

# Convergence Rate of a Distributed Algorithm for Matrix Scaling to Doubly Stochastic Form

Alejandro D. Domínguez-García and Christoforos N. Hadjicostis

**Abstract**—Motivated by matrix scaling applications and, more recently, distributed averaging previous work has considered settings where the interconnections between components in a distributed system are captured by a strongly connected directed graph (digraph) and each component aims to assign assigning weights on its outgoing edges (based on the weights on its incoming edges) so that the corresponding set of weights forms a doubly stochastic matrix. In particular, it has been shown that the system components can obtain a set of weights that form a doubly stochastic matrix via a variety of distributed algorithms. In this paper, we establish that the convergence rate of one such distributed algorithm is linear with rate between zero and one.

## I. INTRODUCTION AND BACKGROUND

In this paper we study a problem that we left open in [1], namely the rate of convergence of a particular distributed matrix scaling algorithm, which allows the components of a multi-component system to obtain a set of weights that forms a doubly stochastic matrix. The primary motivation for this distributed matrix scaling algorithm was asymptotic average consensus—a special case of the distributed consensus problem (see, e.g., [2] and the references therein)—where all nodes (components) converge asymptotically to the average of their initial values. More specifically, a popular approach to asymptotic average consensus is to use a linear iteration, where each node repeatedly updates its value as a weighted linear combination of its own previous value and the previous values of its neighbors. In such case, one deals with an autonomous discrete-time linear system with a (time-invariant) transition matrix, also referred to as *weight matrix*, that is defined by the coefficients (weights) used in the linear updates. It is well-known that if this weight matrix is primitive and doubly stochastic, then the nodes will asymptotically converge to the average of their initial values (see, e.g., [2]–[5] and references therein).

Apart from allowing the nodes to asymptotically reach average consensus, a weight choice that forms a primitive doubly stochastic matrix is also important for a number of other tasks. For example, one of the early motivations for studying (centralized) algorithms for the matrix scaling problem was the desire to start from the stochastic matrix of a Markov chain and obtain a scaled version of it that is doubly stochastic and adheres to the sparsity structure of the original one (see, e.g., [6] and the references therein).

Alejandro D. Domínguez-García is with the ECE Department at the University of Illinois, Urbana, IL 61801, USA. E-mail: aledan@illinois.edu

C. N. Hadjicostis is with the ECE Department at the University of Cyprus, Nicosia, Cyprus; and also with the ECE Department at the University of Illinois, Urbana, IL 61801, USA. E-mail: chadjic@ucy.ac.cy

Many applications of matrix scaling can also be found in economics, urban planning, statistics, and demography (see the discussions in [7]) as well as estimation problems with some underlying averaging operation [8], distributed optimization problems [9], and distributed power control in wireless ad-hoc networks [10].

Starting from a very general setting in which the exchange of information between components (nodes) in a multi-component system is captured by a directed graph—or *digraph*—we proposed in [1] several algorithms that allow the nodes to choose their weights, in a distributed manner and subject to the sparsity structure imposed by the *digraph*, so that the resulting weight matrix is primitive and doubly stochastic. In this paper, we tackle a problem that we left open in [1], namely the analysis of the convergence rate of Algorithm 3 in [1]; we show that this rate is linear.

## II. PRELIMINARIES

### A. Notation and Graph Theoretic Notions

Real vectors are denoted by small letters, e.g.,  $x = [x_1, x_2, \dots, x_n]^T$ ,  $x_j \in \mathbb{R}$ , where the superscript “ $T$ ” denotes matrix/vector transposition. Matrices are defined over the field of real numbers and are denoted by capital letters, i.e.,  $A = [a_{ji}] \in \mathbb{R}^{n \times n}$ , and  $A^T = [a_{ij}]$ . We denote the  $l_p$  norm of  $A \in \mathbb{R}^{n \times n}$  (or  $x \in \mathbb{R}^n$ ) by  $\|A\|_p$  (or  $\|x\|_p$ ). The symbol  $I_n$  denotes the  $n \times n$  identity matrix, whereas  $\mathbf{1}_n$  denotes the all-ones  $n$ -dimensional column vector, and  $\mathbf{0}_n$  denotes the all-zeros  $n$ -dimensional column vector. We reserve the symbol  $k$  to index discrete time (iterations) and explicitly show the time dependence of a vector or a matrix as  $x[k]$  or  $A[k]$ , respectively. We use the notation  $\text{diag}(a_1, a_2, \dots, a_n)$  (or  $\text{diag}(a)$  where  $a = [a_1, a_2, \dots, a_n]^T$ ) to denote a diagonal matrix, with  $a_j$  as its  $(j, j)$  diagonal entry. More generally, the notation  $\text{diag}(A_1, A_2, \dots, A_n)$  denotes a block diagonal matrix, with the square matrix  $A_j$  as its  $(j, j)$  diagonal block.

*Definition 1:* Consider a sequence of real values  $x_1, x_2, \dots, x_k, \dots$  that converges to some limit value  $L$ . We say that the sequence converges *linearly* to  $L$  if there exist a real number  $\mu \in (0, 1)$  such that

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - L|}{|x_k - L|} = \mu.$$

In such case we say that the rate of convergence is  $\mu$ .

The exchange of information between system components (nodes) is described by a digraph  $\mathcal{G}_d = \{\mathcal{V}, \mathcal{E}\}$ , where  $\mathcal{V} = \{1, 2, \dots, n\}$  is the vertex set (each vertex corresponds to a node) and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  is the set of *directed edges*.

Edge  $(j, i) \in \mathcal{E}$  if node  $j$  can receive information from node  $i$ . By convention, we assume no self-loops in  $\mathcal{G}_d$  (i.e.,  $(j, j) \notin \mathcal{E}$  for all  $j \in \mathcal{V}$ ). All nodes that can transmit (receive) information to (from) node  $j$  are said to be in-neighbors (out-neighbors) of node  $j$  and are represented by the set  $\mathcal{N}_j^- = \{i \in \mathcal{V} \mid (j, i) \in \mathcal{E}\}$  ( $\mathcal{N}_j^+ = \{l \in \mathcal{V} : (l, j) \in \mathcal{E}\}$ ). The number of in-neighbors (out-neighbors) of  $j$  is called the in-degree (out-degree) of  $j$  and is denoted by  $\mathcal{D}_j^- = |\mathcal{N}_j^-|$  ( $\mathcal{D}_j^+ = |\mathcal{N}_j^+|$ ). The  $n \times n$  matrix  $A$ , with  $a_{ji} = 1$  if  $(j, i) \in \mathcal{E}$  and  $a_{ji} = 0$  otherwise, is called the adjacency matrix. We clearly have  $\mathcal{D}_j^+ = \sum_l a_{lj}$  and  $\mathcal{D}_j^- = \sum_i a_{ji}$ . A digraph is *strongly connected* if for any pair of vertices  $i$  and  $j$ ,  $i \neq j$ , there exists a path that starts in  $i$  and ends in  $j$ , i.e., a sequence of edges  $(i_1, i), (i_2, i_1), \dots, (i_t, i_{t-1}), (j, i_t)$  all of which belong in  $\mathcal{E}$ .

