Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Asymmetric Randomized Gossip Algorithms for Consensus

F. Fagnani * **S. Zampieri** **

* *Dipartimento di Matematica, Politecnico di Torino, C.so Duca degli Abruzzi, 24, 10129 Torino, Italy, (email: fabio.fagnani@polito.it).*
** *Dipartimento di Ingegneria dell'Informazione, Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy, (email: zampi@dei.unipd.it)*

Abstract: Distributed averaging is a relevant problem in several application areas, such as decentralized computation, sensor networks, clock synchronization and coordinated control of mobile autonomous agents. Gossip randomized consensus algorithms provide a particular simple and efficient solution of such a problem. These algorithms however need bidirectional communication among agents and this can be a rather restrictive hypothesis in some contexts. In this contribution we analyze two important examples of asymmetric randomized consensus algorithms which do not need bidirectional communication and exhibit a speed of convergence comparable to the symmetric gossip. However, differently from the symmetric gossip, these algorithms do not converge to the average. We complete our analysis showing that under rather mild hypotheses, the displacement of the their final state from the average goes to zero as the number of agents goes to infinity.

1. INTRODUCTION

In several application areas it is important to have an efficient algorithm able to compute the average of many quantities in a distributed way. Indeed, the distributed averaging problem appears in a number of different contexts since the early 80's (decentralized computation Tsitsiklis (1984), load balancing Cybenko (1989)) and, more recently, has attracted much attention for possible applications to sensor networks (data fusion problems Kempe et al. (2003); L. Xiao (2005), clock syncronization Li and Rus (2006)) and to coordinated control of mobile autonomous agents Jadbabaie et al. (2003); Olfati-Saber and Murray (2004). Precisely, the distributed averaging problem can be described as follows. Suppose we have a (directed) graph $\mathcal{G}$ with set of nodes $V = \{1, \dots, N\}$ and a measure x_i for every node $i \in V$. The average consensus problem consists in computing the average $x_A = N^{-1} \sum_i x_i$ in an iterative and distributed way, exchanging information among nodes exclusively along the available edges in $\mathcal{G}$.

Several algorithms for the distributed averaging have been proposed. Average consensus algorithm is one of them which is particularly convenient in terms of the amount of required communication and computation. Deterministic (time-invariant and time-varying) consensus algorithms have been studied in many papers. Starting from the pioneering work Tsitsiklis (1984), many variations can be found in above cited literature. Most of the papers study the same algorithm. Every node runs a first order linear dynamical system to update its estimation and the systems are coupled through the available communication edges. The problems typically considered in the literature concern necessary and sufficient conditions for convergence, speed of convergence, optimization issues. On the other hand, random linear schemes have been studied for instance in Kempe et al. (2003); Boyd et al. (2006); Dymakis et al. (2006) under the name of gossip algorithms. In this case the evolution low of the algorithm changes randomly at every clock step. Convergence is now considered in a probabilistic sense and performance is studied in mean square sense or in terms of a sort of contraction time. The algorithms studied in the literature assume symmetric communication graphs and lead in general to symmetric evolution matrices which preserve the global average over time. However, symmetry may be an undesirable feature in situations in which communication is asymmetric. Also, the related property of achieving the exact average can be relaxed in some contexts in which it may be sufficient to converge to some value close to the average.

In this paper we will focus on random consensus algorithms proposed in Boyd et al. (2006). However, differently from Boyd et al. (2006) we will not restrict to symmetric consensus algorithms. Indeed, we will consider two examples of asymmetric consensus algorithms. As in Fagnani and Zampieri (2008) we adopt a mean square analysis of the algorithm behavior. In this way we will show that, although these algorithms do not ensure convergence to the average, the displacement of their final state from the average goes to zero as the number of agents goes to infinity.

1.1 Mathematical preliminaries

In the sequel we will give some preliminary definitions and results concerning the notion of stochastic matrix. Here, for reasons which will be clear in the sequel, we will need to adopt an abstract approach.

Given any finite set S, $\mathbb{R}^S$ denotes the real vector space of functions form S to $\mathbb{R}$. If $x \in \mathbb{R}^S$ and $s \in S$, x_s denotes the component of x indexed by s. We assume $\mathbb{R}^S$ to be equipped with the canonical inner product: for $x, y \in \mathbb{R}^S$ we define $< x, y >:= \sum_s x_s y_s$.

978-1-1234-7890-2/08/$20.00 © 2008 IFAC

10.3182/20080706-5-KR-1001.2306

If S and T are finite sets and $\psi : \mathbb{R}^S \to \mathbb{R}^T$ is a linear mapping, then $\psi^* : \mathbb{R}^T \to \mathbb{R}^S$ denotes the adjoint with respect to the inner product $< \cdot, \cdot >$. Of course linear mappings like ψ can be naturally identified with matrices in $\mathbb{R}^{T\times S}$. This identification will be used whenever needed. In this way 'column' vectors $x \in \mathbb{R}^S$ can be interpreted as linear maps $x : \mathbb{R} \to \mathbb{R}^S$. The corresponding row vector as the adjoint $x^* : \mathbb{R}^S \to \mathbb{R}$.

