_v^+$,

$$b_{v \to w}(t) = \begin{cases} ? & \text{w.p. } \varepsilon \\ a_v(t) & \text{w.p. } 1 - \varepsilon. \end{cases}$$

Here ε is some erasure probability which, for simplicity, is assumed to remain constant in t, v and w.[2] Distributedness of the computation algorithm is then modeled by constraining the transmitted signal $a_v(t)$ to be a function of the local information available to agent v at the end of the $(t-1)$-th round of communication, and the estimate $\hat{y}_v(t)$ to be a function of the information available to agent v at the end of the t-th round of communication. We consider two different local information structures, corresponding to the cases when there is no communication feedback, and when there is causal communication feedback, respectively. When there is no communication feedback, the local information available to agent v at the end of the t-th round of communication consists of its initial observation, as well as of the signals received by v up to time t:

$$i_v(t) := \{\theta_v, b_v(s) : 1 \leq s \leq t\}.$$

On the other hand, when there is causal communication feedback, the local information available to agent v at the end of the t-th round of communication also includes all the signals received so far from v by its out-neighbors:

$$i_v'(t) := \{\theta_v, b_v(s), b_v'(s) : 1 \leq s \leq t\}.$$

The assumption of noiseless communication feedback may be reasonable, for example, to describe a simple situation of variable-rate quantized transmission, where each agent is allowed to broadcast one bit noiselessly to its neighbors with probability $1 - \varepsilon$, and cannot broadcast any bit with probability ε. Observe that the case $\varepsilon = 0$ reduces to one-bit-quantized transmission, which has been already considered in the literature; see (Frasca et al., 2009; Li et al., in press).

The communication setting outlined above can be conveniently described by a directed graph $\mathcal{G}_\varepsilon = (\mathcal{V}, \mathcal{E})$, whose vertices are the agents, and such that an ordered pair (v, w) with $v \neq w$ belongs

---

[1] This causes no loss of generality, as the case of general bounded interval Θ can be easily reduced to the unitary one by means of an affine transformation, with the error estimates continuing to hold modulo a rescaling of the constants by the length of Θ.

[2] It is not necessary, for the validity of our results, to assume mutual independence of the received signals $\{b_{v \to w}(t)\}_w$ given $a_v(t)$. On the other hand, the assumption that the channel is memoryless remains crucial.

to $\mathcal{E}$ if and only if $w \in \mathcal{N}_v^+$ (or, equivalently, if $v \in \mathcal{N}_w^-$), i.e., if v transmits to w with erasure probability $\varepsilon < 1$. Throughout the paper, we assume that the graph $\mathcal{G}_\varepsilon$ is strongly connected, i.e., that there exists a directed path connecting any pair of its vertices. A distributed computation algorithm on the communication graph $\mathcal{G}_\varepsilon = (\mathcal{V}, \mathcal{E})$ is specified by a pair $\mathcal{A} = (\Phi, \Psi)$ of double-indexed families of maps $\Phi = \{\phi_v^{(t)} : v \in \mathcal{V}, t \in \mathbb{N}\}$, and $\Psi = \{\psi_v^{(t)} : v \in \mathcal{V}, t \in \mathbb{N}\}$. When there is no communication feedback, one has

$$\phi_v^{(t)} : \Theta \times \{0, 1, ?\}^{\mathcal{N}_v^- \times [t-1]} \to \{0, 1\},$$

$$\psi_v^{(t)} : \Theta \times \{0, 1, ?\}^{\mathcal{N}_v^- \times [t]} \to \Theta,$$

and $a_v(t) = \phi_v^{(t)}(i_v(t-1))$, $\hat{y}_v(t) = \psi_v^{(t)}(i_v(t))$. On the other hand, in the case when causal communication feedback is available, one has

$$\phi_v^{(t)} : \Theta \times \{0, 1, ?\}^{(\mathcal{N}_v^- \cup \mathcal{N}_v^+) \times [t-1]} \to \{0, 1\},$$

$$\psi_v^{(t)} : \Theta \times \{0, 1, ?\}^{(\mathcal{N}_v^- \cup \mathcal{N}_v^+) \times [t]} \to \Theta,$$

and $a_v(t) = \phi_v^{(t)}(i_v'(t-1))$, $\hat{y}_v(t) = \psi_v^{(t)}(i_v'(t))$.

In the rest of this paper, we shall propose and study some distributed computation algorithms that can be framed in the above general setting. In order to analyze their performance, we will study the distance of the estimates $\hat{y}_v(t)$ from the average of the initial values y:

$$\mathbf{e}(t) = \hat{\mathbf{y}}(t) - y\mathbf{1}.$$

Namely, we define two complexity figures, the *communication complexity* and the *computational complexity*. The communication complexity of a distributed algorithm $\mathcal{A}$ on a graph $\mathcal{G}_\varepsilon$ is measured in terms of the function

$$\tau(\delta) := \inf\{t \in \mathbb{N} : n^{-1}\mathbb{E}\left[\|\mathbf{e}(s)\|^2\right] \le \delta, \ \forall s \ge t\},$$

where $\delta \in]0, 1]$. In other words, $\tau(\delta)$ denotes the minimum number of binary transmissions each agent has to perform in order to guarantee that the average mean squared estimation error does not exceed δ. Instead, the computational complexity of an algorithm $\mathcal{A}$ on a graph $\mathcal{G}_\varepsilon$ is measured as follows. For every $t \in \mathbb{N}$, and $v \in \mathcal{V}$, we denote by $\kappa_v(t)$ the minimum number of binary operations required by agent v to evaluate the functions $\phi_v^{(t)}(\cdot)$ and $\psi_v^{(t)}(\cdot)$. Then, we define

$$\kappa(\delta) := \max\left\{\sum_{t=1}^{\tau(\delta)} \kappa_v(t) : v \in \mathcal{V}\right\}, \quad \delta \in]0, 1].$$

Hence, $\kappa(\delta)$ denotes the maximum, over all agents $v \in \mathcal{V}$, of the total number of binary operations required to be performed, in order to achieve an average mean squared estimation error not exceeding δ.