### B. A Distributed Matrix Scaling Algorithm

Given a strongly connected digraph  $\mathcal{G}_d = \{\mathcal{V}, \mathcal{E}\}$ , we consider a setting where each node can observe but cannot control the (likely different) values on the edges from its in-neighbors (note that it is not required that the node can identify the sender node associated with each value). Based on these (incoming) values, each node chooses its self-weight and the weights on the edges to its out-neighbors. Specifically, node  $j$  chooses its self-weight  $p_{jj}[k+1]$  and the weights  $\{p_{lj}[k+1] \mid l \in \mathcal{N}_j^+\}$  based on the weights  $\{p_{ji}[k] \mid i \in \mathcal{N}_j^- \cup \{j\}\}$ . While updating these weights, node  $j$  also imposes the following constraints:

**[C1]** weights are nonnegative;

**[C2]** the *same* weight  $p_{lj}[k+1]$  is used for all edges to out-neighbors of node  $j$  (this is an attractive feature in the case of wireless networks, because a single broadcast transmission by a node is received by all of its out-neighbors);

**[C3]** the sum of the weights on the edges of the out-neighbors of node  $j$  is equal to one at all times, i.e.,  $\sum_{l \in \mathcal{N}_j^+ \cup \{j\}} p_{lj}[k+1] = 1$  (or, equivalently,  $\mathcal{D}_j^+ p_{lj}[k] + p_{jj}[k] = 1$ ).

Under Constraints **C1–C3**, the weight matrix  $P[k] = [p_{ij}[k]]$  at each iteration is a column stochastic matrix that can be parameterized as

$$P[k] = \bar{P}\Delta[k] + (I_n - \Delta[k]), \quad (1)$$

where  $\Delta[k] = \text{diag}(\delta[k])$ ,  $\delta[k] = [\delta_1[k], \delta_2[k], \dots, \delta_n[k]]^T$ , is a diagonal matrix, and  $\bar{P} = [\bar{p}_{lj}]$  is a matrix with all its diagonal entries being zero; and, for each  $j \in \mathcal{V}$ ,  $\bar{p}_{lj} = 1/\mathcal{D}_j^+$ ,  $\forall l \in \mathcal{N}_j^+$ , and zero otherwise. It is clear that  $\delta_j[k]$  determines node  $j$ 's self-weight and the weights on the edges to its out-neighbors as  $p_{jj}[k] = 1 - \delta_j[k]$  and  $p_{lj}[k] = \delta_j[k]/\mathcal{D}_j^+$ ,  $\forall l \in \mathcal{N}_j^+$ . [Note that, as long as  $0 \leq \delta_j[k] \leq 1$ , Constraints **C1–C3** will be satisfied.]

The algorithm proceeds as follows. Initially, each node  $j$ ,  $j \in \mathcal{V}$ , chooses its self-weight and the weights on the edges to its out-neighbors to be  $p_{lj}[0] = 1/(1 + \mathcal{D}_j^+)$ ,  $\forall l \in \mathcal{N}_j^+ \cup \{j\}$ , by setting  $\delta_j[0] = \mathcal{D}_j^+/(1 + \mathcal{D}_j^+)$ . Subsequently, at each iteration  $k$ , each node  $j$  updates its  $\delta_j[k+1]$  (and, along with it, all the weights it controls, i.e., all  $p_{lj}[k]$ ,  $\forall l \in \mathcal{N}_j^+ \cup \{j\}$ ,

according to the parametrization described in (1) as follows:

$$\delta_j[k+1] = \begin{cases} \delta_j[k]e_j[k], & \text{if } e_j[k] \leq 1, \\ 1 - \frac{1}{e_j[k]}(1 - \delta_j[k]), & \text{if } e_j[k] > 1, \end{cases} \quad (2)$$

where  $e_j[k] = \sum_{i \in \mathcal{N}_j^- \cup \{j\}} p_{ji}[k]$ . Intuitively, the update of each  $\delta_j$  in (2) ensures that when the  $j$ th row sum of column stochastic matrix  $P[k]$  is greater than unity, then the  $j$ th diagonal entry is decreased (and vice-versa), while ensuring that the matrix  $P[k+1]$  remains column stochastic.

A formal description of the iterative procedure described above is provided in Algorithm 1 (Algorithm 3 in [1]). This matrix scaling algorithm can be used to perform distributed averaging in directed graphs [1].

---

### Algorithm 1: Distributed Matrix Scaling

---

**1 Input:** Each node  $j$  sets  $\delta_j[0] = \frac{\mathcal{D}_j^+}{1 + \mathcal{D}_j^+}$   
**for**  $k \geq 0$ , each node  $j$  does the following:  
**2**     **Output:**  $p_{jj}[k] = 1 - \delta_j[k]$   
**3**              $p_{lj}[k] = \frac{1}{\mathcal{D}_j^+} \delta_j[k]$ ,  $\forall l \in \mathcal{N}_j^+$   
**4**     **Receive:**  $p_{ji}[k]$  from all  $i \in \mathcal{N}_j^-$   
**5**     **Compute:**  
             $e_j[k] = \sum_{i \in \mathcal{N}_j^- \cup \{j\}} p_{ji}[k]$   
**6**              $\delta_j[k+1] = \begin{cases} \delta_j[k]e_j[k], & \text{if } e_j[k] \leq 1 \\ 1 - \frac{1}{e_j[k]}(1 - \delta_j[k]), & \text{if } e_j[k] > 1 \end{cases}$   
**7**     **Broadcast:**  $p_{lj}[k+1] = \frac{1}{\mathcal{D}_j^+} \delta_j[k+1]$  to all  $l \in \mathcal{N}_j^+$

---

### III. LINEAR CONVERGENCE

During the execution of Algorithm 1, we obtain a sequence of column stochastic matrices  $P[0], P[1], \dots, P[k], \dots$ , with  $P[k] = [p_{ji}[k]]$ ,  $\forall k$ . Matrix  $P[k]$  can be parameterized as in (1). We use the following quantity as a measure of the closeness of matrix  $P[k]$  to being doubly stochastic.

*Definition 2:* Given an  $n \times n$  column stochastic matrix  $P[k] = [p_{ji}[k]]$  parameterized as in (1), we define

$$V[k] =: V(P[k]) =: V(\delta[k]) = \sum_j |\epsilon_j[k]|,$$

where  $\epsilon_j[k] = -1 + \sum_{i \in \mathcal{N}_j^- \cup \{j\}} p_{ji}[k] = -\delta_j[k] + \sum_{i \in \mathcal{N}_j^-} \frac{1}{\mathcal{D}_j^+} \delta_i[k]$ .

