Convergence in Multiagent Coordination, Consensus, and Flocking

Vincent D. Blondel, Julien M. Hendrickx, Alex Olshevsky, and John N. Tsitsiklis

***Abstract*— We discuss an old distributed algorithm for reaching consensus that has received a fair amount of recent attention. In this algorithm, a number of agents exchange their values asynchronously and form weighted averages with (possibly outdated) values possessed by their neighbors. We overview existing convergence results, and establish some new ones, for the case of unbounded intercommunication intervals.**

I. Introduction

We consider a set $N = \{1, \ldots, n\}$ of agents that try to reach agreement on a common scalar value by exchanging tentative values and combining them by forming convex combinations. The motivation for such a scheme comes from a variety of contexts involving distributed systems. For example, a number of sensors may wish to combine individual estimates of a certain variable or form an aggregate statistic; or a number of vehicles may wish to align their directions of motion through interaction with their neighbors.

The "agreement algorithm" considered here and its original analysis is due to Tsitsiklis et al. [14]. The complete proof is in [13], and a simplified version is presented in the text [3]. A related algorithm was later proposed by Vicsek et al. [15], as a model of cooperative behavior. The subject has attracted considerable recent interest, within the context of flocking and multiagent coordination ([8], [4], [11], [9], [1], [12], [10]). A further special case, concerns the computation of the exact average of the agents' values (as opposed to reaching consensus on some intermediate value); see, e.g., [5] and references therein.

The remainder of this paper is organized as follows. In Section 2, we present the basic model of interest. In Section 3, we present convergence results in the absence of communication delays. In Section 4, we allow for communication delays and establish a new result: convergence, even with unbounded intercommunication intervals, as long as some weak form of symmetry is present. Section 5 provides some brief concluding comments.

This is a slightly modified version of a paper that appeared in the Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 2005, with a corrected proof of Theorem 4. This research was partially supported by Communauté francaise de Belgique - Actions de Recherche Concertées, by the HYCON Network of Excellence (contract number FP6-IST-511368), by the National Science Foundation under a Graduate Research Fellowship and grant ECS-0312921, and by the ARO under grant DAAD10-00-1-0466. Julien Hendrickx is a F.N.R.S. fellow.

V. Blondel and J. M. Hendrickx are with Department of Mathematical Engineering, Université catholique de Louvain, Avenue Georges Lemaitre 4, B-1348 Louvain-la-Neuve, Belgium; hendrickx@inma.ucl.ac.be, blondel@inma.ucl.ac.be

A. Olshevsky and J. N. Tsitsiklis are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; alex_o@mit.edu, jnt@mit.edu

II. The Agreement Algorithm.

In the absence of communication delays, the algorithm is as follows. Each agent i starts with a scalar value $x_i(0)$. The vector $x(t) = (x_1(t), \ldots, x_n(t))$ with the values held by the agents at time t, is updated according to the equation $x(t+1) = A(t)x(t)$, or

$$x_i(t+1) = \sum_{j=1}^{n} a_{ij}(t)x_j(t),$$

where $A(t)$ is a nonnegative matrix with entries $a_{ij}(t)$, and and where the updates are carried out at some discrete set of times which we will take, for simplicity, to be the nonnegative integers. We will assume that the row-sums of $A(t)$ are equal to 1, so that $A(t)$ is a stochastic matrix. In particular, $x_i(t+1)$ is a weighted average of the values $x_j(t)$ held by the agents at time t. We are interested in conditions that guarantee the convergence of each $x_i(t)$ to a constant, independent of i.

Throughout, we assume the following.

Assumption 1. There exists a positive constant α such that:
(a) $a_{ii}(t) \geq \alpha$, for all i, t.
(b) $a_{ij}(t) \in \{0\} \cup [\alpha, 1]$, for all i, j, t.
(c) $\sum_{j=1}^{n} a_{ij}(t) = 1$, for all i, t.

Intuitively, whenever $a_{ij}(t) > 0$, agent j communicates its current value $x_j(t)$ to agent i. Each agent i updates its own value, by forming a weighted average of its own value and the values it has just received from other agents.

The communication pattern at each time step can be described in terms of a directed graph $G(t) = (N, E(t))$, where $(j, i) \in E(t)$ if and only if $a_{ij}(t) > 0$. A minimal assumption, which is necessary for consensus to be reached, requires that following an arbitrary time t, and for any i, j, there is a sequence of communications through which agent i will influence (directly or indirectly) the value held by agent j.

