Load balancing for Markov chains with a specified directed graph

Steve Kirkland*
Hamilton Institute
National University of Ireland Maynooth
Ireland

August 8, 2013

Abstract

Given a strongly directed graph D, let Σ_D be the set of stochastic matrices whose directed graph is a spanning subgraph of D. We consider the problem of finding the infimum of $||x||_\infty$ as x ranges over the set of stationary distribution vectors of irreducible matrices in Σ_D. Using techniques from nonlinear programming, combinatorial matrix theory, and nonnegative matrix theory, we find this infimum, which is given in terms the cardinality of a certain collection of vertex–disjoint cycles in D. The situation that the infimum is attained as a minimum for the stationary distribution vector of some irreducible matrix in Σ_D is also considered.

Keywords: Stochastic matrix; Directed graph; Stationary distribution vector.
AMS Classifications: 15A18, 15B48, 15B51, 60J10.

1 Introduction and motivation

A square matrix S is *stochastic* if all of its entries are nonnegative, and in addition $S\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is denotes an all–ones vector of the appropriate order. Stochastic matrices arise as the centrepiece of the theory of discrete time, time homogeneous Markov chains on a finite state space, and consequently, such matrices have received a good deal of attention over the last century.

Recall that a stochastic matrix S of order n is *irreducible* if it has the property that for each pair of indices i, j between 1 and n, there is a $k \in \mathbb{N}$ such that the (i, j)

*stephen.kirkland@nuim.ie Research supported in part by the Science Foundation Ireland under Grant Numbers SFI/07/SK/I1216b and 11/RFP.1/MTH/3185.

entry of S^k is positive. Further, if there is an $m \in \mathbb{N}$ such that S^m has all positive entries, then we say that S is *primitive*. From the Perron–Frobenius theorem, it follows that for any irreducible stochastic matrix S, there is a unique *stationary distribution vector* – that is, and entrywise positive vector x such that $x^T S = x^T$ and $x^T \mathbf{1} = 1$. The stationary distribution vector is of particular interest in the theory of Markov chains, for the following reason: in the case that S is primitive, for any initial distribution vector, the iterates of a Markov chain with transition matrix S converge to the corresponding stationary distribution vector for S. Thus, each entry in the stationary distribution vector can be interpreted as the long–term probability that the Markov chain is in the corresponding state.

Associated with the any $n \times n$ stochastic matrix S we have a corresponding *directed graph*, $\mathcal{D}(S)$. The vertices of $\mathcal{D}(S)$ are labelled $1, \ldots, n$, and there is an arc $i \to j$ in $\mathcal{D}(S)$ if and only if $s_{ij} > 0$. Observe that $\mathcal{D}(S)$ captures qualitative information about the Markov chain with transition matrix S, as $\mathcal{D}(S)$ records which transitions are possible in one step of the Markov chain. The connections between the structure of directed graph $\mathcal{D}(S)$ and the irreducibility or primitivity of the stochastic matrix S are well–known: S is irreducible if and only if $\mathcal{D}(S)$ is strongly connected, and S is primitive if and only $\mathcal{D}(S)$ is strongly connected, and in addition, the greatest common divisor of the cycle lengths of $\mathcal{D}(S)$ is equal to 1 (see [1]). In view of this connection between the combinatorial properties of $\mathcal{D}(S)$ and the analytic properties of S, it is natural to wonder whether the qualitative information contained in $\mathcal{D}(S)$ can be used to unearth further quantitative information about S. There is an existing body of work in that direction that focuses on the eigenvalues of the stochastic matrix in question; see for example [2], [3] and [4]. In this paper, we adopt a related perspective by considering the influence of the directed graph $\mathcal{D}(S)$ on the structure of the corresponding stationary distribution vector for S.

Markov chains are used extensively to model such diverse phenomena as the succession of species in mathematical ecology (see [5]), the configuration of molecules in mathematical chemistry (see [6]), and the levels of vehicle congestion or pollutants in road networks (see [7] and [8], respectively). In a number of these settings, the transition matrices in question are equipped with an underlying combinatorial structure that is determined by the application. For instance in the vehicle traffic application, where the states of the Markov chain represent road segments, the structure of the road network determines which road segments can be reached from a given road segment in one time step, thus placing restrictions on the directed graph associated with the Markov chain's transition matrix.

In this paper, we consider the following class of stochastic matrices. Given a strongly connected directed graph D on vertices $1, \ldots, n$, let

$$\Sigma_D = \{S \in \mathbb{R}^{n \times n} | S \geq 0, S\mathbf{1} = \mathbf{1}, \mathcal{D}(S) \subseteq D\}.$$

In other words, Σ_D is the set of all stochastic matrices whose directed graph is a spanning subgraph of D. Given an irreducible matrix $S \in \Sigma_D$ having stationary distribution vector x, what sorts of constraints on x are imposed by the directed graph

D? That question is the primary focus of this paper. The following preliminary result provides a little insight into this line of inquiry.

Proposition 1.1. *Suppose that S is an irreducible stochastic matrix of order $n \geq 2$, and having stationary distribution vector x. Fix and i between 1 and n, and let g_i denote the length of a shortest cycle in $\mathcal{D}(S)$ that passes through vertex i. Then*

$$x_i \leq \frac{1}{g_i}. \tag{1}$$

Further, equality holds in (1) if and only if $g_i \geq 2$, every cycle in $\mathcal{D}(S)$ passes through vertex i, and every cycle in $\mathcal{D}(S)$ has length g_i.

Proof. Without loss of generality, we may assume that $i = n$. Note further that if $g_n = 1$, then clearly $x_n < 1 = \frac{1}{g_n}$, so that the conclusion certainly holds in that case. Henceforth we assume that $g_n \geq 2$.

Partition out the last row and column of S as

$$S = \left[\begin{array}{c|c} \tilde{S} & (I - \tilde{S})\mathbf{1} \\ \hline r^T & 1 - r^T\mathbf{1} \end{array}\right].$$

Using the eigen–equation $x^T S = x^T$ and the fact that $x^T\mathbf{1} = 1$, we deduce readily that

$$x_n = \frac{1}{1 + r^T(I - \tilde{S})^{-1}\mathbf{1}}.$$

We have $(I - \tilde{S})^{-1}\mathbf{1} = \sum_{k=0}^{\infty} \tilde{S}^k\mathbf{1}$. For each $i = 1, \dots, n-1$, let p_i denote the length of a shortest path in $\mathcal{D}(S)$ from vertex i to vertex n. We have $(I - \tilde{S})^{-1}\mathbf{1} = \sum_{k=0}^{\infty} \tilde{S}^k\mathbf{1}$. For each $i = 1, \dots, n-1$, let p_i denote the length of a shortest path in $\mathcal{D}(S)$ from vertex i to vertex n. We find that for each $i = 1, \dots, n-1, e_i^T\tilde{S}^k\mathbf{1} = 1$ if and only if $p_i > k$. Observe that $r_i > 0$ only if $n \to i$ in $\mathcal{D}(S)$, and that since $g_n \geq 2$, the (n, n) entry of S is zero, so that necessarily $r^T\mathbf{1} = 1$.

Consider the quantity $r^T(I - \tilde{S})^{-1}\mathbf{1}$. We have $r^T(I - \tilde{S})^{-1}\mathbf{1} = \sum_{i=1}^{n-1}\sum_{k=0}^{\infty} r_i e_i^T\tilde{S}^k\mathbf{1} = \sum_{n\to i}\sum_{k=0}^{\infty} r_i e_i^T\tilde{S}^k\mathbf{1} \geq \sum_{n\to i}\sum_{k=0}^{p_i-1} r_i e_i^T\tilde{S}^k\mathbf{1} = \sum_{n\to i} r_i p_i$. Note that for each vertex i such that $n \to i$, we have $1 + p_i \geq g_n$. Consequently, we find that $\sum_{n\to i} r_i p_i \geq \sum_{n\to i} r_i(g_n - 1) = g_n - 1$. It now follows that

$$x_n = \frac{1}{1 + r^T(I - \tilde{S})^{-1}\mathbf{1}} \leq \frac{1}{g_n},$$

establishing (1).