3. Reliable transmission of continuous information through digital noisy channels

When the communication graph is complete, with all the agents connected through binary erasure broadcast channels, the problem reduces to that of reliable transmission of continuous information through digital noisy channels, which has been recently addressed in Como et al. (2010). While referring to Como et al. (2010) for general information-theoretical limits and complexity versus performance trade-offs, we give here some results which will be used in what follows.

Let θ be a random variable taking values in the unitary interval $\Theta = [0, 1]$, according to some a priori probability law. Consider a memoryless binary erasure channel with erasure probability $\varepsilon \in (0, 1)$. At each time $t \in \mathbb{N}$, the channel has input $a_t \in \{0, 1\}$ and output $b_t \in \{0, 1, ?\}$, with b_t conditionally independent

from x, $\{a_s, b_s : 1 \le s \le t-1\}$, given a_t, and such that $b_t = a_t$ with probability $1 - \varepsilon$, and $b_t = ?$ with probability ε. The goal is to design a sequence of encoders $\Upsilon = (\Upsilon_t : \Theta \to \{0, 1\})_{t \in \mathbb{N}}$, and of decoders $\Lambda = (\Lambda_t : \{0, 1, ?\}^t \to \Theta)_{t \in \mathbb{N}}$, such that, if $a_t = \Upsilon_t(x)$, b_t is the corresponding channel output, and $\hat{\theta}_t := \Lambda_t(b_1, \ldots, b_t)$ the current estimate, the mean squared error $\mathbb{E}[(\theta - \hat{\theta}_t)^2]$ is minimized. The computational complexity of the sequential coding scheme (Υ, Λ) is measured, for every time horizon $\ell \in \mathbb{N}$, in terms of the total number k_ℓ of binary operations required to compute $\Upsilon_t(x)$ and $\Lambda_t(b_1, \ldots, b_t)$ for all $1 \le t \le \ell$.

Here, in particular, we consider two specific classes of sequential transmission schemes described and analyzed in Como et al. (2010). The first class is that of random linear tree codes, referred to by the superscript L. These codes have exponential convergence rates with respect to the number of channel uses, and computational complexity proportional to the cube of the number of channel uses. The second class is that of irregular repetition codes (superscript R). Such codes have linear computational complexity, but subexponential converge rates. The performance of these two classes of codes is summarized in the following lemmas.

Lemma 1 (*Como et al., 2010, Corollary 6.2*). *There exist a sequence of linear encoders Υ^L, and a sequence of decoders Λ^L, such that, if $\hat{\theta}_\ell = \Lambda_\ell^L(b_1, \ldots, b_\ell)$, then, for all $\ell \ge 0$,*

$$\mathbb{E}\left[(\theta - \hat{\theta}_\ell)^2\right] \le \beta_L^{2\ell}, \qquad k_\ell^L \le B\ell^3, \tag{1}$$

where $\beta_L \in (0, 1)$, and $B > 0$ are constants depending only on the erasure probability ε.

Lemma 2 (*Como et al., 2010, Proposition 5.1*). *There exist a sequence of linear encoders Υ^R, and a sequence of decoders Λ^R, such that, if $\hat{\theta}_\ell = \Lambda_\ell^R(b_1, \ldots, b_\ell)$, then, for all $\ell \ge 0$,*

$$\mathbb{E}\left[(\theta - \hat{\theta}_\ell)^2\right] \le \beta_R^{2\sqrt{\ell}}, \qquad k_\ell^R \le 2\ell, \tag{2}$$

where $\beta_R \in (0, 1)$ is a constant depending only on the erasure probability ε.

4. Distributed averaging without communication feedback

In this section, we present two iterative distributed averaging algorithms, working on a strongly connected graph $\mathcal{G}_\varepsilon$, without explicit communication feedback. Both algorithms are based on a sequence of transmission phases, indexed by $j \ge 1$, alternated to averaging steps. Each agent $v \in \mathcal{V}$ maintains a scalar state $x_v(j)$, $j \ge 0$, which is initialized to the original observation θ_v. The state $x_v(j)$ has to be thought as v's estimate of y at the beginning of the $(j+1)$-th phase. During the j-th transmission phase, each agent broadcasts ℓ_j binary signals to its out-neighbors. These binary signals represent an encoding of the state $x_v(j-1)$. At the end of the j-th phase, each agent estimates each of its in-neighbors' states from the signals received from it, and it updates its state to a convex combination of these estimates and its own current state. The process is then iterated.

4.1. Algorithms

We now provide a formal description of the algorithms. Let P be a doubly stochastic, irreducible matrix adapted to $\mathcal{G}_\varepsilon$, with non-zero diagonal entries. Let $(\ell_j)_{j \in \mathbb{N}}$ be a sequence of positive integers, each ℓ_j representing the length of the j-th transmission phase, and define $h_j := \sum_{i \le j} \ell_i$, for all $j \in \mathbb{N}$ and $h_0 = 0$. Further, let Υ and Λ be sequences of encoders and decoders as introduced in Section 2. Then, the proposed distributed algorithms consist of the following steps. First of all, each agent $v \in \mathcal{V}$ initializes its state, setting

$x_v(0) = \theta_v$. Then, for all $j \in \mathbb{N}$ and $v \in \mathcal{V}$:

Communication phase: v broadcasts an encoded version of its state $x_v(j-1)$ to its out-neighbors; namely, for all $h_{j-1} < t \le h_j$, it transmits the binary signal

$$a_t = \Upsilon_k(x_v(j-1)), \quad k = t - h_{j-1}. \tag{3}$$

State update: at the end of the j-th communication phase, v estimates the state of all its in-neighbors, based on the received signals $\{\mathbf{b}_v(t)\}_{t=h_{j-1}+1}^{h_j}$; for each $w \in \mathcal{N}_v^-$, let $\hat{x}_w^{(v)}(j-1)$ be the estimate of $x_w(j-1)$ built by agent v; then

$$\hat{x}_w^{(v)}(j-1) = \Lambda_{\ell_j}\left(b_{w \to v}(h_{j-1}+1), \ldots, b_{w \to v}(h_j)\right). \tag{4}$$