Assumption 2. (Connectivity) The graph $(N, \cup_{s \geq t} E(s))$ is strongly connected for all $t \geq 0$.

We note various special cases of possible interest.

Fixed coefficients: There is a fixed matrix A, with entries a_{ij} such that, for each t, we have $a_{ij}(t) \in \{0\} \cup \{a_{ij}\}$ (depending on whether there is a communication from j to i at that time). This is the case presented in [3].

Symmetric model: If $(i,j) \in E(t)$ then $(j,i) \in E(t)$. That is, whenever i communicates to j, there is a simultaneous communication from j to i.

Equal neighbor model: Here,

$$a_{ij}(t) = \begin{cases} 1/n_i(t), & \text{if } j \in N_i(t), \\ 0, & \text{if } j \notin N_i(t), \end{cases}$$

where $N_i(t) = \{j \mid (j,i) \in E(t)\}$ is the set of agents j whose value is taken into account by i at time t, and $n_i(t)$ is its cardinality.

This model is a linear version of a model considered by Vicsek et al. [15]. Note that here the constant α of Assumption 1 is equal to $1/n$.

Pairwise averaging model ([5]): This is the special case of both the symmetric model and of the equal neighbor model in which, at each time, there is a set of disjoint pairs of agents who communicate (bidirectionally) with each other. If i communicates with j, then $x_i(t+1) = x_j(t+1) = (x_i(t) + x_j(t))/2$. Note that the sum $x_1(t) + \cdots + x_n(t)$ is conserved; therefore, if consensus is reached, it has to be on the average of the initial values of the nodes.

The assumption below is referred to as "partial asynchronism" in [3]. We will see that it is sometimes necessary for convergence.

Assumption 3. (Bounded intercommunication intervals) If i communicates to j an infinite number of times [that is, if $(i,j) \in E(t)$ infinitely often], then there is some B such that, for all t, $(i,j) \in E(t) \cup E(t+1) \cup \cdots \cup E(t+B-1)$.

III. Convergence results in the absence of delays.

We say that the agreement algorithm *guarantees asymptotic consensus* if the following holds: for every $x(0)$, and for every sequence $\{A(t)\}$ allowed by whatever assumptions have been placed, there exists some c such that $\lim_{t\to\infty} x_i(t) = c$, for all i.

Theorem 1. Under Assumptions 1, 2 (connectivity), and 3 (bounded intercommunication intervals), the agreement algorithm guarantees asymptotic consensus.

Theorem 1 subsumes the special cases of symmetry or of the equal neighbor model, and therefore subsequent convergence results and proofs for those cases.

Theorem 1 is presented in [14] and is proved in [13]; a simplified proof, for the special case of fixed coefficients can be found in [3]. The main idea, which applies to most results of this type, is as follows. Let $m(t) = \min_i x_i(t)$ and $M(t) = \max_i x_i(t)$. Since each $A(t)$ is stochastic, it is straightforward to verify that $m(t)$ and $M(t)$ are nondecreasing and nonincreasing, respectively. It then suffices to verify that the difference $M(t) - m(t)$ is reduced by a constant factor over a sufficiently large time interval; the interval is chosen so that every agent gets to influence (indirectly) every other agent; by tracing the chain of such influences, and using the assumption that each influence has a nontrivial "strength" (our assumption that whenever $a_{ij}(t)$ is nonzero, it is bounded below by $\alpha > 0$), the result follows.

In the absence of the bounded intercommunication interval assumption, the algorithm does not guarantee asymptotic consensus, as shown by Example 1 below (Exercise 3.1, in p. 517 of [3]). In particular, convergence to consensus fails even in the special case of the equal neighbor model. The main idea is that the agreement algorithm can closely emulate a nonconvergent algorithm that keeps executing the three instructions $x_1 := x_3$, $x_3 := x_2$, $x_2 := x_1$, one after the other.