Suppose now that equality holds in (1), and observe that necessarily $g_n \geq 2$ in that case. From the argument above, it follows that for each i such that $n \to i$, we have $p_i = g_n - 1$, and $e_i\tilde{S}^k\mathbf{1} = 0$ for all $k \geq g_n - 1$. This last condition is equivalent to the statement that if $n \to i$ in $\mathcal{D}(S)$, every walk in $\mathcal{D}(S)$ starting from vertex i and having length $k \geq g_n - 1$ must pass through vertex n. We claim that necessarily every cycle in $\mathcal{D}(S)$ must pass through vertex n. To see the claim, suppose to the contrary that there is some cycle C that does not pass through vertex n. Since S is irreducible, there is some shortest path from n to C in $\mathcal{D}(S)$ that includes the arc $n \to i_0$, say. It now follows that there are walks from vertex i_0 of arbitrarily long length that do not pass through vertex n, contrary to our hypothesis. Thus, every cycle in $\mathcal{D}(S)$ passes through vertex n, as claimed. Finally, since $p_i = g_n - 1$ for each i such that $n \to i$, and since every path from such an i back to vertex n must have length $g_n - 1$, it follows that every cycle through vertex n has length g_n.

Conversely, if every cycle in $\mathcal{D}(S)$ passes through vertex n and has length g_n, it is readily verified that equality holds in (1).

□

We note in passing that (1) can also be established by appealing to the standard fact (see [9]) that the mean first return time to state i is given by $\frac{1}{x_i}$, and then observing that since a shortest path from vertex i back to vertex i has length g_i, it must be the case that the mean first return time to state i is at least g_i, so that $\frac{1}{x_i} \geq g_i$. Note however, that while this line of reasoning establishes (1), it does not readily yield the characterisation of equality in (1).

The following is immediate from Proposition 1.1.

Corollary 1.1. *Suppose that S is an irreducible stochastic matrix of order $n \geq 2$. Let x denote the stationary vector for S, and let g denote the length of a shortest cycle in $\mathcal{D}(S)$. Then for each $j = 1, \ldots, n$ we have $x_j \leq \frac{1}{g}$.*

While our focus on this paper is on stationary vectors for stochastic matrices, it turns out that our results have implications for more general nonnegative matrices. Specifically, suppose that we have an $n \times n$ irreducible nonnegative matrix M, and denote its Perron value by r. Let u and v be right and left Perron vectors, respectively, for M, normalised so that $v^T u = 1$. Letting U be the diagonal matrix of order n whose i–th diagonal entry is $u_i, i = 1, \ldots, n$, it is readily established that the matrix $S = \frac{1}{r} U^{-1} M U$ is irreducible and stochastic, and that $\mathcal{D}(S) = \mathcal{D}(M)$. Moreover, denoting the stationary distribution of S by x, we find that $x = Uv$, so that in particular, for each $i = 1, \ldots, n$, we have $x_i = u_i v_i$. It is known (see [11]) that for each $i = 1, \ldots, n$, the derivative of the Perron value of M with respect to its i–th diagonal entry is given by the quantity $u_i v_i$. Consequently, the results in this paper also have implications for the sensitivity of the Perron value with respect to the diagonal entries of M.

The following is immediate from Proposition 1.1, Corollary 1.1, and the preceding remarks.

Corollary 1.2. *Suppose that M is an irreducible nonnegative matrix of order $n \geq 2$ with right and left Perron vectors u and v, respectively, normalised so that $v^T u = 1$. For each i between 1 and n, let g_i denote the length of a shortest cycle in $\mathcal{D}(M)$ that passes through vertex i. Then for any $i \in \{1, \ldots, n\}, u_i v_i \leq \frac{1}{g_i}$, with equality holding if and only if $g_i \geq 2$, every cycle in $\mathcal{D}(M)$ passes through vertex i, and every cycle* *in $\mathcal{D}(M)$ has length g_i. In particular, letting g be the length of a shortest cycle in $\mathcal{D}(M)$, we have $u_j v_j \le \frac{1}{g}, j = 1, \ldots, n$.*

Observe that while Proposition 1.1 uses the directed graph of a transition matrix in order to provides a bound on a particular entry in the corresponding stationary distribution vector, it does so without considering where that entry of the stationary distribution vector sits in relation to the remaining entries in that vector. In the remainder of the paper, we fix our attention on the maximum entry of the stationary distribution vector, with a view to understanding when that maximum entry can be made as small as possible. To be precise, we address the following problem:
given a strongly connected directed graph D on n vertices, find

$$\inf\{||x||_\infty | x \text{ is the stationary distribution vector for some irreducible } S \in \Sigma_D\}.$$

Note that this problem can be thought of as a 'load balancing' problem, in the sense that we seek to make the maximum entry in the stationary distribution vector as small as possible, subject to the constraint that the directed graph of the corresponding transition matrix is a strongly connected spanning subgraph of D. Referring to the application of Markov chain techniques to model vehicle traffic in a road network, and noting that in that setting, the entries in the stationary distribution vector are interpreted as the traffic congestion levels on the corresponding road segments, we see that the problem of minimising the maximum entry in the stationary distribution corresponds to ensuring that the maximum congestion level is as small as possible. In this setting at least, the term 'load balancing' seems to be an appropriate one for the problem under consideration.

We note that the so–called Matrix Tree Theorem for Markov chains (see [10]) provides an expression for an entry in the stationary distribution vector of an irreducible stochastic matrix, and that this expression is based in part on the structure of the directed graph for the corresponding transition matrix. That said, we see no straightforward way of using the Matrix Tree Theorem for Markov chains to establish the results in the sequel.

Throughout the sequel, we will freely use standard notions and results on stochastic matrices, directed graphs, and combinatorial matrix theory. We refer the interested reader to [9], [12] and [1], respectively, for the necessary background material.

2 Minimising the maximum entry in the stationary distribution

Recall that the *term rank* of a square matrix A is the minimum number of lines (i.e. rows and columns) that contain all of the nonzero entries of A. By König's theorem, the term rank of A coincides with the maximum number of nonzero entries of A no two of which lie in the same line of A. We refer the interested reader to [1] for further details on the term rank and König's theorem. The following result, whose proof employs the term rank, will be useful in establishing our main results; it may also be of independent interest. Here, for a directed graph D, we denote its vertex set by $V(D)$.

Proposition 2.1. *Let S be a stochastic matrix of order $n \geq 2$, and suppose that we have a vector $x \in \mathbb{R}^n$ such that $x \geq 0, x^T\mathbf{1} = 1$, and $x^T S = x^T$. Let $\mathcal{I} = \{i | x_i < ||x||_\infty\}, \mathcal{J} = \{j | x_j = ||x||_\infty\}$ and let K be the diagonal matrix such that $k_{i,i} = 1$ for all $i \in \mathcal{I}$ and $k_{j,j} = 0$ for all $j \in \mathcal{J}$. We have the following conclusions.*
a) The matrix $S + K$ has term rank equal to n.
b) There is a collection of vertex–disjoint directed cycles in $\mathcal{D}(S)$, say $C_1, \ldots, C_k$, such that $\mathcal{J} \subseteq \cup_{l=1}^k V(C_l)$.

Proof. a) Suppose first that $|\mathcal{J}| = n$. Then all entries of x are equal, and so S is doubly stochastic. (Observe also that K is the zero matrix.) From Birkhoff's theorem (see [1]) it now follows that S has term rank n.

Next we suppose that $|\mathcal{J}| = m \leq n - 1$, and without loss of generality we assume that $\mathcal{J} = \{1, \ldots, m\}$. Partition out the first m rows and columns of S as $S = \left[\begin{array}{c|c} S_{1,1} & S_{1,2} \\ \hline S_{2,1} & S_{2,2} \end{array}\right]$, and partition x conformally with S as $x = \left[\begin{array}{c} x_1\mathbf{1} \\ \hline y \end{array}\right]$; note that each entry of y is less than x_1. Let Δ be the diagonal matrix of order $n - m$ whose diagonal entries are given by the corresponding entries of $\frac{1}{x_1}y$.