Then, v updates its own state according to the following *consensus-like step*:

$$x_v(j) = \sum_{w \in \mathcal{N}_v^-} P_{vw}\hat{x}_w^{(v)}(j-1) + P_{vv}x_v(j-1). \tag{5}$$

Observe that the above-described algorithms can be framed in the general setting described in Section 2. Indeed, for all $j \ge 1$, one has

$$\phi_{h_{j-1}+k}^{(v)}(i_v(h_{j-1}+k)) = \Upsilon_i(x_v(j-1)) \quad 0 < k \le \ell_j,$$

$$\psi_{h_{j-1}+k}^{(v)}(i_v(h_{j-1}+k)) = x_v(j-1) \quad 0 \le k < \ell_j.$$

Notice that state $x_v(j-1)$ represents the estimate that agent v has of y along all j-th phase, i.e.,

$$\hat{y}_v(t) = x_v(j-1), \quad \forall h_{j-1} \le t < h_j. \tag{6}$$

In what follows, we consider two implementations of the algorithm. In the first implementation, referred to as algorithm $\mathcal{A}_L$, we use linear tree codes $\Upsilon = \Upsilon^L$, $\Lambda = \Lambda^L$, and phase lengths $\ell_j^L = S_L j$ for some $S_L \in \mathbb{N}$. In the second implementation, referred to as algorithm $\mathcal{A}_R$, we use repetition codes $\Upsilon = \Upsilon^R$, $\Lambda = \Lambda^R$, and phase lengths $\ell_j^R = S_R j^2$, for some $S_R \in \mathbb{N}$. Observe that, thanks to (1), one has, for algorithm $\mathcal{A}_L$,

$$\mathbb{E}\left[\left(\hat{x}_w^{(v)}(j-1) - x_w(j-1)\right)^2\right] \le \alpha_L^{2j}, \tag{7}$$

for every $j \in \mathbb{N}$, $v \in \mathcal{V}$, and $w \in \mathcal{N}_v^-$, where $\alpha_L := \beta_L^{S_L}$. Similarly, for algorithm $\mathcal{A}_R$, Eq. (2) guarantees that

$$\mathbb{E}\left[\left(\hat{x}_w^{(v)}(j-1) - x_w(j-1)\right)^2\right] \le \alpha_R^{2j}, \tag{8}$$

for every $j \in \mathbb{N}$, $v \in \mathcal{V}$, and $w \in \mathcal{N}_v^-$, where $\alpha_R := \beta_R^{\sqrt{S_R}}$.

It should be mentioned that other choices could have been made for the communication phase lengths, as well as for the coding schemes used during each of them. For instance, block codes of different lengths could have been used during each phase. Our choice of using the same anytime transmission scheme for every agent during each communication phase has the advantage of fewer memory requirements (only one transmission scheme has to be memorized by each agent), anonymity (each agent uses the same transmission scheme, and the state updating rules only depend on its position in the graph), and adaptiveness with respect to the erasure probability ε. In fact, it is not required to know the actual value of ε in order to design Υ and Λ; see Remarks 3 and 5 in Como et al. (2010).

4.2. Performance analysis

We now present results characterizing the performance of the algorithms $\mathcal{A}_L$ and $\mathcal{A}_R$ introduced in Section 4.1. Throughout, we assume that $\mathcal{G}_\varepsilon$ is a strongly connected graph, and that P is

a doubly stochastic, irreducible matrix which is adapted to $\mathcal{G}_\varepsilon$, and has positive diagonal entries. Notice that this implies that P^*P is doubly stochastic and irreducible. It then follows from the Perron–Frobenius theorem that P^*P has the eigenvalue 1 with multiplicity 1 and corresponding eigenvector $\mathbf{1}$, and all its other eigenvalues have modulus strictly smaller than 1. Hence, P has largest singular value equal to 1 and all other singular values strictly smaller than 1. We denote by $\rho := \rho(P) < 1$ the second largest singular value of P, and assume that $\rho \ge \underline{\rho}$, where $\underline{\rho} > 0$ is some a priori constant.[3]

Observe that the vector of the estimation errors on y made by the different agents, $\mathbf{e}(t) = \hat{\mathbf{y}}(t) - y\mathbf{1}$, is constant during each transmission phase, i.e.,

$$\mathbf{e}(t) = \mathbf{e}(h_j), \quad \forall h_j \le t < h_{j+1} \tag{9}$$

for any $j \ge 0$. To analyze the performance of our algorithms, it is useful to introduce a suitable decomposition of $\mathbf{e}$; for all $j \ge 0$, we can write that

$$\mathbf{e}(h_j) = \mathbf{z}(j) + \zeta(j)\mathbf{1},$$

where

$$\mathbf{z}(j) = \mathbf{x}(j) - (n^{-1}\mathbf{1}^*\mathbf{x}(j))\mathbf{1} \tag{10}$$

represents the difference between the current estimates and the average of the current states, whereas

$$\zeta(j) = n^{-1}\mathbf{1}^*\mathbf{x}(j) - y = n^{-1}\mathbf{1}^*(\mathbf{x}(j) - \mathbf{x}(0)) \tag{11}$$

accounts for the distance between the current average of the estimates and the average of the initial conditions. Now, observe that the state dynamics (5) may be rewritten in the following compact form:

$$\mathbf{x}(j+1) = P\mathbf{x}(j) + (P \odot \Delta(j+1))\mathbf{1}, \tag{12}$$

where $\mathbf{x}(0) = \boldsymbol{\theta}$, and where $\Delta(j) = (\Delta_{vw}(j))_{v,w \in \mathcal{V}}$ is defined, for all $j \in \mathbb{N}$, by

$$\Delta_{vw}(j) := \begin{cases} \hat{x}_w^{(v)}(j-1) - x_w(j-1) & \text{if } w \in \mathcal{N}_v^- \\ 0 & \text{if } w \notin \mathcal{N}_v^-. \end{cases}$$

Notice that, in general, $\Delta_{vw}(j)$ has non-zero mean, and it is not independent from $x_w(j)$, and therefore from the errors introduced by the previous transmission phases $\{\Delta(i) : 1 \le i < j\}$. We have the following result.