The symbol $e_s \in \mathbb{R}^S$ denotes the vector with all elements equal to 0, except the s-th component equal to 1. We put $\mathbb{1}_S = \sum_s e_s$. Notice that, if we consider $\mathbb{R}^{S\times S}$, we have the relations: $e_{st} = e_s e_t^*$, $\mathbb{1}_{S\times S} = \mathbb{1}_S \mathbb{1}_S^*$. Whenever this will not create confusion, we will simply write $\mathbb{1}$ for $\mathbb{1}_S$.

A linear mapping $\psi : \mathbb{R}^S \to \mathbb{R}^S$ is said to be positive if for any $x \in \mathbb{R}^S$ such that $x_s \geq 0$ for any $s \in S$, it holds $\psi(x)_s \geq 0$ for any $s \in S$. It is said to be stochastic if it is positive and $\psi(\mathbb{1}_S) = \mathbb{1}_S$. A stochastic ψ is called doubly stochastic if in addition $\psi^*(\mathbb{1}_S) = \mathbb{1}_S$.

If ψ is stochastic, there always exists a probability vector $\pi \in \mathbb{R}^S$ ($\pi_s \geq 0$ for any s and $\sum_s \pi_s = 1$) such that $\psi^*\pi = \pi$. The vector π is said to be an invariant probability for ψ. Of course if it is doubly stochastic, a possible invariant probability vector is $\pi = |S|^{-1}\mathbb{1}_S$, where $|S|$ means the cardinality of the set S.

A stochastic linear mapping ψ is said to be aperiodic if it admits just one probability vector π and it holds:

$$\lim_{t\to+\infty} (\psi^*)^t \xi = \pi \,,$$

for any probability vector $\xi \in \mathbb{R}^S$. This is equivalent to ask that 1 is a simple eigenvalue for ψ and that all other eigenvalues have norm strictly smaller than 1.

Given any linear mapping ψ we will denote by $\mathrm{sr}(\psi)$ the spectral radius of ψ (the max-norm of the eigenvalues). If ψ is stochastic aperiodic, $\mathrm{esr}(\psi)$ will denote the essential spectral radius, namely, the max norm of the eigenvalues different from 1.

2. GOSSIP CONSENSUS ALGORITHMS

Suppose we have N agents labelled by the elements of a set V with $|V| = N$. Each agent $i \in V$ makes a measure $x_i \in \mathbb{R}$. The goal of every agent is to evaluate the global average $x_A = N^{-1}\sum_i x_i$. We assume that no supervision and no leader is available, so that x_A has to be evaluated by exchanging information among the various agents. We assume that communication among agents is possible according to a given communication graph. In other words, we fix a directed graph $\mathcal{G} = (V, E)$, where $E \subseteq V \times V$ denotes the set of edges. If $(i, j) \in E$ it means that agent i can send data to agent j. We will always assume that $\mathcal{G}$ is strongly connected (for any two vertices $i, j \in V$ there always exists a path in $\mathcal{G}$ connecting i to j). It is rather clear that, without this assumption, no algorithm can succeed to make all agents evaluate x_A. Finally, a matrix $W \in \mathbb{R}^{V\times V}$ is said to be *adapted* to $\mathcal{G}$ if $W_{ij} = 0$ whenever $(j, i) \notin E$.

Various strategies have been proposed in the literature for the solution of this problem. In this paper will deal with the so-called randomized consensus algorithms or randomized gossip algorithms. They consist in random exchange of information among agents followed by a local averaging. The have been studied in many papers Muthukrishnan et al. (1998); Boyd et al. (2006). Their main interest is due to the low complexity and low computation profile. The classical gossip algorithm presented in Boyd et al. (2006) requires an undirected communication graph (e.g. $(i, j) \in E$ iff $(j, i) \in E$) and a symmetric communication protocol. This is briefly recalled below.

Example 1. **The symmetric gossip** Fix a real number $q \in (0, 1)$, a strongly connected undirected graph $\mathcal{G} = (V, E)$ and a symmetric matrix $W \in \mathbb{R}^{V\times V}$ positive ($W_{ij} \geq 0$) and adapted to $\mathcal{G}$ such that $\mathbb{1}^*W\mathbb{1} = 1$. At every time instant t the edge (j, i) is activated with probability W_{ij} and nodes i and j exchange their states and produce a new states according to the equations

$$x_i(t+1) = (1-q)x_i(t) + qx_j(t)$$
$$x_j(t+1) = qx_i(t) + (1-q)x_j(t)$$

The other states remain unchanged. The evolution obtained by iterating these equations starting form the initial condition $x_i(0) = x_i$ is described by a linear random dynamical system which is known Boyd et al. (2006); Fagnani and Zampieri (2008) to yield probabilistic average consensus, namely, for every $i \in V$, it holds

$$x_i(t) \to x_A \,, \quad \text{almost surely} \tag{1}$$

The second example is a randomized gossip algorithm which does not require the communication graph to be directed and a symmetric communication protocol.

