A Sinkhorn-Knopp Fixed Point Problem *

Kristen Anderson Ashley Burt Will Cousins Brent Hancock David Strong †

May 6, 2010

Abstract

We consider a fixed point problem that results from the Sinkhorn-Knopp Algorithm for transforming matrices into doubly stochastic matrices. We analytically find solutions for simpler cases, including the circulant case, we develop and prove a number of properties for the general case, and we examine how solutions to the fixed point problem relate to those of the Sinkhorn-Knopp Algorithm.

1 Introduction

Stochastic matrices have many useful and interesting mathematical properties and arise in a variety of applications, such as interpreting economic data [1], ranking sports teams [3], ranking webpages (for example, in Google searches) [5], studying traffic flow [7], and preconditioning sparse matrices [8]. In [10] Sinkhorn showed that any positive square matrix A is diagonally equivalent to a doubly stochastic matrix, $S = D_1AD_2$, and that the diagonal matrices D_1 and D_2 are unique up to a scaling factor. He also found that the product D_1AD_2 can be found as the limit of a sequence of matrices generated by alternately normalizing the rows and the columns of A. In [11] Sinkhorn and Knopp introduced a simple algorithm for finding these diagonal matrices and the corresponding doubly stochastic matrix, and gave conditions which guarantee convergence. (Brualdi, Parter and Schneider also independently obtained some of the same results around the same time [2]). Their algorithm is known by multiple names to researchers in various disciplines, but it is best known as the Sinkhorn-Knopp Algorithm. In this paper we will study a fixed point problem that naturally arises from the Sinkhorn-Knopp Algorithm.

1.1 Notation

We use the following two notations: for $\vec{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{C}^n$,

$$diag(\vec{x}) = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix} \quad \text{and} \quad \vec{x}^{(-1)} = \begin{bmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{bmatrix}.$$

Also, for any vector or matrix, by *positive* and *nonzero* we mean all entries are positive (and real) and nonzero, respectively. For example, for $\vec{x}^{(-1)}$ to exist, $\vec{x}$ must be nonzero.

1.2 Outline

In Section 2 we describe the Sinkhorn-Knopp Algorithm and the resulting fixed point problem, and we discuss the relationship between the two. In Section 3 we consider the general 2×2 case of the fixed point problem. In Section 4 we prove and discuss some of the basic properties of fixed point solutions for general matrices. In Section 5 we examine the 3×3 circulant case. We conclude with some open questions.

*This research was supported by NSF grant DMS-0636648 through a CURM (Center for Undergraduate Research in Mathematics) mini-grant.

†When work for this paper was done, Anderson, Burt, Cousins and Hancock were undergraduates at Pepperdine University, where Strong is an associate professor of mathematics in the Natural Science Division.

2 The Sinkhorn-Knopp Algorithm and Fixed Point Problem

Given a matrix A, we can divide each of its rows by the (necessarily non-zero) sum of the elements in that row to produce a new matrix with all row sums of 1. If $\vec{x} = (1, 1, \ldots, 1)$, then $A\vec{x}$ is the vector of row sums of A. Thus $diag((A\vec{x})^{(-1)})\, A$ is a transformation of A that has row sums of 1. Similarly, the product $A\ diag((A^T\vec{x})^{(-1)})$ has column sums of 1. As we show in Section 2.3, it turns out that for *any* nonzero $\vec{x}$ for which $A\vec{x}$ is entirely nonzero (necessitated by the (-1) operator), the product

$$diag((A\vec{x}\,)^{(-1)})\ A\ diag(\vec{x})$$

has all row sums 1, and similarly, for any $\vec{x}$ for which $A^T\vec{x}$ is entirely nonzero,

$$diag(\vec{x})\ A\ diag((A^T\vec{x}\,)^{(-1)})$$

has all column sums of 1. Obviously to have row sums and column sums of 1, A may have no rows or columns of entirely 0s. We make that assumption throughout this paper.

With these motivating results, Sinkhorn and Knopp proposed an iterative procedure to produce a doubly stochastic matrix, D_1AD_2 (where $D_1 = D_2$ if A is symmetric [9]), by alternately normalizing the rows and the columns of A. We present their algorithm below, in a slightly different form than originally given in [11].

