A Lower Bound on Convergence of a Distributed Network Consensus Algorithm

M. Cao1
Dept. of Electrical Engineering
Yale University
m.cao@yale.edu

D. A. Spielman2
Dept. of Computer Science
Yale University
spielman@cs.yale.edu

A. S. Morse1
Dept. of Electrical Engineering
Yale University
as.morse@yale.edu

Abstract—This paper gives a lower bound on the convergence rate of a class of network consensus algorithms. Two different approaches using directed graphs as a main tool are introduced: one is to compute the “scrambling constants” of stochastic matrices associated with “neighbor shared graphs” and the other is to analyze random walks on a sequence of graphs. Both approaches prove that the time to reach consensus within a dynamic network is logarithmic in the relative error and is in worst case exponential in the size of the network.

I. INTRODUCTION

Distributed algorithms where processors reach consensus by exchanging locally computed results have been studied extensively in the field of distributed and parallel computations [1], [2]. In recent years, network consensus algorithms have attracted more attention from control scientists and engineers interested in distributed coordination of groups of mobile autonomous agents. Some simple local rules can cause a group of agents to behave cooperatively without the existence of a central controller. The cooperative behaviors include moving in the same direction [3], [4], aggregating in swarms [5], and rendezvousing at the same location [6], [7]. Two salient features make the convergence of the consensus algorithms difficult to analyze in the context of distributed coordination within multi-agent systems: First, the final consensus value cannot be determined a priori and depends on the initial value of each agent as well as the dynamic interaction between them; and secondly, because agents move, the underlying network topology may not be fixed during system evolutions.

A special class of network consensus algorithms for flocking via distributed averaging was first studied in [8], and convergence was proved under mild connectivity assumptions in [3] by using switching linear system theory and algebraic graph theory. Since then various works have been done to develop this class of algorithms in continuous time [9], on weakly connected graphs [4], and in asynchronous scenarios [10]. However, so far few if any specific convergence rates have been derived for this class of algorithms. This is due to the fact that tools such as joint spectrum theory cannot compute an explicit bound for this particular class of switched linear systems without equilibriums that can be determined beforehand.

This paper will study the convergence rate for the flocking process based on the encountered directed graphs. A lower bound on the convergence rate is presented using “scrambling constants” for “neighbor shared graphs”. It will be shown that the time needed for all the agents’ headings to converge to a $\varepsilon$-neighborhood of the steady state is logarithmic in $\varepsilon$ and in worst case exponential in the total number of agents. To better understand the factors that influence the convergence rate, a separate approach using random walks on graphs is also discussed.

The rest of the paper is organized as follows. In section 2, the flocking algorithm based on distributed averaging is introduced. In section 3, the convergence is proved for the case when graphs encountered during the flocking process are all “neighbor shared”. In section 4, we study the more general case when graphs encountered are all rooted. In section 5, a separate approach using random walks on graphs is introduced.

II. FLOCKING VIA DISTRIBUTED AVERAGING

Consider a system that consists of $n$ mobile autonomous agents, labelled 1 through $n$, that try to reach agreement on a scalar variable which will be called the heading. Each agent’s heading $\theta_i$ is updated using
a simple local rule based on the average of its own heading and the headings of its “neighbors”. Agent \( i \)'s neighbors at time \( t \) are those agents, including itself, whose value at time \( t \) is available to agent \( i \). Because of motion or other possible causes, agent \( i \)'s neighbors may change with time. Let \( N_i(t) \) and \( n_i(t) \) denote the set of labels and the number of agent \( i \)'s neighbors at time \( t \) respectively. Then agent \( i \)'s heading evolves in discrete-time in accordance with a model of the form

\[
\theta_i(t + 1) = \frac{1}{n_i(t)} \sum_{j \in N_i(t)} \theta_j(t)
\]

where \( t \) is a discrete-time index taking values in the non-negative integers \( \{0, 1, 2, \ldots \} \). We assume that the time between such updates is bounded below by a positive number \( \tau_B \) called a dwell time, and that all agents update their values synchronously.

