2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Contractions for Consensus Processes*

J. Liu, A. S. Morse, B. D. O. Anderson, and C. Yu

***Abstract*— Many distributed control algorithms of current interest can be modeled by linear recursion equations of the form $x(t+1) = M(t)x(t),\ t \geq 1$ where each $M(t)$ is a real-valued "stochastic" or "doubly stochastic" matrix. Convergence of such recursions often reduces to deciding when the sequence of matrix products $M(1), M(2)M(1), M(3)M(2)M(1), \dots$ converges. Certain types of stochastic and doubly stochastic matrices have the property that *any* sequence of products of such matrices of the form $S_1, S_2S_1, S_3S_2S_1, \dots$ converges exponentially fast. We explicitly characterize the largest classes of stochastic and doubly stochastic matrices with positive diagonal entries which have these properties. The main goal of this paper is to find a "semi-norm" with respect to which matrices from these "convergability classes" are contractions. For any doubly stochastic matrix S such a semi-norm is identified and is shown to coincide with the second largest *singular* value of S.**

I. Introduction

Many distributed control algorithms of current interest can be modeled by linear recursion equations of the form

$$x(t+1) = M(t)x(t), \quad t \geq 1 \tag{1}$$

where each $M(t)$ is a real-valued "stochastic" or "doubly stochastic" matrix. Among these are consensus and flocking algorithms [2]–[8], distributed averaging algorithms [9]–[11], and certain types of gossiping algorithms [12]–[14]. Recursion equations like this have their roots in the literature on nonhomogeneous Markov chains [15]. While much is known at this point about conditions on the $M(t)$ for solutions to converge to a limit point, considerably less is known about the rates at which such solutions converge. There are classical concepts such as the coefficient of ergodicity [15] which are helpful in deriving convergence rates, but these are limited to only certain types of processes. The convergence rate question has been studied recently in [1], [11], [16], [17]. In [9], [11] convergence rate results are derived for distributed averaging algorithms. In [12] the question is addressed for probabilistic gossiping algorithms. A modified gossiping algorithm intended to speed up convergence is proposed in [18] without proof of correctness, but with convincing experimental results. The algorithm has recently been analyzed in [19]. Recent results concerning convergence rates appear in [13], [20]–[22] for periodic gossiping and in [1], [11], [23] for deterministic aperiodic gossiping.

Certain types of stochastic and doubly stochastic matrices have the property that *any* sequence of products of such matrices of the form $S_1, S_2S_1, S_3S_2S_1, \dots$ converges exponentially fast. In Section II we explicitly characterize the largest classes of stochastic and doubly stochastic matrices with positive diagonal entries which have these properties. We call these classes "convergable". The main goal of this paper is to find a "semi-norm" with respect to which matrices from these convergability classes are contractions. The role played by semi-norms in characterizing convergence rates is explained in Section III. Three different types of semi-norms are considered. Each is compared to the well known coefficient of ergodicity which plays a central role in the study of convergence rates for nonhomogeneous Markov chains [15]. Somewhat surprisingly, for doubly stochastic matrices it turns out that a particular Euclidean semi-norm on $\mathbb{R}^{n\times n}$ has the required property - namely that in this semi-norm, any doubly stochastic matrix S in the convergability class of all doubly stochastic matrices is a contraction. This particular semi-norm turns out to be the second largest singular value of S.

A. *Stochastic and Doubly Stochastic Matrices*

The type of matrices typically encountered in a consensus process [4] modeled by (1) have only nonnegative entries and row sums all equal one. Matrices with these properties are called *stochastic*. *Doubly stochastic* matrices are stochastic matrices with the additional property that their column sums are also all equal to one. Doubly stochastic matrices are typically encountered when (1) represents a distributed averaging [9] or gossiping [12] process. It is easy to see that a nonnegative matrix S is stochastic if and only if $S\mathbf{1} = \mathbf{1}$ where $\mathbf{1} \in \mathbb{R}^n$ is a column vector whose entries are all ones. Similarly a nonnegative matrix S is doubly stochastic if and only if $S\mathbf{1} = \mathbf{1}$ and $S'\mathbf{1} = \mathbf{1}$. Using these characterizations it is easy to prove that the class of stochastic matrices in $\mathbb{R}^{n\times n}$ is compact and closed under multiplication as is the class of doubly stochastic matrices

Some of the research in this paper was presented at a workshop entitled "A Celebration of the Field of Systems and Control" which was held in Stockholm, Sweden, September, 2009; that research was subsequently published in shortened form in [1]. J. Liu and A. S. Morse are with Yale University, USA ({ji.liu, as.morse}@yale.edu). B. D. O. Anderson and C. Yu are with the Australian National University and National ICT Australia Ltd., Australia ({brian.anderson, brad.yu}@anu.edu.au). C. Yu is also with Shandong Computer Science Center, Jinan, China. The research of the first two authors is supported by the US Army Research Office, the US Air Force Office of Scientific Research, and the National Science Foundation. B. D. O. Anderson is supported by Australian Research Council's Discovery Project DP-110100538 and National ICT Australia-NICTA. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. C. Yu is supported by the Australian Research Council through a Queen Elizabeth II Fellowship and DP-110100538 and Overseas Expert Program of Shandong Province. B. D. O. Anderson and C. Yu are also supported by the US Air Force Research laboratory grant number FA2386-10-1-4102.