*Theorem 1:* Consider a strongly connected digraph  $\mathcal{G}_d = \{\mathcal{V}, \mathcal{E}\}$  (where  $\mathcal{V} = \{1, 2, \dots, n\}$  is the set of nodes and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V} - \{(j, j) \mid j \in \mathcal{V}\}$  is the set of directed edges), in which the nodes execute Algorithm 1, so that they obtain a sequence of matrices  $P[0], P[1], \dots, P[k], \dots$ , parameterized by (1) and updated according to (2). Then, the corresponding sequence of values  $V[0], V[1], \dots, V[k], \dots$  (where  $V[k] = V(P[k])$  as in Definition 2) converges linearly to zero with rate of converge  $\mu \in (0, 1)$ .

We first outline the main steps of the proof and we subsequently discuss each one in detail.

**[S1]** We establish that for the  $V[k]$  in Definition 2 we have  $\lim_{k \rightarrow \infty} V[k] = 0$ .

**[S2]** We then write the  $n$ -dimensional *nonnegative* vector  $\delta[k] = [\delta_1[k], \delta_2[k], \dots, \delta_n[k]]^T$  as  $\delta[k] = \alpha[k] v + E[k]$ , where (i)  $v$  is the unique<sup>1</sup> positive vector that satisfies  $v = \bar{P}v$  and is normalized so that  $\sum_j v_j = 1$ , (ii)  $\alpha[k]$  is positive a scalar that satisfies  $\alpha[k] = \sum_{j=1}^n \delta_j[k]$ , and (iii)  $E[k] = \delta[k] - \alpha[k]v$  is an  $n$ -dimensional “error” vector. We argue that this decomposition is unique and the vector  $E[k] = [E_1[k], E_2[k], \dots, E_n[k]]^T$  satisfies  $\sum_{j=1}^n E_j[k] = 0$  and  $\lim_{k \rightarrow \infty} E[k] = 0$ .

**[S3]** The next step is to argue that for some  $K$ , we have that  $\alpha[k] \geq \alpha^* > 0$  for all  $k \geq K$ ; we use this fact to establish that  $\delta_{\min}[k] := \min_j \{\delta_j[k]\} \geq \delta_{\min}^* > 0, \forall k \geq K$ .

**[S4]** We then establish that  $V[k+n] \leq \gamma^* V[k]$ , where  $0 \leq \gamma^* < 1$ . This means that every  $n$  steps the value of  $V[k]$  is guaranteed to decrease by a factor of  $\gamma^*$ .

**[S5]** The last step is to establish that for each  $j \in \mathcal{V}$ , we have linear convergence of  $\epsilon_j[k]$  to zero, which in turn implies that the sequence  $\delta_j[k]$  is Cauchy, i.e.,  $\delta_j[k]$  converges (linearly in our case) to some limit value  $\delta_j^*$ . Thus, the vector  $\delta[k]$  converges linearly to some limit value  $\delta^* = [\delta_1^*, \delta_2^*, \dots, \delta_n^*]^T$  and  $P[k]$  converges linearly to some limit matrix  $P^* = \bar{P} \text{diag}(\delta^*) + (I_n - \text{diag}(\delta^*))$  which is primitive doubly stochastic.

#### A. Step S1: Convergence of $V[k]$

Here we use the LaSalle Invariance Principle for discrete-time dynamical systems (provided below as stated in [12, Theorem 1.19]) to show that  $\lim_{k \rightarrow \infty} \sum_j |\epsilon_j[k]| = 0$ , where  $\epsilon_j[k]$  was provided in Definition 2.

**Theorem 2 (LaSalle Invariance Principle):** Consider a discrete-time dynamical system described by  $\delta[k+1] = f(\delta[k])$ , where  $f : \mathcal{X} \rightarrow \mathcal{X}$ ,  $\mathcal{X} \subseteq \mathbb{R}^n$ . Assume that the following assumptions hold:

- [L1]** there exists a closed set  $\mathcal{W} \subseteq \mathcal{X}$  that is positively invariant for  $\delta[k+1] = f(\delta[k])$ ,  $\delta[k] \in \mathcal{X}$ ;
- [L2]** there exists a function  $V : \mathcal{X} \rightarrow \mathbb{R}$  that is non-increasing along  $f(\cdot)$  on  $\mathcal{W}$ ;
- [L3]** all evolutions of  $\delta[k+1] = f(\delta[k])$ ,  $\delta[k] \in \mathcal{X}$ , with initial conditions in  $\mathcal{W}$  are bounded; and
- [L4]**  $f(\cdot)$  and  $V(\cdot)$  are continuous on  $\mathcal{W}$ .

Then, each evolution with initial conditions in  $\mathcal{W}$  approaches a set  $\mathcal{E} = V^{-1}(c) \cap \mathcal{S}$ , where  $V^{-1}(c)$  is the level set of  $V(\cdot)$  that corresponds to a real constant  $c$ , and  $\mathcal{S}$  is the largest positively invariant set contained in  $\mathcal{F} = \{\delta \in \mathcal{W} \mid V(f(\delta)) = V(\delta)\} = \{\delta \in \mathcal{W} \mid \bar{P}\delta = \delta\}$ .

The evolution of  $\delta[k]$  during the execution of the matrix scaling algorithm can be written as  $\delta[k+1] = f(\delta[k])$ , with the  $j^{\text{th}}$  component of the vector-valued function  $f(\cdot)$  given

<sup>1</sup>Note that  $\bar{P}$  is a column stochastic matrix that has the same sparsity structure as the adjacency matrix of the strongly connected digraph  $\mathcal{G}_d$ ; thus,  $\bar{P}$  is irreducible. Then, according to [11, Theorem 8.4.4],  $\bar{P}$  has a unique (up to scaling) positive eigenvector  $v$  at eigenvalue 1.

(according to (2)) by

$$f_j(\delta[k]) = \begin{cases} \delta_j[k] e_j[k], & \text{if } e_j[k] \leq 1, \\ 1 - \frac{1}{e_j[k]} (1 - \delta_j[k]), & \text{if } e_j[k] > 1, \end{cases} \quad (3)$$

where  $e_j[k] = \sum_{i \in \mathcal{N}_j^- \cup \{j\}} p_{ji}[k] = 1 - \delta_j[k] + \sum_{i \in \mathcal{N}_j^-} \frac{1}{D_i^+} \delta_i[k]$ . Note that  $V[k]$  in Definition 2 is non-negative and becomes zero if and only if  $\delta_j[k] = \sum_{i \in \mathcal{N}_j^-} \frac{1}{D_i^+} \delta_i[k], \forall j$ . In matrix notation, this is equivalent to  $\bar{P}\delta[k] = \delta[k]$  and implies that  $\delta[k] = \alpha[k]v$ , where  $\alpha[k]$  is a nonnegative constant, and  $v$  is the unique<sup>1</sup> positive vector that satisfies  $v = \bar{P}v$  and  $\sum_j v_j = 1$ .