Proposition 3. *Consider the stochastic system (12), driven by a noise process $\{\Delta(j) : j \ge 1\}$ satisfying*

$$\mathbb{E}[\Delta_{vw}(j)^2] \le \alpha^{2j}, \quad j \ge 1,$$

for some $0 < \alpha < \rho$. Then, for all $j \ge 0$,

$$\mathbb{E}[\zeta^2(j)] \le \alpha^2(1-\alpha)^{-2}, \tag{13}$$

$$n^{-1}\mathbb{E}[\|\mathbf{z}(j)\|^2] \le \rho^{2j}(1-\alpha/\rho)^{-2}. \tag{14}$$

Proof. See the Appendix. $\square$

The following result characterizes the performance of both algorithms $\mathcal{A}_L$ and $\mathcal{A}_R$.

Theorem 4 (*No Communication Feedback*)**.** *For any choice of the initial phase length S_L (respectively, S_R), there exists a real-valued random variable $\hat{y}$ such that*

$$\mathbb{E}\left[(y-\hat{y})^2\right] \le \alpha^2(1-\alpha)^{-2}, \tag{15}$$

---

[3] This may be enforced without using global information, by assuming that $P_{vv} \ge (1+\underline{\rho})/2$. Note that this assumption is for purposes of analysis only, and the agents do not need to know ρ to run the algorithms. The assumption entails a minimal loss of generality in that it rules out the case $\rho = 0$: related results which cover this case can be found in Carli, Como, Frasca, and Garin (2009).

where $\alpha = \beta_L^{S_L}$ (respectively, $\alpha = \beta_R^{\sqrt{S_R}}$) and such that the estimates of algorithm $\mathcal{A}_L$ (respectively, $\mathcal{A}_R$) satisfy, with probability 1,

$$\lim_{t \to \infty} \hat{y}_v(t) = \hat{y}, \quad \forall v \in \mathcal{V}. \tag{16}$$

Moreover, it is possible to choose the initial phase length S_L (respectively, S_R) in such a way that the algorithm $\mathcal{A}_L$ (respectively, $\mathcal{A}_R$) has communication and computational complexities satisfying

$$\tau_L(\delta) \leq C_1 + C_2 \frac{\log^3 \delta^{-1}}{\log^2 \rho^{-1}}, \qquad \kappa_L(\delta) \leq C_3 + C_4 \frac{\log^7 \delta^{-1}}{\log^4 \rho^{-1}},$$

and, respectively,

$$\tau_R(\delta) \leq C_5 + C_6 \frac{\log^5 \delta^{-1}}{\log^3 \rho^{-1}}, \qquad \kappa_R(\delta) \leq C_7 + C_8 \frac{\log^5 \delta^{-1}}{\log^3 \rho^{-1}},$$

for all $\delta \in]0, 1]$, where $\{C_i : i = 1, \ldots, 8\}$ are positive constants depending only on ε.

Proof. See the Appendix. $\square$

Observe that, by (15), the mean squared distance between the asymptotic estimate $\hat{y}$ and the actual value y is upper bounded by a constant which, quite remarkably, is independent of either the size of the network or the consensus matrix P, and depends only on the length of the first transmission phase. Moreover, Theorem 4 shows that both the algorithms $\mathcal{A}_L$ and $\mathcal{A}_R$ have communication and computational complexities growing at most poly-logarithmically in the desired precision. The bounds on the communication (respectively, computation) complexities suggest that for the agents it may be sufficient to use fewer channel transmissions in order to achieve a desired precision when running algorithm $\mathcal{A}_L$ than when running $\mathcal{A}_R$, and that the opposite happens if the number of computations is considered. This behavior has been confirmed in a number of simulations we have run implementing the algorithms, an example of which is reported in Carli et al. (2009). Furthermore, in Theorem 4, both complexities depend on ρ, the second largest singular value of the matrix P. As the matrix P is adapted to the communication graph $\mathcal{G}_\varepsilon$, the dependence of the bounds on ρ captures the effect of the network topology.

5. Distributed averaging with communication feedback

In this section, we discuss how to efficiently modify the algorithms of Section 4 when there is communication feedback. The key point is that, in the presence of noiseless communication feedback, it is possible to modify the algorithms $\mathcal{A}_L$ and $\mathcal{A}_R$ and make them average preserving. These modified algorithms will be shown to converge to average consensus with probability 1, and to have lower communication and computational complexities than their feedbackless counterparts.

We consider distributed averaging algorithms with the iterative structure described in Section 4. Specifically, we use the same communication phase rule (3) of Section 4.1, and modify the state update step as follows. Observe that, at the end of the j-th communication phase, not only can agent v estimate the state of all its in-neighbors as in (4), but it can also use its knowledge of the signals $\{b_{v \to w}(t)\}_{t=h_{j-1}+1}^{h_j}$ received by its out-neighbors $w \in \mathcal{N}_v^+$ in order to compute their estimates $\hat{x}_v^{(w)}(j-1)$ of its own current state. Then, in the presence of communication feedback, the state update step (5) can be replaced by the following one:

$$x_v(j+1) = x_v(j) - \sum_{w \in \mathcal{N}_v^+} P_{wv} \hat{x}_v^{(w)}(j) + \sum_{w \in \mathcal{N}_v^-} P_{vw} \hat{x}_w^{(v)}(j).$$

Notice that this requires every agent v to know not only the entries of the v-th row of the matrix P, but also those of the v-th column

of P. Clearly, such algorithms can be framed in the setting described in Section 2. Indeed, for all $j \geq 1$, one has

$$\phi_{h_{j-1}+k}^{(v)}(i_v'(h_{j-1}+k)) = \Upsilon_k(x_v(j-1)), \quad 0 < k \leq \ell_j$$

$$\psi_{h_{j-1}+k}^{(v)}(i_v'(h_{j-1}+k)) = x_v(j-1), \quad 0 \leq k < \ell_j.$$

The above state update equation may be written in the compact form

$$\mathbf{x}(j+1) = P\mathbf{x}(j) + [(P \odot \Delta(j+1)) - (P \odot \Delta(j+1))^*]\mathbf{1}.$$