Example 1. Let $n = 3$, and suppose that $x(0) = (0,0,1)$. Let ϵ_1 be a small positive constant. Consider the following sequence of events. Agent 3 communicates to agent 1; agent 1 forms the average of its own value and the received value. This is repeated t_1 times, where t_1 is large enough so that $x_1(t_1) \geq 1 - \epsilon_1$. Thus, $x(t_1) \approx (1,0,1)$. We now let agent 2 communicates to agent 3, t_2 times, where t_2 is large enough so that $x_3(t_1+t_2) \leq \epsilon_1$. In particular, $x(t_1+t_2) \approx (1,0,0)$. We now repeat the above two processes, infinitely many times. During the kth repetition, ϵ_1 is replaced by ϵ_k (and t_1, t_2 get adjusted accordingly). Furthermore, by permuting the agents at each repetition, we can ensure that Assumption 2 is satisfied. After k repetitions, it can be checked that $x(t)$ will be within $1 - \epsilon_1 - \cdots - \epsilon_k$ of a unit vector. thus, if we choose the ϵ_k so that $\sum_{k=1}^{\infty} \epsilon_k < 1/2$, asymptotic consensus will not be obtained.

On the other hand, in the presence of symmetry, the bounded intercommunication interval assumption is unnecessary. This result is proved in [9] and [4] for the special case of the symmetric equal neighbor model and in [11], [7], for the more general symmetric model. A more general result will be established in Theorem 4 below.

Theorem 2. Under Assumptions 1 and 2, and for the symmetric model, the agreement algorithm guarantees asymptotic consensus.

IV. Products of stochastic matrices and convergence rate

Theorem 1 and 2 can be reformulated as results on the convergence of products of stochastic matrices.

Corollary 1. Consider an infinite sequence of stochastic matrices $A(0), A(1), A(2), \ldots$, that satisfies Assumptions 1 and 2. If either Assumption 3 (bounded intercommunication intervals) is satisfied, or if we have a symmetric model, then there exists a nonnegative vector d such that

$$\lim_{t\to\infty} A(t)A(t-1)\cdots A(1)A(0) = \mathbf{1}d^T.$$

(Here, $\mathbf{1}$ is a column vector whose elements are all equal to one.)

According to Wolfowitz's Theorem ([16]) convergence occurs whenever the matrices are all taken from a finite set of ergodic matrices, and the finite set is such that any finite product of matrices in that set is again ergodic. Corollary 1 extends Wolfowitz' theorem by not requiring the matrices

A(t) to be ergodic, though it is limited to matrices with positive diagonal entries.

The presence of long matrix products suggests that convergence to consensus in the linear iteration

$$x(t+1) = A(t)x(t),$$

with $A(t)$ stochastic, might be characterized in terms of a joint spectral radius. The joint spectral radius $\rho(\mathcal{M})$ of a set of matrices $\mathcal{M}$ is a scalar that measures the maximal asymptotic growth rate that can be obtained by forming long products of matrices taken from the set $\mathcal{M}$:

$$\rho(\mathcal{M}) = \limsup_{k\to\infty} \sup_{M_{i_1}, M_{i_2}, \ldots, M_{i_k} \in \mathcal{M}} ||M_{i_1} M_{i_2} \ldots M_{i_k}||^{\frac{1}{k}}.$$

This quantity does not depend on the norm used. Moreover, for any $q > \rho(\mathcal{M})$ there exists a C for which

$$||M_{i_k} \ldots M_{i_1} y|| \leq C q^k \, ||y||$$

for all y and $M_{i_j} \in \mathcal{M}$.

Stochastic matrices satisfy $||Ax||_\infty \leq ||x||_\infty$ and $A\mathbf{1} = \mathbf{1}$, and so they have a spectral radius equal to one. The product of two stochastic matrices is again stochastic and so the joint spectral radius of any set of stochastic matrices is equal to one. To analyze the convergence rate of products of stochastic matrices, we consider the dynamics induced by the matrices on a space of smaller dimension.

Consider a matrix $P \in \Re^{(n-1)\times n}$ defining an orthogonal projection on the space orthogonal to $\text{span}\{\mathbf{1}\}$. We have $P\mathbf{1} = 0$, and $||Px||_2 = ||x||_2$ whenever $x^T\mathbf{1} = 0$. Associated to any $A(t)$, there is a unique matrix $A'(t) \in \Re^{(n-1)\times(n-1)}$ that satisfies $PA(t) = A'(t)P$. The spectrum of $A'(t)$ is the spectrum of $A(t)$ after removing one multiplicity of the eigenvalue 1. Let $\mathcal{M}'$ be the set of all matrices $A'(t)$.

