Average consensus on digital noisy networks ⋆

Ruggero Carli [*] **Giacomo Como** [**] **Paolo Frasca** [***]
Federica Garin [****]

[*] *University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: carlirug@engineering.ucsb.edu).*
[**] *Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: giacomo@mit.edu).*
[***] *DIIMA, Università di Salerno, and Istituto per le Applicazioni del Calcolo (CNR), Roma, Italy (e-mail: paolo.frasca@gmail.com).*
[****] *DEI, Università di Padova, Italy (e-mail: garin@dei.unipd.it).*

Abstract: We propose a class of distributed algorithms for computing arithmetic averages (average consensus) over networks of agents connected through digital noisy broadcast channels. Our algorithms do not require the agents to have knowledge of the network structure, nor do they assume any noiseless feedback to be available. We prove convergence to consensus, with both number of channel uses and computational complexity which are poly-logarithmic in the desired precision.

Keywords: Control with communication constraints; Consensus problems; Decentralized algorithms for computation over sensor networks

1. INTRODUCTION

As large-scale networks have emerged –characterized by the lack of centralized access to information, and possibly time-varying topologies–, the last few years have witnessed an increasing research interest in problems of distributed computation. In these scenarios, large collections of agents –each having access to some partial information– aim at computing an application-specific function of the global information. The computation must be completely distributed, i.e. each agent can rely only on local observations, while iteratively processing the available information and communicating with the other agents. The main challenge in the design of such distributed computation systems is posed by the scarce energetic autonomy of the agents, which severely constrains both their computational and communication capabilities.

A special instance, which has been the object of recent extensive work, is the average consensus problem, in which a large number of agents aims at computing the arithmetic average of initial scalar measurements, in a distributed fashion While most of the literature 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, Nedic et al. (2007); Aysal et al. (2008); Kar and Moura (2007); Frasca et al. (2009), packet losses Fagnani and Zampieri (2009), or transmission affected by additive noise Huang and Manton (2009); Rajagopal and Wainwright (2008). However, at our knowledge, there is no contribution yet toward the design of consensus algorithms on networks in which the communication links are modeled as digital noisy channels. For such networks, information-theoretic bounds on the performance of distributed computation algorithms have been established in Ayaso et al. (2008); Como and Dahleh (2009).

In the present paper, we shall present and analyze distributed algorithms for average consensus on networks with digital noisy communication channels. The algorithms we propose combine the classical iterative linear consensus algorithm with coding schemes for the reliable transmission of real numbers on noisy channels, recently proposed in Como et al. (2008). Our algorithms are fully distributed, in the sense that they do not require the agents to have any knowledge of the network structure. The main results consist in showing that such algorithms drive the agents arbitrarily close to average consensus, at the price of using a number of channel transmissions and in-node computations at most poly-logarithmic in the desired precision.

The poly-logarithmic growth in the desired precision of both communication and computational complexity of the consensus algorithms for networks with digital noisy channels has to be compared with the logarithmic growth characterizing the algorithms proposed for quantized transmission channels Carli et al. (2009). It can be argued that such a performance gap has mainly to be attributed to the different availability of feedback information in the two problems. Indeed, deterministic channels inherently include perfect noiseless feedback, as the transmitter knows exactly what the receiver is going to observe: such a feedback information is critical in the design of consensus algorithms for networks with quantized communication links, as the agents use it to coordinate among themselves. In contrast, when noiseless channel feedback is not available–as in the problem addressed in the present paper–, coordination of the different agents becomes a harder task.

General results on coding for interacting communication Schulman (1996); Rajagopalan and Schulman (1994) may suggest logarithmic times to possibly be achievable by

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

embedding an efficient quantized consensus algorithm in a global error correcting coding scheme. However, as it has also been argued in Giridhar and Kumar (2006), the tree-code constructions proposed in Schulman (1996); Rajagopalan and Schulman (1994) suffer from high computational complexity which likely prevents their practical implementation. Moreover, their global design requires each agent to have knowledge of the whole network topology, an assumption which contrasts the reconfigurability requirements. In contrast, the algorithms we shall propose do not require the agents to have any knowledge of the network topology, and their computational complexity can also be kept tractable.

The rest of the paper is organized as follows. In Section 2 we formally state the problem and present our solution. Section 3 we present the main convergence result and the bounds on the convergence times. Simulations are presented in Section 4, comparing our algorithm with a decreasing gains heuristic.