Example 2. **The asymmetric-gossip** In this case we start from a real number $q \in (0, 1)$, a strongly connected directed graph $\mathcal{G} = (V, E)$ and a matrix $W \in \mathbb{R}^{V\times V}$ with nonnegative entries adapted to $\mathcal{G}$ such that $\mathbb{1}^*W\mathbb{1} = 1$. At every time instant t the edge (j, i) is activated with probability W_{ij} and node j sends its state to i and i produces a new state according to the equation

$$x_i(t+1) = (1-q)x_i(t) + qx_j(t)$$

The other states remains unchanged. As before we consider the initial condition $x_i(0) = x_i$. This algorithm yields probabilistic consensus Fagnani and Zampieri (2008), but does not converge to the average. Namely, there exists a random variable α such that for every $i \in V$,

$$x_i(t) \to \alpha \,, \quad \text{almost surely} \,. \tag{2}$$

We add here another example of an asymmetric algorithm. This is motivated by the need that occur in some practical implementations of parallelizing a number of the gossip averaging steps in order to make use of the power of the large scale network.

Example 3. **The syncronous asymmetric gossip** We start from a real number $q \in (0, 1)$, a strongly connected directed graph $\mathcal{G} = (V, E)$ and a stochastic matrix $W \in \mathbb{R}^{V\times V}$ adapted to $\mathcal{G}$. At every time instant t the N edges $(j_i, i) \in E$ for $i = 1, 2, \ldots, N$ are activated each with probability W_{i,j_i}. For each i, the node j_i sends its state to i which produces a new state according to the equation

$$x_i(t+1) = (1-q)x_i(t) + qx_{j_i}(t)$$

This algorithm also yields probabilistic consensus Fagnani and Zampieri (2008), but does not converge to the average.

The issues we want to address in this paper regard the speed of convergence in all the examples we introduced and the distance between the consensus point α and the average x_A in the asymmetric cases.

2.1 A unified approach

The three algorithms can be described in a unified way. They are both iterative algorithms producing for every agent i a sequence of values $x_i(t) \in \mathbb{R}$ (which will be called the state of the i-th agent at time t). If we assemble the various $x_i(t)$ in a vector $x(t) \in \mathbb{R}^V$, the evolution of the three algorithms can be expressed in the form

$$x(t+1) = P(t)x(t)\,, \tag{3}$$

where $P(t) \in \mathbb{R}^{V\times V}$is a sequence of i.i.d. matrix valued random variables such that $P(t)_{ij} \geq 0$, $P(t)\mathbb{1} = \mathbb{1}$ (in other words each realization of $P(t)$ is a stochastic matrix). The initial condition $x(0) \in \mathbb{R}^V$ coincides with the vector of measures $x_i(0) = x_i$ for every $i \in V$. The solution $x(t)$ is thus a stochastic process.

The statistical description of the matrices $P(t)$ in the three examples is quite simple. In the symmetric gossip case we let, for every $(j,i) \in E$,

$$S^{ij} = I - q(e_i - e_j)(e_i - e_j)^*\,,$$

Moreover, $P(t)$ is concentrated on these matrices and

$$\mathbb{P}[P(t) = R^{ij}] = W_{ij}\,.$$

Instead, in the asymmetric case, for every $(j,i) \in E$, we let

$$A^{ij} = I - qe_i(e_i - e_j)^*\,,$$

and

$$\mathbb{P}[P(t) = A^{ij}] = W_{ij}\,.$$

Finally, in the synchronous asymmetric case we fix for every $\mathbf{j} = (j_1,\dots,j_N) \in V^N$ the matrix

$$R^{\mathbf{j}} = (1-q)I + q\sum_i e_i e_{j_i}^*$$

and we put

$$\mathbb{P}[P(t) = R^{\mathbf{j}}] = \prod_{i=1}^{N} W_{i,j_i}$$

Any algorithm like (3) described by a sequence of random stochastic matrices $P(t)$ is said to achieve *(average) probabilistic consensus* if there exists a random variable α ($\alpha = x_A$) such that

$$x(t) \to \alpha\mathbb{1}\,, \quad \text{almost surely}\,. \tag{4}$$

Since $P(t)\mathbb{1} = \mathbb{1}$, such algorithms have the property that if $x(0) = \alpha\mathbb{1}$, for some $\alpha \in \mathbb{R}$, then $x(t) = x(0)$ for every $t \in \mathbb{N}$. Once consensus is achieved, agents do not change anymore their states. If $P(t)$ is doubly stochastic for all t, we also have that $\mathbb{1}^*P(t) = \mathbb{1}^*$ for any t. In this case consensus automatically implies average consensus. This is what happens in the symmetric gossip algorithm.

Let

$$Q(t) = P(t-1)\cdots P(0)\,, \tag{5}$$

so that we can write $x(t) = Q(t)x(0)$. The random variable α in (4) is a linear function of the initial condition $x(0)$ so that we can write $\alpha = \rho^*x(0)$ for some random variable ρ taking values in $\mathbb{R}^V$ and such that $\mathbb{1}^*\rho = 1$. Therefore probabilistic consensus can be equivalently expressed by saying that there exists a random variable ρ taking values in $\mathbb{R}^V$ such that

$$\lim_{t\to\infty} Q(t) = \mathbb{1}\rho^*\,, \quad \text{almost surely}\,. \tag{6}$$

Notice that $\mathbb{1}\rho^*$ is a matrix whose rows are all equal to ρ^* and that $x(\infty) = \rho^*x(0)$. Hence, we have probabilistic average consensus exactly when $\rho = N^{-1}\mathbb{1}$ almost surely.