Let $\gamma = \mathbf{1}^T x(t)/n$ be the mean value of the entries of $x(t)$, then

$$\begin{aligned} Px(t) - P\gamma\mathbf{1} &= Px(t) \\ &= PA(t)A(t-1)\ldots A(0)x(0) \\ &= A'(t)A'(t-1)\ldots A'(0)Px(0). \end{aligned}$$

Since $(x(t) - \gamma\mathbf{1})^T\mathbf{1} = 0$, we have

$$||x(t) - \gamma\mathbf{1}||_2 = ||P(x(t) - \gamma\mathbf{1})||_2 \leq C q^t \, ||x(0)||_2$$

for some C and for any $q > \rho(\mathcal{M}')$.

Assume now that $\lim_{t\to\infty} x(t) = c\mathbf{1}$ for some scalar c. Because all matrices are stochastic, c must belong to the convex hull of the entries of $x(t)$ for all t. We therefore have

$$||x(t) - c\mathbf{1}||_\infty \leq 2\, ||x(t) - \gamma\mathbf{1}||_\infty \leq 2\, ||Px(t) - P\gamma\mathbf{1}||_2,$$

and we may then conclude that

$$||x(t) - c\mathbf{1}||_\infty \leq 2Cq^t \, ||x(0)||_2.$$

The joint spectral radius $\rho(\mathcal{M}')$ therefore gives a measure of the convergence rate of $x(t)$ towards its limit value $c\mathbf{1}$. However, for this bound to be nontrivial, all of the matrices in $\mathcal{M}$ need to be ergodic; indeed, in the absence of an ergodicity condition, the convergence of $x(t)$ need not be geometric, and will depend in general on the particular sequence of elements of $\mathcal{M}$.

V. Convergence in the presence of delays.

The model considered so far assumes that messages from one agent to another are immediately delivered. However, in a distributed environment, and in the presence of communication delays, it is conceivable that an agent will end up averaging its own value with an *outdated* value of another processor. A situation of this type falls within the framework of distributed asynchronous computation developed in [3].

Communication delays are incorporated into the model as follows: when agent i, at time t, uses the value x_j from another agent, that value is not necessarily the most recent one, $x_j(t)$, but rather an outdated one, $x_j(\tau_j^i(t))$, where $0 \leq \tau_j^i(t) \leq t$, and where $t - \tau_j^i(t))$ represents communication and possibly other types of delay. In particular, $x_i(t)$ is updated according to the following formula:

$$x_i(t+1) = \sum_{j=1}^{n} a_{ij}(t) x_j(\tau_j^i(t)). \qquad (1)$$

We make the following assumption on the $\tau_j^i(t)$.

Assumption 4. (Bounded delays) (a) If $a_{ij}(t) = 0$, then $\tau_j^i(t) = t$.
(b) $\lim_{t\to\infty} \tau_j^i(t) = \infty$, for all i, j.
(c) $\tau_i^i(t) = t$, for all i, t.
(d) There exists some $B > 0$ such that $t - B + 1 \leq \tau_j^i(t) \leq t$, for all i, j, t.

Assumption 4(a) is just a convention: when $a_{ij}(t) = 0$, the value of $\tau_j^i(t)$ has no effect on the update. Assumption 4(b) is necessary for any convergence result: it requires that newer values of $x_j(t)$ get eventually incorporated in the updates of other agents. Assumption 4(c) is quite natural, since an agent generally has access to its own most recent value. Finally, Assumption 4(d) requires delays to be bounded by some constant B,

The next result, from [13], [14], is a generalization of Theorem 1. The idea of the proof is similar to the one outlined for Theorem 1, except that we now define $m(t) = \min_i \min_{s=t,t-1,\ldots,t-B+1} x_i(s)$ and $M(t) = \max_i \max_{s=t,t-1,\ldots,t-B+1} x_i(s)$. For convenience, we will adopt the definition that $x_i(t) = x_i(0)$ for all negative t. Once more, one shows that the difference $M(t) - m(t)$ decreases by a constant factor after a bounded amount of time.

Theorem 3. Under Assumptions 1-4 (connectivity, bounded intercommunication intervals, and bounded delays), the agreement algorithm with delays [cf. Eq. (1)] guarantees asymptotic consensus.

Theorem 3 assumes bounded intercommunication intervals and bounded delays. The example that follows (Example 1.2, in p. 485 of [3]) shows that Assumption 4(d) (bounded delays) cannot be relaxed. This is the case even for a symmetric model, or the further special case where $E(t)$ has exactly four arcs (i,i), (j,j), (i,j), and (j,i) at any given time t, and these satisfy $a_{ij}(t) = a_{ji}(t) = 1/2$, as in the pairwise averaging model.

