Xiaoming Hu, Ulf Jonsson, Bo Wahlberg, and Bijoy K. Ghosh (Eds.)

Three Decades of Progress in Control Sciences

Dedicated to Chris Byrnes and Anders Lindquist

Springer
The Contraction Coefficient of a Complete Gossip Sequence*, †

J. Liu¹, A.S. Morse¹, B.D.O. Anderson², and C. Yu²

¹ Yale University, New Haven, CT 06511, USA
² Australian National University and National ICT Australia, Canberra ACT 0200, Australia

Summary. A sequence of allowable gossips between pairs of agents in a group is complete if the gossip graph which the sequence generates contains a tree spanning the graph of all allowable gossip pairs. The stochastic matrix of a sequence of allowable gossips is shown to be a contraction for an appropriately defined Euclidian seminorm, if and only if the gossip sequence is complete. The significance of this result in determining the convergence rate of an infinite aperiodic sequence of gossips is explained.

19.1 Introduction

There has been considerable interest recently in developing algorithms for distributing information among the members of a group of sensors or mobile autonomous agents via local interactions. Notable among these are those algorithms intended to cause such a group to reach a consensus in a distributed manner [9, 12, 2, 11, 13]. In a typical consensus seeking process, the agents in a given group are all trying to agree on a specific value of some quantity. Each agent initially has only limited information available. The agents then try to reach a consensus by communicating what they know to their neighbors either just once or repeatedly, depending on the specific problem of interest. One particular consensus problem which has received much attention is called "gossiping." In a typical gossiping problem each agent has control over a real-valued scalar "gossiping" variable. What distinguishes gossiping

*The authors wish to thank Shaoshuai Mou, Ming Cao and Sekhar Tatikonda for useful discussions which have contributed to this work. The research of the first two authors is supported by the US Army Research Office, the US Air Force Office of Scientific, and the National Science Foundation. B.D.O. Anderson is supported by Australian Research Council's Discovery Project DP-0877562 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 an Australian Postdoctoral Fellowship under DP-0877562.
†This paper is dedicated to Chris Byrnes and Anders Lindquist.
from more general consensus seeking is that during a gossiping process only two agents are allowed to communicate with each other at any one clock time. A pair of agents gossip by updating the current values of their gossip variables to new values which are both equal to the average of their current values. Generally not every pair of agents is allowed to gossip. The edges of a given, undirected "allowable gossip graph" specify which gossip pairs are allowable. The actual sequence of gossip pairs which occurs during a specific gossip sequence might be determined either probabilistically [4] or deterministically, depending on the problem of interest. It is the latter type of problem to which this paper is addressed.

Of particular interest is the rate at which a sequence of agent gossip variables converge to a common value. The convergence rate question for more general deterministic consensus problems has been studied recently in [6, 1]. In [4, 3, 15] the convergence rate question is addressed for gossiping algorithms in which the sequence of gossip pairs under consideration is determined probabilistically. A modified gossiping algorithm intended to speed up convergence is proposed in [7] without proof of correctness, but with convincing experimental results. The algorithm has recently been analyzed in [10].

A typical gossiping process can be modeled as a discrete time linear system of the form \( x(t+1) = M(t)x(t) \), \( t = 0, 1, \ldots \) where \( x \) is a vector of agent gossip variables and each value of \( M(t) \) is a specially structured doubly stochastic matrix (see section 19.2). A sequence of allowable gossips pairs is complete if the gossip graph which the sequence generates contains a tree spanning the graph of all allowable gossip pairs. (see section 19.6). The specific goal of this paper is to find a seminorm with respect to which the state transition matrix of any such complete sequence is a contraction. The role played by seminorms in characterizing convergence rate is explained in section 19.5. Three different types of seminorms are considered in section 19.5. 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 [14]. Somewhat surprisingly, it turns out that a particular Euclidean seminorm on \( \mathbb{R}^{n \times n} \) has the required property - namely that in this seminorm, the state transition matrix of any complete gossip sequence is a contraction. The value of the seminorm for a given gossip sequence is what is meant by the "contraction coefficient" of the sequence.