2.1 The algorithm

Let A be a positive matrix, or a non-negative matrix with total support as defined in [11]. Let $\vec{x}_0 = (1, 1, \ldots, 1)$. Then for $k \geq 0$,

$$\begin{aligned} \vec{y}_k &:= (A\,\vec{x}_k\,)^{(-1)} && (1)\\ \vec{x}_{k+1} &:= (A^T\vec{y}_k\,)^{(-1)} && (2)\\ S_k &:= diag(\vec{y}_k)\ A\ diag(\vec{x}_k). && (3) \end{aligned}$$

Sinkhorn and Knopp showed that under these conditions, the sequence $S_0, S_1, S_2, \ldots$ converges to a doubly stochastic matrix, which we denote as S. It is consequently also true that the sequence $\vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$ produced by (1) and (2) converges to a limiting vector, which we denote as $\vec{x}$. We can thus combine (1) and (2) as

$$\vec{x}_{k+1} := \left(A^T(A\,\vec{x}_k)^{(-1)}\right)^{(-1)}, \tag{4}$$

which converges to $\vec{x}$. So (3) can be written in terms of $\vec{x}_k$ as

$$S_k := diag((A\vec{x}_k)^{(-1)})\ A\ diag(\vec{x}_k), \tag{5}$$

2.2 The Sinkhorn-Knopp Fixed Point Problem

Our work in this paper focuses on solutions to the problem

$$\vec{x} = \left(A^T(A\vec{x}\,)^{(-1)}\right)^{(-1)}, \tag{6}$$

the fixed point formulation of (4). Equation (5) can similarly be written as

$$S = S(\vec{x}) := diag((A\vec{x}\,)^{(-1)})\ A\ diag(\,\vec{x}\,). \tag{7}$$

Since $\left(\vec{x}^{(-1)}\right)^{(-1)} = \vec{x}$, an equivalent and sometimes more convenient formulation of (6) is

$$\vec{x}^{(-1)} = A^T(A\vec{x}\,)^{(-1)}. \tag{8}$$

So $\vec{x}$ is a solution to (6) iff $\vec{x}$ is a solution to (8).

We refer to (6), or equivalently (8), as the Sinkhorn-Knopp Fixed Point Problem. Restated in the context of this fixed point problem, Sinkhorn and Knopp showed in [11] that for any positive A there is a *unique* (up to a scalar multiple, as we show in Section 4.1) *positive* fixed point $\vec{x}$ which results in (7) being doubly stochastic. It is the single stable fixed point to which the sequence generated by (4) will converge.[1]

It turns out that in general there are other non-real and/or non-positive solutions to (6). These are unstable fixed points, points to which (4) would not converge. In the balance of the paper we investigate solutions to (6) of all types, including solutions with negative and/or complex entries, and we examine how these solutions relate to the original problem of trying to produce a doubly stochastic matrix from a given matrix A.

2.3 Relationship between the two Sinkhorn-Knopp Problems

We show that for any $\vec{x}$ for which $A\vec{x}$ is entirely nonzero, the rows of the matrix S in (7) have sum 1, and if $\vec{x}$ is the solution to (6), or equivalently (8), then the columns of (7) also have sum 1. Let $\vec{e}$ be the vector of all 1s. Then

$$S\vec{e} = diag((A\vec{x}\,)^{(-1)})\, A\, diag(\vec{x})\,\vec{e} = diag((A\vec{x}\,)^{(-1)})\, A\vec{x} = \vec{e}.$$

That is, the sum of each row in S is 1. Now if $\vec{x}$ is a solution to (8), then

$$S^T\vec{e} = diag(\vec{x})\, A^T\, diag((A\vec{x}\,)^{(-1)})\,\vec{e} = diag(\vec{x})\ A^T(A\vec{x}\,)^{(-1)} = diag(\vec{x})\,(\vec{x})^{(-1)} = \vec{e}.$$

That is, the sum of each column in S is 1.