2. PROBLEM STATEMENT AND PROPOSED ALGORITHM

2.1 Problem statement

We shall consider a set $\mathcal{V}$ of N agents, possibly representing sensors in a wireless network, each having access to some partial information consisting in the observation of a scalar value, which for instance may represent the measurement of some physical quantity. The observation of agent v will be denoted by $\theta_v \in [0,1]$, while $\boldsymbol{\theta} = (\theta_v)$ will indicate the full vector of observations. The goal is for the network to compute the arithmetic average of such values, $\theta_{\mathrm{ave}} = \frac{1}{N}\sum_v \theta_v$ by exchanging information among themselves 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 bit $a_v(t) \in \{0,1\}$ to a subset of agents, to be denoted by $\mathcal{N}_v^+ \subset V$. Every agent $w \in \mathcal{N}_v^+$ receives a possibly erased version $b_{v,w}(t) \in \{0,1,?\}$ of $a_v(t)$. The set of in-neighbors of agent w will be denoted by $\mathcal{N}_w^- = \{v : w \in \mathcal{N}_v^+\}$, whereas $b_w(t) = (b_{v,w}(t))_{v\in\mathcal{N}_w^-}$ will denote the vector of signals received by agent w at time t. We shall assume that, conditioned on the broadcasted bits $(a_v(t))_v$, the received signals $b_{v,w}(t)$ are mutually independent, with $b_{v,w}(t) = a_v(t)$ with probability $1-\varepsilon$, and $b_{v,w}(t) =?$ with probability ε, where $\varepsilon > 0$ is some erasure probability which -for the sake of simplicity- is assumed to remain constant in t, v and w. Furthermore, erasures will be assumed independent in time. Distributedness of the computation/communication algorithm is then enforced by requiring that the bit $a_v(t)$ to be a function of the local information available to agent v at time t, i.e. of its initial observation θ_v, as well as the signals $\{b_v(s)\}_{1\le s<t}$ received by agent v up to time $t-1$.

The communication setting outlined above can be conveniently described by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (the communication graph), whose vertices are the agents, and such that an ordered pair (i,j) with $i \neq j$ belongs to $\mathcal{E}$ if and only if $j \in \mathcal{N}_i^+$ (equivalently $i \in \mathcal{N}_j^-$), i.e., if i transmits to j with erasure probability $\varepsilon < 1$. Throughout the paper, we shall assume that the graph $\mathcal{G}$ is strongly connected, i.e. that any two nodes u, v are connected by a directed path. We will also assume that $\mathcal{G}$ has self-loops on each vertex; this represents the fact that each node has access to its own information, which is equivalent to assume a noiseless channel available from i to itself.

2.2 Proposed algorithm

Our idea is to use a traditional linear average-consensus algorithm, combined with a technique for transmission of real numbers over noisy digital broadcast channels, a joint source and channel coding proposed in Como et al. (2008).

Consider the above-described scenario, specified by a set of N vertices $\mathcal{V}$, a strongly connected communication graph $\mathcal{G}$, and an erasure probability ε. The ingredients of our algorithm are:

- a consensus matrix P, i.e., a doubly-stochastic primitive matrix adapted to $\mathcal{G}$, with non-zero diagonal entries [1];
- an increasing sequence of positive integers $\{\tau(k)\}_{k\in\mathbb{N}}$, such that $\lim_{k\to\infty}\tau(k) = +\infty$; $\tau(k)$ represents the number of bits that each node transmits at k-th iteration of the algorithm;
- a sequence of encoders, i.e., maps
$$\phi^{(k)} : [0,1] \to \{0,1\}^{\tau(k)};$$
- a sequence of decoders, i.e., maps
$$\psi^{(k)} : \{0,1,?\}^{\tau(k)} \to [0,1].$$

The usual linear average-consensus algorithm is:
$$\boldsymbol{x}(0) = \boldsymbol{\theta}, \qquad \boldsymbol{x}(k+1) = P\boldsymbol{x}(k)$$
i.e., at k-th iteration, node i receives from its in-neighbors the numbers $x_j(k)$, $j \in \mathcal{N}_i^-$, and updates its state:
$$x_i(k+1) = \textstyle\sum_{j\in\mathcal{N}_i^-} P_{ij}x_j(k).$$
It is well-known (Perron-Frobenius theory) that, for primitive doubly-stochastic P, from any initial condition $\boldsymbol{\theta}$, each entry of $\boldsymbol{x}(t)$ converges to the average θ_{ave}.

We propose to adapt this algorithm, in a way that takes into account the necessity to transmit the real values $x_j(k)$ along digital noisy channels. The initialization of the algorithm is unchanged: $\boldsymbol{x}(0) = \boldsymbol{\theta}$. Between iterations k and $k+1$ of our consensus-like algorithm, we allow each node j to broadcast $\tau(k)$ bits to its neighbors:

- the bits transmitted by node j at iteration k are the message $a_j(k) = \phi^{(k)}(x_j(k))$, i.e., a suitably encoded version of the state $x_j(k)$;
- each $i \in \mathcal{N}_j^+$ will receive a corrupted version of $a_j(k)$, $b_{ij}(k)$, and will use the decoder $\psi^{(k)}$ to recover an estimate $\hat{x}_{ij}(k) = \psi^{(k)}(b_{ij}(k))$

Then, the next consensus iteration will take place, where node i will use $\hat{x}_{ij}(k)$ to replace the exact state $x_j(k)$ which he can not know exactly:
$$x_i(k+1) = P_{ii}x_i(k) + \textstyle\sum_{j\in\mathcal{N}_i^-} P_{ij}\hat{x}_{ij}(k)$$
Clearly, we can write $\hat{x}_{ij}(k) = x_j(k) + v_{ij}(k)$, and we might think at $v_{ij}(k)$ as a residual noise which could not be removed by the error-correction performed by the decoder. Notice that $v_{ij}(k)$ in general does not have zero-mean, and depends on $x_j(t)$ (and thus depends on all past noises). A suitable choice of the encoder/decoder pairs and of the transmission phases allows to obtain a noise deacreasing