Consider the matrix $\widehat{S}$ given by

$$\widehat{S} = \left[\begin{array}{c|c} S_{1,1} & S_{1,2} \\ \hline \Delta S_{2,1} & \Delta S_{2,2} + (I - \Delta) \end{array}\right],$$

and note that the zero-nonzero pattern of $\widehat{S}$ coincides with that of $S + K$. It is straightforward to determine that $\widehat{S}$ is doubly stochastic. Again applying Birkhoff's theorem, we find that $\widehat{S}$, and hence $S + K$, has term rank equal to n.

b) Denote the directed graph of the matrix $S + K$ by $\widehat{D}$. From part a) it follows that there is a spanning subgraph of $\widehat{D}$ that consists of a vertex–disjoint collection of cycles. Noting that the only arcs in $\widehat{D}$ that are not arcs of $\mathcal{D}(S)$ are arcs of the form $i \rightarrow i$, where $i \in \mathcal{I}$, the conclusion now follows readily.

$\square$

Our next result adopts the perspective of nonlinear programming in order to address the problem of minimising the maximum entry in the stationary distribution vector.

Proposition 2.2. *Suppose that $\overline{S}$ is a stochastic matrix of order n, and that $\overline{x} \in \mathbb{R}^n$ with $\overline{x} > 0, \overline{x}^T\mathbf{1} = 1$ and $\overline{x}^T\overline{S} = \overline{x}^T$. Suppose further that for some index k between 1 and $n - 1$ we have $\overline{x}_j = \overline{x}_1$ for $j = 1, \ldots, k$ and $\overline{x}_j < \overline{x}_1$ for $j = k + 1, \ldots, n$.*

Finally, suppose that $\overline{x}$ is a solution to the following optimisation problem:

$$\begin{aligned}
&\text{minimise } x_1 \text{ subject to:}\\
&x \geq 0, x^T\mathbf{1} = 1, x_j = x_1, j = 1, \ldots, k, x_j \leq x_1, j = k+1, \ldots, n,\\
&x^T S = x^T \text{ for some stochastic matrix } S \text{ with } \mathcal{D}(S) \subseteq \mathcal{D}(\overline{S}).
\end{aligned} \tag{2}$$

Then there is a collection of vertex disjoint cycles $C_1, \ldots, C_p$ in $\mathcal{D}(\overline{S})$ such that $1 \in \cup_{l=1}^p V(C_l)$ and

$$\overline{x}_1 = \frac{1}{\sum_{l=1}^p |V(C_l)|}.$$

Proof. Denote $\mathcal{D}(\overline{S})$ by $\overline{\mathcal{D}}$ and for each i, j between 1 and n such that $i \to j$ in $\overline{\mathcal{D}}$ we denote the arc $i \to j$ by $a_{i,j}$. Next, observe that $\overline{x}$ is a solution to the following nonlinear programming problem.

$$\begin{aligned}
&\text{Minimise } x_1 \text{ subject to:}\\
&x_1 \geq x_j, j = k+1, \ldots, n,\\
&kx_1 + \sum_{j=k+1}^{n} x_j = 1,\\
&x_1 \sum_{i=1,\ldots,k, a_{i,j} \in \overline{\mathcal{D}}} s_{i,j} + \sum_{i=k+1,\ldots,n, a_{i,j} \in \overline{\mathcal{D}}} x_i s_{i,j} - x_1 = 0, j = 1, \ldots, k,\\
&x_1 \sum_{i=1,\ldots,k, a_{i,j} \in \overline{\mathcal{D}}} s_{i,j} + \sum_{i=k+1,\ldots,n, a_{i,j} \in \overline{\mathcal{D}}} x_i s_{i,j} - x_j = 0, j = k+1, \ldots, n,\\
&\sum_{j=1,\ldots,n, a_{i,j} \in \overline{\mathcal{D}}} s_{i,j} = 1, i = 1, \ldots, n,\\
&x_i \geq 0, i = 1, i = k+1, \ldots, n,\\
&s_{i,j} \geq 0 \text{ for all } a_{i,j} \in \overline{\mathcal{D}}.
\end{aligned} \tag{3}$$

We note here that in (3), we are thinking of the variables over which the minimum is taken as $x_1, x_{k+1}, \ldots, x_n$ and $s_{i,j}$ for each i, j such that $a_{i,j} \in \overline{\mathcal{D}}$.

From the hypothesis, we find that $\overline{x}_1, \overline{x}_{k+1}, \ldots, \overline{x}_n$ along with the elements $\overline{s}_{i,j}$ such that $a_{i,j} \in \overline{\mathcal{D}}$, are all positive, and that taken together they yield an absolute minimum to (3). Consequently, the Karush–Kuhn–Tucker necessary conditions apply (see [13]). In order to formulate those conditions, we require a little extra notation. Suppose that $\overline{\mathcal{D}}$ has m arcs, and select an ordering of those arcs. Construct the $m \times n$ $(0, 1)$ matrices $E(\text{out})$ and $E(\text{in})$ as follows: for each arc $a_{p,q} \in \overline{\mathcal{D}}$, we set

$$E(\text{out})_{\mathrm{a_{p,q},r}} = \begin{cases} 1 & \text{if } r = p, \\ 0 & \text{if } r \neq p \end{cases}, \mathrm{E(in)}_{\mathrm{a_{p,q},r}} = \begin{cases} 1 & \text{if } r = q, \\ 0 & \text{if } r \neq q \end{cases}.$$

That is, $E(\text{out})$ is an $m \times n$ incidence matrix for the initial vertices of the arcs in $\overline{\mathcal{D}}$, while $E(\text{in})$ is an $m \times n$ incidence matrix for the terminal vertices of the arcs in $\overline{\mathcal{D}}$. Next, we let Δ be the diagonal matrix of order m such that for each arc $a_{p,q} \in \overline{\mathcal{D}}$, the corresponding diagonal entry of Δ is $\overline{x}_p$.

With this notation in place, we can now formulate the Karush–Kuhn–Tucker necessary conditions. We order the variables by putting $x_1, x_{k+1}, \ldots, x_n$ first, then following them with the $s_{i,j}$s where the latter are ordered according to the ordering of the arcs in $\overline{\mathcal{D}}$ selected above. Computing the necessary gradient vectors, we find from the Karush–Kuhn–Tucker conditions that there exist scalars $\lambda_k, \ldots, \lambda_{n-1}, \lambda_n, \lambda_{n+1}, \ldots, \lambda_{2n}, \lambda_{2n+1}, \ldots, \lambda_{3m}$ such that

$$\left[\begin{array}{c} e_1 \\ \hline 0 \end{array}\right] + \left[\begin{array}{c|c|c|c|c} (e_1 - e_{k+1}) & (e_1 - e_{k+2}) & \ldots & (e_1 - e_n) & ((k-1)e_1 + \mathbf{1}) \\ \hline 0 & 0 & \ldots & 0 & 0 \end{array}\right] \begin{bmatrix} \lambda_k \\ \vdots \\ \lambda_{n-1} \\ \lambda_n \end{bmatrix}$$

$$+ \left[\begin{array}{c} U(\overline{S} - I) \\ \hline \Delta E(\text{in}) \end{array}\right] \begin{bmatrix} \lambda_{n+1} \\ \vdots \\ \lambda_{2n-1} \\ \lambda_{2n} \end{bmatrix} + \left[\begin{array}{c} 0 \\ \hline E(\text{out}) \end{array}\right] \begin{bmatrix} \lambda_{2n+1} \\ \vdots \\ \lambda_{3n-1} \\ \lambda_{3n} \end{bmatrix} = 0, \quad (4)$$

where U is the $(n-k+1) \times n$ matrix

$$U = \left[\begin{array}{c|c} \mathbf{1}_k^T & 0_{n-k}^T \\ \hline 0 & I_{n-k} \end{array}\right]$$