2.4 Fixed point solutions vs. doubly stoubly stochastic matrices

As just shown, if the vector $\vec{x}$ in (7) is a solution to (6), then the product in (7) will have all row and column sums of 1. However, this does not guarantee that the product is doubly stochastic: it may have row and column sums 1, but with some negative or complex entries. As mentioned above, Sinkhorn and Knopp showed that there is at most one positive solution to (6), and exactly one if A is positive. For example, if

$$A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix},$$

then (6) has two solutions, $\vec{x}_1 = (2, 1)$ and $\vec{x}_2 = (2, -1)$, which in (7) result in

$$S(\vec{x}_1) = \begin{bmatrix} 2/6 & 4/6 \\ 4/6 & 2/6 \end{bmatrix} \quad \text{and} \quad S(\vec{x}_2) = \begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}.$$

The first is doubly stochastic and results from $\vec{x}_1$, the unique positive vector the Sinkhorn-Knopp Algorithm in (4) would generate. $\vec{x}_2$ is also a solution to (6) for the given A, but could not be found by iterating (4), as it is an unstable fixed point. The matrix $S(\vec{x}_2)$ corresponding to this unstable fixed point also has row and column sums of 1, but of course is not doubly stochastic.

Consider a second example, with A not entirely positive:

$$A = \begin{bmatrix} 1 & -4 \\ 1 & 1 \end{bmatrix}.$$

For this matrix, (6) has two solutions, $\vec{x}_1 = (2i, 1)$ and $\vec{x}_2 = (-2i, 1)$, which in (7) result in

$$S(\vec{x}_1) = \begin{bmatrix} 1/5 - 2/5\ i & 4/5 + 2/5\ i \\ 4/5 + 2/5\ i & 1/5 - 2/5\ i \end{bmatrix} \quad \text{and} \quad S(\vec{x}_2) = \begin{bmatrix} 1/5 + 2/5\ i & 4/5 - 2/5\ i \\ 4/5 - 2/5\ i & 1/5 + 2/5\ i \end{bmatrix}.$$

While the row and column sums of both matrices are 1, neither is doubly stochastic. We reiterate that Sinkhorn and Knopp's work guarantees a single stochastic matrix, $S(\vec{x})$ in (7), if matrix A is positive, which is not the case here. Our discussion next leads to the general solution for 2×2 matrices, as well as solutions for certain types of $n \times n$ matrices.

[1] In their algorithm, Sinkhorn and Knopp's work assumed a starting vector $\vec{x}_0$ of all 1s, but our numerical experimentation suggests that any real $\vec{x}_0$ for which $A\vec{x}_k$, $k \geq 0$, is entirely nonzero also leads to convergence of (4).

3 Solutions for 2×2 matrices

In the previous section we gave examples of two 2×2 matrices and the corresponding solutions to (6). We now find the solution in the general 2×2 case. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{9}$$

where a, b, c and d are nonzero complex values. Then (6) becomes

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \dfrac{(ax_1 + bx_2)(cx_1 + dx_2)}{2acx_1 + (ad + bc)x_2} \\ \dfrac{(ax_1 + bx_2)(cx_1 + dx_2)}{(ad + bc)x_1 + 2bdx_2} \end{bmatrix}$$

Solving for x_1 and x_2 results in two solutions,

$$\vec{x} = \begin{bmatrix} \pm\sqrt{\dfrac{bd}{ac}}\, x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} \pm\sqrt{\dfrac{bd}{ac}} \\ 1 \end{bmatrix}. \tag{10}$$

We see that any non-zero multiple of this vector is also a solution. If either $b = 0$ or $c = 0$, but not both (that is, if A is upper or lower triangular, but not diagonal), then there is no solution. If A is diagonal with all nonzero diagonal entries, it is easy to show that any nonzero vector is a solution to (6). We are currently studying the general upper and lower triangular cases.

We found the solutions for the two examples in Section 2.4 using (10), although the positive solution $\vec{x}_1 = (2, 1)$ in the first example could also have been found by iterating (4). We give one final example, one with complex entries in A. Let

$$A = \begin{bmatrix} 1 & -4i \\ 1 & 1 \end{bmatrix}.$$

Then the two solutions given by (10) and the corresponding resulting matrix products in (7) are

$$\vec{x}_1 = \begin{bmatrix} 2\sqrt{-i} \\ 1 \end{bmatrix} \implies S(\vec{x}_1) = \begin{bmatrix} \dfrac{2\sqrt{-i}}{2\sqrt{-i} - 4i} & \dfrac{-4i}{2\sqrt{-i} - 4i} \\ \dfrac{2\sqrt{-i}}{2\sqrt{-i} + 1} & \dfrac{1}{2\sqrt{-i} + 1} \end{bmatrix},$$

$$\vec{x}_2 = \begin{bmatrix} -2\sqrt{-i} \\ 1 \end{bmatrix} \implies S(\vec{x}_2) = \begin{bmatrix} \dfrac{-2\sqrt{-i}}{-2\sqrt{-i} - 4i} & \dfrac{-4i}{-2\sqrt{-i} - 4i} \\ \dfrac{-2\sqrt{-i}}{-2\sqrt{-i} + 1} & \dfrac{1}{-2\sqrt{-i} + 1} \end{bmatrix}.$$