[1] A matrix $P \in \mathbb{R}^{N\times N}$ is *doubly stochastic* if it has non-negative entries, and all its rows and columns sum to one. Its induced graph $\mathcal{G}_P$ has N vertices and an edge (i,j) if and only if $P_{ij} \neq 0$. P is adapted to $\mathcal{G}$ if $\mathcal{G}_P \subseteq \mathcal{G}$. P is primitive if $\mathcal{G}_P$ is strongly connected and aperiodic (i.e., the greatest common divisor of cycle lengths is 1).

with respect to k, with a speed which will be discussed in Section 3.2. To this effect, the assumption that the transmission lengths $\tau(k)$ are increasing in k is essential, because the coding gain is asymptotic in the length of codewords. This remarks leads us to name our proposed algorithm 'Increasing Precision Algorithm' (IPA).

We can summarize the evolution of the state $\boldsymbol{x}(k)$ in the IPA algorithm in the following compact form: define $V(k)$ to be the matrix with entries $V(k)_{ij} = v_{ij}(k)$ if $j \in \mathcal{N}_i^-$, and 0 otherwise; then:

$$\boldsymbol{x}(k+1) = P\boldsymbol{x}(k) + (V(k) \odot P(k))\,\mathbf{1}\,, \qquad (1)$$

where $\mathbf{1}$ denotes a vector with entries all equal to 1 and $\odot$ denotes entry-wise product between two matrices.

Notice that k takes into account the number of consensus-like updates. However, it is more reasonable to measure the elapsed time in terms of the number of transmissions: state $\boldsymbol{x}(k)$ is reached after $t(k) = \sum_{h\leq k}\tau(h)$ transmissions. Another way of analyzing our algorithm is to take into account also computation time. In fact, the decoding process can be computationally relevant, in particular for codes with best error-correction performance. For this reason we introduce the two following definitions: the computation time per channel use $\kappa(\tau)$, which depends on the choice of the encoder/decoder pair and on the length τ used for transmission, and the total transmission/computation time $T(k) = \sum_{h\leq k}\max(\tau(h), \kappa(\tau(h)))$. The choice of transmission time t or total transmission/computation time T as the relevant notion of time step depends if the focus is on the number of channel uses (which are power-expensive), or the actual time (particularly relevant in the asymptotic regime).

3. PERFORMANCE ANALYSIS

3.1 Average quadratic error

To analyze the performance of our algorithm, we will study the distance from the average of initial values θ_{ave}, at k-th iteration of (1):

$$\boldsymbol{e}(k) = \boldsymbol{x}(k) - \tfrac{1}{N}\left(\mathbf{1}^T\boldsymbol{x}(0)\right)\mathbf{1}\,.$$

We have that $\boldsymbol{e}(k)$ can be decomposed as $\boldsymbol{e}(k) = \boldsymbol{z}(k) + \zeta(k)\mathbf{1}$, where $\boldsymbol{z}(k) = \boldsymbol{x}(k) - \left(\frac{1}{N}\mathbf{1}^T\boldsymbol{x}(k)\right)\mathbf{1}$ represents the distance from consensus (distance from average of current states), whereas $\zeta(k) = \frac{1}{N}\mathbf{1}^T\left(\boldsymbol{x}(k) - \boldsymbol{x}(0)\right)$ accounts for the distance between the current average and the average of the initial conditions. Our aim is to give a bound on $\frac{1}{N}\mathbb{E}[\|\boldsymbol{e}(k)\|^2] = \mathbb{E}[\|\zeta(k)\|^2] + \frac{1}{N}\mathbb{E}[\|\boldsymbol{z}(k)\|^2]$.

We will use the following classical result (Perron-Frobenius theorem): a primitive stochastic matrix P has the eigenvalue 1 with multiplicity 1 and eigenvector $\mathbf{1}$, and all other eigenvalues with modulus strictly smaller than 1. If P is a doubly-stochastic primitive matrix with positive diagonal elements, then also P^TP is doubly-stochastic and primitive. Hence, P has largest singular value equal to 1 and all other singular values strictly smaller than 1. Define $\rho(P)$ to be the second largest singular value of P, and notice that for all vectors $\boldsymbol{v} \perp \mathbf{1}$, $\|P\boldsymbol{v}\| \leq \rho(P)\|\boldsymbol{v}\|$, where $\|\cdot\|$ denotes Euclidean norm.

This fact is essential for our main result, which is the following theorem.