**Lemma 1 (See the proof of Theorem 5 in [1]):** Consider a strongly connected digraph  $\mathcal{G}_d = \{\mathcal{V}, \mathcal{E}\}$ , where each node  $j \in \mathcal{V}$  executes the distributed matrix scaling algorithm and updates variable  $\delta_j[k]$  according to (2). Then, we have  $V[k+1] \leq V[k]$ . Furthermore,  $V[k+1] = V[k]$  if and only if  $\delta[k] \in \mathcal{F} = \{\delta \mid \bar{P}\delta = \delta, 0 \leq \delta_j \leq 1, \forall j\}$ .

**Theorem 3:** Consider a strongly connected digraph  $\mathcal{G}_d = \{\mathcal{V}, \mathcal{E}\}$ , where each node  $j, j \in \mathcal{V}$ , executes the distributed matrix scaling algorithm and updates variable  $\delta_j[k]$  according to (2), with  $\delta_j[0] = \frac{D_j^+}{1+D_j^+}$ . Then, as  $k \rightarrow \infty$ ,  $\delta[k] = [\delta_1[k], \delta_2[k], \dots, \delta_n[k]]^T$  will approach the set  $\mathcal{F} = \{\delta \mid \bar{P}\delta = \delta\}$ .

**Proof:** In order to establish that, as  $k \rightarrow \infty$ ,  $\delta[k] = [\delta_1[k], \delta_2[k], \dots, \delta_n[k]]^T$  approaches the set  $\mathcal{F} = \{\delta \mid \bar{P}\delta = \delta\}$ , we will use the LaSalle Invariance Principle in Theorem 2. Assumptions **L1–L3** are easy to establish and we omit the discussion due to space limitations. Regarding **L4**, we can check that whenever  $0 \leq \delta_j[k] \leq 1, \forall j$ , and  $e_j[k] < 1$  or  $e_j[k] > 1$ , the  $f_j(\cdot)$ 's are continuous (with respect to the  $\delta_j[k]$ 's) and so is  $f(\cdot)$ . Additionally, whenever  $e_j[k] \uparrow 1$ , it follows from the first line of (3) that  $f_j(\delta[k]) = \delta_j[k]$ , and whenever  $e_j[k] \downarrow 1$ , it follows from the second line of (3) that  $f_j(\delta[k]) = \delta_j[k]$ ; therefore, the  $f_j(\cdot)$ 's are continuous everywhere on  $\mathcal{W}$  and so is  $f(\cdot)$ . Finally, since  $\epsilon_j(\delta[k]) = -\delta_j[k] + \sum_{i \in \mathcal{N}_j^-} \frac{1}{D_i^+} \delta_i[k], \forall j$ , are continuous on  $\mathcal{W}$ , it follows that  $V(\delta[k]) = \sum_{j=1}^n |\epsilon_j(\delta[k])|$  is continuous on  $\mathcal{W}$ . Thus, since  $f(\cdot)$  and  $V(\cdot)$  are both continuous on  $\mathcal{W}$ , Assumption **L4** in Theorem 2 holds.

Since Assumptions **L1–L4** in Theorem 2 hold for the system  $\delta[k+1] = f(\delta[k])$  whenever  $0 \leq \delta_j[0] \leq 1, \forall j$ , we know that each system evolution will approach a set  $\mathcal{E} = V^{-1}(c) \cap \mathcal{S}$ , where  $V^{-1}(c)$  is the level set of  $V(\cdot)$  that corresponds to a real constant  $c$ , and  $\mathcal{S}$  is the largest positively invariant set contained in  $\mathcal{F} = \{\delta \in \mathcal{W} \mid V(f(\delta)) = V(\delta)\} = \{\delta \in \mathcal{W} \mid \bar{P}\delta = \delta\}$ . The set  $\mathcal{F}$  is also positively invariant for the system: this follows from the fact that if  $\delta[0] \in \mathcal{F}$ , we have  $e_j[0] = (1 - \delta_j[0]) + \sum_{i \in \mathcal{N}_j^-} \frac{1}{D_i^+} \delta_i[0] = 1, \forall j$ , which implies that each node  $j$  obtains  $\delta_j[1] = \delta_j[0], \forall j$ ; thus,  $e_j[1] = (1 - \delta_j[1]) + \sum_{i \in \mathcal{N}_j^-} \frac{1}{D_i^+} \delta_i[1] = 1, \forall j$ , therefore  $\delta[1] \in \mathcal{F}$ . By induction, we can establish that  $\delta[k] \in \mathcal{F}, k > 0$ , if  $\delta[0] \in \mathcal{F}$ . Thus, the largest subset  $\mathcal{S}$  of  $\mathcal{F}$  that is positively invariant is indeed  $\mathcal{F}$ , i.e.,  $\mathcal{S} = \mathcal{F}$ .

The final step is to determine the real constant  $c$  that defines  $\mathcal{E} = V^{-1}(c) \cap \mathcal{S}$ ; we will establish that the set  $\mathcal{E} = V^{-1}(c) \cap \mathcal{S}$  is nonempty only when  $c = 0$ . Since  $\mathcal{S} \equiv \mathcal{F}$ , it follows from the definition of  $\mathcal{F}$  that  $\mathcal{S} = \mathcal{S}_1 \cap \mathcal{S}_2 \cap \dots \cap \mathcal{S}_n$ , where  $\mathcal{S}_j = \{\delta \mid (1 - \delta_j) + \sum_{i \in \mathcal{N}_j^-} \frac{1}{\mathcal{D}_i^+} \delta_i = 1\}$ . Thus, for the set  $\mathcal{E} = V^{-1}(c) \cap \mathcal{S}$  to be nonempty, we need each set  $\mathcal{E}_j = V^{-1}(c) \cap \mathcal{S}_j$  also to be nonempty. Since  $V(\delta[k]) = \sum_{j=1}^n |\epsilon_j(\delta[k])|$ , if  $c > 0$ , there exists at least one  $l \in \mathcal{V}$  for which  $|\epsilon_l(\delta[k])| > 0$ , thus  $(1 - \delta_l[k]) + \sum_{i \in \mathcal{N}_l^-} \frac{1}{\mathcal{D}_i^+} \delta_i[k] \neq 1$ ; therefore  $\mathcal{E}_l = V^{-1}(c) \cap \mathcal{S}_l = \emptyset$ , which implies that  $\mathcal{E}$  is empty whenever  $c > 0$ . Since  $V(\delta[k]) = 0$  only if  $(1 - \delta_j[k]) + \sum_{i \in \mathcal{N}_j^-} \frac{1}{\mathcal{D}_i^+} \delta_i[k] = 1, \forall j$ , we have that  $V^{-1}(0) = \{\delta \mid (1 - \delta_j) + \sum_{i \in \mathcal{N}_j^-} \frac{1}{\mathcal{D}_i^+} \delta_i = 1, \forall j\}$ , i.e.,  $V^{-1}(0) \equiv \mathcal{S}$ ; thus  $\mathcal{E} \equiv \mathcal{S} \equiv \mathcal{F}$ . Now, since for all executions of the matrix scaling algorithm, we have that  $\delta_j[0] = \frac{\mathcal{D}_j^+}{1 + \mathcal{D}_j^+}, \forall j$ , it holds that  $\delta[0] \in \mathcal{W}$ ; therefore, the LaSalle Invariance Principle establishes that  $\delta[k]$  approaches  $\mathcal{E} \equiv \mathcal{S} \equiv \mathcal{F}$ , i.e., as  $k \rightarrow \infty$ , we have that  $e_j[k] = (1 - \delta_j[k]) + \sum_{i \in \mathcal{N}_j^-} \frac{1}{\mathcal{D}_i^+} \delta_i[k] = 1, \forall j$ , i.e.,  $|\epsilon_j[k]| \rightarrow 0$  as  $k \rightarrow \infty$ . ■