The explicit form of the update equations determined by (1) depends on the neighbor relationships which exist at time \( t \). These relationships can be conveniently described by a directed graph \( G \) with vertex set \( V = \{1, 2, \ldots, n\} \) and arc set \( A(G) \subset V \times V \) which is defined in such a way so that \((i, j)\) is an arc or directed edge from \( i \) to \( j \) just in case agent \( i \) is a neighbor of agent \( j \). Thus \( G \) is a directed graph on \( n \) vertices with at most one arc from any vertex to another and with exactly one self-arc at each vertex. We write \( G \) for the set of all such graphs. We use the symbol \( P \) to denote a suitably defined set indexing \( G \). Thus \( G = \{G_p : p \in P\} \) where for \( p \in P \), \( G_p \) denotes the \( p \)th graph in \( G \). It is natural to call a vertex \( i \) a neighbor of vertex \( j \) in \( G \) if \((i, j)\) is an arc in \( G \). In addition we sometimes refer to a vertex \( k \) as an observer of vertex \( j \) in \( G \) if \((j, k)\) is an arc in \( G \). Thus every vertex of \( G \) can observe its neighbors, which with the interpretation of vertices as agents, is precisely the kind of relationship \( G \) is supposed to represent.

The set of agent heading update rules defined by (1) can be written in state form. Toward this end, for each \( p \in P \), define the flocking matrix

\[
F_p = D_p^{-1} A_p'
\]

where \( A_p' \) is the transpose of the adjacency matrix of the graph \( G_p \) and \( D_p \) is the diagonal matrix whose \( j \)th diagonal element is the in-degree of vertex \( j \) within the graph\(^1\). Then

\[
\theta(t + 1) = F_{\sigma(t)} \theta(t), \quad t \in \{0, 1, 2, \ldots\}
\]

where \( \theta \) is the state vector \( \theta = [\theta_1 \theta_2 \cdots \theta_n]' \) and \( \sigma : \{0, 1, \ldots\} \rightarrow P \) is a switching signal whose value at time \( t \) is the index of the graph representing the agents’ neighbor relationships at time \( t \).

We will show for a large class of switching signals that all the agents will reach the same steady state value \( \theta_{ss} \) given any initial set of agent headings. Convergence of the \( \theta_i \) to \( \theta_{ss} \) is equivalent to the state vector \( \theta \) converging to a vector of the form \( \theta_{ss} \cdot 1 \) where \( 1 \) is the \( n \times 1 \) vector \([1 \ 1 \ \cdots \ 1]'\). In addition, we will also give a sharp lower bound on the rate of this convergence process. The main challenge lies in how to define and analyze the rate of convergence for a multi-agent system under changing neighbor relationships.

## III. Neighbor shared graphs

We will first give a convergence result for the case where the graphs encountered along a trajectory of (3) are all “neighbor shared”. This requirement can be relaxed significantly as explained in the next section. Let us call \( G \in G \) neighbor shared if each set of 2 distinct vertices share a common neighbor. It is now possible to state the following elementary convergence result.

**Theorem 1:** Let \( Q \) denote the subset of \( P \) consisting of those indices \( q \) for which \( G_q \in G \) is neighbor shared. Let \( x(0) \) be fixed and let \( \sigma : \{0, 1, 2, \ldots\} \rightarrow P \) be a switching signal satisfying \( \sigma(t) \in Q, \ t \in \{0, 1, \ldots\} \). Then there is a constant steady state heading \( \theta_{ss} \) depending only on \( \theta(0) \) and \( \sigma \) for which

\[
\lim_{t \to \infty} \theta(t) = \theta_{ss} \cdot 1
\]

where the limit is approached exponentially fast.

In order to explain why this theorem is true, we will make use of certain structural properties of \( F_p \). As defined, each \( F_p \) is square and non-negative, where by a non-negative matrix is meant a matrix whose entries are all non-negative. Each \( F_p \) also has the property that its row sums all equal 1 \( \{i.e., F_p \cdot 1 = 1\} \). Matrices with these two properties are called \{row\} stochastic. Because each vertex of each graph in \( G \) has a self-arc, the \( F_p \) have the additional property that their diagonal elements are all non-zero.