Theorem 1. Consider the IPA algorithm, and let $\rho = \rho(P)$, $\boldsymbol{z}(k)$ and $\zeta(k)$ be defined as above. Assume that the sequence of matrices $\{V(k)\}_{k\in\mathbb{N}}$ satisfies the property that for all $k \in \mathbb{N}$, $\mathbb{E}[V_{n,m}(k)^2] \leq \alpha^{2k}$ for some $0 < \alpha < \rho$. Then, for all $k \in \mathbb{N}$,

$$\mathbb{E}[\zeta^2(k)] \leq \frac{\alpha^2}{(1-\alpha)^2}$$

and

$$\frac{1}{N}\mathbb{E}[\|\boldsymbol{z}(k)\|^2] \leq \rho^{2k}\frac{1}{\left(1-\frac{\alpha}{\rho}\right)^2}$$

□

Proof. Let us first consider $\zeta(k) = \frac{1}{N}\mathbf{1}^T(\boldsymbol{x}(k) - \boldsymbol{x}(0))$. Observe that it satisfies the recursion

$$\zeta(0) = 0; \qquad \zeta(k+1) = \zeta(k) + \xi(k+1)\,,$$

where $\xi(k) = \frac{1}{N}\mathbf{1}^T(P \odot V(k))\mathbf{1}$ is some noise whose second moment can be bounded by

$$\begin{aligned}
\mathbb{E}\left[\xi(k)^2\right] &= \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{n,m}P_{n,m}V_{n,m}(k)\right)^2\right]\\
&= \frac{1}{N^2}\sum_{h,l}\sum_{m,n}P_{h,l}P_{m,n}\mathbb{E}\left[V_{h,l}(k)V_{m,n}(k)\right]\\
&\leq \frac{1}{N^2}\sum_{h,l}\sum_{m,n}P_{h,l}P_{m,n}\sqrt{\mathbb{E}\left[V_{h,l}(k)^2\right]\mathbb{E}\left[V_{m,n}(k)^2\right]}\\
&\leq \frac{1}{N^2}\left(\sum_{h,l}P_{h,l}\alpha^k\right)^2 = \alpha^{2k}\,.
\end{aligned}$$

It then follows that

$$\begin{aligned}
\mathbb{E}[\zeta^2(k)] &= \mathbb{E}\Big[\Big(\sum_{1\leq s\leq k}\xi(s)\Big)^2\Big]\\
&= \sum_{1\leq s,r\leq k}\mathbb{E}\left[\xi(s)\xi(r)\right]\\
&\leq \sum_{1\leq s,r\leq k}\mathbb{E}\left[\xi(s)^2\right]^{1/2}\mathbb{E}\left[\xi(r)^2\right]^{1/2}\\
&\leq \left(\sum_{1\leq s\leq k}\alpha^s\right)^2\\
&\leq \frac{\alpha^2}{(1-\alpha)^2}.
\end{aligned} \qquad (2)$$

Now, let us focus on the orthogonal component $\boldsymbol{z}(k)$ of the error. It satisfies the recursion

$$\boldsymbol{z}(0) = \boldsymbol{w}(0)\,, \qquad \boldsymbol{z}(k+1) = P\boldsymbol{z}(k) + \boldsymbol{w}(k+1)$$

where

$$\boldsymbol{w}(0) = \boldsymbol{x}(0) - \frac{1}{N}\mathbf{1}^T\boldsymbol{x}(0)\mathbf{1}, \qquad \boldsymbol{w}(k) = (P\odot V(k))\mathbf{1} - \xi(k)\mathbf{1}$$

is a noise vector sequence satisfying $\mathbf{1}^T\boldsymbol{w}(k) = 0$, so that $\|P\boldsymbol{w}(k)\| \leq \rho\|\boldsymbol{w}(k)\|$, where $\rho = \rho(P)$. The noise $\boldsymbol{w}(k)$ also satisfies the following bound:

$$
\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{w}(k)\|^2\right] =&\mathbb{E}\left[\|(P\odot V(k))\mathbf{1}\|^2\right] - N\mathbb{E}\left[\xi(k)^2\right]\\
\leq& \sum_l \mathbb{E}\left[\left(\sum_n P_{l,n}V_{l,n}(k)\right)^2\right]\\
=&\sum_l\sum_{m,n} P_{l,m}P_{l,n}\mathbb{E}\left[V_{l,m}(k)V_{l,n}(k)\right]\\
\leq&\sum_l\sum_{m,n} P_{l,m}P_{l,n}\sqrt{\mathbb{E}\left[V_{l,m}(k)^2\right]\mathbb{E}\left[V_{l,n}(k)^2\right]}\\
\leq&\sum_l\left(\sum_m P_{l,m}\alpha^k\right)^2\\
\leq& N\alpha^{2k}.
\end{aligned}
\tag{3}
$$