(here the subscripts on $\mathbf{1}, 0$ and I denote the orders of the all ones vector, zero vector, and identity matrix, respectively). Also, since $x_j < x_1$ for $j = k+1, \ldots, n$, we find, again from the Karush–Kuhn–Tucker conditions, that $\lambda_j = 0, j = k, \ldots, n-1$. We rewrite (4) in two sets of equations as

$$U(I - \overline{S}) \begin{bmatrix} \lambda_{n+1} \\ \vdots \\ \lambda_{2n} \end{bmatrix} = e_1 + \lambda_n \begin{bmatrix} k \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad (5)$$

$$\Delta E(\text{in}) \begin{bmatrix} \lambda_{n+1} \\ \vdots \\ \lambda_{2n} \end{bmatrix} + \mathrm{E(out)} \begin{bmatrix} \lambda_{2n+1} \\ \vdots \\ \lambda_{3n} \end{bmatrix} = 0. \quad (6)$$

Suppose that there are indices i, j_1, j_2 such that the arcs a_{i,j_1}, a_{i,j_2} are both in $\overline{\mathcal{D}}$. Considering the entry of (6) corresponding to the arc a_{i,j_1}, we see that $x_i\lambda_{n+j_1} + \lambda_{2n+i} = 0$. Similarly, considering the entry of (6) corresponding to the arc a_{i,j_2}, yields $x_i\lambda_{n+j_2} + \lambda_{2n+i} = 0$. Thus we have $x_i\lambda_{n+j_1} + \lambda_{2n+i} = 0 = x_i\lambda_{n+j_2} + \lambda_{2n+i}$, from which we conclude that $\lambda_{n+j_1} = \lambda_{n+j_2}$.

Next, we consider (5). Since $\left[\begin{array}{cccc}\overline{x}_1 & \overline{x}_{k+1} & \ldots & \overline{x}_n\end{array}\right]U = \overline{x}^T$, it follows from (5) that $0 = \overline{x}_1 + \lambda_n(k\overline{x}_1 + \sum_{j=k+1}^n \overline{x}_j) = \overline{x}_1 + \lambda_n$. Hence we have $\lambda_n = -x_1$. Applying Proposition 2.1 b) to $\overline{S}$ and $\overline{x}$, we find that there is a collection of vertex disjoint cycles, $C_1, \ldots, C_p$ in $\overline{\mathcal{D}}$ such that $\{1, \ldots, k\} \subseteq \cup_{l=1}^p V(C_l)$. Without loss of generality, we may take $\cup_{l=1}^p V(C_l) = \{1, \ldots, m\}$.

Now construct a stochastic matrix $\tilde{S}$ of order n as follows: for each $i = 1, \ldots, m$, set $\tilde{s}_{i,j} = 1$ if $i \to j$ is an arc in $\cup_{l=1}^p C_l$ and $\tilde{s}_{i,j} = 0$ otherwise; for each $i = m+1, \ldots, n$, set $e_i^T\tilde{S}$ equal to $e_i^T\overline{S}$. Observe that the leading $m \times m$ principal submatrix of $\tilde{S}$ is a permutation matrix. In particular, we have $\left[\begin{array}{c|c}\mathbf{1}_m^T & 0_{n-m}^T\end{array}\right](I - \tilde{S}) = 0^T$. Further, from the fact, established above, that $\lambda_{n+j_1} = \lambda_{n+j_2}$ whenever there is an index i such that $i \to j_1, j_2$ in $\overline{\mathcal{D}}$, it follows that

$$(I - \tilde{S})\begin{bmatrix}\lambda_{n+1}\\ \vdots \\ \lambda_{2n}\end{bmatrix} = (I - \overline{S})\begin{bmatrix}\lambda_{n+1}\\ \vdots \\ \lambda_{2n}\end{bmatrix}.$$

Multiplying both sides of (5) from the left by the vector $y^T = \left[\begin{array}{c|c|c}1 & \mathbf{1}_{m-k}^T & 0_{n-m}^T\end{array}\right]$, we find that

$$y^TU(I - \tilde{S})\begin{bmatrix}\lambda_{n+1}\\ \vdots \\ \lambda_{2n}\end{bmatrix} = y^TU(I - \overline{S})\begin{bmatrix}\lambda_{n+1}\\ \vdots \\ \lambda_{2n}\end{bmatrix} = y^T\left(e_1 + \lambda_n\begin{bmatrix}k\\ 1\\ \vdots \\ 1\end{bmatrix}\right).$$

Since $y^TU(I - \tilde{S}) = \left[\begin{array}{c|c}\mathbf{1}_m^T & 0_{n-m}^T\end{array}\right](I - \tilde{S}) = 0^T$, it now follows that $0 = 1 + m\lambda_n$. But then we have $\lambda_n = \frac{-1}{m}$, from which we deduce that

$$x_1 = \frac{1}{m} = \frac{1}{\sum_{l=1}^p |V(C_l)|},$$

as desired.

□

Next, we show by example that the conclusion of Proposition 2.2 may fail to hold if $\overline{x}$ is not an optimal solution to (2).

Example 2.1. Fix an $a \in (0, 1)$, and consider the matrix $\overline{S} = \begin{bmatrix}a & 1-a\\ 1 & 0\end{bmatrix}$. The stationary distribution vector for $\overline{S}$ is given by $\overline{x} = \frac{1}{2-a}\begin{bmatrix}1\\ 1-a\end{bmatrix}$, so that $\overline{x}_1 > \overline{x_2}$. However, observe that since $\frac{1}{2} < \overline{x}_1 < 1$, the conclusion of Proposition 2.2 fails to hold. Evidently $\overline{x}$ is not an optimal solution to (2).

Next, we introduce a little notation that will be useful in the sequel. Let D be a strongly connected directed graph on vertices $1, \ldots, n$, and fix an index i

between 1 and n. For each collection $\mathcal{C}$ of vertex disjoint cycles in D that passes through i, let $m_{\mathcal{C}} = |V(\mathcal{C})|$. Denote the maximum of all such $m_{\mathcal{C}}$'s by $\overline{m}_i(D)$. Finally, let $\overline{m}(D) = \max\{\overline{m}_i(D)|i = 1, \ldots, n\}$. Propositions 2.1 and 2.2 will assist in establishing the following theorem, which is one of our main results.

Theorem 2.1. *Let D be a strongly connected directed graph on n vertices, and consider the set $P_i = \{x \geq 0|x^T\mathbf{1} = 1, x_i = ||x||_\infty$, and $x^T S = x^T$ for some $S \in \Sigma_D\}$. Then*

$$\min\{x_i|x \in P_i\} = \frac{1}{\overline{m}_i(D)}.$$

Proof. We begin with a claim that $P_i \neq \emptyset$. To see the claim, let S be the $(0,1)$ adjacency matrix of a spanning subgraph of D containing a single directed cycle, passing through vertex i. It is readily seen that the corresponding left Perron vector of S, normalised so that its entries sum to one, is an element of P_i. Suppose now that x is a vector in P_i. By Proposition 2.2, there is a collection $\mathcal{C}$ of vertex disjoint cycles in D such that $x_i \geq \frac{1}{m_{\mathcal{C}}}$. Consequently we have $x_i \geq \frac{1}{\overline{m}_i(D)}$ whenever $x \in P_i$.