### B. Step S2: Convergence of $E[k]$

Since  $\mathcal{G}_d$  is strongly connected, the column stochastic matrix  $\bar{P}$ , which has the same sparsity structure as the adjacency matrix of  $\mathcal{G}_d$ , is irreducible; thus, it has a unique positive eigenvector  $v$  corresponding to eigenvalue 1 (i.e.,  $v$  satisfies  $\bar{P}v = v$ ) that we take, without loss of generality, to be normalized so that  $\sum_j v_j = 1$  (see, e.g., [11, Theorem 8.4.4]). The Jordan decomposition of  $\bar{P}$  ensures that we can write

$$\begin{aligned} \bar{P} &= UJU^{-1}, \\ U &= [u_1 \ u_2 \ \dots \ u_n], \\ J &= \text{diag}(J_1(\lambda_1), J_2(\lambda_2), \dots, J_l(\lambda_l)), \end{aligned}$$

where  $J_\ell, \ell = 1, 2, \dots, l$ , is the  $\ell$ th Jordan block and  $u_1, u_2, \dots, u_n$  (the columns of the invertible matrix  $U$ ) are the generalized eigenvectors of  $\bar{P}$ . The matrix  $\bar{P}$  has  $l$  eigenvalues  $\{\lambda_\ell \mid \ell = 1, 2, \dots, l\}$ , each of them associated with a Jordan block of size  $n_\ell$ , such that  $\sum_\ell n_\ell = n$ . Without loss of generality, we assume that the first Jordan block is of form  $[J_1] = [1]$  (i.e., there is a single eigenvalue at 1) and all other eigenvalues have values different from 1 (and, since  $\bar{P}$  is column stochastic, magnitude  $|\lambda_\ell| \leq 1$ ). Moreover, we have  $u_1 = v$  (one can easily see that  $u_1$  is the right eigenvector that corresponds to  $\lambda_1$  by writing  $PU = UJ$  which implies that  $Pu_1 = u_1\lambda_1$ ) and that the first row of matrix  $U^{-1}$  is the left eigenvector that corresponds to  $\lambda_1 = 1$  (this can be seen by writing  $U^{-1}P = JU^{-1}$  and focusing on the first row), i.e., the first row of  $U^{-1}$  has to be  $\mathbf{1}_n^T$  (so that  $\mathbf{1}_n^T v = 1$ ).

Clearly, we can write  $\delta[k] = Uc[k]$  for a coefficient vector  $c[k] = U^{-1}\delta[k] := [c_1[k], c_2[k], \dots, c_n[k]]^T$ . Specifically, the above decomposition implies that we can write  $\delta[k] = c_1[k]v + \tilde{E}[k]$ , where  $v = u_1$ , and  $\tilde{E}[k] = U\tilde{c}[k]$  with  $\tilde{c}[k] = [0, c_2[k], \dots, c_n[k]]^T$ . We observe that

$\mathbf{1}_n^T \tilde{E}[k] = \mathbf{1}_n^T U \tilde{c}[k] = [1, 0, 0, \dots, 0] \tilde{c}[k] = 0$  (because  $\mathbf{1}_n^T$  is the first row of matrix  $U^{-1}$ ). Thus, we have  $c_1[k] = \alpha[k]$  and  $\tilde{E}[k] = E[k]$  (i.e., the above decomposition is identical to the decomposition introduced in Step S2).

It is easy to verify that the vector  $\epsilon[k] := [\epsilon_1[k], \epsilon_2[k], \dots, \epsilon_n[k]]^T$  (where the  $\epsilon_j[k]$  were introduced in Definition 2) can be written as

$$\epsilon[k] = (\bar{P} - I_n)\delta[k] = U(J - I_n)c[k].$$

Thus,

$$\underbrace{\begin{bmatrix} 0 & \mathbf{0}_{n-1}^T \\ \mathbf{0}_{n-1} & \tilde{Q} \end{bmatrix}}_{J - I_n} \underbrace{\begin{bmatrix} 0 \\ c_2[k] \\ c_3[k] \\ \vdots \\ c_n[k] \end{bmatrix}}_{\tilde{c}[k]} = U^{-1}\epsilon[k],$$

which allows us to conclude that

$$\begin{aligned} |\tilde{c}[k]|_\infty &\leq \|\tilde{Q}^{-1}\|_1 \|U^{-1}\|_1 |\epsilon[k]|_\infty \\ |E[k]|_\infty &\leq \|U\|_1 |\tilde{c}[k]|_\infty \leq \underbrace{\|U\|_1 \|\tilde{Q}^{-1}\|_1 \|U^{-1}\|_1}_a |\epsilon[k]|_\infty \end{aligned} \quad (4)$$

(note that  $\tilde{Q}$  is invertible because it is an upper-triangular matrix with nonzero entries on the diagonal—since  $\lambda_\ell \neq 1$  for all  $\ell \neq 1$ , we have that  $\lambda_\ell - 1 \neq 0, \ell \neq 1$ ). Thus, since  $V[k]$  goes to zero with increasing  $k$  (and so does  $\epsilon[k]$ ), we have that  $c_2[k], c_3[k], \dots, c_n[k]$  also go to zero. Therefore, we have established that in the decomposition  $\delta[k] = \alpha[k]v + E[k]$ , the vector  $E[k]$  goes to zero.