Of course it can be verified that (6) is satisfied by both $\vec{x}_1$ and $\vec{x}_2$. It is easy to see that for both resulting matrices the row sums are 1, and straightforward to show that the column sums are 1. Obviously neither S matrix is doubly stochastic.

4 Basic properties of fixed point solutions

There are a number of interesting and useful properties of the fixed point problem (6) and its solutions. In this section we discuss a few of the more important ones.

4.1 Scalar multiples of matrices and solutions

If $\vec{x}$ is a solution to (6), that is, if $\left(A^T (A\vec{x})^{(-1)}\right)^{(-1)} = \vec{x}$, then it is easy to show that any nonzero multiple of $\vec{x}$ is also a solution to (6) for the same A. That is, for $c \neq 0$,

$$\left(A^T(A(c\vec{x}))^{(-1)}\right)^{(-1)} = c\vec{x}.$$

Similarly, if $\vec{x}$ is a solution to (6) for a given matrix A, then $\vec{x}$ is also a solution to (6) for cA, $c \neq 0$, since

$$\left((cA)^T\,(\,(cA)\vec{x}\,)^{(-1)}\right)^{(-1)} = \vec{x}.$$

Because every solution $\vec{x}$ of (6) must consist of entirely nonzero entries, due to the entry-wise inverse (-1) in (6), we may treat one of the entries of $\vec{x}$ as a fixed constant. For example, for given a vector $\vec{x} = (x_1, x_2, \ldots, x_n)$, we can instead used the scalar multiple

$$\frac{1}{x_n}(x_1, x_2, \ldots, x_n) = (\frac{x_1}{x_n}\,,\ \frac{x_2}{x_n}\,,\ \ldots\,,\ 1).$$

Thus, in finding a solution, we can simply assume $\vec{x}$ is of the form $\vec{x} = (x_1, x_2, \ldots, 1)$. This strategy of eliminating one variable has been useful in simplifying some of the work we have done in numerically finding or verifying solutions. Of course any non-zero multiple of any found solution is also a solution, as just found.

4.2 Diagonal and constant matrices

There are two types of matrices for which it is easy to find the general solution for $n \times n$ matrix A: diagonal and constant matrices. As mentioned above, it is easy to show that for any $n \times n$ nonzero diagonal matrix A, any nonzero vector $\vec{x}$ is a solution to (6). A bit less trivial is the case of constant matrix. Since, as just shown, $\vec{x}$ is a solution to (6) for a given A iff $\vec{x}$ is a solution to (6) for cA, $c \neq 0$, we can simply consider the constant case when A is the $n \times n$ matrix of all 1's, in which case (6) becomes

$$(x_1, x_2, \ldots, x_n) = (\sum_{i=1}^{n} x_i\,/\,n\ ,\ \sum_{i=1}^{n} x_i\,/\,n\ ,\ \ldots\ ,\ \sum_{i=1}^{n} x_i\,/\,n\,),$$

from which we can see that any solution must satisfy $x_1 = x_2 = \cdots = x_n$. Since (1, 1, ..., 1) does in fact satisfy (6), then it is the unique (up to a scalar multiple) solution. If we recall the original problem of finding a vector $\vec{x}$ so that (7) is doubly stochastic, it is not surprising that $\vec{x} = (1, 1, \ldots, 1)$ is a solution—although it is not obvious that it was necessarily the only solution—since in this case $diag((A\vec{x})^{(-1)})$ is the diagonal matrix whose entries are the inverses of the row (and column) sums of A, and $diag(\vec{x})$ is the identity.