### 19.2 Gossiping

The type of gossiping we want to consider involves a group of \( n \) agents labeled 1 to \( n \). Each agent \( i \) has control over a real-valued scalar quantity \( x_i \) called a gossip variable which the agent is able to update at discrete clock times \( t = 1, 2, \ldots \). A gossip occurs at time \( t \) between agents \( i \) and \( j \) if the values of both agents variables at time \( t+1 \) equal the average of their values at time \( t \). In other words \( x_i(t+1) = x_j(t+1) = \frac{1}{2}(x_i(t) + x_j(t)) \). Generally not every pair of agents are allowed to gossip. The edges of a given simple directed \( n \)-vertex graph \( A \), called an allowable gossip graph, specify which gossip pairs are allowable. In other words a gossip between
agents \(i\) and \(j\) is allowable if \((i, j)\) is an edge in \(A\). In this paper we will stipulate that at most one allowable gossip can occur at each clock time; the value of the gossip variable of any agent which does not gossip at time \(t\), does not change at time \(t + 1\). The goal of gossiping is for the \(n\) agents to reach a consensus in the sense that all \(n\) gossip variable ultimately reach the same value in the limit as \(t \to \infty\). For this to be possible, no matter what the initial values of the gossiping variables are, it is clearly necessary that \(A\) be a connected graph, an assumption we henceforth make.

A gossiping process can be conveniently modeled as a discrete time linear system of the form \(x(t + 1) = M(t)x(t), \quad t = 0, 1, \ldots\), where \(x \in \mathbb{R}^n\) is a state vector of gossiping variable and \(M(t)\) is a matrix characterizing how \(x\) changes as the result of the gossip between two given agents at time \(t\). Because there are only a finite number of allowable gossips, there are only a finite number of values that each \(M(t)\) can take on. In particular, if agents \(i\) and \(j\) gossip at time \(t\), then \(M(t) = S_{ij}\) where \(S_{ij}\) is the \(n \times n\) matrix for which \(s_{ii} = s_{jj} = s_{ij} = s_{ji} = \frac{1}{2}, s_{kk} = 1, k \notin \{i, j\}\) and all remaining entries equal zero. Thus \(S_{ij}\) is a nonnegative matrix whose row sums and column sums are equal one. Matrices with these two properties are called doubly stochastic. Note that the type of doubly stochastic matrix which characterizes a gossip \(i.e., a\) gossip matrix \(\{\}\) has two additional properties - it is symmetric and its diagonal entries are all positive. Mathematically, reaching a consensus by means of an infinite sequence of gossips modeled by a corresponding infinite sequence of gossip matrices \(M(1), M(2), \ldots\), means that the sequence of matrix products \(M(1), M(2)M(1), M(3)M(2)M(1), \ldots\) converges to a matrix of the form \(1c\) where \(1 \in \mathbb{R}^n\) is a vector whose entries are all ones. It turns out that if convergence occurs, the limit matrix \(1c\) is also a doubly stochastic matrix which means that \(c = \frac{1}{n}1\).

### 19.3 Stochastic Matrices

Doubly stochastic matrices are special types of "stochastic matrices" where by a stochastic matrix is meant a nonnegative \(n \times n\) matrix whose row sums all equal one. It is easy to see that a nonnegative matrix \(S\) is stochastic if and only if \(SI = 1\). Similarly a nonnegative matrix \(S\) is doubly stochastic if and only if \(SI = 1\) and \(S'1 = 1\). Using these characterizations it is easy to prove that the class of stochastic matrices in \(\mathbb{R}^{n \times n}\) is closed under multiplication as is the class of doubly stochastic matrices in \(\mathbb{R}^{n \times n}\). It is also true that the class of nonnegative matrices in \(\mathbb{R}^{n \times n}\) with positive diagonals is closed under multiplication.

#### 19.3.1 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 non-negative 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 \(m_{ij} \neq 0\); 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.
19.3.2 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 $G$, a root of $G$ if for each other vertex $j$ of $G$, there is a directed path from $i$ to $j$. Thus $i$ is a root of $G$, if it is the root of a directed spanning tree of $G$. We will say that $G$ is rooted at $i$ if $i$ is in fact a root. Thus $G$ is rooted at $i$ just in case each other vertex of $G$ is reachable from vertex $i$ along a directed path within the graph. $G$ is strongly rooted at $i$ if each other vertex of $G$ is reachable from vertex $i$ along a directed path of length 1. Thus $G$ is strongly rooted at $i$ if $i$ is a neighbor of every other vertex in the graph. By a rooted graph $G$ 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.