Next, let $\mathcal{C}_0$ be a collection of vertex disjoint cycles in D passing through vertex i such that $m_{\mathcal{C}_0} = \overline{m}_i(D)$, and let S be a $(0,1)$ matrix in Σ_D whose principal submatrix corresponding to $V(\mathcal{C}_0)$ is the adjacency matrix of $\mathcal{C}_0$. Consider the vector x such that

$$x_j = \begin{cases} \frac{1}{\overline{m}_i} & \text{if } j \in V(\mathcal{C}_0), \\ 0 & \text{if } j \notin V(\mathcal{C}_0). \end{cases}$$

Evidently $x \in P_i$ and $x_i = \frac{1}{\overline{m}_i(D)}$, and so the conclusion follows readily. □

From Theorem 2.1, we find that there is an $x \in P_i$ such that $x_i = \frac{1}{\overline{m}_i(D)}$, and $x^T S = x^T$ for some $S \in \Sigma_D$. Note however that the matrix S may be reducible. The following result discusses a certain limit of stationary distribution vectors for irreducible matrices in Σ_D. We introduce the following notation: given an $n \times n$ matrix A and a set $W \subseteq \{1, \ldots, n\}$, we let $A(W)$ denote the principal submatrix of A on the rows and columns indexed by W.

Proposition 2.3. *Let D be a strongly connected graph on n vertices, and let $C_1, \ldots, C_k$ be a collection of vertex disjoint cycles in D. Denote $\cup_{j=1}^k V(C_j)$ by W, let $w = |W|$, and let u be the indicator vector for W (that is, $u_i = 1$ if $i \in W$, and $u_i = 0$ if $i \notin W$). Then there is a sequence of irreducible matrices $S_p \in \Sigma_D$ such that:*
a) the sequence $S_p(W)$ converges to the adjacency matrix of $\cup_{j=1}^k C_j$ as $p \to \infty$; and
b) denoting the stationary vector of S_p by $x(p)$ for $p \in \mathbb{N}$, , we have $x(p) \to \frac{1}{w}u$ as $p \to \infty$.

Proof. We proceed by induction on n, and note that the case that $n = 2$ is easily established. Suppose now that $n \geq 3$, and suppose without loss of generality that $W = \{1, \ldots, w\}$. Consider the directed graph $\hat{D}$ on $n - w + k$ vertices formed as follows:

i) for each $j = 1, \ldots, k$, identify the vertices of C_j with a single vertex j;
ii) for each $j = k+1, \ldots, n-w+k$, identify vertex j of $\hat{D}$ with vertex $j+w-k$ of D;
iii) for $i, j = 1, \ldots, k$, $i \to j$ is an arc in $\hat{D}$ provided that in D there are vertices $a \in C_i, b \in C_j$ such that $a \to b$;
iv) for each $i = 1, \ldots, k, j = k+1, \ldots n-w+k$, $i \to j$ is an arc in $\hat{D}$ provided that in D there is a vertex $a \in C_i$, such that $a \to j+w-k$, and $j \to i$ is an arc in $\hat{D}$ provided that in D there is a vertex $b \in C_i$, such that $j+w-k \to b$;
v) for each $i, j = k+1, \ldots, n-w+k$, $i \to j$ is an arc in $\hat{D}$ provided that $i+w-k \to j+w-k$ in D.
It is straightforward to show that since D is strongly connected, then so is $\hat{D}$. Further, by deleting arcs from $\hat{D}$ if necessary, we may assume that $\hat{D}$ is a minimally strong directed graph. (Recall that a strongly connected directed graph is *minimally strong* if it has the property that the deletion of any arc yields a directed graph that is no longer strongly connected.) Since $\hat{D}$ is minimally strong, it has no loops. Further, we find from Lemma 3.3.2 of [1] that there is a vertex j of $\hat{D}$ having indegree 1 and outdegree 1. We consider the following two cases.

Case 1, $k+1 \le j \le n-w+k$:
Without loss of generality we take $j = n-w+k$. In that case we may consider $n \times n$ irreducible matrices in Σ_D of the following form

$$\left[\begin{array}{c|c} A & te_l \\ \hline e_m^T & 0 \end{array}\right].$$

Observe that for such a matrix, $A + te_l e_m^T$ is irreducible and stochastic, and that $\cup_{r=1}^k C_r \subseteq \mathcal{D}(A + te_l e_m^T)$. Suppose first that the arc $l \to m$ is not among the arcs of $\cup_{r=1}^k C_r$. Let D_0 be the subgraph of D induced by vertices $1, \ldots, n-1$. Applying the induction hypothesis, we find that there is a sequence of matrices of the form $B_p \equiv A_p + t_p e_l e_m^T \in \Sigma_{D_0 \cup \{l \to m\}}$ whose sequence of stationary vectors $x(p)$ converges to the appropriately normalised indicator vector of W in $\mathbb{R}^{n-1}$ and such that $B_p(W)$ converges to the adjacency matrix of $\cup_{r=1}^k C_r$. Note in particular that since $l \to m$ is not among the arcs of $\cup_{r=1}^k C_r$, we have $t_p \to 0$ as $p \to \infty$. For each $p \in \mathbb{N}$, consider the irreducible matrix

$$S_p = \left[\begin{array}{c|c} A_p & t_p e_l \\ \hline e_m^T & 0 \end{array}\right],$$

which has stationary distribution given by

$$\left[\begin{array}{c|c} \frac{1}{1+t_p x(p)_l} x(p)^T & \frac{t_p x(p)_l}{1+t_p x(p)_l} \end{array}\right]^T.$$

It now follows that the sequence S_p has the desired properties.

On the other hand, if the arc $l \to m$ is one of the arcs of $\cup_{r=1}^k C_r$, we again consider the sequence of matrices A_p and scalars t_p above. Note that t_p may not converge to zero in this case. For each $p \in \mathbb{N}$, let

$$\tilde{S}_p = \left[\begin{array}{c|c} A_p + \frac{p-1}{p} t_p e_l e_m^T & \frac{1}{p} t_p e_l \\ \hline e_m^T & 0 \end{array}\right].$$

It follows that the corresponding stationary vectors are given by

$$\left[\begin{array}{c|c} \frac{p}{p+t_p \tilde{x}(p)_l} x(p)^T & \frac{t_p x(p)_l}{p+t_p x(p)_l} \end{array}\right]^T,$$

and hence the sequence $\tilde{S}_p$ has the desired properties.

Case 2, $1 \le j \le k$:
If $k = 1$, let A be a $(0,1)$ matrix in Σ_D whose directed graph contains a single cycle, namely C_1. Since the irreducible matrices in Σ_D are dense in that set, we may find a sequence of irreducible matrix in Σ_D converging to A, and such a sequence is readily seen to satisfy the desired properties.

Henceforth we assume that $k \ge 2$. Without loss of generality we take $j = 1$, and suppose for concreteness that the cycle C_1 has length q. By applying a permutation similarity transformation (if necessary), we may consider irreducible matrices in Σ_D of the following form

$$S = \left[\begin{array}{c|c} M_s & s e_1 e_i^T \\ \hline t e_l e_m^T & A \end{array}\right],$$

where the $q \times q$ matrix M_s has the form

$$M_s = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 1-s \\ 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & \dots & 0 & 1 & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Observe that for such an S, the matrix $A + te_l e_i^T$ is irreducible, stochastic, and that $\cup_{r=2}^k C_r \subseteq \mathcal{D}(A)$. From the induction hypothesis, there is a sequence of matrices A_p and scalars $t_p > 0$ such that for each p, $A_p + t_p e_l e_i^T$ is irreducible and stochastic, $(A_p + t_p e_l e_i^T)(W \setminus V(C_1))$ converges to the adjacency matrix of $\cup_{r=2}^k C_r \subseteq \mathcal{D}(A)$, and the stationary vectors $x(p)$ of $A_p + t_p e_l e_i^T$ converge to the appropriate scalar multiple of the indicator vector for $V(\cup_{r=2}^k C_r)$.