Observe that $\mathbf{1}^*[(P \odot \Delta(j)) - (P \odot \Delta(j))^*]\mathbf{1} = 0$, so that, since P is a doubly stochastic matrix, one has $\mathbf{1}^*\mathbf{x}(j+1) = \mathbf{1}^*P\mathbf{x}(j) = \mathbf{1}^*\mathbf{x}(j)$. It follows that $n^{-1}\mathbf{1}^*\mathbf{x}(j) = n^{-1}\mathbf{1}^*\mathbf{x}(0) = y$ for any j. Hence,

$$\zeta(j) = 0, \qquad \mathbf{e}(h_j) = \mathbf{z}(j), \quad \forall j \geq 0. \tag{17}$$

Now, we consider two implementations of the above-described algorithms. Such implementations have increasing communication phase lengths, analogously to those introduced in Section 4. In the first implementation, referred to as algorithm $\mathcal{A}_L'$, linear codes are used in the communication phase, and the length of the j-th phase is $\ell_j = S_L j$ for some $S_L \in \mathbb{N}$. The second implementation, named $\mathcal{A}_R'$, uses repetition codes in the communication phase, and the length of the j-th phase is $\ell_j = S_R j^2$ for some $S_R \in \mathbb{N}$. The following result characterizes the performance of the algorithms $\mathcal{A}_L'$ and $\mathcal{A}_R'$, showing that, with probability 1, the estimates of all the agents converge to the actual value y, and estimating the communication and computational complexities.

Theorem 5 (*With Communication Feedback*)**.** *For any choice of the initial phases length S_L (respectively, S_R), the estimates of the algorithm $\mathcal{A}_L'$ and $\mathcal{A}_R'$ satisfy, with probability 1,*

$$\lim_{t \to \infty} \hat{y}_v(t) = y, \quad \forall v \in \mathcal{V}.$$

Moreover, it is possible to choose the initial phase length S_L (S_R, respectively) in such a way that algorithm $\mathcal{A}_L'$ (respectively, $\mathcal{A}_R'$) has communication and computational complexities satisfying

$$\tau_L'(\delta) \leq C_1' + C_2' \frac{\log^2 \delta^{-1}}{\log^2 \rho^{-1}}, \qquad \kappa_L'(\delta) \leq C_3' + C_4' \frac{\log^4 \delta^{-1}}{\log^4 \rho^{-1}},$$

and, respectively,

$$\tau_R'(\delta) \leq C_5' + C_6' \frac{\log^3 \delta^{-1}}{\log^3 \rho^{-1}}, \qquad \kappa_R'(\delta) \leq C_7' + C_8' \frac{\log^3 \delta^{-1}}{\log^3 \rho^{-1}},$$

for all $\delta \in]0, 1]$, where $\{C_i' : i = 1, \ldots, 8\}$ are positive constants depending only on ε.

Proof. See the Appendix. $\square$

The bounds in Theorem 5 exhibit a better dependence on the desired precision δ with respect to their analogues in Theorem 4. On the other hand, the dependence on ρ is the same. The reason lies in the average-preserving property which can be guaranteed when communication feedback is available. In this case, as shown by Theorem 5, it is not necessary to determine the initial phase length as a function of final desired precision, since the estimates produced by both $\mathcal{A}_L'$ and $\mathcal{A}_R'$ converge to y with probability 1. In contrast, when communication feedback is not available, it is not possible to guarantee that the average of the agents' estimates is preserved. This is the reason why, in Section 4, we had to adjust the initial phase length as a function of the desired precision δ, inducing a worse dependence on δ of the bounds on communication and computational complexities of the algorithms $\mathcal{A}_L$ and $\mathcal{A}_R$ shown in Theorem 4.

6. Conclusion

In this paper, for the first time we have considered the averaging problem on networks of digital links, and established suitable performance figures to evaluate its algorithmic solutions, in terms of communication and computation complexities. The

main contribution of the paper has consisted in proposing and analyzing a family of average consensus algorithms, based on encoding/decoding schemes with precision increasing with time: the increase is meant to compensate the effect of errors in digital communications. Depending on the application, one might prefer to avoid such increase, and to compensate the accumulation of errors by applying a decreasing gain strategy, as proposed in Huang and Manton (2009) and Rajagopal and Wainwright (2008) for networks whose links support the transmission of a real number affected by additive noise. Compared to ours, these results show almost sure convergence to average consensus, with mean square error decreasing as the inverse of time, under slightly more stringent assumptions on the noise (mainly, independence of the additive noise). This guarantees communication and computation complexities growing polynomially in the desired precision, as opposed to the poly-logarithmic dependence of our algorithms. Also, we have investigated how to make use of communication feedback, when available, in order to make the algorithms average preserving, and improve their performance. The question is open whether a logarithmic algorithm can be designed for average consensus on digital networks, and how much global information it would require to be run by the agents.