19.3.3 Composition

Since we will be interested in products of stochastic matrices, we will be interested in graphs of such products and how the are related to the graphs of the matrices comprising the products. For this we need the idea of "composition" of graphs. Let $G_p$ and $G_q$ be two directed graphs with vertex set $\mathcal{V}$. By the composition of $G_p$ with $G_q$, written $G_q \circ G_p$, is meant the directed graph with vertex set $\mathcal{V}$ and arc set defined in such a way so that an arc of the composition just in case there is a vertex $k$ such that $(i,k)$ is an arc of $G_p$ and $(k,j)$ is an arc of $G_q$. Thus $(i,j)$ is an arc in $G_q \circ G_p$ if and only if $i$ has an observer in $G_p$ which is also a neighbor of $j$ in $G_q$. Note that composition is an associative binary operation; because of this, the definition extends unambiguously to any finite sequence of directed graphs $G_1, G_2, \ldots, G_k$ 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, $\gamma(M_2 M_1) = \gamma(M_2) \circ \gamma(M_1)$.

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 $G_p$ and $G_q$ are arcs of $G_q \circ G_p$; the converse is false. The definition of composition also implies that if $G_p$ has a directed path from $i$ to $k$ and $G_q$ has a directed path from $k$ to $j$, then $G_q \circ G_p$ 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
\( G_1, G_2, \ldots, G_k \) 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.

19.4 Convergability

It is of obvious interest to have a clear understanding of what kinds of stochastic ma-
trices 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 diagonals.
Call a compact subset \( \mathcal{M} \subset \mathcal{S} \) convergable if for each infinite sequence of matrices
\( M_1, M_2, \ldots, M_i, \ldots \) from \( \mathcal{M} \), the sequence of products \( M_1 M_1, M_2 M_2 M_1 \cdots \)
converges exponentially fast to a matrix of the form 1c. Convergability can be characterized as
follows.

**Theorem 19.4.1.** Let \( \mathcal{D} \) denote 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 \( \mathcal{M} \subset \mathcal{D} \).

The theorem implies that \( \mathcal{D} \) is the largest subset of \( n \times n \) stochastic matrices with
positive diagonals whose compact subsets are all is convergable, \( \mathcal{D} \) itself is not con-
vergable because it is not closed and thus not compact.

**Proof of Theorem 19.4.1:** The fact that any compact subset of \( \mathcal{D} \) is convergable is
an immediate consequence of Proposition 11 of [5]. 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 [5], that \( \mathcal{M} \) must be a subset of \( \mathcal{D} \). \( \square \)

Although doubly stochastic matrices are stochastic, convergability for classes of dou-
by 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 19.4.2.** Let \( \mathcal{W} \) denote set of all matrices in \( \mathcal{D} \) with weakly 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 diagonals whose compact subsets are all is convergable. Like \( \mathcal{D}, \mathcal{W} \) is
not convergable because it is not compact.

An interesting set of stochastic matrices in \( \mathcal{S} \) whose compact subsets are known
be convergable, is the set of all “scrambling matrices.” A matrix \( S \in \mathcal{S} \) is scram-
bbling if for each distinct pair of integers \( i, j \), there is a column \( k \) of \( S \) for which \( s_{ik} \)
and \( s_{jk} \) are both nonzero [14]. 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 [5]. Convergability of compact subsets of scrambling matrices is tied up with the concept of the coefficient of ergodicity [14] which for a given stochastic matrix $S \in \mathcal{S}$ is defined by the formula

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

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

$$\tau(S) < 1$$

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

$$\tau(S_2 S_1) \leq \tau(S_2) \tau(S_1), \quad S_1, S_2 \in \mathcal{S} \quad (19.1)$$

It can be shown that 19.1 and 19.2 are sufficient conditions to ensure that any compact subset of scrambling matrices is convergability. 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 $S_i S_{i-1} \cdots S_2 S_1$ of scrambling matrices $S_i \in \mathcal{C}$ converges as to a matrix of the form $1c$ as fast as $\lambda^i$ 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 19.1 and 19.2 be found for the set of stochastic matrices with rooted graphs or perhaps for the set of doubly stochastic matrices with weakly 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 seminorms of stochastic matrices.