Example 2. We have two agents who initially hold the values $x_1(0) = 0$ and $x_2(0) = 1$, respectively. Let t_k be an increasing sequence of times, with $t_0 = 0$ and $t_{k+1} - t_k \to \infty$. If $t_k \le t < t_{k+1}$, the agents update according to

$$\begin{aligned} x_1(t+1) &= (x_1(t) + x_2(t_k))/2, \\ x_2(t+1) &= (x_1(t_k) + x_2(t))/2. \end{aligned}$$

We will then have $x_1(t_1) = 1 - \epsilon_1$ and $x_2(t_1) = \epsilon_1$, where $\epsilon_1 > 0$ can be made arbitrarily small, by choosing t_1 large enough. More generally, between time t_k and t_{k+1}, the absolute difference $|x_1(t) - x_2(t)|$ contracts by a factor of $1 - 2\epsilon_k$, where the corresponding contraction factors $1 - 2\epsilon_k$ approach 1. If the ϵ_k are chosen so that $\sum_k \epsilon_k < \infty$, then $\prod_{k=1}^{\infty}(1 - 2\epsilon_k) > 0$, and the disagreement $|x_1(t) - x_2(t)|$ does not converge to zero.

According to the preceding example, the assumption of bounded delays cannot be relaxed. On the other hand, the assumption of bounded intercommunication intervals can be relaxed, in the presence of symmetry, leading to the following generalization of Theorem 2, which is a new result.

Theorem 4. Under Assumptions 1, 2 (connectivity), and 4 (bounded delays), and for the symmetric model, the agreement algorithm with delays [cf. Eq. (1)] guarantees asymptotic consensus.

Proof. Let

$$\begin{aligned} M_i(t) &= \max\{x_i(t), x_i(t-1), \ldots, x_i(t-B+1)\}, \\ M(t) &= \max_i M_i(t), \\ m_i(t) &= \min\{x_i(t), x_i(t-1), \ldots, x_i(t-B+1)\}, \\ m(t) &= \min_i m_i(t). \end{aligned}$$

Recall that we are using the convention that $x_i(t) = x_i(0)$ for all negative t. An easy inductive argument, as in p. 512 of [3], shows that the sequences $m(t)$ and $M(t)$ are are nondecreasing and nonincreasing, respectively. The convergence proof rests on the following lemma.

Lemma 1: If $m(\tau - B) = 0$ and $M(\tau) = 1$, then there exists a time $\tau' \ge \tau$ such that $M(\tau') - m(\tau' - B) \le 1 - \alpha^{nB}$.

Given Lemma 1, the convergence proof is completed as follows. Using the linearity of the algorithm, there exists a time τ_1 such that $M(\tau_1) - m(\tau_1 - B) \le (1 - \alpha^{nB})(M(B) - m(0))$. By applying Lemma 1, with τ replaced by τ_1, and using induction, we see that for every k there exists a time τ_k such that $M(\tau_k) - m(\tau_k - B) \le (1 - \alpha^{nB})^k(M(B) - m(0))$, which converges to zero. This, together with the monotonicity properties of $m(t)$ and $M(t)$, implies that $m(t)$ and $M(t)$ converge to a common limit, which is equivalent to asymptotic consensus.

Proof of Lemma 1: For $k = 1, \ldots, n$, we say that "Property P_k holds at time t" if there exist at least k indices i for which $m_i(t) \ge \alpha^{kB}$.

Since $m(\tau - B) = 0$, it follows that $m(t) \ge 0$ for all $t \ge \tau - B$ by the monotonicity of $m(t)$. In turn, this implies that $x_i(t+1) \ge \alpha x_i(t)$ for all i and all $t \ge \tau - B$.

Since $M(\tau) = 1$, there exists some i and some $t' \in \{\tau - B + 1, \tau - B + 2, \ldots, \tau\}$ such that $x_i(t') = 1$. Using the inequality $x_i(t+1) \ge \alpha x_i(t)$, we obtain $m_i(t' + B) \ge \alpha^B$. This shows that there exists a time at which property P_1 holds.

We continue inductively. Suppose that $k < n$ and that Property P_k holds at some time t. Let S be a set of cardinality k containing indices i for which $m_i(t) \ge \alpha^{kB}$, and let S^c be the complement of S. Let t' be the first time, greater than or equal to t, at which $a_{ij}(t') \ne 0$, for some $j \in S$ and $i \in S^c$ (i.e., an node j in S gets to influence the value of an node i in S^c). Such a time exists by the connectivity assumption (Assumption 2).