Acknowledgements

Appendix. Proofs

Proof of Proposition 3. Let us first consider the quantity $\zeta(j)$ defined in (11). It is straightforward to verify that $\zeta(0) = 0$, and the recursion $\zeta(j+1) = \zeta(j) + \xi(j+1)$ is satisfied, with $\xi(j) := n^{-1}\mathbf{1}^*(P \odot \Delta(j))\mathbf{1}$. For $j \geq 0$, $\xi(j)$ is a random variable whose second moment can be upper bounded using the Cauchy–Schwarz inequality:

$$\mathbb{E}[\xi(j)^2] = n^{-2}\mathbb{E}\left[\left(\sum_{v,w} P_{vw}\Delta_{vw}(j)\right)^2\right]$$

$$= n^{-2}\sum_{v',w'}\sum_{v,w}\Delta_{v'w'}\Delta_{vw}\mathbb{E}[\Delta_{v'w'}(j)\Delta_{vw}(j)]$$

$$\leq n^{-2}\sum_{v',w'}\sum_{v,w}P_{v'w'}P_{vw}\sqrt{\mathbb{E}[\Delta_{v'w'}(j)^2]\mathbb{E}[\Delta_{vw}(j)^2]}$$

$$\leq n^{-2}\left(\sum_{v,w}P_{vw}\alpha^j\right)^2 = \alpha^{2j}. \tag{A.1}$$

It follows again from the Cauchy–Schwarz inequality that $\mathbb{E}[\xi(s)\xi(r)] \leq \mathbb{E}[\xi(s)^2]^{1/2}\mathbb{E}[\xi(r)^2]^{1/2} \leq \alpha^{s+r}$, for all $1 \leq s, r \leq j$. Therefore,

$$\mathbb{E}[\zeta^2(j)] = \sum_{1 \leq s,r \leq j}\mathbb{E}[\xi(s)\xi(r)] \leq \sum_{1 \leq s,r \leq j}\alpha^{r+s}$$

$$= \left(\sum_{1 \leq s \leq j}\alpha^s\right)^2 \leq \frac{\alpha^2}{(1-\alpha)^2}.$$

Now, consider $\mathbf{z}(j)$ defined in (10). Observe that $\mathbf{z}(0) = \mathbf{u}(0)$, and the recursion $\mathbf{z}(j+1) = P\mathbf{z}(j) + \mathbf{u}(j+1)$ is satisfied with $\mathbf{u}(0) := \mathbf{x}(0) - n^{-1}\mathbf{1}^T\mathbf{x}(0)\mathbf{1}$, and, for $j \geq 1$, $\mathbf{u}(j) := (P \odot \Delta(j))\mathbf{1} - \xi(j)\mathbf{1}$. Notice that $\|P\mathbf{x}\| \leq \rho\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{1}^*\mathbf{x} = 0$. Since $\mathbf{1}^*\mathbf{u}(j) = 0$, we have

$$\|P\mathbf{u}(j)\| \leq \rho\|\mathbf{u}(j)\|. \tag{A.2}$$

On the other hand, again from the Cauchy–Schwarz inequality, for all $u, v, w \in \mathcal{V}$, we have that

$$\mathbb{E}[\Delta_{vw}(j)\Delta_{vu}(j)] \leq \mathbb{E}[\Delta_{vw}(j)^2]^{\frac{1}{2}}\mathbb{E}[\Delta_{vu}(j)^2]^{\frac{1}{2}} \leq \alpha^{2j},$$

so the random vector $\mathbf{u}(j)$, for $j \geq 1$, satisfies the following bound:

$$\mathbb{E}[\|\mathbf{u}(j)\|^2] = \mathbb{E}[\|(P \odot \Delta(j))\mathbf{1}\|^2] - n\mathbb{E}[\xi(j)^2]$$

$$\leq \sum_v\mathbb{E}\left[\left(\sum_w P_{vw}\Delta_{vw}(j)\right)^2\right]$$

$$= \sum_v\sum_{w,w'}P_{vw}P_{vw'}\mathbb{E}[\Delta_{vw}(j)\Delta_{vw'}(j)]$$

$$\leq n\alpha^{2j}. \tag{A.3}$$

Moreover, recall that $\theta_v \in \Theta$ for any $v \in \mathcal{V}$, where Θ is an interval of unitary length. As a consequence, one has $|\theta_v - y| \leq 1$ for any $v \in \mathcal{V}$, so

$$\mathbb{E}[\|\mathbf{u}(0)\|^2] = \mathbb{E}[\|\mathbf{z}(0)\|^2] = \sum_v\mathbb{E}[(\theta_v - y)^2] \leq n. \tag{A.4}$$

Consider now $\mathbb{E}[\|\mathbf{z}(j)\|^2] = \mathbb{E}\left[\|\sum_{0 \leq s \leq j}P^{j-s}\mathbf{u}(s)\|^2\right]$. By successively applying the Cauchy–Schwarz inequality, (A.2)–(A.4), we get

$$\mathbb{E}[\|\mathbf{z}(j)\|^2] \leq \sum_{0 \leq s,r \leq j}\sqrt{\mathbb{E}[\|P^{j-s}\mathbf{u}(s)\|^2]\mathbb{E}[\|P^{j-r}\mathbf{u}(r)\|^2]}$$

$$\leq \left(\sum_{0 \leq s \leq j}\rho^{(j-s)}\sqrt{\mathbb{E}[\|\mathbf{u}(s)\|^2]}\right)^2$$

$$\leq n\left(\sum_{0 \leq s \leq j}\rho^{(j-s)}\alpha^s\right)^2$$

$$\leq n\left(\rho^j\sum_{s \geq 0}\left(\frac{\alpha}{\rho}\right)^s\right)^2 = n\rho^{2j}\left(1 - \frac{\alpha}{\rho}\right)^{-2},$$

which completes the proof. □

Proof of Theorem 4. We begin by estimating the difference $\mathbf{x}(j+1) - \mathbf{x}(j)$, for $j \geq 0$. Toward this goal, let $\xi(j) := n^{-1}\mathbf{1}^*(P \odot \Delta(j))\mathbf{1}$. Then, we may rewrite

$$\mathbf{x}(j+1) - \mathbf{x}(j)$$
$$= \mathbf{x}(j+1) - n^{-1}\mathbf{1}^*\mathbf{x}(j+1)\mathbf{1} + n^{-1}\mathbf{1}^*\mathbf{x}(j+1)\mathbf{1} - \mathbf{x}(j)$$
$$= \mathbf{x}(j+1) - n^{-1}\mathbf{1}^*\mathbf{x}(j+1)\mathbf{1} + n^{-1}\mathbf{1}^*\mathbf{x}(j)\mathbf{1} + \xi\mathbf{1} - \mathbf{x}(j)$$
$$= \mathbf{z}(j+1) - \mathbf{z}(j) + \xi(j)\mathbf{1}.$$