### 19.5 Seminorms

Let $|| \cdot ||_p$ be the induced $p$ norm on $\mathbb{R}^{m \times n}$. We will be interested in $p = 1, 2, \infty$. Note that

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

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

$$|M|_p = \min_{c \in \mathbb{R}^{1 \times n}} ||M - 1c||_p$$
As defined, $|\cdot|_p$ is nonnegative and $|M|_p \leq ||M||_p$. Let $M_1$ and $M_2$ be matrices in $\mathbb{R}^{n \times n}$ and let $c_0, c_1$ and $c_2$ denote values of $c$ which minimize $||M_1 + M_2 - 1c||_p, ||M_1 - 1c||_p$ and $||M_2 - 1c||_p$ respectively. Note that

$$||M_1 + M_2||_p = ||M_1 + M_2 - 1c_0||_p \leq ||M_1 + M_2 - 1(c_1 + c_2)||_p$$

$$\leq ||M_1 - 1c_1||_p + ||M_2 - 1c_2||_p$$

$$= ||M_1||_p + ||M_2||_p.$$

Thus the triangle inequality holds which makes $|\cdot|_p$ a seminorm. $|\cdot|_p$ behaves much like a norm. For example, if $N$ is a sub-matrix 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 $M = 1c$ for some row vector $c$ which minimizes $||M - 1c||_p$. For our purposes, $|\cdot|_p$ has a particularly important property:

Lemma 19.5.1. Suppose $\mathcal{M}$ is a subset of $\mathbb{R}^{n \times n}$ such that $M1 = 1$ for all $M \in \mathcal{M}$. Then

$$||M_2M_1||_p \leq ||M_2||_p||M_1||_p \quad (19.3)$$

We say that $|\cdot|_p$ is sub-multiplicative on $\mathcal{M}$.

Proof of Lemma 19.5.1: Let $c_0, c_1$ and $c_2$ denote values of $c$ which minimize $||M_2M_1 - 1c||_p, ||M_1 - 1c||_p$ and $||M_2 - 1c||_p$ respectively. Then

$$||M_2M_1||_p = ||M_2M_1 - 1c_0||_p$$

$$\leq ||M_2M_1 - 1(c_2M_1 + c_1 - c_21c_1)||_p$$

$$= ||M_2M_1 - 1c_2M_1 - M_21c_1 + 1c_21c_1||_p$$

$$= ||(M_2 - 1c_2)(M_1 - 1c_1)||_p$$

$$\leq ||M_2 - 1c_2||_p||M_1 - 1c_1||_p$$

$$= ||M_2||_p||M_1||_p$$

Thus 19.3 is true. $\square$

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 19.5.1, the product of semi-contractive matrices in $\mathcal{M}$ is thus semi-contractive. The importance of these ideas lies in the following fact.

Proposition 19.5.1. Suppose $\mathcal{M}$ is a subset of $\mathbb{R}^{n \times n}$ such that $M1 = 1$ for all $M \in \mathcal{M}$. Let $p$ be fixed and let $\mathcal{M}$ be a compact set of semi-contractive matrices in $\mathcal{M}$. Let

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

Then for each infinite sequence of matrices $M_i \in \mathcal{M}$, $i \in \{1, 2, \ldots \}$, the matrix product

$$M_1M_{i-1} \cdots M_1$$

converges as fast as $\lambda_i$ to a rank one matrix of the form $1c$.

Proof of Proposition 19.5.1: To be given in the full length version of this paper.
19.5.1 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 seminorm $|M|_1$ of a nonnegative matrix $M \in \mathbb{R}^{n \times n}$.

**Proposition 19.5.2.** Let $q$ be the unique integer quotient of $n$ divided by 2. Let $M \in \mathbb{R}^{n \times m}$ be a non-negative 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 $j$th column of $M$.

This result is a direct consequence of the following lemma and the definition of $|\cdot|_1$.

**Lemma 19.5.2.** Let $q$ denote the unique integer quotient of $n$ divided by 2. Let $y$ be a non-negative $n$ vector and write $\mathcal{L}$ and $\mathcal{S}$ for the row indices of the $q$ largest and $q$ smallest entries in $y$ respectively. Then

$$|y|_1 = \sum_{i \in \mathcal{L}} y_i - \sum_{i \in \mathcal{S}} y_i \quad (19.4)$$

where $y_i$ is the $i$th entry in $y$.

**Proof of Lemma 19.5.2:** To be given in the full length version of this paper.

Consider now the case when $M$ is a doubly stochastic matrix $S$. Then the column sums of $S$ are all equal to 1. This implies that $|S|_1 \leq 1$ because $|S|_1 \leq \|S\|_1 \leq 1$. The column sums all equaling one also implies that

$$\sum_{i \in \mathcal{L}} s_{ij} + rs_{(q+r)j} + \sum_{i \in \mathcal{S}} a_{ij} = 1, \quad j \in \{1, 2, \ldots, n\}$$

Therefore

$$|S|_1 = \max_{i \in \{1, 2, \ldots, n\}} \left\{ 2 \sum_{i \in \mathcal{L}} s_{ij} + rs_{(q+r)j} - 1 \right\}$$

This means that $S$ is semi-contractive in the one-norm just in case

$$\sum_{i \in \mathcal{L}} s_{ij} + \frac{r}{2}s_{(q+r)j} < 1, \quad j \in \{1, 2, \ldots, n\}$$

We are led to the following result.