Suppose first that the arc $l \to i$ is not an arc of $\cup_{r=2}^k C_r$. From this it follows that $t_p \to 0$ as $p \to \infty$. For each $p \in \mathbb{N}$, let

$$s_p = \frac{q t_p x(p)_l (w-q)}{q + (q-m) t_p x(p)_l (w-q)},$$

and note that for all sufficiently large p, $0 < s_p < 1$; without loss of generality we take $s_p \in (0,1)$ for all $p \in \mathbb{N}$. Let $y(p) \in \mathbb{R}^q$ be given by

$$y(p)^T = \frac{1}{q-(q-m)s_p} \left[\ \mathbf{1}_m \mid (1-s_p)\mathbf{1}_{q-m} \ \right].$$

Finally we consider the sequence of matrices

$$S_p = \left[\begin{array}{c|c} M_{s_p} & s_p e_1 e_i^T \\ \hline t_p e_l e_m^T & A_p \end{array} \right].$$

Evidently each S_p is an irreducible matrices in Σ_D, and a straightforward verification shows that the stationary vector for S_p is given by

$$z(p)^T = \left[\ \tfrac{q}{w} y(p)^T \mid (1-\tfrac{q}{w}) x(p)^T \ \right].$$

Evidently $S_p(W)$ converges to the adjacency matrix of $\cup_{r=1}^k C_r$, while $z(p)$ converges to $\frac{1}{w}u$.

Now we suppose that the arc $l \to i$ is an arc of $\cup_{r=2}^k C_r$. Again we consider the sequence of matrices A_p scalars t_p above, and note that t_p may not converge to zero in this case. For each $p \in \mathbb{N}$, set

$$\tilde{s}_p = \frac{q t_p x(p)_l (w-q)}{pq + (q-m) t_p x(p)_l (w-q)},$$

$$\tilde{y}(p)^T = \frac{1}{q-(q-m)\tilde{s}_p} \left[\ \mathbf{1}_m \mid (1-\tilde{s}_p)\mathbf{1}_{q-m} \ \right],$$

and

$$\tilde{S}_p = \left[\begin{array}{c|c} M_{\tilde{s}_p} & \tilde{s}_p e_1 e_i^T \\ \hline \frac{t_p}{p} e_l e_m^T & A_p + \frac{p-1}{p} t_p e_l e_i^T \end{array} \right].$$

It can be verified that the stationary vector for $\tilde{S}_p$ is equal to

$$\left[\ \tfrac{q}{w} \tilde{y}(p)^T \mid (1-\tfrac{q}{w}) x(p)^T \ \right]^T,$$

and it now follows that the sequence $\tilde{S}_p$ has the desired properties. This completes the proof of the induction step. $\square$

We now apply Theorem 2.1 and Proposition 2.3 to obtain the following.

Corollary 2.1. *Let D be a strongly connected graph on n vertices, and let*

$$P = \{x | x \text{ is the stationary vector for an irreducible matrix in } \Sigma_D\}. \tag{7}$$

Then

$$\inf\{||x||_\infty | x \in P\} = \frac{1}{\overline{m}(D)}.$$

Proof. We find from Theorem 2.1 that for each $x \in P, ||x||_\infty \geq \frac{1}{\overline{m}(D)}$. Let $\mathcal{C}$ be a vertex disjoint union of cycles in D such that $|V(\mathcal{C})| = \overline{m}(D)$, and let u be the indicator vector for $V(\mathcal{C})$. From Proposition 2.3, we find that there is a sequence of irreducible matrices $S_p \in \Sigma_D$ with stationary vectors $x(p)$ that converges to $\frac{1}{\overline{m}(D)}u$, as $p \to \infty$ so that $\lim_{p\to\infty} ||x(p)||_\infty = \frac{1}{\overline{m}(D)}$. □

The following is immediate from Corollary 2.1 and the remarks preceding Corollary 1.2. Here, for vectors $u, v \in \mathbb{R}^n$, we denote their entrywise product by $u \circ v$.

Corollary 2.2. *Let D be a strongly connected graph on n vertices, and let*

$$Q = \{u \circ v | \exists \text{ an irreducible } M \geq 0 \text{ with } \mathcal{D}(M) \subseteq D, \text{ and with right and left Perron vectors } u, v \text{ respectively, normalised so that } v^T u = 1\}.$$

Then

$$\inf\{||y||_\infty | y \in Q\} = \frac{1}{\overline{m}(D)}.$$

3 Equality cases and examples

Corollary 2.1 establishes the infimum of $||x||_\infty$ taken over the set P of (7), but leaves open the question of whether or not the infimum is attained – i.e. whether there is an irreducible matrix $S \in \Sigma_D$ such that for the corresponding stationary distribution vector x we have $||x||_\infty = \frac{1}{\overline{m}(D)}$. In this section we address that question, which appears to be quite difficult. We begin our discussion with a class of highly structured examples.

Example 3.1. Suppose that D is a strongly connected directed graph on n vertices that is periodic with period $p \geq 2$ – that is, the greatest common divisor of the cycle lengths in D is p. It follows that there is a partition of $V(D)$ as $\cup_{k=1}^p A_k$ such that $i \to j$ is an arc in D only if for some index k between 1 and $p, i \in A_k$ and $j \in A_{k+1}$, where we take the subscripts modulo p (see [9]). (We refer to the sets $A_1, \ldots, A_p$ as the *cyclically transferring classes* of the periodic directed graph D.) Suppose further that at least one A_k has cardinality 1. By relabelling the vertices of D if necessary, we may take $A_1 = \{1\}$. From our hypothesis that $A_1 = \{1\}$, it follows that a) each cycle in D passes through vertex 1, and b) each cycle in D has length p. (Note that we have already seen this type of directed graph in the context of Proposition 1.1.) We conclude then that $\overline{m}(D) = p$.

Suppose now that $S \in S_D$ and that S is irreducible. It follows from standard results then that S is permutationally similar to a matrix $\hat{S}$ of the form

$$\hat{S} = \left[\begin{array}{c|c|c|c|c|c} 0 & M_1 & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & M_2 & 0 & \dots & 0 \\ \hline \vdots & & \ddots & \ddots & & \vdots \\ \hline 0 & 0 & \dots & 0 & M_{p-2} & 0 \\ \hline 0 & 0 & \dots & 0 & 0 & M_{p-1} \\ \hline M_p & 0 & 0 & \dots & 0 & 0 \end{array}\right],$$

where the first diagonal block in the partitioned form for $\hat{S}$ is 1×1. It is straightforward to verify that the vector x given by

$$x^T = \frac{1}{p}\left[\ 1 \mid M_1 \mid M_2 M_2 \mid \dots \mid M_1 \dots M_{p-1}\ \right]$$

serves as the stationary distribution vector for $\hat{S}$. In particular, note that $x_1 = ||x||_\infty = \frac{1}{p}$. Thus we find that for any irreducible matrix in Σ_D, the corresponding stationary distribution vector attains the infimum in Corollary 2.1.

Our next result continues with the theme of periodic matrices, and considers (5) in a special case.

Proposition 3.1. *Let S be an irreducible stochastic matrix of order $n \geq 3$. Suppose that there is an integer $m \geq 2$ and a vector $\lambda \in \mathbb{R}^n$ such that*

$$(I - S)\lambda = e_1 - \frac{1}{m}\mathbf{1}. \tag{8}$$

Suppose also that for each $i = 1, \dots, n$, we have $\lambda_{j_1} = \lambda_{j_2}$ whenever $s_{i,j_1}, s_{i,j_2} > 0$. Then S is a periodic matrix with period at least m, and in addition $\{1\}$ is one of the cyclically transferring classes of $\mathcal{D}(S)$.

Proof. We proceed by induction on n, and first consider the case that $n = 3$. From (8) we see that λ cannot be a scalar multiple of $\mathbf{1}$. Further, if the entries of λ are distinct, then each vertex of $\mathcal{D}(S)$ has outdegree 1; it then follows that S is a cyclic permutation matrix, for which the conclusion is evident. It remains to consider two cases: i) $\lambda_1 = \lambda_2 \neq \lambda_3$ (without loss of generality), and ii) $\lambda_2 = \lambda_3 \neq \lambda_1$.