By successively applying the triangle inequality, Proposition 3, and (A.1), we get

$$\mathbb{E}[\|\mathbf{x}(j+1) - \mathbf{x}(j)\|^2] \leq \mathbb{E}[(\|\mathbf{z}(j+1)\| + \|\mathbf{z}(j)\| + \|\xi(j)\mathbf{1}\|)^2]$$

$$\leq 3\left(\mathbb{E}[\|\mathbf{z}(j+1)\|^2] + \mathbb{E}[\|\mathbf{z}(j)\|^2] + n\mathbb{E}[\xi(j)^2]\right)$$

$$\leq 3n(\rho^{2(j+1)} + \rho^{2j})(1 - \alpha/\rho)^{-2} + 3n\alpha^{2j}$$

$$\leq 9n(1 - \alpha/\rho)^{-2}(\max\{\rho, \alpha\})^{2j} = 9n(1 - \alpha/\rho)^{-2}\rho^{2j}.$$

Hence, one can estimate the probability of the event $E_j := \left\{\|\mathbf{x}(j+1) - \mathbf{x}(j)\|^2 \geq \rho^{2j}\right\}$ by Markov's inequality, obtaining $\mathbb{P}(E_j) \leq 9n(1 - \alpha/\rho)^{-2}\rho^{2j}$. Therefore, $\sum_{j \geq 0}\mathbb{P}(E_j)$ is finite, and the Borel–Cantelli lemma implies that, with probability 1, the event E_j occurs for finitely many $j \in \mathbb{Z}^+$. This implies that, with probability 1, the sequence $\{\mathbf{x}(j)\}$ is Cauchy, and hence convergent. Hence, there exists a $\mathbb{R}^n$-valued random variable $\mathbf{x}_\infty$ such that $\lim_j \mathbf{x}(j) = \mathbf{x}_\infty$ with probability 1.

On the other hand, define $g(\mathbf{x}) = \mathbf{x} - n^{-1}\mathbf{1}^*\mathbf{x}\mathbf{1}$. Then, it can be deduced from (14), again using Markov's inequality and the Borel–Cantelli lemma, that $g(\mathbf{x}(j)) = \mathbf{z}(j)$ converges to $\mathbf{0}$ with probability 1. Then, from the continuity of g, it follows that $g(\mathbf{x}_\infty) = 0$, i.e., $\mathbf{x}_\infty = \hat{y}\mathbf{1}$ for some scalar random variable $\hat{y}$. In order to verify that (15) holds, observe that $\zeta(j) = n^{-1}\mathbf{1}^*\mathbf{x}(j) - y$ is bounded and convergent to $\hat{y} - y$ with probability 1. Hence, $\lim_j \mathbb{E}[\zeta(j)^2] = \mathbb{E}[(\hat{y} - y)^2]$. Then, from (14) we have $\alpha^2(1-\alpha)^{-2}$

$\geq \lim_j \mathbb{E}[\zeta(j)^2] = \mathbb{E}\left[(\hat{y} - y)^2\right]$. Therefore, (16) follows by simply recalling that $\hat{y}_v(t) = x_v(j)$, for $h_j < t \leq h_{j+1}$.

In order to prove the second part of the claim, first recall that $\alpha_L = \beta_L^{S_L}$, with β_L depending only on ε. Hence, $\alpha_L \leq \rho/2$ for all $S_L \geq (\log \rho^{-1} + \log 2)/\log \beta_L^{-1}$. Then, for $\delta \in]0, 1]$, let $u := \sqrt{\delta/2}$. It follows from (9) and (7), and Proposition 3, that, for

$$n^{-1}\|\mathbf{e}(t)\|^2 \leq \delta, \quad \forall t \geq h_j \tag{A.5}$$

to hold, it is sufficient that

$$\alpha_L(1 - \alpha_L)^{-1} \leq u, \quad \rho^j \leq u/2. \tag{A.6}$$

The leftmost inequality in (A.6) is satisfied provided that $S_L \geq \log(2u^{-1})/\log(\beta_L^{-1})$, and the right inequality is satisfied if $j \geq \log(2u^{-1})/\log(\rho^{-1})$. Now, recall that $h_j = \sum_{1 \leq i \leq j} \ell_i = S_L \sum_{1 \leq i \leq j} i \geq 1/2 S_L j^2$. It follows that

$$h_j \geq \frac{S_L}{2} \frac{\log^2 u}{\log^2 \rho} \geq \frac{1}{2 \log \beta_L} \frac{\log^3 u}{\log^2 \rho}$$

implies (A.5). Then, the upper bound on $\tau_L(\delta)$ easily follows. In order to prove the bound on $\kappa_L(\delta)$, observe that (1) implies that, for every $v \in \mathcal{V}$,

$$\sum_{1 \leq t \leq h_j} \kappa_v(t) \leq \sum_{1 \leq i \leq j} B\ell_i^3 = BS_L^3 \sum_{1 \leq i \leq j} i^3 \leq BS_L^3 j^4.$$

Finally, the bounds on $\tau_L(\delta)$ and $\kappa_L(\delta)$ follow from analogous arguments. $\square$

Proof of Theorem 5. From (17), one has $\zeta(j) = 0$ for all j. On the other hand, in the same way as in the proof of Theorem 4, one can argue that $\lim_j \mathbf{z}(j) = \mathbf{0}$ with probability 1. Hence $\lim_t \mathbf{e}(t) = \mathbf{0}$ with probability 1, which is equivalent to the first part of the claim.