**Theorem 19.5.1.** 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. A doubly stochastic matrix. Then $|S| \leq 1$. Moreover $S$ is a semicontraction in the one norm if and only if the number of nonzero entries in each column of $S$ exceeds $q$. 

19.6 Further Comments

For $p = 1$, the maximum of the seminorm $|\cdot|_1$ is obtained on any matrix with a single nonzero entry.

attains the value $1$.

This completes the proof.

where $M$ is a $2 \times 2$ matrix.

to the following.

**Proposition 19.6.1.** Suppose $M$ is a $2 \times 2$ matrix.

where $M$ is a $2 \times 2$ matrix.

Now consider the case of a $2 \times 2$ matrix.

More generally, for $p = 1$, the maximum of the seminorm $|\cdot|_1$ is obtained on any matrix with a single nonzero entry.

**Lemma 19.6.2.** Suppose $M$ is a $2 \times 2$ matrix.

where $M$ is a $2 \times 2$ matrix.

**Proof of Theorem 19.6.1:** To be given in the full length version of this paper.

where $M$ is a $2 \times 2$ matrix.

We prove this by induction on $n$.

**Theorem 19.6.1.** 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. A doubly stochastic matrix. Then $|S| \leq 1$. Moreover $S$ is a semicontraction in the one norm if and only if the number of nonzero entries in each column of $S$ exceeds $q$. 

where $M$ is a $2 \times 2$ matrix.

The proof is completed.

where $M$ is a $2 \times 2$ matrix.
19.5.2 The Case $p = 2$

For the case when $p = 2$ it is also possible to derive an explicit formula for the seminorm $|M|_2$ 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

$$g(x, c) = x'((M - 1c)'((M - 1c)x = x'M'x - 2x'M'1cx + n(cx)^2$$

attains its minimum with respect to $c$ at

$$c = \frac{1}{n}M'$$

This implies that

$$|M|_2 = \|M - \frac{1}{n}11'M\|_2 = \sqrt{\mu\{(M - \frac{1}{n}11'M)'(M - \frac{1}{n}11'M)\}}$$

where for any symmetric matrix $T$, $\mu\{T\}$ is the largest eigenvalue of $T$. We are led to the following result.

**Proposition 19.5.3.** Let $M \in \mathbb{R}^{n \times n}$ be a nonnegative matrix.

$$|M|_2 = \sqrt{\mu\{M'(I - \frac{1}{n}11')M\}} \quad (19.5)$$

where $I - \frac{1}{n}11'$ is the orthogonal projection on the orthogonal complement of the span of $1$.

Now suppose that $M$ is a doubly stochastic matrix $S$. Then $S'S$ is also doubly stochastic and $1'S = 1'$. The latter and 19.5 imply that

$$|S|_2 = \sqrt{\mu\{S'S - \frac{1}{n}11'\}} \quad (19.6)$$

More can be said:

**Lemma 19.5.3.** If $S$ is doubly stochastic, then $\mu\{S'S - \frac{1}{n}11'\}$ is the second largest eigenvalue of $S'S$.

**Proof:** Since $S'S$ is symmetric it has orthogonal eigenvectors one of which is $1$. Let $1, x_2, \ldots, x_n$ be such a set of eigenvectors with eigenvalues $1, \lambda_2, \ldots, \lambda_n$. Then $S'S1 = 1$ and $S'Sx_i = \lambda_ix_i$, $i \in \{2, 3, \ldots, n\}$. Clearly $(S'S - \frac{1}{n}11')1 = 0$ and $(S'S - \frac{1}{n}11')x_i = \lambda_ix_i$, $i \in \{2, 3, \ldots, n\}$. 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 $S'S - \frac{1}{n}11'$. \(\square\)

We summarize:

**Theorem 19.5.2.** For $p = 2$, the seminorm of a doubly stochastic matrix $S$ is the second largest singular value of $S$. 
We are now in a position to characterize in graph theoretic terms those doubly stochastic matrices with positive diagonals which are semicontractions for \( p = 2 \).

**Theorem 19.5.3.** Let \( S \) be a doubly stochastic matrix with positive diagonals. Then \( |S|_2 \leq 1 \). Moreover \( S \) is a semicontraction in the 2-norm if and only if the graph of \( S \) is weakly connected.