2.2 Performance indices

In this paper, we will evaluate the performance of the algorithm $P(t)$ by considering two indices. The first index we consider is a normalized version of the *distance from the consensus*

$$d(t) = \frac{1}{N}||x(t) - \mathbb{1}\rho^*x(0)||^2$$

The second one is the *average asymptotic displacement* from its initial value

$$\beta = |\rho^*x(0) - x_A|^2$$

Of course in those situations in which $P(t)$ is doubly stochastic, we have that $\beta = 0$. Notice moreover that

$$\frac{1}{N}||x(t) - \mathbb{1}x_A||^2 = d(t) + \beta$$

which shows that the evolution of $d(t)$ and β determines the evolution of $\frac{1}{N}||x(t) - \mathbb{1}x_A||^2$. This coincides with the average distance between $x_i(t)$ and x_A and so it is the most important error parameter to minimize.

In the sequel we will work out a mean analysis for the two quantities defined above.

3. MEAN ANALYSIS

In this section we assume that we have a fixed random algorithm $P(t)$ achieving probabilistic consensus so that (6) is satisfied with a suitable ρ. Here we will focus on a mean analysis. In other terms we will study the mean of the random variable $d(t)$ and β. More refined probabilistic considerations can be carried on (see for instance Fagnani and Zampieri (2008)). In the sequel we first assume that $x(0)$ is a fixed initial condition. However, in the final part of this section, we will make the analysis assuming that also $x(0)$ has a probabilistic distribution: the mean analysis will consequently also be with respect to $x(0)$.

We are interested in studying $\mathbb{E}[d(t)]$ and, in particular, its exponential rate of convergence. In this section $\mathbb{1}$ we always denote $\mathbb{1}_V \in \mathbb{R}^V$. We also let $\Omega := I - N^{-1}\mathbb{1}\mathbb{1}^*$.

We introduce an operator which will play a fundamental role in the sequel. Let $\mathcal{L} : \mathbb{R}^{V\times V} \to \mathbb{R}^{V\times V}$ be given by

$$\mathcal{L}(M) = \mathbb{E}[P(0)^*MP(0)]$$

In Fagnani and Zampieri (2008) a detailed analysis of this operator has been carried on. It is easy to verify that $\mathcal{L}^* : \mathbb{R}^{V\times V} \to \mathbb{R}^{V\times V}$ is given by $\mathcal{L}^*(M) = \mathbb{E}[P(0)MP(0)^*]$. Notice that

$$\mathcal{L}^*(\mathbb{1}\mathbb{1}^*) = \mathbb{E}[P(0)\mathbb{1}\mathbb{1}^*P(0)^*] = \mathbb{1}\mathbb{1}^*\,.$$

It can be shown that both $\mathcal{L}$ and $\mathcal{L}^*$ are positive operators. Therefore $\mathcal{L}^*$ is a stochastic operator. Other properties are recalled below.

Proposition 4. (i) If P achieves probabilistic consensus, then $\mathcal{L}$ is aperiodic and $\mathbb{E}[\rho\rho^*]$ is its eigenvector relative to the eigenvalue 1, namely $\mathcal{L}(\mathbb{E}[\rho\rho^*]) = \mathbb{E}[\rho\rho^*]$.
(ii) If $A \leq B$ (namely $B - A$ is a positive semidefinite matrix), then $\mathcal{L}(A) \leq \mathcal{L}(B)$.
(iii) $\mathcal{L}^t(\Omega) \leq ||\mathcal{L}(\Omega)||^t\Omega$, where $||\cdot||$ is the matrix induced 2-norm.

We can rewrite $\mathbb{E}[d(t)]$ in terms of $\mathcal{L}$, as follows.

Proposition 5.

$$\mathbb{E}[d(t)] = x^*(0)\Delta(t)x(0)$$

where

$$\Delta(t) = \mathcal{L}^t\Big(N^{-1}I - N^{-1}\mathbb{1}\mathbb{E}[\rho^*] - N^{-1}E[\rho]\mathbb{1}^* + \mathbb{E}[\rho\rho^*]\Big).$$

Proof. Notice that

$$\begin{aligned}&\mathbb{E}[d(t)] = N^{-1}\mathbb{E}[(x^*(t) - x(0)^*\rho\mathbb{1}^*)(x(t) - \mathbb{1}\rho^*x(0))]\\ &= x(0)^*[N^{-1}\mathbb{E}[Q(t)^*Q(t)] - N^{-1}\mathbb{E}[\rho\mathbb{1}^*Q(t)]\\ &-N^{-1}\mathbb{E}[Q(t)^*\mathbb{1}\rho^*] + \mathbb{E}[\rho^*\rho]]x(0)\end{aligned}$$

A simple recursive argument shows that $\mathbb{E}[Q(t)^*Q(t)] = \mathcal{L}^t(I)$. On the other hand, using Lebesgue dominated convergence and similar recursive arguments, we can write

$$\begin{aligned}\mathbb{E}[\rho\mathbb{1}^*Q(t)] &= \lim_{s\to+\infty}\mathbb{E}[Q(s)^*Q(t)]\\ &= \lim_{s\to+\infty}\mathbb{E}[Q(t)^*\prod_{r=t}^{s-1}P(r)^*Q(t)]\\ &= \lim_{s\to+\infty}\mathcal{L}^t(\mathbb{E}[\prod_{r=t}^{s-1}P(r)^*]) = \mathcal{L}^t(\mathbb{1}\mathbb{E}[\rho^*]).\end{aligned}$$

This yields the result.