### C. Step S3: Bounding $\delta[k]$ Away From Zero

In this section we establish that there exists a constant  $\delta_{\min}^* > 0$  such that, for some  $K$ , we have  $\delta_j[k] \geq \delta_{\min}^*$  for all  $j \in \mathcal{V}$  and all  $k \geq K$  (i.e.,  $\delta_{\min}[k] := \min_j \{\delta_j[k]\} \geq \delta_{\min}^* > 0$ ). We start by showing that for some  $K$ , the following inequality holds for  $k \geq K$ :  $\alpha[k+1] \geq \alpha[k]$ , if  $\alpha[k]v_{\min} < 1/2$ , where  $v_{\min} = \min_j \{v_j\}$ . To establish this, we re-write the evolution of  $\delta_j[k]$  in terms of  $\epsilon_j[k]$ . Specifically, using (2), we have

$$\delta_j[k+1] = \delta_j[k] + \epsilon'_j[k+1], \quad (6)$$

where

$$\epsilon'_j[k+1] = \begin{cases} \delta_j[k]\epsilon_j[k], & \epsilon_j[k] \leq 0, \\ (1 - \delta_j[k])\frac{\epsilon_j[k]}{1 + \epsilon_j[k]}, & \epsilon_j[k] > 0. \end{cases}$$

Multiplying both sides of (6) on the left by  $\mathbf{1}_n^T$  we obtain that  $\sum_j \delta_j[k+1] = \sum_j \delta_j[k] + \sum_j \epsilon'_j[k]$ , or  $\alpha[k+1] = \alpha[k] + \sum_j \epsilon'_j[k]$ . Focusing on  $\sum_j \epsilon'_j[k]$ , we can write it as follows:

$$\sum_j \epsilon'_j[k] = \sum_{j \in \mathcal{A}[k]} \delta_j[k]\epsilon_j[k] + \sum_{j \in \mathcal{B}[k]} (1 - \delta_j[k])\frac{\epsilon_j[k]}{1 + \epsilon_j[k]},$$

where  $\mathcal{A}[k] := \{j \mid \epsilon_j[k] \leq 0\}$ , and  $\mathcal{B}[k] := \{j \mid \epsilon_j[k] > 0\}$ .

From Step **S1**, we know that for any  $\varepsilon > 0$ , it is always possible to find a large enough  $K$  such that  $V[k] < \varepsilon$  (and thus  $|\epsilon_j[k]| < \varepsilon/2 < \varepsilon$ ) for all  $k \geq K$ . Letting  $\delta_{\max}[k] = \max_j \{\delta_j[k]\}$ , for  $k \geq K$ , we have that

$$\sum_j \epsilon'_j[k] \geq \delta_{\max}[k] \sum_{j \in \mathcal{A}[k]} \epsilon_j[k] +$$

$$\begin{aligned} & \frac{1 - \delta_{\max}[k]}{1 + \varepsilon} \sum_{j \in \mathcal{B}[k]} \epsilon_j[k] \\ & \geq \left( \frac{1 - \delta_{\max}[k]}{1 + \varepsilon} - \delta_{\max}[k] \right) \sum_{j \in \mathcal{B}[k]} \epsilon_j[k]. \end{aligned}$$

[Note that  $-\sum_{j \in \mathcal{A}[k]} \epsilon_j[k] = \sum_{j \in \mathcal{B}[k]} \epsilon_j[k] \geq 0$ .] This implies that  $\sum_j \epsilon'_j[k] \geq 0$  as long as  $\delta_{\max}[k] \leq \frac{1}{2+\varepsilon}$ .

For small values of  $\varepsilon$ , we know from (5) that our choice of  $K$  will ensure that

$$|\delta_{\max}[k] - \alpha[k] v_{\max}| \leq |E[k]|_{\infty} \leq a\varepsilon/2 < a\varepsilon,$$

where  $v_{\max} = \max_j \{v_j\}$ . Therefore, we have  $\sum_j \epsilon'_j[k] \geq 0$  as long as  $\alpha[k] \leq \frac{\frac{1}{2+\varepsilon} - a\varepsilon}{v_{\max}}$ . Specifically, this means that for  $k \geq K$ ,  $\alpha[k+1] \geq \alpha[k]$  if  $\alpha[k] \leq \frac{\frac{1}{2+\varepsilon} - a\varepsilon}{v_{\max}}$  (note that  $\alpha[k+1]$  might decrease if  $\alpha[k] > \frac{\frac{1}{2+\varepsilon} - a\varepsilon}{v_{\max}}$  but it will increase again if it goes below this value). We conclude that

$$\alpha[k] \geq \underbrace{\min\left(\alpha[K], \frac{\frac{1}{2+\varepsilon} - a\varepsilon}{v_{\max}} - n\varepsilon\right)}_{\alpha^* > 0}.$$

Note that  $n\varepsilon$  is the maximum decrease that  $\alpha[k+1]$  can suffer if  $\alpha[k]$  ever goes above  $\frac{\frac{1}{2+\varepsilon} - a\varepsilon}{v_{\max}}$ : to see this, notice from (6) that  $|\epsilon'_j[k+1]| \leq |\epsilon_j[k]| \leq \varepsilon/2 < \varepsilon$  for  $k \geq K$ . Clearly, we can always choose  $\varepsilon$  small enough so that  $\alpha^* > 0$ .

Consider now the parameterization  $\delta[k] = \alpha[k] v + E[k]$ , where  $E[k]$  converges to zero. For any  $\varepsilon_1 > 0$ , we can find  $K_1$  such that for  $k \geq K_1$ , we have  $|E[k]|_{\infty} \leq \varepsilon_1$ . Choose  $\varepsilon_1 < \alpha^* v_{\min}$ ; then, for  $k \geq \max\{K, K_1\}$  we have

$$\delta_{\min}[k] \geq \underbrace{\alpha^* v_{\min} - \varepsilon_1}_{\delta_{\min}^* > 0}.$$

#### D. Step **S4**: Linear Convergence of $V[k]$

Next, we establish that  $V[k]$  converges linearly to zero for  $k \geq \max\{K, K_1\}$ . [Recall that these choices of  $K$  and  $K_1$  imply that  $V[k] \leq \varepsilon$  for  $k \geq \max\{K, K_1\}$ .]

From the already established relationship  $\epsilon[k] = (\bar{P} - I_n)\delta[k]$  and the evolution of  $\delta_j[k]$  in (6), we have

$$\begin{aligned} \epsilon[k+1] &= (\bar{P} - I_n)\delta[k+1] \\ &= (\bar{P} - I_n)(\delta[k] + D[k]\epsilon[k]) \\ &= \epsilon[k] + (\bar{P} - I_n)D[k]\epsilon[k] \\ &= \underbrace{((I_n - D[k]) + \bar{P}D[k])}_{\mathbb{P}[k]} \epsilon[k], \end{aligned} \quad (7)$$

where  $D[k]$  is a diagonal matrix with entries that are either equal to (i)  $\delta_j[k]$  if node  $j \in \mathcal{A}[k]$ , or (ii)  $\frac{1 - \delta_j[k]}{1 + \epsilon_j[k]}$  if  $j \in \mathcal{B}[k]$  (note that in both cases the diagonal entries of  $D[k]$  are between 0 and 1). It is worth pointing out that the matrix  $\mathbb{P}[k]$

as defined in (7) is a column stochastic matrix regardless of the particular choices on the diagonal matrix  $D[k]$ .