To prove this theorem we need several concepts and results. Let \( G \) denote a directed graph and write \( G' \) for that graph which results when the arcs in \( G \) 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, rooted, strongly connected and weakly connected properties are equivalent. Note also that if \( G \) be the graph of a nonnegative matrix \( M \) with positive diagonals, then \( G' \) is the graph of \( M' \) and \( G' \circ G \) is the graph of \( M'M \).

**Lemma 19.5.4.** A directed graph \( G \) with self arcs at all vertices is weakly connected if and only if \( G' \circ G \) is strongly connected.

**Proof of Lemma 19.5.4:** Since \( G \) has self arcs at all vertices so does \( G' \). This implies that the arc set of \( G' \circ G \) contains the arc sets of \( G \) and \( G' \). Thus for any undirected path in \( G \) between vertices \( i \) and \( j \) there must be a corresponding directed path in \( G' \circ G \) between the same two vertices. Thus if \( G \) is weakly connected, \( G' \circ G \) must be strongly connected.

Now suppose that \((i,j)\) is an arc in \( G' \circ G \). Then because of the definition of composition, there must be a vertex \( k \) such that \((i,k)\) is an arc in \( G \) and \((k,j)\) is an arc in \( G' \). This implies that \((i,k)\) and \((j,k)\) are arcs in \( G \). Thus \( G \) has an undirected path from \( i \) to \( j \). Now suppose that \((i,v_1),(v_1,v_2),\ldots,(v_q,j)\) is a directed path in \( G' \circ G \) between \( i \) and \( j \). Between each pair of successive vertices along this path there must therefore be an undirected path in \( G \). Thus there must be an undirected path in \( G \) between \( i \) and \( j \). It follows that if \( G' \circ G \) is strongly connected, then \( G \) is weakly connected. \( \square \)

**Lemma 19.5.5.** Let \( T \) be a stochastic matrix with positive diagonals. If \( T \) has a strongly connected graph then the magnitude of its second largest eigenvalue is less than 1. 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.

**Proof of Lemma 19.5.5:** Suppose that the graph of \( T \) is strongly connected. Then via Theorem 6.2.24 of [8], \( T \) is irreducible. Thus there is an integer \( k \) such that \((I + T)^k > 0 \). Since \( T \) has positive diagonals, this implies that \( T^k > 0 \). Therefore \( T \) is primitive [8]. Thus by the Perron-Frobenius theorem [14], \( T \) can have only one eigenvalue of maximum modulus. Since the spectral radius of \( T \) is 1 and 1 is an eigenvalue, the magnitude of the second largest eigenvalue of \( T \) must be less than 1.

To prove the converse, suppose that \( T \) is a stochastic matrix whose second largest eigenvalue in magnitude is less than 1. Then

\[
\lim_{t \to \infty} T^t = Ic
\]  \hspace{1cm} (19.7)
for some row vector \( c \). Suppose that the graph of \( T \) is not weakly connected. Therefore if \( q \) denotes the number of weakly connected components of the graph, then \( q > 1 \). This implies that \( T = P^{-1}DP \) for some permutation matrix \( P \) and block diagonal matrix \( D \) with \( q \) blocks. Since \( D = PTP' \), \( D \) is also stochastic. Thus each of its \( q \) diagonal blocks is stochastic. Since \( T^i \) converges to \( 1c \), \( D^i \) must converge to a matrix of the form \( 1c \). But this is clearly impossible because \( 1c \) cannot have \( q > 1 \) diagonal blocks. \( \square \)

**Proof of Theorem 19.5.3:** Let \( S \) be a doubly stochastic matrix with positive diagonals. Then 1 is the largest singular value of \( S \) because \( S'S \) is doubly stochastic. From this and theorem 19.5.2 it follows that \( \|S\|_2 \leq 1 \).

Suppose \( S \) is a semicontraction. Then in view of Theorem 19.5.2, the second largest eigenvalue of \( S'S \) is less than 1. Thus by Lemma 19.5.5, the graph of \( S'S \) is weakly connected. But \( S'S \) is symmetric so its graph must be strongly connected. Therefore by Lemma 19.5.4, the graph of \( S \) is weakly connected.