From this result it is possible to determine the exponential rate of convergence of $\mathbb{E}[d(t)]$. Let

$$S = \{\Delta \in \mathbb{R}^{V\times V} \ : \Delta \text{ symmetric and } \Delta\mathbb{1} = 0\}$$

We have to following result.

Corollary 6.

$$\sup_{x(0)}\lim_{t\to+\infty}\mathbb{E}[d(t)]^{1/t} = \mathrm{sr}(\mathcal{L}_{|S}),$$

where, we recall $\mathrm{sr}(\cdot)$ means the spectral radius of a linear operator and where $\mathcal{L}_{|S}$ means the restriction of $\mathcal{L}$ to S.

Proof. $\leq$ immediately follows from Proposition 5.

For proving $\geq$ notice that, from the identity $\Omega x(t) = \Omega(x(t) - \mathbb{1}\rho^*x(0))$, it follows that

$$\begin{aligned}&\mathbb{E}[d(t)] = N^{-1}\mathbb{E}[(x^*(t) - x(0)^*\rho\mathbb{1}^*)(x(t) - \mathbb{1}\rho^*x(0))]\\ &\geq N^{-1}\mathbb{E}[(x^*(t) - x(0)^*\rho\mathbb{1}^*)\Omega(x(t) - \mathbb{1}\rho^*x(0))]\\ &= N^{-1}\mathbb{E}[(x(t)^*\Omega x(t)] = N^{-1}x(0)^*\mathcal{L}^t(\Omega)x(0)\end{aligned}$$

Assume now that $\Delta \in S$ is such that $\mathcal{L}(\Delta) = \lambda\Delta$. Then there exists $c \in \mathbb{R}^+$ such that $\Delta \leq c\Omega$ and $x(0) \in \mathbb{R}^N$ such that $\Delta x(0) \neq 0$. This implies (using (ii) of Proposition 4) that $\mathcal{L}^t(\Delta) \leq c\mathcal{L}^t(\Omega)$ and so

$$\begin{aligned}&c\mathbb{E}[d(t)] \geq cN^{-1}x(0)^*\mathcal{L}^t(\Omega)x(0)\\ &\geq N^{-1}x(0)^*\mathcal{L}^t(\Delta)x(0) = N^{-1}x(0)^*\Delta x(0)|\lambda|^t\end{aligned}$$

Hence $\lim_{t\to+\infty}\mathbb{E}[d(t)]^{1/t} \geq |\lambda|$. This concludes the proof.

The spectral radius $\mathrm{sr}(\mathcal{L}_{|S})$ is thus an important performance parameter. In Fagnani and Zampieri (2008) it is shown that

$$\mathrm{esr}(\overline{P})^2 \leq \mathrm{sr}(\mathcal{L}_{|S}) \leq \mathrm{sr}(\mathcal{L}(\Omega)), \tag{7}$$

where $\overline{P} := \mathbb{E}[P(0)]$ and where $\mathrm{esr}(\cdot)$ means the essential spectral radius of a stochastic linear operator, namely the second maximum absolute value of its eigenvalues.

For what concerns β, it is easy to see that

$$\mathbb{E}[\beta] = x(0)^*Bx(0)$$

where

$$B = \mathbb{E}[\rho\rho^*] - N^{-1}\mathbb{E}[\rho]\mathbb{1}^* - N^{-1}\mathbb{1}\mathbb{E}[\rho]^* + N^{-2}\mathbb{1}\mathbb{1}^* \tag{8}$$

Notice that $B\mathbb{1} = 0$ and that for all x orthogonal to $\mathbb{1}$ we have that $x^*Bx = x^*\mathbb{E}[\rho\rho^*]x$. Let $c = ||\mathbb{E}[\rho]||_\infty$ the infinity norm of $\mathbb{E}[\rho]$. It is immediate to check that $c^{-1}\mathbb{E}[\rho\rho^*]$ is symmetric and sub-stochastic (it is positive and each row sum does not exceed 1). Hence $x^*\mathbb{E}[\rho\rho^*]x \leq c||x||^2$. This implies that $B \leq c\Omega$ and thus also

$$\mathbb{E}[\beta] \leq c||\Omega x||^2 \tag{9}$$

From this estimation of B we can actually also obtain a stronger upper bound on $\mathbb{E}[d(t)]$. Notice indeed that $\Delta(0) = N^{-1}\Omega + B \leq (N^{-1} + c)\Omega$. Using (ii) and (iii) of Proposition 4, we obtain that

$$\Delta(t) \leq (N^{-1} + c)||\mathcal{L}(\Omega)||^t\Omega \tag{10}$$

and so

$$\mathbb{E}[d(t)] \leq (N^{-1} + c)||\mathcal{L}(\Omega)||^t||x(0)||^2 \tag{11}$$

Both bounds (9) and (11) are in terms of c and $||\mathcal{L}(\Omega)||$. For what concerns c, there is an important class of algorithms for which it can be explicitly computed. Notice indeed that $\overline{P} = \mathbb{E}[P(0)]$ is stochastic and using Lebesgue dominated convergence theorem, we have

$$\overline{P}^t = \mathbb{E}[Q(t)] \to \mathbb{1}\mathbb{E}[\rho^*].$$

Hence, $\overline{P}$ is also an aperiodic stochastic matrix with the unique invariant probability vector given by $\mathbb{E}[\rho]$. The case when $\overline{P}$ is doubly stochastic is particularly important, since in this case $\mathbb{E}[\rho] = N^{-1}\mathbb{1}$. This happens of course in the symmetric gossip case but also in many cases of the asymmetric gossip settings, as we will see later on.