Stochastic matrices have been studied extensively in the literature for a long time largely because of their connection with Markov chains \([11]\). One problem of particular relevance here is to describe the asymptotic behavior of products of \( n \times n \) stochastic matrices of the
form
\[ S_j S_{j-1} \cdots S_1 \]
as \( j \) tends to infinity. This is equivalent to looking at
the asymptotic behavior of all solutions to the recursion
equation
\[ x(j + 1) = S_j x(j) \quad (5) \]
because any solution \( x(j) \) can be written as
\[ x(j) = (S_j S_{j-1} \cdots S_1) x(1), \quad j \geq 1 \]
One especially useful idea, which goes back at least
to [12], is to consider the behavior of the scalar-valued non-
negative function \( V(x) = \lceil x \rceil - \lfloor x \rfloor \) along solutions
to (5) where \( x = [x_1 \ x_2 \ \cdots \ x_n]' \) is a non-negative
\( n \times 1 \) vector and \( \lceil x \rceil \) and \( \lfloor x \rfloor \) are its largest and smallest
elements respectively. The key observation is that for
any \( n \times n \) stochastic matrix \( S \), \( |Sx| \geq |x| \) and, consequently, that
\( V(Sx) \leq V(x) \). These inequalities and (5) imply that the sequences
\[ |x(1)|, |x(2)|, \ldots, |x(j)|, \lfloor x(2) \rfloor, \ldots \]
are each monotone. Hence, because each of these se-
quencies is also bounded, the limits
\[ \lim_{j \to \infty} |x(j)|, \quad \lim_{j \to \infty} |x(j)|, \quad \lim_{j \to \infty} V(x(j)) \]
also exist. Note that whenever the limit of \( V(x(j)) \) is
zero, all components of \( x(j) \) must tend to the same value
and moreover this value must be a constant equal to the
limiting value of \( |x(j)| \).

We will make use of the standard partial ordering
\( \geq \) on \( n \times n \) nonnegative matrices by writing \( B \geq A \)
everywhere \( B - A \) is nonnegative. We extend the domain of definitions of \( \lceil \cdot \rceil \) and \( \lfloor \cdot \rfloor \) to the class of all non-
negative \( n \times n \) matrix \( M \), by letting \( \lceil M \rceil \) and \( \lfloor M \rfloor \)
now denote the \( 1 \times m \) row vectors whose \( j \)th entries
are the smallest and largest elements respectively, of the
\( j \)th column of \( M \). For any infinite sequence of \( n \times n \)
stochastic matrices \( S_1, S_2, \ldots \), we henceforth use the
symbol \( \cdots S_j \cdots S_1 \) to denote the limit
\[ \lfloor \cdots S_j \cdots S_2 S_1 \rfloor = \lim_{j \to \infty} |S_j \cdots S_2 S_1| \quad (6) \]
From the preceding discussion it is clear that this limit
exists whether or not the product \( S_j \cdots S_2 S_1 \) itself has
a limit. In the sequel we will be interested in the non-
negative matrix
\[ [S] = 1(|S| - \lfloor S \rfloor) \quad (7) \]
Note that whenever \( [S] = 0 \), all rows of \( S \) must be
equal.

**Lemma 1:** For any two \( n \times n \) stochastic matrices \( S_1 \)
and \( S_2 \),
\[ [S_2 S_1] \leq \mu(S_2) [S_1] \quad (8) \]
where for any \( n \times n \) stochastic matrix \( S \),
\[ \mu(S) = \max_{i,j} \left( 1 - \sum_{k=1}^n \min \{s_{ik}, s_{jk}\} \right) \quad (9) \]
The proof of this lemma will appear in the full length
version of this paper. The quantity \( \mu(S) \) has been widely
studied before [13], [11] and is known as the *scrambling con-
stant* of the stochastic matrix \( S \). Note that since the
row sums of \( S \) all equal 1, \( \mu(S) \) is non-negative. It is
easy to see that \( \mu(S) = 0 \) just in case all the rows of \( S \)
are equal. Let us note that for fixed \( i \) and \( j \), the \( k \)th term
in the sum appearing in (9) will be positive just in case
both \( s_{ik} \) and \( s_{jk} \) are positive. It follows that the sum will
be positive if and only if for at least one \( k \), \( s_{ik} \) and \( s_{jk} \)
are both positive. Thus \( \mu(S) \leq 1 \) if and only if for each
distinct \( i \) and \( j \), there is at least one \( k \) for which \( s_{ik} \)
and \( s_{jk} \) are both positive. Matrices with this property have
been widely studied and are called *scrambling matrices*.
Thus a stochastic matrix \( S \) is a scrambling matrix if and
only if \( \mu(S) < 1 \). It is easy to see that the definition of a
scrambling matrix also implies that \( S \) is scrambling if
and only if its associated graph is neighbor shared.