Then, consider $\mathbb{E}[\|\boldsymbol{z}(k)\|^2] = \mathbb{E}[\|\sum_{s=0}^{k} P^{k-s}w(s)\|^2$: by triangle inequality and Cauchy-Schwarz inequality, we get

$$
\begin{aligned}
\mathbb{E}[\|\boldsymbol{z}(k)\|^2] \leq& \sum_{0\leq s,r\leq k}\sqrt{\mathbb{E}\left[\|P^{k-s}\boldsymbol{w}(s)\|^2\right]\mathbb{E}\left[\|P^{k-r}\boldsymbol{w}(r)\|^2\right]}\\
\leq&\left(\sum_{0\leq s\leq k}\rho^{(k-s)}\sqrt{\mathbb{E}\left[\|\boldsymbol{w}(s)\|^2\right]}\right)^2\\
\leq&\left(\sum_{0\leq s\leq k}\rho^{(k-s)}\sqrt{N}\alpha^s\right)^2\\
\leq&\left(\sqrt{N}\rho^k\sum_{s\geq 0}\left(\frac{\alpha}{\rho}\right)^s\right)^2 = N\rho^{2k}\left(1-\frac{\alpha}{\rho}\right)^{-2} \qquad \blacksquare
\end{aligned}
$$

The assumption $\alpha < \rho$ is not essential in Theorem 1: Equation (2) is true for all $\alpha > 0$, while Equation (3) becomes $\frac{1}{N}\mathbb{E}[\|\boldsymbol{z}(k)\|^2] \leq (k+1)^2\rho^{2k}$ if $\alpha = \rho$, and $\frac{1}{N}\mathbb{E}[\|\boldsymbol{z}(k)\|^2] \leq \alpha^{2k}\left(1-\frac{\rho}{\alpha}\right)^{-2}$ if $\rho < \alpha < 1$.

An important consequence of Theorem 1 is that, as $k \to \infty$, mean square consensus is asymptotically reached. Moreover observe that the mean squared distance between the asymptotic consensus point and the average of the initial conditions, is upper bounded by $\frac{\alpha^2}{(1-\alpha)^2}$; since α depends only on the coding transmission scheme, this bound is, remarkably, independent of either the size of the network or the consensus matrix P.

3.2 Achievable noise decay

Here, we shortly describe two families of codes which allow (at the price of a different complexity) to obtain two levels of decay speed of the error after decoding. More details can be found in Como et al. (2008).

A first family, which we will call class-(a) encoder/decoder pairs, is based on linear random-tree codes. Such coding schemes are guaranteed to have error estimated by $\mathbb{E}[(x-\hat{x})^2] \leq C\beta^{2\tau}$, where τ is the transmission length, and $C > 0$, $\beta \in (0,1)$, are constants depending on the erasure probability ε only. The encoding complexity of these schemes grows quadratically in τ, while the decoding complexity scales like τ^3.

A second family, which will be referred to as class-(b), has both encoding and decoding complexity scaling linearly in τ, and error which can be estimated by $\mathbb{E}[(x-\hat{x})^2] \leq C\beta^{2\sqrt{\tau}}$, for some constants $C > 0$, $\beta \in (0,1)$.

Clearly, performance of the coding schemes is given with respect to the transmission length τ. The speed of convergence of the error to zero with respect to τ suggests the correct choice of time phases $\tau(k)$ to use in the IPA algorithm, in order to achieve exponential decay of the error with respect to iteration number k:

(a) if the encoder belongs to class (a), choose $\tau(k) = Sk$ for $S > 0$;
(b) if the encoder belongs to class (b), choose $\tau(k) = S^2k^2$ for some $S > 0$.

With this choice, the assumptions of Theorem 1 are met with $\alpha = \beta^S$. Notice that α can be made arbitrarily small by increasing S, but at the same time large S corresponds to longer transmission time; in Section 3.3 we will discuss suitable choices of the parameter S.

We define the algorithm IPA-(a) and IPA-(b) respectively to be the IPA algorithm with a sequence of encoder/decoder pairs chosen from class (a) or resp. (b), and with the corresponding choice of $\tau(k)$ defined above. We summarize here the scaling of $t(k)$ and $T(k)$ for such algorithms, for further use in the next section.

IPA-(a) $\tau(k) = Sk$ implies that the total transmission time is $\frac{1}{2}Sk^2 \leq t(k) \leq Sk^2$. The computational complexity is $\kappa(\tau) = K\tau^3$, so that the total transmission/computation time is $T(k) = \Theta(k^4)$

IPA-(b) $\tau(k) = S^2k^2$ implies that the total transmission time is $t(k) = \Theta(k^3)$. The computational complexity is $\kappa(\tau) = K\tau$, so that the total transmission/computation time is $T(k) = \Theta(k^3)$.

3.3 Convergence times

First of all, we can re-write the result in Theorem 1 expressing the decay of the error with respect to transmission time t and total transmission/computation time T. With a slight abuse of notation, we will write $\boldsymbol{z}(t)$ for $\boldsymbol{z}(t(k))$ and $\boldsymbol{z}(T)$ for $\boldsymbol{z}(T(k))$. The results are summarized in the following corollary.