Now, let us consider algorithm $\mathcal{A}'_L$, and notice that the proof of Proposition 3 allows us to obtain the following bound for the case with feedback, analogous to (14):

$$n^{-1}\mathbb{E}[\|\mathbf{z}(j)\|^2] \leq 2\rho^{2j}(1 - \alpha/\rho)^{-2}.$$

Also recall that one has $\alpha_L \leq \rho/2$ for all $S_L \geq \log(2\rho^{-1})/\log \beta_L^{-1}$. Hence, for any such S_L,

$$n^{-1}\mathbb{E}[\|\mathbf{e}(t)\|^2] = n^{-1}\mathbb{E}[\|\mathbf{z}(j)\|^2] \leq 2\rho^{2j}(1 - \alpha/\rho)^{-2} \leq \delta$$

for all $t \geq h_j$, if $j \geq \log(8/\delta)/\log(\rho^{-2})$. The upper bound on $\tau'_L(\delta)$ then follows in view of $\ell_j = S_L j$. The upper bound on $\kappa'_L(\delta)$ follows using (1). The bounds on $\tau'_R(\delta)$ and $\kappa'_R(\delta)$ follow from similar arguments. $\square$

References

Ayaso, O., Shah, D., & Dahleh, M. A. (2010) Information theoretic bounds for distributed computation over networks of point-to-point channels. *IEEE Transaction on Information Theory* (in press).

Aysal, T. C., Coates, M. J., & Rabbat, M. G. (2008). Distributed average consensus with dithered quantization. *IEEE Transactions on Signal Processing*, 56(10), 4905–4918.

Carli, R., Bullo, F., & Zampieri, S. (2010). Quantized average consensus via dynamic coding/decoding schemes. *International Journal of Robust and Nonlinear Control*, 20(2), 156–175.

Carli, R., Como, G., Frasca, P., & Garin, F. (2009). Average consensus on digital noisy networks. In *IFAC workshop on estimation and control of networked systems. Venice, Italy. September*.

Como, G., & Dahleh, M. (2009). Lower bounds on the estimation error in problems of distributed computation. In *Information theory and applications workshop. San Diego, CA. February*.

Como, G., Fagnani, F., & Zampieri, S. (2010). Anytime reliable transmission of real-valued information through digital noisy channels. *SIAM Journal on Control and Optimization*, 48(6), 3903–3924.

Denantes, P., Benezit, F., Thiran, P., & Vetterli, M. (2008). Which distributed averaging algorithm should I choose for my sensor network? In *IEEE conference on computer communications. INFOCOM. April* (pp. 986–994).

Frasca, P., Carli, R., Fagnani, F., & Zampieri, S. (2009). Average consensus on networks with quantized communication. *International Journal of Robust and Nonlinear Control*, 19(16), 1787–1816.

Giridhar, A., & Kumar, P. R. (2006). Towards a theory of in-network computation in wireless sensor networks. *IEEE Communications Magazine*, 44(4), 98–107.

Hendrickx, J. M., Olshevsky, A., & Tsitsiklis, J. N. (2009). Distributed anonymous function computation in information fusion and multiagent systems. In *Allerton conf. on communications, control and computing. Allerton, IL, USA. September*.

Huang, M., & Manton, J. H. (2009). Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. *SIAM Journal on Control and Optimization*, 48(1), 134–161.

Kar, S., & Moura, J. M. F. (2009). Distributed consensus algorithms in sensor networks with imperfect communication: link failures and channel noise. *IEEE Transactions on Signal Processing*, 57(1), 355–369.

Kar, S., & Moura, J. M. F. (2010). Distributed consensus algorithms in sensor networks: quantized data and random link failures. *IEEE Transactions on Signal Processing*, 58(3), 1383–1400.

Kushilevitz, E., & Nisan, N. (1997). *Communication complexity*. Cambridge University Press.

Li, T., Fu, M., Xie, L., & Zhang, J. -F. (2010). Distributed consensus with limited communication data rate. *IEEE Transactions on Automatic Control*, in press (doi:10.1109/TAC.2010.2052384).

Martínez, S., Bullo, F., Cortés, J., & Frazzoli, E. (2007). On synchronous robotic networks—part I: models, tasks and complexity. *IEEE Transactions on Automatic Control*, 52(12), 2199–2213.

Rajagopal, R., & Wainwright, M. J. (2008). Network-based consensus averaging with general noisy channels. May. arXiv:0805.0438.

Ying, L., Srikant, R., & Dullerud, G. E. (2007). Distributed symmetric function computation in noisy wireless sensor networks. *IEEE Transactions on Information Theory*, 53(12), 4826–4833.

Ruggero Carli received his Laurea Degree in Information Engineering from the University of Padova, Padova, Italy in 2004. From February 2008 to July 2010 he was a Post-doctoral Fellow in the Department of Mechanical Engineering, Center for Control, Dynamical Systems, and Computation, University of California at Santa Barbara (USA). He is currently a Post-doctoral Fellow in the Department of Information Engineering, University of Padova, Padova. His research interests include control theory and, in particular, control under communication constraints, cooperative control, and distributed estimation.

Giacomo Como is currently a Post-doctoral Associate at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA. He received his B.Sc., MS, and Ph.D. degrees in Applied Mathematics from Politecnico di Torino in 2002, 2004, and 2008, respectively. In 2006–2007 he was a Visiting Assistant in Research at the Department of Electrical Engineering, Yale University. His current research interests include information theory, coding theory, distributed estimation and control, information dynamics, and flows over networks.

Paolo Frasca received his Ph.D. degree in Applied Mathematics from Politecnico di Torino, Italy, in 2009. In 2008, he was at the Department of Mechnical Engineering, University of California at Santa Barbara. In 2009, he was at the DIIMA, Università di Salerno and at the IAC, CNR, Roma. He is currently at the Dipartimento di Matematica, Politecnico di Torino. His research interests include the mathematical foundations of multi-agent and networked systems: modeling, analysis, control, and optimization.

Federica Garin is presently with the NeCS team at INRIA Rhône-Alpes, Grenoble (France). She received her B.Sc., MS, and Ph.D. degrees in Applied Mathematics from Politecnico di Torino (Italy) in 2002, 2004, and 2008, respectively. In 2006–2007 she was a visiting scholar at CMRR, University of California at San Diego; in 2008 and 2009 she was a post-doctoral researcher at DEI, Università di Padova (Italy). Her research interests are in information and coding theory, control under communication constraints, distributed estimation, and control.