Now suppose that the graph of \( S \) is weakly connected. Then the graph of \( S'S \) is strongly connected because of Lemma 19.5.4. Thus by Lemma 19.5.5, the magnitude of the second largest eigenvalue of \( S'S \) is less than 1. From this and Theorem 19.5.2 it follows that \( S \) is a semicontraction. \( \square \)

**Proof of Theorem 19.4.2:** Let \( \mathcal{M} \) be any compact subset of \( \mathcal{W} \). In view of Theorem 19.5.3, each matrix in \( \mathcal{M} \) is a semicontraction in the two-norm. From this and Proposition 19.5.1, 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 \( 1c \) as \( i \to \infty \). This means that the second largest eigenvalue of \( S \) must be less than 1 in magnitude. Thus by Lemma 19.5.5, \( S \) must have a weakly connected graph. \( \square \)

The importance of Theorem 19.5.3 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 19.5.1, 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 weakly connected graph would have to be a scrambling matrix. The following counterexample shows that this is not the case.

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

In particular, \( S \) doubly stochastic matrix with a weakly connected graph but it is not a scrambling matrix.
19.5.3 The Case \( p = \infty \)

Note that in this case \( |S|_\infty \leq 1 \) for any stochastic matrix because \( |S|_\infty \leq \|S\|_\infty \leq 1 \). Despite this, the derivation of an explicit formula for the seminorm \( |S|_\infty \) has so far eluded us. We suspect that \( |S|_\infty \) is the coefficient of ergodicity of \( S \). This conjecture is prompted by the following.

**Proposition 19.5.4.** For any stochastic matrix \( S \in \mathbb{R}^{n \times n} \),

\[
\tau(S) \leq |S|_\infty \leq 2\tau(S)
\]

(19.8)

where \( \tau(S) \) is the coefficient of ergodicity

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

(19.9)

One implication of Proposition 19.5.4 is that \( M \) cannot be a contraction in \( \infty \) seminorm if \( M \) is not a scrambling matrix. But as noted at the end of the last section, there are doubly stochastic matrices with weakly connected graphs which are not scrambling matrices. There are also plenty of stochastic matrices with rooted graphs which are not scrambling matrices thus there are matrices both \( \mathcal{S} \) and \( \mathcal{W} \) which are not contractions in this norm. For this reason, the \( \infty \) seminorm does not have the property we seek; i.e., providing a seminorm with value less than one for either stochastic matrices with rooted graphs or doubly stochastic matrices with weakly connected graphs.

**Proof of Proposition 19.5.4:** To be given in the full length version of this paper.

19.6 Complete Gossip Sequences

Let \( (i_1, j_1), (i_2, j_2), \ldots, (i_m, j_m) \) be a given sequence of pairs of indices of agent pairs which have gossiped. Let \( S = S_m S_{m-1} \cdots S_1 \) where \( S_i \) is the stochastic matrix associated with the \( i \)th gossip in the sequence. By the sequence's contraction coefficient is meant the seminorm \( |S|_2 \). By the gossip graph of the sequence is meant that undirected graph with vertex set \( \{1, 2, \ldots, n\} \) and edge set equal to the union of all pairs \( \{i_k, j_k\} \) in the sequence. This graph, which does not take into account the sequence's order, is clearly a subgraph of the allowable gossip graph \( \mathcal{A} \) for the gossiping process under consideration. We call such a gossiping sequence complete if its corresponding gossip graph contains a tree spanning \( \mathcal{A} \). Call the directed graph of \( S \), the sequence's composed gossip graph. In general the gossip graph of a given sequence must be a subgraph of the undirected version of the sequence's composed gossip graph.

The main result we want to prove is as follows.

**Theorem 19.6.1.** A gossip sequence \( (i_1, j_1), (i_2, j_2), \ldots, (i_m, j_m) \) is complete if and only if the sequence's contraction coefficient is less than 1.
To prove Theorem 19.6.1 we need several preliminary results.

**Lemma 19.6.1.** Let $\mathbb{G}$ and $\mathbb{H}$ be directed graphs on the same $n$ vertices. Suppose that both graphs have self-arcs at all vertices. If there is an undirected path from $i$ to $j$ in $\mathbb{H} \circ \mathbb{G}$, then there is an undirected path from $i$ to $j$ in the union of $\mathbb{H}$ and $\mathbb{G}$.