4.3 Block diagonal matrices

Equation (8) is more convenient than (6) for the following discussion. Suppose A is block diagonal

$$A = \begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_k \end{bmatrix}.$$

If $\vec{x}_i$ is any solution to (8) for block B_i, $i = 1$ to k, then $\vec{x} = (\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k)$ is a solution to (8) for the block matrix A, since (8) becomes

$$\begin{bmatrix} B_1^T & & & \\ & B_2^T & & \\ & & \ddots & \\ & & & B_k^T \end{bmatrix} \left(\begin{bmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_k \end{bmatrix} \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_k \end{bmatrix} \right)^{(-1)} = \begin{bmatrix} B_1^T(B_1\vec{x}_1)^{(-1)} \\ B_2^T(B_2\vec{x}_2)^{(-1)} \\ \vdots \\ B_k^T(B_k\vec{x}_k)^{(-1)} \end{bmatrix} = \begin{bmatrix} \vec{x}_1^{(-1)} \\ \vec{x}_2^{(-1)} \\ \vdots \\ \vec{x}_k^{(-1)} \end{bmatrix} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_k \end{bmatrix}^{(-1)}.$$

4.4 Permutations of rows and columns

We next consider one of the more interesting and potentially more useful properties of this problem, one that we are currently investigating in more detail: the effect of swappings rows and/or columns of A on the solution to (6), or equivalently, (8). We describe this property with the following theorem:

Theorem 4.1. *Let P and Q be $n \times n$ permutation matrices. Then $\vec{x}$ is a solution to (8) for $n \times n$ A if and only if $Q^T\vec{x}$ is a solution to (8) for PAQ.*

Proof. First, for permuation matrices P and Q, $(P\vec{v})^{(-1)} = P\vec{v}^{(-1)}$ and $(Q^T\vec{v})^{(-1)} = Q^T\vec{v}^{(-1)}$ for any entirely nonzero vector $\vec{v}$. Also, of course, $P^TP = QQ^T = I$. Then we have

$$(PAQ)^T((PAQ)(Q^T\vec{x}))^{(-1)} = Q^TA^TP^T(PA\vec{x})^{(-1)} = Q^TA^TP^TP(A\vec{x})^{(-1)} = Q^TA^T(A\vec{x})^{(-1)}.$$

Thus $(PAQ)^T((PAQ)(Q^T\vec{x}))^{(-1)} = Q^T\vec{x}^{(-1)} = (Q^T\vec{x})^{(-1)}$ if and only if $A^T(A\vec{x})^{(-1)} = \vec{x}^{(-1)}$. $\square$

In the above result, PA is the matrix obtained by permuting the rows of A corresponding to P, thus we see that permuting the rows of A does not change the solution to (6). However, since AQ is the matrix obtained by permuting the columns of A corresponding to Q and $Q^T\vec{x}$ is the vector obtained by permuting the rows of $\vec{x}$ corresponding to the columns of A permuted by AQ, we see that permuting the *columns* of A will result in a solution to (6) whose *rows* (elements) have been permuted in the same way.

A simple corollary of this theorem is the 2×2 case. In (9) if we swap the two *rows*, the solution to (6) given by (10) does not change. In (9) if we swap the two *columns*, the solution (10) changes by swapping its two *rows*.

5 Solutions for circulant matrices

As seen in Section 3, it was straightforward to analytically find the solution to (6) in the general 2×2 case. With the aid of Maple, we have also analytically found the solution to (6) in the general 3×3 case. We do not include that solution here, as nearly six printed pages are required to describe the solution. We do note that it showed that there are up to six solutions, as our numerical experimentation had already strongly suggested. Ideally we could find the analytic solution to the general $n \times n$ case. Unfortunately, for the general 4×4 case, with Maple we were unable to find any sort of analytic solution, including for patterned matrices such as circulant matrices, regardless of how we manipulated or transformed the problem for it to solve, although given our experience with the 3×3 case, any sort of closed form solution would probably have been far too complicated to be of much use anyway. For larger matrices, the most obvious work to pursue is for special types of matrices, which we are currently doing. We conclude this paper by looking at the circulant case.