Corollary 2. (a) For the IPA-(a) algorithm there exists constants $C, \bar{C} > 0$ and $\gamma, \bar{\gamma} \in (0,1)$ such that:

$$\tfrac{1}{N}\mathbb{E}[\|\boldsymbol{z}(t)\|^2] \leq C\gamma^{\sqrt{t}}$$

and

$$\tfrac{1}{N}\mathbb{E}[\|\boldsymbol{z}(T)\|^2] \leq \bar{C}\bar{\gamma}^{\sqrt[4]{T}}$$

(b) For the IPA-(b) algorithm there exists constants $C, \bar{C} > 0$ and $\gamma, \bar{\gamma} \in (0,1)$ such that:

$$\tfrac{1}{N}\mathbb{E}[\|\boldsymbol{z}(t)\|^2] \leq C\gamma^{\sqrt[3]{t}}$$

and

$$\tfrac{1}{N}\mathbb{E}[\|\boldsymbol{z}(T)\|^2] \leq \bar{C}\bar{\gamma}^{\sqrt[3]{T}} \qquad \square$$

Now we investigate how much time is necessary to achieve a specified tolerance on the distance from average consensus. A traditional index to evaluate the performance of the standard consensus algorithm in terms of its speed, is defined as follows. Given $\delta > 0$, we define the *δ-convergence time* as

$$k_\delta = \inf\{k \in \mathbb{N} \mid \tfrac{1}{N}\|\boldsymbol{e}(h)\|^2 \leq \delta,\ \forall h \geq k\}.$$

With this definition, for standard linear average consensus algorithm we have that, for δ small enough,

$$k_\delta \leq \frac{\log\delta^{-1}}{\log\rho^{-1}}.$$

To understand the performance of the proposed digital average consensus algorithm, we need to adapt the above definition. Just considering the number of algorithm steps, irrespective of the number of channel accesses which are used at each step, would not be appropriate. We rather want to consider as performance index the *δ-transmission time* t_δ, defined as

$$t_\delta = \inf\{t(k), k \in \mathbb{N} \mid \tfrac{1}{N}\mathbb{E}[\|\boldsymbol{e}(h)\|^2] \leq \delta, \ \forall h \geq k\}.$$

If additionally one wants to keep into account the time required by computations, a suitable convergence index is the *δ-computation/transmission time* T_δ, defined as

$$T_\delta = \inf\{T(k), k \in \mathbb{N} \mid \tfrac{1}{N}\mathbb{E}[\|\boldsymbol{e}(h)\|^2] \leq \delta, \ \forall h \geq k\}.$$

These two indexes are the object of the next two results.

Corollary 3. Consider algorithm IPA and assume that the assumptions of Theorem 1 are met with $\alpha = \beta^S$. Then, S can be chosen in order to ensure that, for δ small enough,

- for algorithm IPA-(a),

$$t_\delta \leq \frac{1}{8\log\beta^{-1}} \frac{\log^3\delta^{-1}}{\log^2\rho^{-1}};$$

- for algorithm IPA -(b)

$$t_\delta \leq \frac{1}{32\log^2\beta^{-1}} \frac{\log^5\delta^{-1}}{\log^3\rho^{-1}}.$$

□

Proof. A precision δ is obtained if the inequality

$$\frac{\alpha^2}{(1-\alpha)^2} + \rho^{2k}\frac{1}{\left(1-\frac{\alpha}{\rho}\right)^2} \leq \delta \qquad (4)$$

is satisfied. Let $u^2 = \delta/2$, and recall that $\alpha = \beta^S$. Then $\frac{\alpha^2}{(1-\alpha)^2} \leq u^2$ is satisfied if $S \geq \frac{\log\frac{u}{1+u}}{\log\beta}$, and then if $S \geq \frac{\log u^{-1}}{\log\beta^{-1}}$. On the other hand, $\rho^{2k}\frac{1}{\left(1-\frac{\alpha}{\rho}\right)^2} \leq u^2$ if

$$k \geq \frac{\log u}{\log\rho} + \frac{\log\left(1-\frac{\alpha}{\rho}\right)}{\log\rho},$$

and then if $k \geq \frac{\log u^{-1}}{\log\rho^{-1}}$. We conclude that inequality (4) is satisfied, for δ small enough, if both $k \geq \frac{1}{2}\frac{\log\delta^{-1}}{\log\rho^{-1}}$ and $S \geq \frac{1}{2}\frac{\log\delta^{-1}}{\log\beta^{-1}}$. To meet the assumptions of Theorem 1, we have to assume that $S \geq \frac{\log\rho}{\log\beta}$; however, for δ small enough, the latter condition can be disregarded. Hence, for algorithm IPA-(a),

$$t_\delta = \frac{S}{2}k_\delta(k_\delta+1) \leq S k_\delta^2 \leq \frac{1}{8\log\beta^{-1}}\frac{\log^3\delta^{-1}}{\log^2\rho^{-1}}.$$

Instead, for algorithm IPA-(b),

$$t_\delta \leq S^2 k_\delta^3 \leq \frac{1}{32\log^2\beta^{-1}}\frac{\log^5\delta^{-1}}{\log^3\rho^{-1}}.$$

■

Reasoning similarly to the previous derivation, we can also argue the following result, regarding the δ-computation/ transmission time.