*Proofs of some assertions in this paper are omitted due to space limitations; they are available from the first author upon request and will be given in a full length version of this paper.

in $\mathbb{R}^{n\times n}$. It is also true that the class of nonnegative matrices in $\mathbb{R}^{n\times n}$ with positive diagonal entries is closed under multiplication. Stochastic and doubly stochastic matrices with positive diagonal entries are commonly encountered in the study of consensus processes; positive diagonal entries greatly simplify convergence analysis.

Mathematically, reaching a consensus means that the state vector $x(t)$ appearing in (1) converges to a limit vector of the form $\alpha\mathbf{1}$ were α is a number depending on the initial value of x. This will always be the case if the infinite sequence of matrix products $M(1), M(2)M(1), M(3)M(2)M(1), \ldots$ converges to a matrix of the form $\mathbf{1}c$ in which case $\alpha = cx(1)$. It should be clear from what has just been stated that if the $M(t)$ are all doubly stochastic, then $c = \frac{1}{n}\mathbf{1}'$ which means that in this case α is the average of the values of the entries in $x(1)$. Thus to study convergence of the consensus process modeled by (1), it suffices to study the convergence of infinite sequences of products of stochastic and doubly stochastic matrices. Such sequences are closely related to what are called "nonhomogeneous Markov chains" for which there is a substantial literature [15]. Notwithstanding this, the following question remains. What determines the convergence rate of such sequences? This is the question which will be considered in the sequel. We begin with a few basic ideas.

B. Graph of a Stochastic Matrix

Many properties of a stochastic matrix can be usefully described in terms of an associated directed graph determined by the matrix. The *graph* of nonnegative matrix $M \in \mathbb{R}^{n\times n}$, written $\gamma(M)$, is a directed graph on n vertices with an arc from vertex i to vertex j just in case $m_{ji} \neq 0$; if (i, j) is such an arc, we say that i is a *neighbor* of j and that j is an *observer* of i. Thus $\gamma(M)$ is that directed graph whose adjacency matrix is the transpose of the matrix obtained by replacing all nonzero entries in M with ones.

C. Connectivity

There are various notions of connectivity which are useful in the study of the convergence of products of stochastic matrices. Perhaps the most familiar of these is the idea of "strong connectivity". A directed graph is *strongly connected* if there is a directed path between each pair of distinct vertices. A directed graph is *weakly connected* if there is an undirected path between each pair of distinct vertices. There are other notions of connectivity which are also useful in this context. To define several of them, let us agree to call a vertex i of a directed graph $\mathbb{G}$, a *root* of $\mathbb{G}$ if for each other vertex j of $\mathbb{G}$, there is a directed path from i to j. Thus i is a root of $\mathbb{G}$ if it is the root of a directed spanning tree of $\mathbb{G}$. We will say that $\mathbb{G}$ is *rooted at* i if i is in fact a root. Thus $\mathbb{G}$ is rooted at i just in case each other vertex of $\mathbb{G}$ is *reachable* from vertex i along a directed path within the graph. $\mathbb{G}$ is *strongly rooted at* i if each other vertex of $\mathbb{G}$ is reachable from vertex i along a directed path of length 1. Thus $\mathbb{G}$ is strongly rooted at i if i is a neighbor of every other vertex in the graph. By a *rooted graph* is meant a directed graph which possesses at least one root. A *strongly rooted graph* is a graph which has at least one vertex at which it is strongly rooted. Note that a nonnegative matrix $M \in \mathbb{R}^{n\times n}$ has a strongly rooted graph if and only if it has a positive column. Note that every strongly connected graph is rooted and every rooted graph is weakly connected. The converse statements are false. In particular there are weakly connected graphs which are not rooted and rooted graphs which are not strongly connected.

D. Composition

Since we will be interested in products of stochastic matrices, we will be interested in graphs of such products and how they are related to the graphs of the matrices comprising the products. For this we need the idea of "composition" of graphs. Let $\mathbb{G}_p$ and $\mathbb{G}_q$ be two directed graphs with vertex set $\mathcal{V}$. By the *composition* of $\mathbb{G}_p$ with $\mathbb{G}_q$, written $\mathbb{G}_q \circ \mathbb{G}_p$, is meant the directed graph with vertex set $\mathcal{V}$ and arc set defined in such a way so that (i, j) is an arc of the composition just in case there is a vertex k such that (i, k) is an arc of $\mathbb{G}_p$ and (k, j) is an arc of $\mathbb{G}_q$. Thus (i, j) is an arc in $\mathbb{G}_q \circ \mathbb{G}_p$ if and only if i has an observer in $\mathbb{G}_p$ which is also a neighbor of j in $\mathbb{G}_q$. Note that composition is an associative binary operation; because of this, the definition extends unambiguously to any finite sequence of directed graphs $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_k$ with the same vertex set.