From now on we will always assume that $\overline{P}$ is doubly stochastic and so $\mathbb{E}[\rho] = N^{-1}\mathbb{1}$. In this case we have that

$$B = \mathbb{E}[\rho\rho^*] - N^{-2}\mathbb{1}\mathbb{1}^* \tag{12}$$

Moreover, since $c = N^{-1}$, previous estimations become

$$\mathbb{E}[\beta] \leq N^{-1}||\Omega x||^2 \tag{13}$$

$$\mathbb{E}[d(t)] \leq 2N^{-1}||\mathcal{L}(\Omega)||^t||x(0)||^2 \tag{14}$$

We assume now that also the initial conditions $x_i(0)$ are random variables. In the assumption that are all identically distributed with mean m and variance σ^2 (and, of course, independent, of the randomness of the gossip algorithm), we easily obtain

$$\mathbb{E}[d(t)] = \mathbb{1}^*\Delta(t)\mathbb{1}m^2 + \mathrm{trace}(\Delta(t))\sigma^2 = \mathrm{trace}(\Delta(t))\sigma^2$$

where last equality follows from the fact that $\Delta(t)\mathbb{1} = 0$. This (using (10)) leads to the bound

$$\mathbb{E}[d(t)] \leq 2||\mathcal{L}(\Omega)||^t\sigma^2. \tag{15}$$

Similarly, it is not difficult to show that in this case

$$\mathbb{E}[\beta] = \mathrm{trace}(B)\sigma^2 = \mathbb{E}||\rho - N^{-1}\mathbb{1}||^2 \tag{16}$$

Notice that the inequality $B \leq N^{-1}\Omega$ yields $\mathrm{trace}(B) \leq N^{-1}(N-1) \leq 1$. Hence, $\mathbb{E}[\beta] \leq \sigma^2$. This bound says that the average displacement never exceeds the variance noise σ^2 independently on the number of agents. We will see that in many specific situations we will be able to improve a lot this bound showing that actually in these cases $\mathbb{E}[\beta]$ converges to 0 when N goes to ∞.

4. ANALYSIS OF THE ASYNCHRONOUS ASYMMETRIC GOSSIP ALGORITHM

Notice first that since, for all $i \neq j$ we have $\overline{P}_{ij} = qW_{ij}$, it follows that

$$\mathbb{1}^*\overline{P} = \mathbb{1}^* \;\Leftrightarrow\; W\mathbb{1} = W^*\mathbb{1}\,.$$

In particular this is true if W is symmetric (actually in this case $\overline{P}$ is symmetric as well).

More generally, if we start from a directed graph $\mathcal{G} = (V, E)$ possessing the property that, for every node i,

$$|\{j \in V \mid (i,j) \in E\}| = |\{j \in V \mid (j,i) \in E\}|$$

a possible simple choice of a W adapted to $\mathcal{G}$ such that $W\mathbb{1} = W^*\mathbb{1}$ is as follows. Let A be the adjacency matrix of the graph, then $A\mathbb{1} = A^*\mathbb{1}$. Then we put

$$W := (\mathbb{1}^*A\mathbb{1})^{-1}A^*$$

In this way we make all the allowed matrices equally likely. This, in particular, shows that the assumption that $\overline{P}$ is doubly stochastic does cover interesting cases even for asymmetric gossip.

We now study in detail the operator $\mathcal{L}$ in the case of asymmetric gossip algorithms. The only assumption on W is that $W\mathbb{1} = W^*\mathbb{1}$. First, we find explicit expressions for the operator $\mathcal{L}$. The proof is very long and so we omit it due to space limitations.

Proposition 7.

$$\begin{aligned}&\mathcal{L}(\Delta) = \Delta - q(\text{diag }(W\mathbb{1}\mathbb{1}^*)\Delta - \Delta\text{diag }(W\mathbb{1}\mathbb{1}*)) +\\ &+q(W^*\Delta + \Delta W)q^2\text{diag }(W\mathbb{1}\mathbb{1}^*)\text{diag }(\Delta) +\\ &+q^2[\text{diag }(\mathbb{1}\mathbb{1}^*\text{diag }(\Delta)W) - W^*\text{diag }(\Delta) - \text{diag }(\Delta)W]\end{aligned}$$

The following fact is rather easy to prove.

Proposition 8.

$$\begin{aligned}\mathcal{L}(I) &= I + q(1-q)(W + W^* - 2\text{diag }(W\mathbb{1}\mathbb{1}^*))\\ \mathcal{L}(\mathbb{1}\mathbb{1}^*) &= \mathbb{1}\mathbb{1}^* - q^2(W + W^* - 2\text{diag }(W\mathbb{1}\mathbb{1}^*))\end{aligned}$$

This yields the following interesting result:

Corollary 9. It holds:

$$\mathbb{E}[\rho\rho^*] = \frac{1}{qN + (1-q)N^2}[qI + (1-q)\mathbb{1}\mathbb{1}^*]$$

Using now (16), (12), and Corollary 9, we can evaluate exactly $\mathbb{E}[\beta]$:

Corollary 10. It holds:

$$\mathbb{E}[\beta] = \frac{q(N-1)}{(qN + (1-q)N^2)N}\sigma^2$$

Notice that $\mathbb{E}[\beta]$ is infinitesimal in N. Moreover, the formula above only depends on N and q and not at all on the particular structure of W.