We are interested in establishing the rate of convergence of  $V[k] = -2 \sum_{j \in \mathcal{A}[k]} \epsilon_j[k]$ , where we used the fact that  $-\sum_{j \in \mathcal{A}[k]} \epsilon_j[k] = \sum_{j \in \mathcal{B}[k]} \epsilon_j[k]$ . Let  $\epsilon^-[k]$ ,  $\epsilon^+[k]$  be the vectors that satisfy

$$\begin{aligned} \epsilon_j^-[k] &= \begin{cases} \epsilon_j[k], & \text{if } \epsilon_j[k] \leq 0 \\ 0, & \text{otherwise} \end{cases} \\ \epsilon_j^+[k] &= \begin{cases} \epsilon_j[k], & \text{if } \epsilon_j[k] > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where we have  $\sum_j \epsilon_j^-[k] = \sum_{j \in \mathcal{A}[k]} \epsilon_j[k] = -V[k]/2$ , and  $\sum_j \epsilon_j^+[k] = \sum_{j \in \mathcal{B}[k]} \epsilon_j[k] = V[k]/2$ .

Iterating the relationship in (7) for  $\ell$  steps, we have

$$\epsilon[k + \ell] = \underbrace{(\prod_{i=0}^{\ell-1} [I_n - D[k+i] + \bar{P}D[k+i]])}_{\mathbb{P}[k:(k+\ell-1)}} \epsilon[k],$$

where matrix  $\mathbb{P}[k:(k+\ell-1)]$ , being a product of column stochastic matrices, is a column stochastic matrix itself. Moreover, for any  $\ell$  we have

$$\begin{aligned} \epsilon[k + \ell] &:= \epsilon^+[k + \ell] + \epsilon^-[k + \ell] \\ &= \mathbb{P}[k:(k + \ell - 1)](\epsilon^+[k] + \epsilon^-[k]) \\ &= \underbrace{\mathbb{P}[k:(k + \ell - 1)]\epsilon^+[k]}_{\tilde{\varepsilon}^+[k+\ell]} + \underbrace{\mathbb{P}[k:(k + \ell - 1)]\epsilon^-[k]}_{\tilde{\varepsilon}^-[k+\ell]}, \end{aligned}$$

where (since  $\mathbb{P}[k:(k + \ell - 1)]$  is a column stochastic matrix) we have  $\sum_j \tilde{\varepsilon}_j^-[k + \ell] = \sum_j \epsilon_j^-[k] = -V[k]/2$  and  $\sum_j \tilde{\varepsilon}_j^+[k + \ell] = \sum_j \epsilon_j^+[k] = V[k]/2$ . It should also be clear that  $\sum_j \tilde{\varepsilon}_j^+[k + \ell] \leq \sum_j \epsilon_j^+[k + \ell]$  and  $\sum_j \tilde{\varepsilon}_j^-[k + \ell] \geq \sum_j \epsilon_j^-[k + \ell]$ .

Denote by  $j^-$  the entry that has the smallest (negative) value in  $\epsilon^-[k]$ , and by  $j^+$  the entry that has the largest (positive) entry in  $\epsilon^+[k]$ . Clearly, we have  $\frac{V[k]}{2n} \leq \epsilon_{j^+}^+[k] \leq \frac{V[k]}{2}$ , and  $-\frac{V[k]}{2} \leq \epsilon_{j^-}^-[k] \leq -\frac{V[k]}{2n}$ . Consider the two cases below.

**[Case 1]** Suppose that for some  $\ell \leq n$  the following holds true: at time steps  $k+1, k+2, \dots, k+\ell-1$ , node  $j^+$  has a positive  $\epsilon_{j^+}$  and then it becomes negative at time step  $k+\ell$ . This means that the  $(j^+, j^+)$  diagonal entry of matrices  $\mathbb{P}[k+l]$ ,  $l = 0, 1, \dots, \ell-1$ , is of the form

$$\begin{aligned} 1 - D[k+l](j^+, j^+) &= 1 - \frac{1 - \delta_{j^+}[k+l]}{1 + \epsilon_{j^+}[k+l]} = \frac{\delta_{j^+}[k+l] + \epsilon_{j^+}[k+l]}{1 + \epsilon_{j^+}[k+l]} \\ &\geq \frac{\delta_{\min}^*}{1 + V[k]/2} \geq \delta_{\min}^*/2 \end{aligned}$$

(assuming, without loss of generality, that  $V[k] < \varepsilon < 2$  for  $k \geq \max\{K, K_1\}$ ) and we have

$$\tilde{\varepsilon}_{j^+}^+[k + \ell - 1] \geq \left(\frac{\delta_{\min}^*}{2}\right)^{\ell-1} \frac{V[k]}{2n}.$$

Since  $\epsilon_{j^+}[k + \ell] < 0$ , this means that at time step  $k + \ell$ ,

$$\begin{aligned} V[k + \ell] &\leq V[k] - 2 \left(\frac{\delta_{\min}^*}{2}\right)^{\ell-1} \frac{V[k]}{2n} \\ &\leq \left(1 - \frac{(\delta_{\min}^*/2)^{\ell-1}}{n}\right) V[k]. \end{aligned}$$

**[Case 2]** Suppose that at time steps  $k+1, k+2, \dots, k+n$ , node  $j^+$  has a positive  $\epsilon_{j^+}$  (i.e., unlike Case **C1**, it does not become negative). Then, using the same argument as in Case **C1**, we have that

$$\tilde{\epsilon}_{j^+}^+[k+n-1] \geq \left(\frac{\delta_{\min}^*}{2}\right)^{n-1} \frac{V[k]}{2n}.$$

Now consider what happens to node  $j^-$ . Since the digraph is strongly connected, there exists some  $\ell, \ell \leq n-1$ , such that there is path from  $j^-$  to  $j^+$ , say through nodes  $j_0 = j^-, j_1, j_2, \dots, j_{\ell-1}, j_\ell = j^+$ . Suppose that the nodes  $j_0, j_1, \dots, j_{\ell-1}$  satisfy  $\epsilon_{j_0}[k] \leq 0, \epsilon_{j_1}[k+1] \leq 0, \dots, \epsilon_{j_{\ell-1}}[k+\ell-1] \leq 0$  (if this is not the case, the same argument that we develop below goes through using, instead of  $j^+$ , the first node in the path from  $j^-$  to  $j^+$  that violates this assumption). Considering matrix  $\mathbb{P}[k : (k+\ell-1)]$  we see that