Note that between times t and t', the nodes ℓ in the set S only form convex combinations between the values of the nodes in the set S (this is a consequence of the symmetry assumption). Since all of these values are bounded below by α^{kB}, it follows that this lower bound remains in effect, and that $m_\ell(t') \ge \alpha^{kB}$, for all $\ell \in S$.

For times $s \ge t'$, and for every $\ell \in S$, we have $x_\ell(s + 1) \ge \alpha x_\ell(s)$, which implies that $x_\ell(s) \ge \alpha^{kB}\alpha^B$, for $s \in \{t' + 1, \ldots, t' + B\}$. Therefore, $m_\ell(t' + B) \ge \alpha^{(k+1)B}$, for all $\ell \in S$.

Consider now an node $i \in S^c$ for which $a_{ij}(t') \ne 0$. We have

$$x_i(t' + 1) \ge a_{ij}(t')x_j(\tau_j^i(t')) \ge \alpha m_j(t') \ge \alpha^{kB+1}.$$

Using also the fact $x_i(s+1) \ge \alpha x_i(s)$, we obtain that $m_i(t' + B) \ge \alpha^{(k+1)B}$. Therefore, at time $t' + B$, we have $k + 1$ nodes with $m_\ell(t' + B) \ge \alpha^{(k+1)B}$ (namely, the nodes in S, together with node i). It follows that Property P_{k+1} is satisfied at time $t' + B$.

This inductive argument shows that there is a time τ' at which Property P_n is satisfied. At that time $m_i(\tau') \ge \alpha^{nB}$ for all i, which implies that $m(\tau') \ge \alpha^{nB}$. On the other hand, $M(\tau' + B) \le M(0) = 1$, which proves that $M(\tau' + B) - m(\tau') \le 1 - \alpha^{nB}$. **q.e.d.**

The symmetry condition [$(i, j) \in E(t)$ iff $(j, i) \in E(t)$] used in Theorem 4 is somewhat unnatural in the presence of communication delays, as it requires perfect synchronization of the update times. A looser and more natural assumption is the following.

Assumption 5. There exists some $B > 0$ such that whenever $(i, j) \in E(t)$, then there exists some τ that satisfies $|t - \tau| < B$ and $(j, i) \in E(\tau)$.

Assumption 5 allows for protocols such as the following. Agent i sends its value to agent j. Agent j responds by sending its own value to agent i. Both agents update their values (taking into account the received messages), within a bounded time from receiving other agent's value. In a realistic setting, with unreliable communications, even this loose symmetry condition may be impossible to enforce with absolute certainty. One can imagine more complicated protocols based on an exchange of acknowledgments, but fundamental obstacles remain (see the discussion of the "two-army problem" in pp. 32-34 of [2]). A more realistic model would introduce a positive probability that some of the updates are never carried out. (A simple possibility is to assume that each $a_{ij}(t)$, with $i \neq j$, is changed to a zero, independently, and with a fixed probability.) The convergence result that follows remains valid in such a probabilistic setting (with probability 1). Since no essential new insights are provided, we only sketch a proof for the deterministic case.

Theorem 5. Under Assumptions 1, 2 (connectivity), 4 (bounded delays), and 5, the agreement algorithm with delays [cf. Eq. (1)] guarantees asymptotic consensus.

Proof. (Outline) A minor change is needed in the proof of Lemma 1. In particular, we define P_k as the event that there exist at least k indices l for which $m_l(t) \geq \alpha^{2kB}$. It follows that P_1 holds at time $t = 2B$.