Composition and matrix multiplication are closely related. In particular, the graph of the product of two nonnegative matrices $M_1, M_2 \in \mathbb{R}^{n\times n}$ is equal to the composition of the graphs of the two matrices comprising the product. In other words, $\gamma(M_2M_1) = \gamma(M_2) \circ \gamma(M_1)$.

If we focus exclusively on graphs with self-arcs at all vertices, more can be said. In this case the definition of composition implies that the arcs of both $\mathbb{G}_p$ and $\mathbb{G}_q$ are arcs of $\mathbb{G}_q \circ \mathbb{G}_p$; the converse is false. The definition of composition also implies that if $\mathbb{G}_p$ has a directed path from i to k and $\mathbb{G}_q$ has a directed path from k to j, then $\mathbb{G}_q \circ \mathbb{G}_p$ has a directed path from i to j. These implications are consequences of the requirement that the vertices of the graphs in question have self-arcs at all vertices. It is worth emphasizing that the union of the arc sets of a sequence of graphs $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_k$ with self-arcs must be contained in the arc set of their composition. However the converse is not true in general and it is for this reason that composition rather than union proves to be the more useful concept for our purposes.

II. CONVERGABILITY

It is of obvious interest to have a clear understanding of what kinds of stochastic matrices within an infinite product guarantee that the infinite product converges. There are many ways to address this issue and many existing results. Here we focus on just one issue.

Let $\mathcal{S}$ denote the set of all stochastic matrices in $\mathbb{R}^{n\times n}$ with positive diagonal entries. Call a compact subset $\mathcal{M} \subset \mathcal{S}$ *convergable* if for each infinite sequence of matrices $M_1, M_2, M_3, \ldots$ from $\mathcal{M}$, the sequence of products

$M_1, M_2M_1, M_3M_2M_1, \ldots$ converges exponentially fast to a matrix of the form $\mathbf{1}c$. Convergability can be characterized as follows.

Theorem 1: Let $\mathcal{R}$ denote the set of all matrices in $\mathcal{S}$ with rooted graphs. Then a compact subset $\mathcal{M} \subset \mathcal{S}$ is convergable if and only if $\mathcal{M} \subset \mathcal{R}$.

The theorem implies that $\mathcal{R}$ is the largest subset of $n \times n$ stochastic matrices with positive diagonal entries whose compact subsets are all convergable. $\mathcal{R}$ itself is not convergable because it is not closed and thus not compact.

Proof of Theorem 1: The fact that any compact subset of $\mathcal{R}$ is convergable is more or less well known from the work reported in [24]; the statement also follows from Proposition 11 of [25]. To prove the converse, suppose that $\mathcal{M} \subset \mathcal{S}$ is convergable. Then by continuity, every sufficiently long product of matrices from $\mathcal{M}$ must be a matrix with a positive column. Therefore, the graph of every sufficiently long product of matrices from $\mathcal{M}$ must be strongly rooted. It follows from Proposition 5 of [25] that $\mathcal{M}$ must be a subset of $\mathcal{R}$. $\blacksquare$

Although doubly stochastic matrices are stochastic, convergability for classes of doubly stochastic matrices has a different characterization than it does for classes of stochastic matrices. Let $\mathcal{D}$ denote the set of all doubly stochastic matrices in $\mathcal{S}$. In the sequel we will prove the following theorem.

Theorem 2: Let $\mathcal{W}$ denote the set of all matrices in $\mathcal{D}$ with strongly connected graphs. Then a compact subset $\mathcal{M} \subset \mathcal{D}$ is convergable if and only if $\mathcal{M} \subset \mathcal{W}$.

The theorem implies that $\mathcal{W}$ is the largest subset of $n \times n$ doubly stochastic matrices with positive diagonal entries whose compact subsets are all convergable. Like $\mathcal{R}$, $\mathcal{W}$ is not convergable because it is not compact. Results which more or less imply the sufficiency of strong connectivity can be found in [24] and elsewhere. Note that sufficiency is also implied by Theorem 1 because doubly stochastic matrices with strongly connected graphs are stochastic matrices with rooted graphs. It remains therefore, to prove the necessity of Theorem 2. This will be done in the sequel.