5.1 Solutions for 3×3 circulant matrices

Let

$$A = \begin{bmatrix} a & c & b \\ b & a & c \\ c & b & a \end{bmatrix}.$$

We find the following six solutions to the three equations with two unknowns in (6):

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \vec{x}_2 = \begin{bmatrix} 1 \\ -\frac{1}{2} + \frac{\sqrt{3}}{2}i \\ -\frac{1}{2} - \frac{\sqrt{3}}{2}i \end{bmatrix}, \qquad \vec{x}_3 = \begin{bmatrix} 1 \\ -\frac{1}{2} - \frac{\sqrt{3}}{2}i \\ -\frac{1}{2} + \frac{\sqrt{3}}{2}i \end{bmatrix}, \tag{11}$$

$$\text{and} \quad \vec{x}_4 = \begin{bmatrix} \dfrac{a^2-bc}{a} \\ \dfrac{b^2-ac}{b} \\ \dfrac{c^2-ab}{c} \end{bmatrix}, \quad \vec{x}_5 = \begin{bmatrix} \dfrac{c^2-ab}{c} \\ \dfrac{a^2-bc}{a} \\ \dfrac{b^2-ac}{b} \end{bmatrix}, \quad \vec{x}_6 = \begin{bmatrix} \dfrac{b^2-ac}{b} \\ \dfrac{c^2-ab}{c} \\ \dfrac{a^2-bc}{a} \end{bmatrix}. \tag{12}$$

provided $a^2 \neq bc$, $b^2 \neq ac$ and $c^2 \neq ab$. We will consider the general case in more detail for the first three solutions (11) in Section 5.2 and for the other three solutions (12) in Section 5.3.

5.2 Eigenvector solutions

The three solutions in (11) are simply the eigenvectors of A. Below we show that in general the n eigenvectors of an $n \times n$ circulant matrix are solutions to (6). First, recall that the nth roots of unity have the form $\omega_m = e^{(2\pi i/n)m}$ for $m = 0, 1, \ldots, n-1$. So $(\omega_m)^n = 1$ and $1/\omega_m = \omega_{n-m}$. It is well known [4] that an $n \times n$ circulant matrix

$$C = \begin{bmatrix} C_0 & C_{n-1} & \cdots & C_1 \\ C_1 & C_0 & \cdots & C_2 \\ \vdots & \vdots & \ddots & \vdots \\ C_{n-1} & C_{n-2} & \cdots & C_0 \end{bmatrix} \tag{13}$$

has n eigenvectors and corresponding eigenvalues of the form

$$\vec{v}_m = \begin{bmatrix} 1 \\ \omega_m \\ \vdots \\ (\omega_m)^k \\ \vdots \\ (\omega_m)^{n-1} \end{bmatrix}, \ \lambda_m = C_0 + C_{n-1}\omega_m + \cdots + C_{n-k}(\omega_m)^k + \cdots + C_k(\omega_m)^{n-k} + \cdots + C_1(\omega_m)^{n-1} \tag{14}$$

for $m = 0, 1, \ldots, n-1$. The eigenvalues λ_m depend on the entries of C, but the eigenvectors $\vec{v}_m$ do not. To show that all eigenvectors of a circulant matrix are solutions to (6), we first need the following lemma:

Lemma 5.1. *For $m = 0, \ldots, n-1$, if $\vec{v}_m$ is an eigenvector of circulant C, as given in (13) and (14), then $\vec{v}_m$ is an eigenvector of C^T with eigenvalue λ_{n-m}.*

Proof. First note that

$$(\omega_m)^{n-k} = \frac{(\omega_m)^n}{(\omega_m)^k} = \frac{1}{(\omega_m)^k} = \left(\frac{1}{\omega_m}\right)^k = (\omega_{n-m})^k.$$

If C is circulant, then C^T is also circulant, so $\vec{v}_m$ is also an eigenvector of C^T, but with eigenvalue

$$\begin{aligned} & C_0 + C_1\,\omega_m + \ldots + C_k(\omega_m)^k + \ldots + C_{n-k}(\omega_m)^{n-k} + \ldots + C_{n-1}(\omega_m)^{n-1} \\ &= C_0 + C_{n-1}(\omega_m)^{n-1} + \ldots + C_{n-k}(\omega_m)^{n-k} + \ldots + C_k(\omega_m)^k + \ldots + C_1\,\omega_m \\ &= C_0 + C_{n-1}\,\omega_{n-m} + \ldots + C_{n-k}(\omega_{n-m})^k + \ldots + C_k(\omega_{n-m})^{n-k} + \ldots + C_1(\omega_{n-m})^{n-1} \\ &= \lambda_{n-m}. \end{aligned}$$

□

We use this result in the following theorem.