Suppose that \( F_p \) is a flocking matrix for which \( \mathcal{G}_p \)
is neighbor shared. In view of the definition of a flocking
matrix, any non-zero entry in \( F_p \) must be bounded below
by \( \frac{1}{n} \). Fix distinct \( i \) and \( j \) and suppose that \( k \)
is a neighbor that \( i \) and \( j \) share. Then \( f_{ik} \) and \( f_{jk} \)
are both non-zero, so \( \min \{f_{ik}, f_{jk}\} \leq \frac{1}{n} \). This implies
that the sum in (9) must be bounded below by \( \frac{1}{n} \)
and consequently that \( \mu(F_p) \leq 1 - \frac{1}{n} \).

Now let \( F_p \) be that flocking matrix whose graph
\( \mathcal{G}_p \in \mathcal{G} \) is such that vertex 1 has no neighbors
other than itself, vertex 2 has every vertex as a neighbor, and
vertices 3 through \( n \) have only themselves and agent 1
as neighbors. Since vertex 1 has no neighbors other than
itself, \( f_{1k} = 0 \) for all \( k \) and for \( k > 1 \). Thus for all \( i, j \), it
must be true that \( \sum_{k=1}^n \min \{f_{ik}, f_{jk}\} = \min \{f_{i1}, f_{j1}\} \.
Now vertex 2 has \( n \) neighbors, so \( f_{21} = \frac{1}{n} \). Thus
\( \min \{f_{i1}, f_{j1}\} \) attains its lower bound of \( \frac{1}{n} \) when either
\( i = 2 \) or \( j = 2 \). It thus follows that with this \( F_p \), \( \mu(F_p) \)
attains its upper bound of \( 1 - \frac{1}{n} \). We summarize.

**Lemma 2:** Let \( Q \) be the set of indices in \( \mathcal{P} \) for which
\( \mathcal{G}_p \) is neighbor shared. Then
\[ \max_{q \in Q} \mu(F_q) = 1 - \frac{1}{n} \quad (10) \]
Lemma 1 and 2 will be used in the proof of Theorem 1.

**Proof of Theorem 1:** By definition, the graph \( G_p \) of each matrix \( F_p \) in the finite set \( \{ F_p : p \in \mathbb{Q} \} \) is neighbor shared. By assumption, \( F_{\sigma(t)} \in \{ F_p : p \in \mathbb{Q} \} \) for \( t \geq 0 \). Let \( \lambda = 1 - \frac{1}{n} \). In view of Lemma 1 and 2,

\[
[F_{\sigma(t)} \cdots F_{\sigma(0)}] \leq [F_{\sigma(t)}] \cdots [F_{\sigma(0)}] \leq \lambda^{t+1}
\]

and \( \lambda^{t+1} \) approaches zero exponentially fast as \( t \to \infty \). Thus the product \( F_{\sigma(t)} \cdots F_{\sigma(0)} \) converges to \( 1 \cdot \cdots F_{\sigma(t)} \cdots F_{\sigma(0)} \) exponentially fast at a rate no slower than \( \lambda \). But it is clear from (3) that

\[
\theta(t) = F_{\sigma(t-1)} \cdots F_{\sigma(1)} F_{\sigma(0)} \theta(0), \quad t \geq 1
\]

Therefore (4) holds with \( \theta_{\sigma} = [\cdots F_{\sigma(t)} \cdots F_{\sigma(0)}] \theta(0) \) and the convergence is exponential.

Hence, we have proved that \( 1 - \frac{1}{n} \) is a worst case bound on the rate of convergence of products of flocking matrices whose graphs are all neighbor shared.

**IV. Rooted Graphs**

The proof of Theorem 1 depends crucially on the fact that the graphs encountered along a trajectory of (3) are all neighbor shared. The aim of this section is to show that this requirement can be relaxed. To do this we need to have a meaningful way of “combining” sequences of graphs so that only the combined graph need be neighbor shared, but not necessarily the individual graphs making up the combination. Let us agree to say that the definition extend unambiguously to any finite sequence of graph which possesses at least one root.

**Lemma 3:** Each neighbor shared graph in \( G \) is rooted.

**Proof:** In a graph \( G \in G \), we say \( v \) is a root for vertices \( 1, \ldots, k \) if each of vertices \( 1, \ldots, k \) is reachable from \( v \). In a neighbor shared graph, vertices 1 and 2 have a root. One may now prove by induction that if \( 1, \ldots, k \) have a root for an integer \( 2 \leq k < n \), then \( 1, \ldots, k+1 \) do as well: any common neighbor of vertex \( k+1 \) and the root of \( 1, \ldots, k \) will suffice.