An interesting set of stochastic matrices in $\mathcal{S}$ whose compact subsets are known to be convergable, is the set of all "scrambling matrices". A matrix $S \in \mathcal{S}$ is *scrambling* if for each distinct pair of integers i and j, there is a column k of S for which s_{ik} and s_{jk} are both nonzero [15]. In graph theoretic terms S is a scrambling matrix just in case its graph is "neighbor shared" where by *neighbor shared* we mean that each distinct pair of vertices in the graph share a common neighbor [25]. Convergability of compact subsets of scrambling matrices is tied up with the concept of the coefficient of ergodicity [15] which for a given stochastic matrix $S \in \mathcal{S}$ is defined by the formula

$$\tau(S) = \frac{1}{2}\max_{i,j}\sum_{k=1}^{n}|s_{ik} - s_{jk}| \tag{2}$$

It is known that $0 \leq \tau(S) \leq 1$ for all $S \in \mathcal{S}$ and that

$$\tau(S) < 1 \tag{3}$$

if and only if S is a scrambling matrix. It is also known that

$$\tau(S_2S_1) \leq \tau(S_2)\tau(S_1), \quad S_1, S_2 \in \mathcal{S} \tag{4}$$

It can be shown that (3) and (4) are sufficient conditions to ensure that any compact subset of scrambling matrices is convergable. But $\tau(\cdot)$ has another role. It provides a worst case convergence rate for any infinite product of scrambling matrices from a given compact set $\mathcal{C} \subset \mathcal{S}$. In particular, it can be easily shown that as $i \to \infty$, any product $S_iS_{i-1}\cdots S_2S_1$ of scrambling matrices $S_i \in \mathcal{C}$ converges to a matrix of the form $\mathbf{1}c$ as fast as λ^i converges to zero where

$$\lambda = \max_{S\in\mathcal{C}}\tau(S)$$

This preceding discussion suggests the following question. Can analogs of the coefficient of ergodicity satisfying formulas like (3) and (4) be found for the set of stochastic matrices with rooted graphs or perhaps for the set of doubly stochastic matrices with strongly connected graphs? In the sequel we will provide a partial answer to this question for the case of stochastic matrices and a complete answer for the case of doubly stochastic matrices. Our approach will be to appeal to certain types of semi-norms of stochastic matrices.

III. Semi-norms

Let $||\cdot||_p$ be the induced p-norm on $\mathbb{R}^{m\times n}$. In this paper we will be primarily interested in the cases $p = 1, 2, \infty$. Note that for a nonnegative matrix A

$$\begin{aligned}
||A||_1 &= \text{max column sum } A \\
||A||_2 &= \sqrt{\mu(A'A)} \\
||A||_\infty &= \text{max row sum } A
\end{aligned}$$

where $\mu(A'A)$ is the largest eigenvalue of $A'A$; that is, the square of the largest singular value of A. For any integer $p > 0$ and matrix $M \in \mathbb{R}^{m\times n}$ define

$$|M|_p = \min_{c\in\mathbb{R}^{1\times n}} ||M - \mathbf{1}c||_p$$

As defined, $|\cdot|_p$ is nonnegative and $|M|_p \leq ||M||_p$; clearly $|\mu M|_p = |\mu|\,|M|_p$ for all real numbers μ so $|\cdot|_p$ is "positively homogeneous" [26]. Let M_1 and M_2 be matrices in $\mathbb{R}^{m\times n}$ and let c_0, c_1, and c_2 denote values of c which minimize $||M_1+M_2-\mathbf{1}c||_p$, $||M_1-\mathbf{1}c||_p$, and $||M_2-\mathbf{1}c||_p$ respectively. Note that

$$\begin{aligned}
|M_1 + M_2|_p &= ||M_1 + M_2 - \mathbf{1}c_0||_p \\
&\leq ||M_1 + M_2 - \mathbf{1}(c_1 + c_2)||_p \\
&\leq ||M_1 - \mathbf{1}c_1||_p + ||M_2 - \mathbf{1}c_2||_p \\
&= |M_1|_p + |M_2|_p
\end{aligned}$$

Thus the triangle inequality holds. These properties mean that $|\cdot|_p$ is a *semi-norm.* $|\cdot|_p$ behaves much like a norm. For example, if N is a submatrix of M, then $|N|_p \leq |M|_p$. However $|\cdot|_p$ is not a norm because $|M|_p = 0$ does not imply $M = 0$; rather it implies that $M = \mathbf{1}c$ for some row vector c which minimizes $||M - \mathbf{1}c||_p$. For our purposes, $|\cdot|_p$ has a particularly important property:

Lemma 1: Suppose $\mathcal{M}$ is a subset of $\mathbb{R}^{n\times n}$ such that $M\mathbf{1} = \mathbf{1}$ for all $M \in \mathcal{M}$. Then

$$|M_2M_1|_p \leq |M_2|_p|M_1|_p \tag{5}$$

Proof of Lemma 1: Let c_0, c_1, and c_2 denote values of c which minimize $||M_2M_1 - \mathbf{1}c||_p$, $||M_1 - \mathbf{1}c||_p$, and $||M_2 - \mathbf{1}c||_p$ respectively. Then

$$\begin{aligned}
|M_2M_1|_p &= ||M_2M_1 - \mathbf{1}c_0||_p \\
&\leq ||M_2M_1 - \mathbf{1}(c_2M_1 + c_1 - c_2\mathbf{1}c_1)||_p \\
&= ||M_2M_1 - \mathbf{1}c_2M_1 - M_2\mathbf{1}c_1 + \mathbf{1}c_2\mathbf{1}c_1)||_p \\
&= ||(M_2 - \mathbf{1}c_2)(M_1 - \mathbf{1}c_1)||_p \\
&\leq ||(M_2 - \mathbf{1}c_2)||_p||(M_1 - \mathbf{1}c_1)||_p \\
&= |M_2|_p|M_1|_p
\end{aligned}$$

Thus (5) is true. ■

We say that $M \in \mathbb{R}^{n\times n}$ is *semi-contractive* in the p-norm if $|M|_p < 1$. In view of Lemma 1, the product of semi-contractive matrices in $\mathcal{M}$ is thus semi-contractive. The importance of these ideas lies in the following fact.

Proposition 1: Suppose $\mathcal{M}$ is a subset of $\mathbb{R}^{n\times n}$ such that $M\mathbf{1} = \mathbf{1}$ for all $M \in \mathcal{M}$. Let p be fixed and let $\bar{\mathcal{M}}$ be a compact set of semi-contractive matrices in $\mathcal{M}$. Let

$$\lambda = \sup_{\bar{\mathcal{M}}} |M|_p$$

Then for each infinite sequence of matrices $M_i \in \bar{\mathcal{M}}$, $i \in \{1, 2, \ldots\}$, the matrix product $M_iM_{i-1}\cdots M_1$ converges as $i \to \infty$ as fast as λ^i converges to zero, to a rank one matrix of the form $\mathbf{1}\bar{c}$.

A. The case $p = 1$

We now consider in more detail the case when $p = 1$. For this case it is possible to derive an explicit formula for the semi-norm $|M|_1$ of a nonnegative matrix $M \in \mathbb{R}^{n\times n}$.

Proposition 2: Let q be the unique integer quotient of n divided by 2. Let $M \in \mathbb{R}^{n\times n}$ be a nonnegative matrix. Then

$$|M|_1 = \max_{j\in\{1,2,\ldots,n\}} \left\{ \sum_{i\in\mathcal{L}_j} m_{ij} - \sum_{i\in\mathcal{S}_j} m_{ij} \right\}$$

where $\mathcal{L}_j$ and $\mathcal{S}_j$ are respectively the row indices of the q largest and q smallest entries in the jth column of M.

Consider now the case when M is a doubly stochastic matrix S, more can be said:

Theorem 3: Let q be the unique integer quotient of n divided by 2. Let $S \in \mathbb{R}^{n\times n}$ be a doubly stochastic matrix. Then $|S| \leq 1$. Moreover S is a semi-contraction in the one-norm if and only if the number of nonzero entries in each column of S exceeds q.

Note that the doubly stochastic matrix

$$S = \begin{bmatrix} .5 & .125 & .125 & .125 & .125 \\ .5 & .125 & .125 & .125 & .125 \\ 0 & .25 & .25 & .25 & .25 \\ 0 & .25 & .25 & .25 & .25 \\ 0 & .25 & .25 & .25 & .25 \end{bmatrix}$$

has a strongly connected graph but is not a semi-contractions for $p = 1$. Thus this particular semi-norm is not as useful as we would like because there are matrices in $\mathcal{W}$ which are not semi-contractions for $p = 1$.

It is possible to compare this semi-norm with the coefficient of ergodicity. Observe that while the preceding matrix is not a semi-contraction it is a scrambling matrix. Thus for this example, $\tau(S) < |S|_1 = 1$. On the other hand there are also doubly stochastic matrices which are semi-contractions but which are not scrambling matrices. An example of this is the matrix

$$S = \begin{bmatrix} .5 & 0 & 0 & 0 & .5 & 0 \\ 0 & .5 & 0 & 0 & 0 & .5 \\ .125 & .125 & .25 & .25 & .125 & .125 \\ .125 & .125 & .25 & .25 & .125 & .125 \\ .125 & .125 & .25 & .25 & .125 & .125 \\ .125 & .125 & .25 & .25 & .125 & .125 \end{bmatrix}$$

Thus for this example, $|S|_1 < \tau(S) = 1$, which means that there are situations when it may be more advantageous to use the semi-norm $|\cdot|_1$ to compute convergence rates than to appeal to the coefficient of ergodicity.