Corollary 4. Consider algorithm IPA and assume that the assumptions of Theorem 1 are met with $\alpha = \beta^S$. Then, S can be chosen in order to ensure that, for δ small enough,

- for algorithm IPA-(a),

$$T_\delta \leq \frac{K}{128\log^2\beta^{-1}}\frac{\log^7\delta^{-1}}{\log^4\rho^{-1}};$$

- for algorithm IPA-(b)

$$T_\delta \leq \frac{K}{32\log^2\beta^{-1}}\frac{\log^5\delta^{-1}}{\log^3\rho^{-1}}.$$

□

4. SIMULATION RESULTS AND COMPARISON WITH DECREASING GAINS STRATEGY

For implementing our algorithm, we have chosen a very low-complexity strategy: we have considered a particularly simple instance of class-(b) coding scheme, which is a generalization of repetition codes. The encoder $\psi^{(k)}$: $[0,1] \to \{0,1\}^{\tau(k)}$ is constructed as follows. Given $x \in [0,1]$, denote its diadic expansion by $x = \sum_{i\geq 1} c_i 2^{-i}$, $c_i \in \{0,1\}$. Then $\phi^{(k)}(x) \in \{0,1\}^\tau$ consists in transmitting the bits c_i's and some repetitions of them which are more frequent for most significant bits, as follows

$$\phi^{(k)}(x) = (c_1,\, c_1, c_2,\, c_1, c_2, c_3,\, c_1, c_2, c_3, c_4,\, \ldots)\,.$$

The decoder $\psi^{(k)}$ sees a version of such vector where some of the transmitted bits are erased, and constructs a decoded $\hat{x} = \sum_{i\geq 1} d_i 2^{-i}$ as follows. First, notice that all c_i's with $i > \nu(k)$ were not transmitted at all, where $\nu(k)$ is such that $\nu(k)(\nu(k)+1)/2 = \tau(k)$; the decoder will put $d_i = 0$ for $i > \nu(k)$. For bits c_i, $i \leq \nu(k)$, the decoder will put correctly $d_i = c_i$ if at least one of the repeated occurrences of c_i in the transmitted word has been received un-erased, and otherwise will let d_i be 0 or 1 uniformly at random.

To form an instance of IPA-(b) algorithm with such coding scheme, we have chosen transmission lengths $\tau(k) = \frac{k(k+1)}{2} \sim \frac{k^2}{2}$ (in this case, $\nu(k) = k$). Theorem 1 and its corollaries apply and predict convergence to consensus.

With simulations, we want to compare our algorithm with a different strategy which was used in previous literature to compute approximate averages running a consensus algorithm in the presence of noisy communications. We will refer to such family of algorithms as to 'decreasing gain' algorithms, because the key idea is to have time-varying gains, which give decreasing weight to information coming from neighbors, so as to avoid accumulating an amount of error growing to infinity with time. Algorithms exploiting this idea were presented independently by various authors (see Huang and Manton (2009) and Rajagopal and Wainwright (2008)). More precisely, the algorithm is the following. After initializing $\boldsymbol{x}(0) = \boldsymbol{\theta}$, iterate:

$$x_i(k+1) = \big(1-\mu(k)\big)P_{ii}x_i(k) + \mu(k)\sum_{j\in\mathcal{N}_i^-} P_{ij}\tilde{x}_{ij}(k)$$

where $\tilde{x}_{ij}(k) = x_j(k) + w_{ij}(k)$ is the version of $x_j(k)$ received by i, affected by noise, while $\mu(k) \in (0,1)$ are chosen to satisfy $\sum_{k\geq 0}\mu(k) = \infty$ and $\sum_{k\geq 0}\mu^2(k) < \infty$. Such algorithms were designed for channels were real numbers can be transmitted and are affected by an additive noise with zero-mean, and independent from past history as well as from other channels in the network. Under such assumptions, Huang and Manton (2009) and Rajagopal and Wainwright (2008) use techniques of stochastic approximation theory to prove convergence to consensus.

However, we might apply them also to our digital noisy networks, if we replace $\tilde{x}_{ij}(k)$ with the value $\hat{x}_{ij}(k)$ obtained after the process of encoding – transmitting over the channel from i to j – decoding, by some suitable coding scheme. What we want to compare is the strategy of increasing transmission lengths versus that of decreasing gains, where we plug into the decreasing gain algorithm a coding/encoding of fixed length $\bar{\tau}$, not varying with k.

In the example of implementation that we propose, we choose the same simple repetition-like coding scheme described above, with a transmission length $\bar{\tau} = 15$ and we use gains $\mu(k) = \frac{1}{k}$. We consider $N = 30$ agents, and a communication graph which is a strongly connected realization of a two-dimensional random geometric graph, where vertices are 30 points uniformly distributed in the unit square, and there is a pair of edges (i, j) and (j, i) whenever points i, j have a distance smaller than 0.4. The erasure probability on the links is $\varepsilon = 0.5$. The initial condition $\boldsymbol{\theta}$ is randomly chosen inside $[0, 1]^N$.