The behavior of $\mathbb{E}[d(t)]$ can be studied through the estimation (15). We now evaluate $\mathcal{L}(\Omega)$ and in particular its norm. From Proposition 8 we obtain

$$\mathcal{L}(\Omega) = \Omega - 2q[1-q-qN^{-1}](\text{diag }(W\mathbb{1}\mathbb{1}^*) - (W+W^*)/2)\,.$$

Clearly, $\mathcal{L}(\Omega)\mathbb{1} = 0$. On the other hand, on the invariant subspace orthogonal to $\mathbb{1}$, $\mathcal{L}(\Omega)$ is equal to

$$\mathcal{L}(\Omega) = I - 2q[(1-q) - qN^{-1}](\text{diag }(W\mathbb{1}\mathbb{1}^*) - (W+W^*)/2)\,.$$

The matrix diag $(W\mathbb{1}\mathbb{1}^*) - (W+W^*)/2$ is positive definite. If we denote by μ its smallest nonzero eigenvalue, we obtain, for N sufficiently large,

$$\text{sr}(\mathcal{L}(\Omega)) = 1 - 2q[(1-q) - qN^{-1}]\mu\,.$$

In the special case when W is symmetric μ is the smallest nonzero eigenvalue of diag $(W\mathbb{1}\mathbb{1}^*) - W$. On the other hand, since $\overline{P} = qW + (1 - q\text{diag }(W\mathbb{1}\mathbb{1}^*))$, it follows that

$$\text{esr}(\overline{P}) = 1 - q\mu\,.$$

Using (7) we have that, in this case, when μ tends to zero, that both the lower bound upper bound of $\text{sr}(\mathcal{L}_{|S})$ are linear in μ but with different constants.

Example 11. Consider that case in which W is an $N \times N$ symmetric circulant matrix Davis (1979) with first row equal to

$$[0\ 1/2N\ 0\ \cdots\ 0\ 1/2N]$$

In this case diag $(W\mathbb{1}\mathbb{1}^*) - W = N^{-1}I - W$. Its smallest nonzero eigenvalue can be explicitly computed in this case and we obtain

$$\mu = \frac{1}{N}\left(1 - \cos\frac{2\pi}{N}\right) \simeq \frac{2\pi^2}{N^3}$$

which leads to the approximation

$$\text{sr}(\mathcal{L}(\Omega)) \simeq 1 - \frac{4\pi^2 q(1-q)}{N^3}$$

5. ANALYSIS OF THE SYNCHRONOUS ASYMMETRIC GOSSIP ALGORITHM

First notice that in this case

$$\overline{P} = \sum_{\mathbf{j}}\prod_{i=1}^{N} W_{i,j_i}R^{\mathbf{j}} = (1-q)I + qW$$

Hence, $\overline{P}$ is doubly stochastic if and only if W is doubly stochastic. We will make this assumption from now on.

There is an explicit expression for $\mathcal{L}$.

Proposition 12.

$$\begin{aligned}\mathcal{L}(\Delta) = {}&[(1-q)I + W]^*\Delta[(1-q)I + W] +\\ &+ q^2[\text{diag }(\mathbb{1}\mathbb{1}^*\text{diag }(\Delta)W) - W^*\text{diag }(\Delta)W]\end{aligned}$$

Unfortunately, in this case it is not possible to compute explicitly the matrices $\mathbb{E}[\rho\rho^*]$ and B. However a useful estimate can be obtained in the following way.

For what concerns $\mathcal{L}(\Omega)$, we obtain

$$\mathcal{L}(\Omega) = \Omega - q(1-q)[2I - W - W^*] - q^2N^{-1}[I - WW^*]\,.$$

In the special case when W is symmetric, we obtain that the eigenvalues of $\mathcal{L}(\Omega)$ are

$$f(\lambda) = 1 + 2q(1-q)(\lambda - 1) + q^2N^{-1}(\lambda^2 - 1)$$

where λ are the eigenvalues of W. Hence, we have that, if N is sufficiently large,

$$\text{sr}(\mathcal{L}(\Omega)) = f(\mu^+(W))$$

where $\mu^+(W)$ denotes the maximum of the eigenvalues of W. On the other hand, since $\overline{P} = (1-q)I + qW$, it follows that (when $\mu^+(W)$ is sufficiently close to 1),

$$\text{esr}(\overline{P}) = 1 - q + q\mu^+(W)\,.$$

Similarly to previous section, this allows to conclude that, using (7), when $\mu^+(W)$ tends to zero, both the lower bound upper bound of $\mathrm{sr}(\mathcal{L}_{|S})$ are linear in $1-\mu^+(W)$ but with different constants.