B. The case $p = \infty$

Note that in this case $|S|_\infty \leq 1$ for any stochastic matrix because $|S|_\infty \leq ||S||_\infty = 1$. Although not at all obvious, it turns out that $|S|_\infty$ equals the coefficient of ergodicity discussed earlier and defined by (2). This is an immediate consequence of Proposition 3 which is stated below. Unfortunately, the last example in the preceding subsection shows that there are doubly stochastic matrices with strongly connected graphs which are not scrambling matrices. Thus this particular semi-norm is also not as useful as we might hope for.

Proposition 3: Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative matrix. Then

$$|A|_\infty = \frac{1}{2}\max_{i,j}\sum_{k=1}^{n}|a_{ik} - a_{jk}|$$

C. The case $p = 2$

For the case when $p = 2$ it is also possible to derive an explicit formula for the semi-norm $|M|_2$ of a nonnegative matrix $M \in \mathbb{R}^{n\times n}$. Towards this end note that for any $x \in \mathbb{R}^n$, the function $g(x, c) = x'(M - \mathbf{1}c)'(M - \mathbf{1}c)x$ attains its minimum with respect to c at $\frac{1}{n}\mathbf{1}'M$. This implies that

$$|M|_2 = ||PM||_2 = \sqrt{\mu\{M'P'PM\}}$$

where $P = I - \frac{1}{n}\mathbf{1}\mathbf{1}'$ and, for any symmetric matrix T, $\mu\{T\}$ is the largest eigenvalue of T. We are led to the following result.

Proposition 4: Let $M \in \mathbb{R}^{n\times n}$ be a nonnegative matrix. Then $|M|_2$ is the largest singular value of the matrix PM where P is the orthogonal projection on the orthogonal complement of the span of $\mathbf{1}$.

Now suppose that M is a doubly stochastic matrix S. Then $S'S$ is also doubly stochastic and $\mathbf{1}'S = \mathbf{1}'$. The latter and Proposition 4 imply that

$$|S|_2 = \sqrt{\mu\left\{S'S - \frac{1}{n}\mathbf{1}\mathbf{1}'\right\}} \tag{6}$$

More can be said:

Lemma 2: If S is doubly stochastic, then $\mu\{S'S - \frac{1}{n}\mathbf{1}\mathbf{1}'\}$ is the second largest eigenvalue of $S'S$.

Proof of Lemma 2: Since $S'S$ is symmetric it has orthogonal eigenvectors one of which is $\mathbf{1}$. Let $\mathbf{1}, x_2, \ldots, x_n$ be such a set of eigenvectors with eigenvalues $1, \lambda_2, \ldots, \lambda_n$. Then $S'S\mathbf{1} = \mathbf{1}$ and $S'Sx_i = \lambda_i x_i,\ i \in \{2, 3, \ldots, n\}$. Clearly $(S'S - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{1} = 0$ and $(S'S - \frac{1}{n}\mathbf{1}\mathbf{1}')x_i = \lambda_i x_i,\ i \in \{2, 3, \ldots, n\}$. Since 1 is the largest eigenvalue of $S'S$ it must therefore be true that the second largest eigenvalue $S'S$ is the largest eigenvalue of $S'S - \frac{1}{n}\mathbf{1}\mathbf{1}'$. $\blacksquare$

We summarize:

Theorem 4: For $p = 2$ the semi-norm of a doubly stochastic matrix S is the second largest singular value of S.

There is another way to think about what this theorem implies. Prompted by the work in [9] and [11], suppose one wants to measure in the sense of a 2-norm $||\cdot||$, how much closer an n-vector x gets to the average vector $z = \frac{1}{n}\mathbf{1}\mathbf{1}'x$ when it is multiplied by a doubly stochastic matrix S. In other words how does the norm $||Sx - z||$ compare with $||x - z||$? To address this question, note first that $x - z \in \mathcal{O}$ where $\mathcal{O}$ is the orthogonal complement of the span of $\mathbf{1}$. Note next that

$$||Sx - z||^2 = ||S(x - z)||^2 \leq \left(\sup_{y\in\mathcal{O}} \frac{y'S'Sy}{y'y}\right)||x - z||^2$$

But $\sup_{y\in\mathcal{O}} \frac{y'S'Sy}{y'y}$ is the second largest eigenvalue of $S'S$ which in turn is the square of the second largest singular value of S. In other words, $||Sx - z|| \leq |S|_2||x - z||$. Thus Sx is always as close to the average vector z as x is and is even closer if $|S|_2$ is a contraction.

In the light of Theorem 4, we are now in a position to characterize in graph theoretic terms those doubly stochastic matrices with positive diagonal entries which are semi-contractions for $p = 2$.

Theorem 5: Let S be a doubly stochastic matrix with positive diagonal entries. Then $|S|_2 \leq 1$. Moreover S is a semi-contraction in the 2-norm if and only if the graph of S is strongly connected.