$$\begin{aligned} \mathbb{P}[k : (k+\ell-1)] &= (I_n - D_0)(I_n - D_1)\dots(I_n - D_{\ell-1}) \\ &\quad + \overline{P}D_0(I_n - D_1)\dots(I_n - D_{\ell-1}) \\ &\quad + \dots + \overline{P}D_0\overline{P}D_1\overline{P}D_2\dots\overline{P}D_{\ell-1}, \end{aligned}$$

where all components in the above summation are nonnegative matrices (for notational convenience, we wrote  $D_l = D[k+l]$ ). In particular, note that the  $(j^+, j^-)$  entry in the matrix  $\overline{P}D_0\overline{P}D_1\overline{P}D_2\dots\overline{P}D_{\ell-1}$  will have a value that is at least  $\left(\frac{\delta_{\min}^*}{\mathcal{D}_{\max}}\right)^\ell$ , where  $\mathcal{D}_{\max} := \max_j \{\mathcal{D}_j^+\}$ . This follows by considering the path from  $j^-$  to  $j^+$  and realizing that the diagonal entries of the matrices  $D_0, D_1, \dots, D_{\ell-1}$  will satisfy  $D_l(j_l, j_l) = \delta_{j_l}[k+l] \geq \delta_{\min}^*$  for  $l = 0, 1, \dots, \ell-1$ . Thus, we have

$$\tilde{\epsilon}_{j^+}^-[k+\ell] \leq -\left(\frac{\delta_{\min}^*}{\mathcal{D}_{\max}}\right)^\ell \frac{V[k]}{2n}.$$

Since  $\tilde{\epsilon}_{j^+}^+[k+\ell] > (\delta_{\min}^*/2)^\ell \frac{V[k]}{2n}$ , we are guaranteed that

$$\begin{aligned} V[k+\ell] &\leq V[k] - 2\left(\frac{\delta_{\min}^*}{\mathcal{D}_{\max}}\right)^\ell \frac{V[k]}{2n} \\ &\leq \left(1 - \frac{1}{n}\left(\frac{\delta_{\min}^*}{\mathcal{D}_{\max}}\right)^\ell\right) V[k]. \end{aligned}$$

Clearly, the worst case scenario for **Case 2** occurs when  $\ell = n-1$ . Considering the worst case for both **Case 1** and **Case 2** (and since  $\mathcal{D}_{\max} \geq 2$ ) we have

$$V[k+n] \leq \underbrace{\left(1 - \frac{1}{n}\left(\frac{\delta_{\min}^*}{\mathcal{D}_{\max}}\right)^{n-1}\right)}_{\gamma^*} V[k].$$

#### E. Step S5: Establishing Convergence of $\delta[k]$

*Lemma 2:* For  $j \in \mathcal{V}$ , the sequence  $\delta_j[0] = \frac{\mathcal{D}_j^+}{1+\mathcal{D}_j^+}, \delta_j[1], \dots, \delta_j[k], \dots$  is Cauchy and has a limit  $\delta_j^*$ .

*Proof:* From (6) and the update for  $\epsilon'_j[k+1]$  below it, we have  $|\delta_j[k+1] - \delta_j[k]| = |\epsilon'_j[k+1]| \leq |\epsilon_j[k]| \leq V[k]/2$ , where the next to last inequality follows from the fact that  $0 \leq \delta_j[k] \leq 1$ , while the last inequality follows from the definition of  $V[k]$ . To show that  $\delta_j[k]$  is a Cauchy sequence, we need to establish that for any  $\epsilon$ , we can find an  $N$  such that, for any  $k_2 \geq k_1 \geq N$ , we have  $|\delta_j[k_2] - \delta_j[k_1]| < \epsilon$ .

Take  $\epsilon' = \frac{2(1-\gamma^*)}{n}\epsilon$  and let  $N$  be such that  $V[k] \leq \epsilon'$  for  $k \geq N$ . Then for  $k \geq N$  we have  $|\delta_j[k+1] - \delta_j[k]| \leq \epsilon'/2$ . For any  $k_2 \geq k_1 \geq N$ , we have

$$\begin{aligned} |\delta_j[k_2] - \delta_j[k_1]| &\leq \sum_{k=k_1}^{k_2-1} |\delta_j[k+1] - \delta_j[k]| \\ &\leq \sum_{k=k_1}^{\infty} |\delta_j[k+1] - \delta_j[k]| \\ &\leq \sum_{j=0}^{\infty} \sum_{k=k_1+j}^{k_1+(j+1)n-1} |\delta_j[k+1] - \delta_j[k]| \\ &\leq \sum_{j=0}^{\infty} (\gamma^*)^j \left(n\frac{\epsilon'}{2}\right) = \frac{1}{1-\gamma^*} n\epsilon'/2 = \epsilon. \end{aligned}$$

Thus,  $\delta_j[k]$  is a Cauchy sequence and has a limit  $\delta_j^*$ . ■

We conclude that  $\delta[k]$  has limit  $\delta^*$  and  $P[k]$  has limit  $P^* = \overline{P}\text{diag}(\delta^*) + (I_n - \text{diag}(\delta^*))$ .

#### IV. CONCLUDING REMARKS

In this paper we tackle a problem that we left open in [1]; specifically, we analyze the convergence of a distributed algorithm for matrix scaling to doubly stochastic form from [1]. Our analysis establishes that the rate of convergence of this algorithm is linear.

The relevance of such matrix scaling was demonstrated in [1], where we showed that the output of this algorithm at each iteration, i.e., the set of nonnegative (time-varying) weights, can be used in a linear time-varying algorithm that allows the nodes to asymptotically reach average consensus in a directed graph.

An interesting follow-up problem is to study the convergence rate of the aforementioned time-varying average consensus algorithm.

#### REFERENCES

- [1] A. D. Domínguez-García and C. N. Hadjicostis, "Distributed matrix scaling and application to average consensus in directed graphs," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 667–681, March 2013.
- [2] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.
- [3] S. Sundaram and C. N. Hadjicostis, "Distributed function calculation and consensus using linear iterative strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 650–660, May 2008.
- [4] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
- [5] R. Olfati-Saber and R. M. Murray, "Agreement problems in networks with directed graphs and switching topology," in *Proc. of IEEE Conference on Decision and Control*, vol. 4, Jun. 2003, pp. 4123–4132.
- [6] B. N. Parlett and T. L. Landis, "Methods for scaling to doubly stochastic form," *Linear Algebra and its Applications*, vol. 48, pp. 53–79, 1982.
- [7] M. H. Schneider and S. A. Zenios, "A comparative study of algorithms for matrix balancing," *Operations Research*, vol. 38, no. 3, pp. 439–455, May–Jun. 1990.
- [8] M. Mesbahi and M. Egerstedt, *Graph-Theoretic Methods in Multiagent Networks*, ser. Applied Mathematics Series. Princeton University Press, 2010.
- [9] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.
- [10] G. J. Foschini, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 641–646, Nov. 1993.
- [11] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1990.
- [12] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*, ser. Applied Mathematics Series. Princeton University Press, 2009.