It is worth noting that although neighbor shared graphs are rooted, the converse is not necessarily true. The reader may wish to construct a three vertex example which illustrates this. Although rooted graphs in \( G \) need not be neighbor shared, it turns out that the composition of any \( n-1 \) rooted graphs in \( G \) is.

**Lemma 4:** The composition of any set of \( m \geq n-1 \) rooted graphs in \( G \) is neighbor shared.

The proof of this lemma will appear in the full length version of this paper.

It is also possible to derive a worst case convergence rate for products of flocking matrices which have rooted rather than neighbor-shared graphs. We will use the following result without providing its proof.

**Lemma 5:** Let \( b \) be a positive number less than 1. Let \( S_m^n, m \geq 1 \), denote the set of all \( m \)-term matrix products \( S = S_m S_{m-1} \cdots S_1 \) where each \( S_i \) is an \( n \times n \) stochastic matrix with a rooted graph in \( G \) and all nonzero entries bounded below by \( b \). Then

\[
\max_{S \in S^{n-1}} \mu(S) = 1 - b^{(n-1)}
\]

It turns out that this bound is actually attained if all the \( S_i \) are the same taking the form

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & \cdots & 0 \\
b & 1-b & 0 & 0 & \cdots & 0 \\
0 & b & 1-b & 0 & \cdots & 0 \\
0 & 0 & 0 & b & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1-b & 0 \\
0 & 0 & 0 & \cdots & b & 1-b
\end{bmatrix}
\]

(11)

It is possible to apply at least part of the preceding to the case when the \( S_i \) are flocking matrices. Towards this end, let \( G_{p_1}, G_{p_2}, \ldots, G_{p_{n-1}} \) be any sequence of
Let $\mathcal{G}$ denote the subset of $\mathcal{P}$ consisting of those indices $q$ for which $\mathcal{G}_q \in \mathcal{G}$ is rooted. Let $\theta(0)$ be fixed and let $\sigma : \{0, 1, 2, \ldots\} \rightarrow \mathcal{P}$ be a switching signal satisfying $\sigma(t) \in \mathcal{Q}$, $t \in \{0, 1, \ldots\}$. Then there is a constant steady state heading $\theta_s$, depending only on $\theta(0)$ and $\sigma$ for which

$$
\lim_{t \to \infty} \theta(t) = \theta_s 1
$$

where the limit is approached exponentially fast at a rate no slower than

$$
\lambda = \left\{ 1 - \left(\frac{1}{n}\right)^{(n-1)} \right\}^\frac{1}{m-1}
$$

It is possible to develop analogous results for strongly connected graphs, where by a strongly connected graph we mean a directed graph that has a path from each vertex to every other vertex. We will state the following result without proof.

**Corollary 1:** Under the hypotheses of Theorem 2, and the additional assumption that $\sigma$ takes values only in the subset of $\mathcal{Q}$ composed of those indices for which $\mathcal{G}_p$ is strongly connected, convergence of $\theta(t)$ to $\theta_s 1$ is exponential at a rate no slower than

$$
\lambda = \left\{ 1 - \left(\frac{1}{n}\right)^m \right\}^{\frac{1}{m}}
$$

where $m$ is the integer quotient of $n$ divided by 2.

V. Random walks on directed graphs

We need some more ideas in this section. By the reverse graph of $\mathcal{G} \in \mathcal{G}$, written $\mathcal{G}'$, we mean the graph in $\mathcal{G}$ which results when the directions of all arcs in $\mathcal{G}$ are reversed. It is clear that $\mathcal{G}$ is closed under the reverse operation. It is also clear that $(\mathcal{G}_q \circ \mathcal{G}_p)' = \mathcal{G}_q' \circ \mathcal{G}_p', p, q \in \mathcal{P}$. For any flocking matrix $F_p$, $p \in \mathcal{P}$, let $\gamma(F_p)$ denote that graph $\mathcal{G} \in \mathcal{G}$ corresponding to $F_p$. A different approach to analyzing the convergence of the flocking process is to multiply $F_p$ by row vectors from the left. In this approach, we are looking at the random walk [14] where at each time step we apply a different graph $G_{\gamma(t)}$, $\gamma : \{0, 1, 2, \ldots\} \rightarrow \mathcal{P}$. Then the random walk will converge to some fixed distribution if the flocking process converges.

First, we will consider walks that begin at just one vertex, that is to begin with a distribution given by an elementary $1 \times n$ vector $e_i$, with $e_i(i) = 1$ and $e_i(j) = 0$ for $j \neq i$. In the sequel, let $\mathcal{Q}$ denote the set of $p \in \mathcal{P}$ for which $\mathcal{G}_p$ is rooted.