By induction, let P_k hold at time t, and let S be the set of cardinality k containing indices l for which $m_l(t) \geq \alpha^{2kB}$. Furthermore, let τ be the first time after time t that $a_{ij}(\tau) \neq 0$ where exactly one of i, j is in S. Along the same lines as in the proof of Lemma 1, $m_l(\tau) \geq \alpha^{2kB}$ for $l \in S$; since $x_l(t+1) \geq \alpha x_l(t)$, it follows that $m_l(\tau + 2B) \geq \alpha^{2(k+1)B}$ for each $l \in S$. By our assumptions, exactly one of i,j is in S^c. If $i \in S^c$, then $x_i(\tau+1) \geq a_{ij}(\tau) x_j(\tau_j^i(\tau)) \geq \alpha^{2kB+1}$ and consequently $x_i(\tau+2B) \geq \alpha^{2B-1}\alpha^{2kB+1} = \alpha^{2(k+1)B}$. If $j \in S^c$, then there must exist a time $\tau_j \in \{\tau+1, \tau+2, \ldots, \tau+B-1\}$ with $a_{ji}(\tau_j) > 0$. It follows that:

$$\begin{aligned} m_j(\tau+2B) &\geq \alpha^{\tau+2B-(\tau_j+1)} x_j(\tau_j+1) \\ &\geq \alpha^{\tau+2B-\tau_j-1} \alpha x_i(\tau_j) \\ &\geq \alpha^{\tau+2B-\tau_j-1} \alpha \alpha^{\tau_j-\tau} \alpha^{2kB} \\ &= \alpha^{2(k+1)B} \end{aligned}$$

Therefore, P_{k+1} holds at time $\tau + 2B$ and the induction is complete. **q.e.d.**

VI. CONCLUDING REMARKS

Many variations of the available convergence results and of the new ones presented here are possible, by considering additional sources of asynchronism, as well as probabilistic (rather than deterministic) assumptions. The proof technique introduced in [3] (based on the contraction of the difference $M(t) - m(t)$) has so far been able to handle such variations.

One particular variation that has been investigated in the recent literature is one where strong connectivity is relaxed: some agents act as "leaders" and influence the values of the other agents (the "followers") but not vice versa. This is similar to the setting considered in Chapter 6 of [3] where leaders and followers correspond to the "computing" and "noncomputing" processors of [3].

Let us also note that there is a related algorithm for distributed load balancing [6], for which similar convergence results are available (see Section 7.4 of [3]). The latter algorithm has some commonalities with the pairwise averaging model: the sum of the agents' entries/loads is a long-term invariant, although in the load balancing algorithm, some of the load can be temporarily "in transit." In particular, the load balancing algorithm guarantees convergence to the exact average of the initial values, even in the presence of asynchronism, time delays, and dynamic topology changes.

REFERENCES

[1] D. Angeli, P.-A. Bliman, "Stability of leaderless multi-agent systems, Extension of a result by Moreau", arXiv:math.OC/0411338.

[2] D. P. Bertsekas and R. G. Gallager, *Data Networks*, Prentice Hall, 2nd edition, 1991.

[3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall, 1989.

[4] M. Cao, A. S. Morse, B. D. O. Anderson, "Coordination of an Asynchronous, Multi-Agent System via Averaging," preprint, 2004.

[5] S. P. Boyd, A. Ghosh, B. Prabhakar, D. Shah, "Gossip Algorithms: Design, Analysis and Applications," *Proceedings of Infocom 05*.

[6] G. Cybenko. "Dynamic load balancing for distributed memory multiprocessors," *Journal of Parallel and Distributed Computing*, Vol. 7, no. 2, 1989, pp. 279 - 301.

[7] J. M. Hendrickx and V. D. Blondel, "Convergence of different linear and non-linear Vicsek models," Université catholique de Louvain, CESAME research report, RR 2005.57, 2005.

[8] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control,* vol. 48, no. 6, 2003, pp. 988-1001.

[9] S. Li, H. Wang, "Multi-agent coordination using nearest-neighbor rules: revisiting the Vicsek model"; http://arxiv.org/abs/cs.MA/0407021

[10] C. C. Moallemi and B. Van Roy, "Consensus Propagation," preprint, 2005.

[11] L. Moreau, "Stability of multi-agent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, Vol. 50, No. 2, 2005, pp. 169-182.

[12] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, Vol. 49, No. 9, 2004, pp. 15201533.

[13] J. N. Tsitsiklis, *Problems in decentralized decision making and computation,* Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1984; http://web.mit.edu/jnt/www/PhD-84-jnt.pdf

[14] J. N. Tsitsiklis, D. P. Bertsekas and M. Athans, "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," *IEEE Transactions on Automatic Control,* Vol. 31, No. 9, 1986, pp. 803-812.

[15] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Schochet, "Novel type of phase transitions in a system of self-driven particles," *Physical Review Letters,* Vol. 75, No. 6, 1995, pp. 1226-1229.

[16] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proceedings of the American Mathematical Society,* Vol. 15, 1963, pp. 733-737.