Suppose first that i) $\lambda_1 = \lambda_2 \neq \lambda_3$. Observe that $\mathcal{D}(S)$ cannot contain any of the arcs $1 \to 1, 1 \to 2, 2 \to 1, 2 \to 2$, for otherwise (8) yields $0 = \frac{m-1}{m}$ in the first two cases, and $0 = \frac{-1}{m}$ in the last two cases. Consequently, it must be the case that $1 \to 3$ and $2 \to 3$ in $\mathcal{D}(S)$. Again referring to (8), we then find that $\lambda_1 - \lambda_3 = \frac{m-1}{m}$ and $\lambda_2 - \lambda_3 = \frac{-1}{m}$, contrary to the hypothesis that $\lambda_1 = \lambda_2$. So, we see in fact that i) cannot hold. Next, suppose that ii) $\lambda_2 = \lambda_3 \neq \lambda_1$. Observe that $\mathcal{D}(S)$ cannot contain any of the arcs $2 \to 2, 2 \to 3, 3 \to 2, 3 \to 3$, otherwise (8) yields the equation $\lambda_2 - \lambda_3 = 0 = \frac{-1}{m}$, a contradiction. Similarly, $\mathcal{D}(S)$ cannot contain the arc $1 \to 1$, otherwise (8) yields $\lambda_1 - \lambda_1 = 0 = \frac{m-1}{m}$, another contradiction. We deduce that the only possible arcs in $\mathcal{D}(S)$ are $2 \to 1, 3 \to 1, 1 \to 2$ and $1 \to 3$. Referring again to (8), it follows that $\lambda_2 = \lambda_3 = \lambda_1 - \frac{1}{m}$, which in turn yields that $\frac{1}{m} = \frac{m-1}{m}$. Hence $m = 2$. But note that S is necessarily a periodic matrix with period 2 and in addition that $\{1\}$ is one of the cyclically transferring classes of $\mathcal{D}(S)$. Hence the conclusion holds for the case that $n = 3$.

Next we suppose that $n \geq 4$, that the induction hypothesis holds for $n-1$, and that S is of order n. We consider the following three cases.

Case 1, $\lambda_i \neq \lambda_j$ whenever $i \neq j$:
From the hypothesis, we find that each vertex of $\mathcal{D}(S)$ has outdegree 1. Hence S is an $n \times n$ cyclic permutation matrix, and the conclusion follows readily.

Case 2, $\lambda_i \neq \lambda_j$ whenever $i, j \geq 2$ and $i \neq j$, and $\lambda_1 = \lambda_i$ for some $i \geq 2$:
For concreteness we suppose that $\lambda_1 = \lambda_2$. Observe that $\mathcal{D}(S)$ does not contain the arc $2 \to 1$, for otherwise (8) yields $\lambda_2 - \lambda_1 = 0 = \frac{-1}{m}$, a contradiction. Consider a shortest path from 2 to 1 in $\mathcal{D}(S)$, say $2 \to i_1 \to \ldots \to i_{d-1} \to 1$, where $i_j \in \{3, \ldots, n\}, j = 1, \ldots, d-1$. From (8), we find that $\lambda_2 - \lambda_{i_1} = \frac{-1}{m}, \lambda_{i_j} - \lambda_{i_{j+1}} = \frac{-1}{m}, j = 1, \ldots, d-2$, and $\lambda_{i_{d-1}} - \lambda_1 = \frac{-1}{m}$. Consequently, $\lambda_2 - \lambda_1 = \frac{-d}{m}$, a contradiction. Thus, we find that case 2 cannot occur.

Case 3, $\lambda_i = \lambda_j$ for some $i, j \geq 2$:
Without loss of generality, we suppose that $\lambda_{n-1} = \lambda_n$. Write the matrix S as

$$S = \left[\begin{array}{c|c} S_{1,1} & S_{1,2} \\ \hline S_{2,1} & S_{2,2} \end{array} \right],$$

where $S_{2,2}$ is 2×2. Note that from the hypothesis, if $\mathcal{D}(S)$ contains the arcs $n \to i$ and $n-1 \to j$, then by (8) we have $\lambda_n - \lambda_i = \frac{-1}{m}, \lambda_{n-1} - \lambda_j = \frac{-1}{m}$, so that necessarily $\lambda_i = \lambda_j$. Observe also that $S_{2,2} = 0$, otherwise we find that $0 = -\frac{1}{m}$, which is impossible. Set

$$\tilde{S} = \left[\begin{array}{c|c} S_{1,1} & S_{1,2}\mathbf{1} \\ \hline \frac{1}{2}\mathbf{1}^T S_{2,1} & \frac{1}{2}\mathbf{1}^T S_{2,2}\mathbf{1} \end{array} \right],$$

and note that from (8), we have

$$(I - \tilde{S}) \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_{n-1} \end{bmatrix} = e_1 - \frac{1}{m}\mathbf{1}.$$

Further, for each $i \in \{1, \ldots n-1\}$, $\lambda_{j_1} = \lambda_{j_2}$ whenever $\tilde{s}_{i,j_1}, \tilde{s}_{i,j_2} > 0$. Consequently, the matrix $\tilde{S}$ of order $n-1$ satisfies the induction hypothesis. Hence, $\tilde{S}$ is periodic with period $p \geq m$, and $\{1\}$ is one of the cyclically transferring classes of $\mathcal{D}(\tilde{S})$. Thus there is a partition of $\{1, \ldots, n-1\}$ as $A_1 = \{1\}, A_2, \ldots, A_p$, such that $a \to b$ in $\mathcal{D}(\tilde{S})$ only if there is an index j such that $a \in A_j, b \in A_{j+1}$. (Here we take the subscripts on the A_js modulo p.) Observe that the arcs of $\mathcal{D}(\tilde{S})$ are of one of three types: $i \to j$ where $1 \le i, j \le n-2$ and $i \to j \in \mathcal{D}(S)$; $i \to n-1$ where $i \le n-2$ and either $i \to n-1$ or $i \to n$ in $\mathcal{D}(S)$; and $n-1 \to j$ where $j \le n-2$ and either $n-1 \to j$ or $n \to j$ in $\mathcal{D}(S)$. Let q denote the index such that $n-1 \in A_q$. Considering the partitioning of $\{1, \dots, n\}$ given by $A_1, \dots, A_{q-1}, A_q \cup \{n\}, A_{q+1}, \dots, A_p$, it now follows that S is periodic with period at least $p \ge m$.

This completes the proof of the induction step, and the conclusion now follows. □

Corollary 3.1. *Let D be a strongly connected directed graph on $n \ge 3$ vertices. Suppose that $S \in \Sigma_D$ is irreducible with stationary vector x. If $x_1 = \frac{1}{\overline{m}_1(D)} > x_j, j = 2, \dots, n$, then necessarily S is periodic with period $\overline{m}_1(D)$. Further, there is a unique cyclically transferring class of $\mathcal{D}(S)$ having cardinality one, namely $\{1\}$.*

Proof. From the hypothesis, we find that the vector x and the matrix S satisfy the hypotheses of Proposition 2.2, where the parameter k in that proposition is equal to 1. Following the proof of Proposition 2.2, we see that the matrix S satisfies (5), where the matrix U in that equation is an identity matrix, and the parameter λ_n is equal to $-\frac{1}{\overline{m}_1(D)}$. Hence S satisfies the hypothesis of Proposition 3.1; the conclusion now follows readily. □

Example 3.1 shows that for an irreducible periodic stochastic matrix whose directed graph has a cyclically transferring class of cardinality one, the corresponding stationary distribution vector necessarily attains the infimum in Corollary 2.1. From Corollary 3.1, we see that for an irreducible stochastic matrix whose stationary distribution vector attains the infimum in Corollary 2.1 and has a unique maximum entry, the matrix must be of the form in Example 3.1. However, there are examples of primitive stochastic matrices for which the infimum of Corollary 2.1 is attained. Our next example presents just such a family of matrices.