To prove this theorem we need several concepts and results. Let $\mathbb{G}$ denote a directed graph and write $\mathbb{G}'$ for that graph which results when the arcs in $\mathbb{G}$ are reversed; i.e., the *dual graph*. Call a graph *symmetric* if it is equal to its dual. Note that in the case of a symmetric graph, the three properties of being rooted, strongly connected, and weakly connected are equivalent. Note also that if $\mathbb{G}$ is the graph of a nonnegative matrix M with positive diagonal entries, then $\mathbb{G}'$ is the graph of M' and $\mathbb{G}' \circ \mathbb{G}$ is the graph of $M'M$.

Lemma 3: A directed graph $\mathbb{G}$ with self-arcs at all vertices is weakly connected if and only if $\mathbb{G}' \circ \mathbb{G}$ is strongly connected.

Lemma 4: Let T be a stochastic matrix with positive diagonal entries. If T has a strongly connected graph, then the magnitude of its second largest eigenvalue is less than one. If, on the other hand, the magnitude of the second largest eigenvalue of T is less than one, then the graph of T is weakly connected.

Lemma 5: The graph $\mathbb{G}$ of a doubly stochastic matrix D is strongly connected if and only if it is weakly connected.[1]

The proof of Lemma 5 which follows is based on ideas from [15] and [27]. Let $\mathbb{G}$ be a directed graph with vertex set $\mathcal{V} = \{1, 2, \ldots, n\}$. Call a vertex j is *reachable* from i if either $j = i$ or if there is a directed path from i to j. Call a vertex i *essential* if i is reachable from all vertices which are reachable from i.

Lemma 6: Every directed graph has at least one essential vertex.

To proceed, let us say that vertices i and j are *mutually reachable* if each is reachable from the other. Mutual reachability is clearly an equivalence relation on $\mathcal{V}$ which partitions $\mathcal{V}$ into the disjoint union of a finite number of equivalence classes. Note that if i is an essential vertex of $\mathbb{G}$, then every vertex in the equivalence class of i is also essential. Thus every directed graph possesses at least one mutually reachable equivalence class whose members are all essential.

Proof of Lemma 5: Strong connectivity clearly implies weak connectivity. We prove the converse. Suppose $\mathbb{G}$ is weakly connected. In view of the proceeding, $\mathbb{G}$ has at least one mutually reachable equivalence class $\mathcal{E}$ whose members are all essential. If $\mathcal{E} = \mathcal{V}$, then $\mathbb{G}$ is obviously strongly connected. Thus to prove the lemma, it is enough to show that $\mathcal{E} = \mathcal{V}$. Suppose the contrary, namely that $\mathcal{E} = \{i_1, i_2, \ldots, i_m\}$ is a strictly proper subset of $\mathcal{V}$. Let π be any permutation map for which $\pi(i_j) = j,\ j \in \{1, 2, \ldots, m\}$ and let P be the corresponding permutation matrix. Then clearly

$$P'DP = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

and $P'DP$ is doubly stochastic. Since $P'DP$ is doubly stochastic, the column sums of A must all equal one as must the row sums of the submatrix $[\,A \quad B\,]$. But the transformation $D \longmapsto P'DP$ corresponds to a relabeling of the vertices of $\mathbb{G}$, so the graph of $P'DP$ must also be weakly connected. Thus means that B cannot be the zero matrix. Therefore the sum of the row sums of A must be less than m. But this contradicts the fact that the sum of the column sums of A equals m. Therefore $\mathcal{E} = \mathcal{V}$. $\blacksquare$

Proof of Theorem 5: Let S be a doubly stochastic matrix with positive diagonal entries. Then 1 is the largest singular value of S because $S'S$ is doubly stochastic. From this and Theorem 4 it follows that $|S|_2 \leq 1$.

Suppose S is a semi-contraction. Then in view of Theorem 4, the second largest eigenvalue of $S'S$ is less than 1. Thus by Lemma 4, the graph of $S'S$ is weakly connected. But

[1] It is clear that strong connectivity of $\mathbb{G}$ implies weak connectivity of $\mathbb{G}$. The converse was conjectured by John Tsitsiklis in a private communication.

$S'S$ is symmetric so its graph must be strongly connected. Therefore by Lemma 3, the graph of S is weakly connected. In view of Lemma 5, the graph of S is strongly connected.

Now suppose that the graph of S is strongly connected. Then S is weakly connected so the graph of $S'S$ is strongly connected because of Lemma 3. Thus by Lemma 4, the magnitude of the second largest eigenvalue of $S'S$ is less than 1. From this and Theorem 4 it follows that S is a semi-contraction. ■

Proof of Theorem 2: Let $\mathcal{M}$ be any compact subset of $\mathcal{W}$. In view of Theorem 5, each matrix in $\mathcal{M}$ is a semi-contraction in the two-norm. From this and Proposition 1, it follows that $\mathcal{M}$ is convergable.