Let $Z := \mathcal{L}(\mathbb{1}\mathbb{1}^*) - \mathbb{1}\mathbb{1}^*$. Notice that $Z \geq 0$ in the sense that Z is a positive semidefinite matrix. Notice moreover that the operator $\mathcal{L}$ preserve the order between matrices given by the positive semidefinite condition. More precisely we have that $A \leq B$ implies that $\mathcal{L}(A) \leq \mathcal{L}(B)$. This implies that $\mathcal{L}^t(Z) \geq 0$ for all t. Observe moreover that

$$\begin{aligned} B &= \mathbb{E}[\rho\rho^*] - N^{-2}\mathbb{1}\mathbb{1}^* = \\ &= \mathbb{E}[\rho\rho^*] - N^{-2}\mathcal{L}^{t+1}(\mathbb{1}\mathbb{1}^*) + N^{-2}\sum_{i=0}^{t}\mathcal{L}^i(Z) \\ &\leq N^{-2}\sum_{i=0}^{\infty}\mathcal{L}^i(Z) \end{aligned}$$

Finally notice that from the fact that $Z \leq q^2\Omega$, it follows that $\mathcal{L}^t(Z) \leq q^2||\mathcal{L}(\Omega)||^t\Omega$ and hence we have that

$$B \leq N^{-2}\sum_{i=0}^{\infty}\mathcal{L}^i(Z) \leq q^2N^{-2}\sum_{i=0}^{\infty}||\mathcal{L}(\Omega)||^t\Omega \leq \frac{q^2}{gN}\frac{1}{N}\Omega$$

where $g := 1 - \mathrm{sr}(\mathcal{L}(\Omega))$. Notice that in many situations g goes to zero as N tends to infinity (as $1 - \mathrm{esr}(W)$ if W is symmetric). Only in case in which g tends to zero less quickly then $1/N$ we have that the previous bound is more tight than the general bound $B \leq N^{-1}\Omega$ we found above. In fact when g tends to zero less quickly then $1/N$, since

$$\mathbb{E}[\beta] = \sigma^2\mathrm{trace}(B) \leq \sigma^2\frac{q^2}{gN}$$

then $\mathbb{E}[\beta]$ goes to zero as N tends to infinity. These situations will be illustrated by the following examples.

Example 13. Consider that case in which W is a $N \times N$ symmetric circulant matrix Davis (1979) with first row equal to

$$[0 \ 1/2 \ 0 \ \cdots \ 0 \ 1/2]$$

The corresponding graph has a circular structure. We obtain that the max eigenvalue of W is given by

$$\mu^+(W) = \cos\frac{2\pi}{N}$$

For $N \to +\infty$, this leads to the approximation (valid for N sufficiently large)

$$\mathrm{sr}(\mathcal{L}(\Omega)) \simeq 1 - \frac{4\pi^2q(1-q)}{N^2}, \quad g \simeq \frac{4\pi^2q(1-q)}{N^2}$$

We can argue that in this case the bound proposed above does not ensure that $\mathbb{E}[\beta]$ goes to zero as N tends to infinity.

Example 14. Consider now is a $N \times N$ matrix $W \in \mathbb{R}^{H\times H}$ where $H = \{0,1\}^n$ such that

$$W_{ij} = \begin{cases} n^{-1} & \text{if } i \text{ and } j \text{ differs in one digit} \\ 0 & \text{otherwise} \end{cases}$$

In this case we have that $N = |H| = 2^n$. Also in this case the eigenvalues of W can be computed explicitly. It is easy to see that we obtain $\mu^+(W) = n^{-1}(n-2)$. Hence

$$\mathrm{sr}(\mathcal{L}(\Omega)) \simeq 1 - \frac{4q(1-q)}{n}, \quad g \simeq \frac{4q(1-q)}{n}$$

Therefore, since $gN \simeq 4q(q-1)n^{-1}2^n$ tends to infinity as N tends to infinity, we can ague that in this case the bound proposed above ensures that $\mathbb{E}[\beta]$ goes to zero as N tends to infinity.

Notice that, by using the structure of the matrices in the example it is possible to show that in both cases $\mathbb{E}[\beta]$ goes to zero as N tends to infinity. The fact that this does not appear in Example 13 shows that the bound we proposed is not tight in general.

REFERENCES

S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. *Journal of parallel and distributed computing*, 7:279–301, 1989.

P. J. Davis. *Circulant matrices.* A Wiley-Interscience Publication, Pure and Applied Mathematics. John Wiley & Sons, New York-Chichester-Brisbane, 1979.

A. Dymakis, A.D. Sarwate, and M.J. Wainwright. Geographic gossip: efficient aggregation on sensor networks. Technical report, EECS Berkeley, 2006.

F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. *IEEE J. Selected Areas Comm.*, 2008. to appear.

A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 1–10, 2003.

S. Lall L. Xiao, S. Boyd. A scheme for robust distributed sensor fusion based on average consensus. In *Proc. Int. Conf. Information Processing in Sensor Networks*, 2005.

Q. Li and D. Rus. Global clock syncronization in sensor networks. *IEEE Transaction on Computers*, 55(2):214–226, 2006.

S. Muthukrishnan, B. Ghosh, and M. Schultz. First and second order diffusive methods for rapid, coarse, distributed load balancing. *Theory of computing systems*, 31:331–354, 1998.

R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9): 1520–1533, 2004.

J. Tsitsiklis. *Problems in decentralized decision making and computation.* PhD thesis, Department of EECs, MIT, 1984.