Example 3.2. Consider the directed graph D pictured in Figure 1. An inspection reveals that D contains a vertex–disjoint union of cycles involving five vertices, but no such collection of cycles involving six vertices. In particular, we have $\overline{m}_i(D) = 5, i = 1, \dots, 6$. For each $0 < a < 1$, consider the matrix

$$S_a = \begin{bmatrix} 0 & a & 0 & 0 & 0 & (1-a) \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1-a) & a \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & (1-a) & a & 0 & 0 \end{bmatrix},$$

and note that $\mathcal{D}(S_a) = D$. The stationary distribution vector for S_a is given by $\frac{1}{5}\begin{bmatrix} 1 & a & 1 & 1 & (1-a) & 1 \end{bmatrix}^T$, and so we find that for $i = 1, 3, 4, 6$, there is an

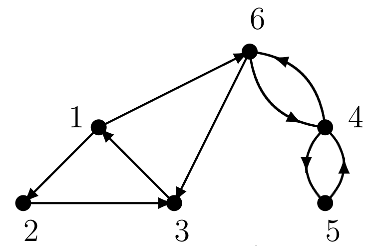

Figure 1: The graph for Example 3.2

irreducible matrix in Σ_D for which the i–th entry in the stationary distribution is simultaneously the largest in the stationary distribution vector, and equal to $\frac{1}{\overline{m}_i(D)}$.

Observe also that if S is any irreducible matrix in Σ_D, then there are scalars $r, s, t \in (0, 1)$ such that

$$S = \begin{bmatrix} 0 & r & 0 & 0 & 0 & (1-r) \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1-s) & s \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & (1-t) & t & 0 & 0 \end{bmatrix}.$$

From the eigen–equation $x^T S = x^T$, it follows that for the stationary distribution vector x we have $x_2 = rx_1 < x_1$ and $x_5 = (1-s)x_4 < x_4$, so that neither x_2 nor x_5 can be the maximum entry in the stationary distribution vector.

In Example 3.2 we saw that vertices 2 and 5 cannot yield the maximum entry in the stationary distribution vector for any irreducible matrix S in Σ_D. That observation prompts the following question.

Open Problem 3.1. Let D be a strongly connected directed graph on n vertices, and fix an index i between 1 and n. Determine necessary and sufficient conditions in order that there is an irreducible matrix $S \in \Sigma_D$ whose stationary distribution vector x has the property that $x_i = ||x||_\infty$.

Example 3.3. In this example, we consider an application of our results to a family of matrices arising in mathematical ecology. A standard stage–classified model for population growth leads to the analysis of the so–called population projection matrix for a particular species. See [14, Chapter 4] for an extensive discussion of this model. This stage–classified population model leads us to consider an entrywise nonnegative matrix M of order n whose Perron value, say $r(M)$ represents the asymptotic growth rate of the population under consideration. As noted in section 1, if M is irreducible with right Perron vector u, then letting U denote the diagonal matrix with $u_{i,i} = u_i, i = 1, \ldots, n$, it follows that the matrix $S = \frac{1}{r(M)} U^{-1} M U$ is irreducible, stochastic, and has the same directed graph as M. Further, for each index i between 1 and n, the i–th entry in the stationary distribution vector of S coincides with the derivative of $r(M)$ with respect to the i–th diagonal entry of M.

Such derivatives arise in the sensitivity analysis of the corresponding population model, and can be used to inform strategies for species management, among other things. Chapter 9 of [14] discusses sensitivity analysis in further detail.

As a specific example, we consider the population projection matrix for the North Atlantic right whale. According to [15], the directed graph for the corresponding population projection matrix (which is known as the life cycle graph) is given in Figure 2.

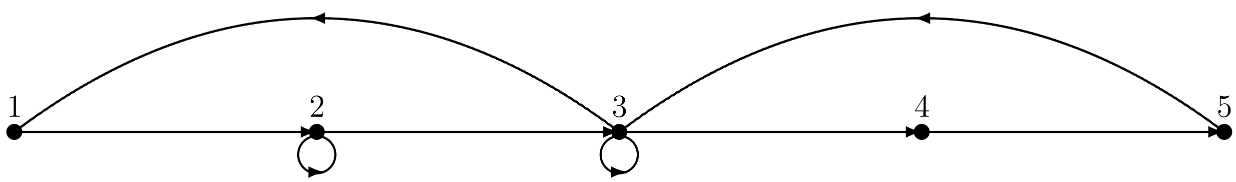

Figure 2: Life cycle graph for the North American right whale (female)

Consider an irreducible stochastic matrix S whose directed graph is given by the graph D of Figure 2. Then S can be written as

$$S = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & s & 1-s & 0 & 0 \\ t & 0 & r & 1-t-r & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

for some choice of the parameters r, s, t such that $0 \le r, 0 \le s < 1, 0 < t, t + r < 1$. It is straightforward to verify that the stationary distribution vector x for S is given by

$$x^T = \frac{1}{3 - r + \frac{st}{1-s}} \left[\; t \quad \tfrac{t}{1-s} \quad 1 \quad (1-t-r) \quad (1-t-r) \; \right].$$

Observe that for any such r, s and t, the maximum entry is either x_2 or x_3.

Inspecting D, we find that it has a vertex–disjoint union of cycles passing through four vertices, but no such union of cycles passing through all five vertices. It now follows from Corollary 2.1 that if x is the stationary distribution vector for any stochastic matrix with directed graph D, then $\max\{x_2, x_3\} = ||x||_\infty \ge \frac{1}{4}$. In particular, we deduce that for a population projection matrix M with directed graph D, the maximum of the sensitivities of the Perron value with respect to the $(2, 2)$ and $(3, 3)$ entries is bounded below by $\frac{1}{4}$. Thus we see that the qualitative information contained in the life cycle graph leads to quantitative information about certain sensitivities of the asymptotic growth rate of the population being modelled.

Acknowledgement: The author expresses his thanks to an anonymous referee, whose comments on an earlier version of this paper resulted in improvements to the exposition. In particular, both the presentation and the proof of Proposition 2.1 were greatly improved by the referee.

References

[1] R. Brualdi, H. Ryser. *Combinatorial Matrix Theory*. Cambridge (UK): Cambridge University Press; 1991.

[2] M. Akelbek, S. Kirkland. Coefficients of ergodicity and the scrambling index. *Linear Algebra Appl.*. 2009; 430: 1099–1110.

[3] S. Kirkland. Girth and subdominant eigenvalues for stochastic matrices. *Electronic J. Linear Algebra*. 2004/2005; 12: 25–41.

[4] S. Kirkland. Fastest expected time to mixing for a Markov chain on a directed graph. *Linear Algebra Appl.*. 2010; 433: 1988–1996.

[5] H. Caswell. Sensitivity analysis of discrete Markov chains via matrix calculus. *Linear Algebra Appl.*. 2011; doi:10.1016/j.laa.2011.07.046.

[6] P. Deuflhard, W. Huisinga, A. Fischer, C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra Appl.*. 2000; 315: 39–59.

[7] E. Crisostomi, S. Kirkland, R. Shorten. A Google–like model of road network dynamics and its application to regulation and control. *International J. Control*. 2011; 84: 633–651.

[8] E. Crisostomi, S. Kirkland, A. Schlote, R. Shorten. Markov chain–based emissions models: a precursor for green control. In J. Kim, M. Lee, editors. *Green IT: Technologies and Applications*. New York, NY: Springer–Verlag; 2011. p. 381–400.

[9] E. Seneta. *Non–Negative Matrices and Markov Chains, 2nd edition*. New York, NY: Springer; 1981.

[10] M.I. Freĭdlin, A.D. Wentzell. *Perturbations of Stochastic Dynamic Systems*. New York, NY: Springer–Verlag; 1984.

[11] G.W. Stewart. *Introduction to Matrix Computations*. New York, NY: Academic Press; 1973.

[12] G. Chartrand, L. Lesniak. *Graphs and Digraphs, 4th edition*. Boca Raton, FL: Chapman and Hall; 2005.

[13] M.S. Bazarra, H.D. Sherali, C.M. Shetty. *Nonlinear Programming: Theory and Algorithms, 3rd edition*. Hoboken, NJ: Wiley; 2006.

[14] H. Caswell. *Matrix Population Models, 2nd edition*. Sunderland, MA: Sinauer; 2001.

[15] H. Caswell. Stage, age and individual stochasticity in demography. *Oikos* 2009; 118: 1763–1782.