Now suppose that $\mathcal{M}$ is convergable and let S be a matrix in $\mathcal{M}$. Then S^i converges to a matrix of the form $\mathbf{1}c$ as $i \to \infty$. This means that the second largest eigenvalue of S must be less than 1 in magnitude. Thus by Lemma 4, S must have a weakly connected graph. By Lemma 5, the graph of S must be strongly connected. ■

The importance of Theorem 5 lies in the fact that the matrices in every convergable set of doubly stochastic matrices are contractions in the 2-norm. In view of Proposition 1, this enables one to immediately compute a rate of convergence for any infinite product of matrices from any given convergable set. The coefficient of ergodicity mentioned earlier does not have this property. If it did, then every doubly stochastic matrix with a strongly connected graph would have to be a scrambling matrix. The following counterexample shows that this is not the case:

$$S = \begin{bmatrix} .5 & .25 & 0 & 0 & 0 & .25 \\ .25 & .5 & 0 & 0 & 0 & .25 \\ 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & .5 & .25 & 0 & .25 \\ 0 & 0 & 0 & 0 & .875 & .125 \\ .25 & .25 & 0 & .25 & .125 & .125 \end{bmatrix}$$

In particular, S is a doubly stochastic matrix with a strongly connected graph but it is not a scrambling matrix.

IV. Concluding Remarks

In this paper we have identified the largest "alphabets" of stochastic and doubly stochastic matrices with positive diagonal entries whose "words" converge exponentially fast as word length increases. In the case of double stochastic matrices, each matrix in the corresponding alphabet is shown to be a semi-contraction in the two-norm. In the case of stochastic matrices which are not doubly stochastic, we were not similarly successful and the problem of discovering a suitable semi-contraction for this case remains unresolved.

References

[1] J. Liu, A. S. Morse, B. D. O. Anderson, and C. Yu. The contraction coefficient of a complete gossip sequence. In *Three Decades of Progress in Control Sciences*, pages 275–290. Springer, 2010.

[2] J. N. Tsitsiklis. *Problems in decentralized decision making and computation*. PhD thesis, Department of Electrical Engineering and Computer Science, MIT, 1984.

[3] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

[4] 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.

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

[6] V. D. Blondel, J. M. Hendrichkx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 2996–3000, 2005.

[7] L. Moreau. Stability of multi-agent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005.

[8] W. Ren and R. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, 50(5):655–661, 2005.

[9] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53(1):65–78, 2004.

[10] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proceedings of the 4th International Conference on Information Processing in Sensor Networks*, pages 63–70, 2005.

[11] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis. On distributed averaging algorithms and quantization effects. *IEEE Transactions on Automatic Control*, 54(11):2506–2517, 2009.

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

[13] B. D. O. Anderson, C. Yu, and A. S. Morse. Convergence of periodic gossiping algorithms. In *Perspectives in Mathematical System Theory, Control, and Signal Processing*, pages 127–138. Springer, 2010.

[14] J. Liu, S. Mou, A. S. Morse, B. D. O. Anderson, and C. Yu. Request-based gossiping. In *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011. to be published.

[15] E. Seneta. *Non-negative matrices and Markov chains*. Sringer, 2006.

[16] M. Cao, D. A. Spielman, and A. S. Morse. A lower bound on convergence of a distributed network consencus algorithm. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 2356–2361, 2005.

[17] A. Olshevsky and J. N. Tsitsiklis. Convergence rates in distributed consensus and averaging. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 3387–3392, 2006.

[18] M. Cao, D. A. Spielman, and E. M. Yeh. Accelerated gossip algorithms for distributed computation. In *Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computing*, pages 952–959, 2006.

[19] J. Liu, B. D. O. Anderson, M. Cao, and A. S. Morse. Analysis of accelerated gossip algorithms. In *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 871–876, 2009.

[20] S. Mou, C. Yu, B. D. O. Anderson, and A. S. Morse. Deterministic gossiping with a periodic protocol. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 5787–5791, 2010.

[21] F. He, A. S. Morse, J. Liu, and S. Mou. Periodic gossiping. In *Proceedings of the 18th IFAC World Congress*, 2011. to be published.

[22] C. Yu, B. D. O. Anderson, S. Mou, J. Liu, F. He, and A. S. Morse. Gossiping periodically. *IEEE Transactions on Automatic Control*, 2011. submitted.

[23] A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal Control and Optimization*, 48(1):33–55, 2009.

[24] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proceedings of the American Mathematical Society*, 15(5):733–737, 1963.

[25] M. Cao, A. S. Morse, and B. D. O. Anderson. Reaching a consensus in a dynamically changing environment: a graphical approach. *SIAM Journal on Control and Optimization*, 47(2):575–600, 2008.

[26] R. C. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[27] R. G. Gallager. *Discrete Stochastic Processes*. Kluwer Academic Publishers, 1996.