**Lemma 6:** For any sequence $p(1), p(2), \ldots, p(n-1)$ of indices in $\mathcal{Q}$, let $R_i$ denote the set of indices of those vertices that are reachable from vertex $i$ after a walk on
a sequence of $n-1$ graphs $\gamma(F_{p(n-1)})', \ldots, \gamma(F_{p(1)})'$. Then for all $i$ and $j \in R_i$

$$(e_iF_{p(n-1)} \cdots F_{p(1)}) (j) \geq (\frac{1}{n})^{(n-1)}$$

Proof: The vector $e_i$ can be thought of as a unit positive mass at vertex $i$. From the definition of $R_i$, we know vertex $j$ can be reached from vertex $i$ by a walk of $n-1$ steps on the sequence of $n-1$ graphs. Since in each step at least $\frac{1}{n}$ fraction of the mass is propagated along an arc, we know vertex $j$ has at least $(\frac{1}{n})^{(n-1)}$ fraction of the unit mass.

Now we will consider the difference between any pair of rows of the matrix $F_{p(n-1)} \cdots F_{p(1)}$.

Lemma 7: For any sequence $p_1, p_2, \ldots, p_{n-1}$ of indices in $Q$ and for each $i \neq j$

$$\|e_i - e_j\|F_{p(n-1)} \cdots F_{p(1)}\|_1 \leq 2 - 2(\frac{1}{n})^{(n-1)}$$

where for a vector $a = [a_1 \ a_2 \ \ldots \ a_n]'$, $\|a\|_1 = \sum_i |a_i|$.

Proof: The initial vector $e_i - e_j$ can be thought of as unit positive mass at vertex $i$ and unit negative mass at vertex $j$. As the walk progresses, the sum of positive mass minus negative mass remains zero. When a positive mass meets a negative mass, they cancel each other out. Since $\gamma(F_{p(n-1)} \cdots F_{p(1)})$ by Lemma 4 is neighbor shared, the pair of vertices $i$ and $j$ can both reach some vertex $k$ in $\gamma(F_{p(n-1)} \cdots F_{p(1)})'$. Hence, after $n-1$ steps of the random walk starting from vertices $i$ and $j$, in view of Lemma 6 we will cancel at least $(\frac{1}{n})^{(n-1)}$ mass at vertex $k$.

Now we are in a position to give a lower bound on the rate of convergence of the random walk process.

Lemma 8: For any sequence $p_1, p_2, \ldots, p_{n-1}$ of indices in $Q$ and for every row vector $x$ such that $\sum_i x(i) = 0$,

$$\|x F_{p(n-1)} \cdots F_{p(1)}\|_1 \leq \bar{\lambda} \|x\|_1$$

where $\bar{\lambda} = 1 - (\frac{1}{n})^{(n-1)}$.

Proof: Let $i$ be the index of the minimal non-zero value $|x_i|$. Assume without loss of generality that $x_i$ is positive. Then pick some $j$ such that $x_j$ is negative. Lemma 7 says that at least a $1 - (\frac{1}{n})^{(n-1)}$ fraction of the mass propagated from $x_i$ will be cancelled by mass propagated from $x_j$. Now, remove $x_i$ from consideration, and $x_j$ of the negative mass at $x_j$. Continuing in this way, we can pair up positive masses with negative masses, and by observing the fraction of cancellation, we have

$$\|xF_{p(n-1)} \cdots F_{p(1)}\|_1 \leq (1 - (\frac{1}{n})^{(n-1)}) \|x\|_1$$

Using Lemma 8, we can arrive at the same convergence result as that in Theorem 2. According to our experience, we can gain insight into the convergence rate by constructing special sequences of graphs on which the random walk progresses.

VI. CONCLUDING REMARKS

It is possible to relax still further the conditions under which the flocking process converges. We only require the sequence of graphs encountered during the flocking process be repeatedly “jointly rooted” where a finite sequence of directed graphs $\mathcal{G}_{p_1}, \ldots, \mathcal{G}_{p_k}$ in $\mathcal{G}$ is jointly rooted if the composition $\mathcal{G}_{p_k} \circ \cdots \circ \mathcal{G}_{p_1}$ is rooted.

In the future, we are interested in designing consensus algorithms with faster convergence rates once we gain deeper insight into the factors that affect the convergence process.

REFERENCES