The communication graph is undirected, in the sense that $\mathcal{N}_i^- = \mathcal{N}_i^+$ for all $i \in \mathcal{V}$. So we choose to use, for both algorithms that we are comparing, a consensus matrix built according to the Metropolis weights rules for undirected graphs, illustrated in Xiao et al. (2005), which can be constructed distributedly, using only information on neighbors, as follows:

$$P_{ij} = \begin{cases} \dfrac{1}{1 + \max\{\deg(i), \deg(j)\}} & \text{if } (i, j) \in \mathcal{E} \\ 1 - \displaystyle\sum_{k \in \mathcal{N}_i^-} P_{ik} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

where $\deg(i)$ is the number of neighbors of node i.

Figures 1 and 2 show the decay of $\frac{1}{N}\mathbb{E}[\|\boldsymbol{z}(t)\|^2]$ and $\frac{1}{N}\mathbb{E}[\|\boldsymbol{e}(t)\|^2]$ respectively, with respect to the number of transmissions t. All our simulations show a similar behavior, where IPA algorithm significantly outperforms the decreasing gain strategy.

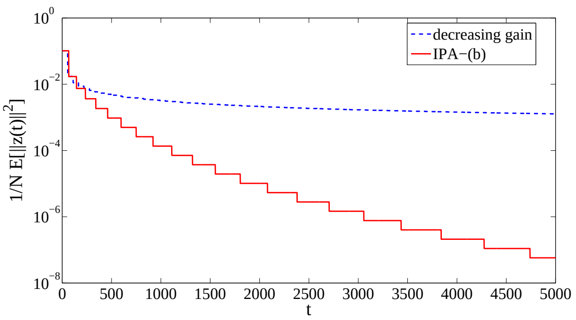

Fig. 1. $\frac{1}{N}\mathbb{E}[\|\boldsymbol{z}(t)\|^2]$ vs. number of transmissions t.

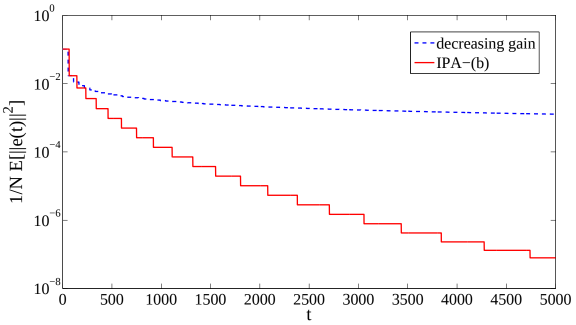

Fig. 2. $\frac{1}{N}\mathbb{E}[\|\boldsymbol{e}(t)\|^2]$ vs. number of transmissions t.

5. CONCLUSION

The main contribution of the paper has consisted in proposing a family of average consensus algorithms designed for digital networks, based on encoding/decoding schemes with precision increasing with time (IPA). Estimates of the effort required to achieve a prescribed precision have been given, in terms of both the number of transmissions and of the number of computations. We showed that the convergence time of (IPA) is polylogarithmic in the prescribed precision. The question is open whether a logarithmic algorithm can be designed for average consensus on digital networks.

ACKNOWLEDGEMENTS

The authors are grateful to Fabio Fagnani and Sandro Zampieri for their encouragement, as well as for many suggestions and discussions on the topics of this paper.

REFERENCES

Ayaso, O., Shah, D., and Dahleh, M. (2008). Information theoretic bounds for distributed computation. Subm.

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

Carli, R., Bullo, F., and Zampieri, S. (2009). Quantized average consensus via dynamic coding/decoding schemes. *Int. J. Robust and Nonlinear Control*. In press.

Como, G. and Dahleh, M. (2009). Lower bounds on the estimation error in problems of distributed computation. In *ITA Workshop*, 70–76. San Diego, CA.

Como, G., Fagnani, F., and Zampieri, S. (2008). Anytime reliable transmission of real-valued information through digital noisy channels. Subm.

Fagnani, F. and Zampieri, S. (2009). Average consensus with packet drop communication. *SIAM J. Control and Optimization*, 48(1), 102–133.

Frasca, P., Carli, R., Fagnani, F., and Zampieri, S. (2009). Average consensus on networks with quantized communication. *Int. J. Robust and Nonlinear Control*. In press.

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

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

Kar, S. and Moura, J.M.F. (2007). Distributed consensus algorithms in sensor networks: Quantized data. Subm.

Nedic, A., Olshevsky, A., Ozdaglar, A., and Tsitsiklis, J.N. (2007). On distributed averaging algorithms and quantization effects. Technical Report 2778, LIDS–MIT.

Rajagopal, R. and Wainwright, M.J. (2008). Network-based consensus averaging with general noisy channels. Technical Report 751, Dept. of Statistics, UC Berkeley.

Rajagopalan, S. and Schulman, L.J. (1994). A coding theorem for distributed computation. In *Annual ACM Symposium on Theory of Computing*, 790–799. Montréal, Québec, Canada.

Schulman, L.J. (1996). Coding for interactive communication. *IEEE Trans. on Inf. Th.*, 42(6), 1745–1756.

Xiao, L., Boyd, S., and Lall, S. (2005). A scheme for asynchronous distributed sensor fusion based on average consensus. In *Intern. Conf. on Information Processing in Sensor Networks (IPSN'05)*, 63–70. Los Angeles, CA.