**Proof of Lemma 19.6.1:** First suppose that there is an undirected path of length one between vertices $i$ and $j$ in $\mathbb{H} \circ \mathbb{G}$. Then either $(i, j)$ or $(j, i)$ must be an arc in $\mathbb{H} \circ \mathbb{G}$. Without loss of generality suppose that $(i, j)$ is an arc in $\mathbb{H} \circ \mathbb{G}$. Then because of the definition of composition, there must be a vertex $k$ such that $(i, k)$ is an arc in $\mathbb{G}$ and $(k, j)$ is an arc in $\mathbb{H}$. This implies that $(i, k)$ and $(k, j)$ are arcs in $\mathbb{H} \cup \mathbb{G}$. Thus $\mathbb{H} \cup \mathbb{G}$ has a directed path from $i$ to $j$. Therefore $\mathbb{H} \cup \mathbb{G}$ has an undirected path from $i$ to $j$. Now suppose that $(i, v_1), (v_1, v_2), \ldots, (v_q, j)$ is a undirected path in $\mathbb{H} \circ \mathbb{G}$ between $i$ and $j$. Between each pair of successive vertices along this path there must therefore be an undirected arc in $\mathbb{H} \circ \mathbb{G}$. Therefore between each pair of successive vertices along this path there must therefore be an undirected arc in $\mathbb{H} \cup \mathbb{G}$. Thus there must be an undirected path in $\mathbb{H} \cup \mathbb{G}$ between $i$ and $j$. □

It is obvious that the preceding lemma applies to a finite set of directed graphs. In the sequel we appeal to this extension without special mention.

**Lemma 19.6.2.** A gossip sequence is complete if and only if its composed gossip graph is weakly connected.

**Proof of Lemma 19.6.2:** Because all vertices of all directed graphs under consideration have self-arcs, a gossip graph is always a subgraph of the undirected version of the composed gossip graph. Thus if the gossip sequence is complete then the undirected version of the composed gossip graph must contain a subgraph which is a spanning tree in $\mathbb{A}$. Thus the composed gossip graph must be weakly connected.

Now suppose that the composed gossip graph is weakly connected. Then the union of the one-gossip graphs comprising the composition must be weakly connected because of Lemma 19.6.1. But the union of the undirected versions of one-gossip graphs is the gossip graph of the sequence. Thus gossip graph of the sequence must be connected and therefore must contain a subgraph which is a spanning tree of $\mathbb{A}$. Thus the gossip sequence must be complete. □

**Proof of Theorem 19.6.1:** Suppose that the gossip sequence generating the sequence of matrices $S_1, S_2, \ldots, S_m$ is complete. In view of Lemma 19.6.2, the composed gossip graph is weakly connected. Therefore by Theorem 19.5.3, the matrix $S_m \cdots S_1$ is a semicontraction so the contraction coefficient of the sequence is less than one.

Now suppose that the contraction coefficient is less than one. Therefore matrix $S = S_m \cdots S_1$ is a semicontraction. Then by Theorem 19.5.3, the composed gossip graph is weakly connected. Thus by Lemma 19.6.2, the gossip sequence must be complete. □

### 19.7 Concluding Remarks

Let $\mathbb{A}$ be a given allowable gossip graph. Let us call a complete gossip sequence *minimal* if there is no shorter sequence of allowable gossips which is compete. It is
easy to see that a gossip sequence will be minimal if and only if its gossip graph is a minimal spanning tree of $A$. For a given allowable gossip graph there can be many complete minimal sequences. Moreover, there can be differing largest singular values for the different doubly stochastic matrices associated with different complete minimal sequences. A useful challenge then would be to determine those complete minimal sequences whose associated singular values are as small as possible. This issue will be addressed in a future paper.

The definition of a contraction coefficient proposed in this paper is appropriate for aperiodic gossip sequences. For example, suppose $(i_1, j_1), (i_2, j_2), \ldots$ is an infinite gossip sequence composed of successive complete subsequences which are each of length at most $m$. Suppose in addition that $\lambda < 1$ is a positive constant which bounds from above the contraction coefficients for each successive subsequence. Then it is easy to see that the entire sequence converges at a rate no slower than $\lambda^\frac{i}{m}$.

For certain applications it is useful to consider gossip sequences $(i_1, j_1), (i_2, j_2), \ldots$ which are periodic in the sense that for some integer $m$ subsequence $(i_1, j_1), (i_2, j_2), \ldots, (i_m, j_m)$ repeats itself over and over. In this case a more useful notion of a contraction coefficient might be the second largest eigenvalue of the doubly stochastic matrix $S$ defined by the product of the $m$ stochastic matrices corresponding to the $m$ gossips in the sequence. As with the notion of a contraction coefficient for aperiodic sequences considered in this paper, it would be interesting to determine those complete minimal sequences whose associated contraction coefficients are as small as possible.

Results similar to the main results of this paper for gossip sequences have recently appeared in [16]. A careful comparison of findings will be made in a future paper on deterministic gossiping.

References