Theorem 5.2. *If $n \times n$ matrix C is circulant, then each eigenvector $\vec{v}_m$, for $m = 0, \ldots, n-1$, is a solution to (6).*

Proof. Where $\vec{v}_m$ is as in (14), $\vec{v}_m^{(-1)} = \vec{v}_{n-m}$, since $1/(\omega_m)^k = (\omega_{n-m})^k$. Thus

$$\begin{aligned}
\left(C^T(C\vec{v}_m)^{(-1)}\right)^{(-1)} &= \left(C^T(\lambda_m\vec{v}_m)^{(-1)}\right)^{(-1)} \\
&= \left(C^T\frac{1}{\lambda_m}(\vec{v}_m)^{(-1)}\right)^{(-1)} \\
&= \left(\frac{1}{\lambda_m}C^T\vec{v}_{n-m}\right)^{(-1)} \\
&= \left(\frac{1}{\lambda_m}\lambda_{n-(n-m)}\vec{v}_{n-m}\right)^{(-1)} \qquad \text{by Lemma 5.1} \\
&= (\vec{v}_{n-m})^{(-1)} \\
&= \vec{v}_{n-(n-m)} \\
&= \vec{v}_m.
\end{aligned}$$

□

This result is true for any size matrix. One corollary of the above result is the 2×2 circulant case for which the two eigenvectors of A are $(1, 1)$ and $(1, -1)$, precisely the two solutions given by (10).

5.3 Non-eigenvector solutions

We next consider the other three solutions given in (12) in a bit more detail. We first note that these three vectors would form the columns of a circulant matrix, which we denote as A_1:

$$A_1 = \begin{bmatrix} \dfrac{a^2-bc}{a} & \dfrac{c^2-ab}{c} & \dfrac{b^2-ac}{b} \\ \dfrac{b^2-ac}{b} & \dfrac{a^2-bc}{a} & \dfrac{c^2-ab}{c} \\ \dfrac{c^2-ab}{c} & \dfrac{b^2-ac}{b} & \dfrac{a^2-bc}{a} \end{bmatrix} = \begin{bmatrix} a_1 & c_1 & b_1 \\ b_1 & a_1 & c_1 \\ c_1 & b_1 & a_1 \end{bmatrix}$$

The second form of A_1, with a_1, b_1 and c_1, is given so that we can more easily refer to it and use it below. With this notation, we will refer to the original matrix A as A_0. So the columns of A_1 are the three non-eigenvector solutions (12) for A_0.

The solutions to (6) for A_1, a circulant matrix itself, are again the six we found for A_0: the same three eigenvectors in (11), and the other three circulant solutions in (12), but replacing a, b and c with a_1, b_1 and c_1. We look at the first term of $\vec{x}_4$ in (12) for A_1.

$$\frac{a_1^2 - b_1c_1}{a_1} = \frac{\left(\dfrac{a^2-bc}{a}\right)^2 - \left(\dfrac{b^2-ac}{b}\right)\left(\dfrac{c^2-ab}{c}\right)}{\dfrac{a^2-bc}{a}} = \frac{a^3b^3 + a^3c^3 + b^3c^3 - 3a^2b^2c^2}{abc(a^2-bc)}$$

Similarly, the second and third elements of $\vec{x}_4$ for A_1 are

$$\frac{a^3b^3 + a^3c^3 + b^3c^3 - 3a^2b^2c^2}{abc(b^2-ac)} \quad \text{and} \quad \frac{a^3b^3 + a^3c^3 + b^3c^3 - 3a^2b^2c^2}{abc(c^2-ab)}.$$

We factor the $(a^3b^3 + a^3c^3 + b^3c^3 - 3a^2b^2c^2)/abc$ term from each element of $\vec{x}_4$ for A_1, as well as from the other two solutions, $\vec{x}_5$ and $\vec{x}_6$, and use these three solutions as the columns for another matrix

$$A_2 = \begin{bmatrix} \dfrac{1}{a^2-bc} & \dfrac{1}{c^2-ab} & \dfrac{1}{b^2-ac} \\ \dfrac{1}{b^2-ac} & \dfrac{1}{a^2-bc} & \dfrac{1}{c^2-ab} \\ \dfrac{1}{c^2-ab} & \dfrac{1}{b^2-ac} & \dfrac{1}{a^2-bc} \end{bmatrix}.$$

One final iteration of this process produces an interesting result. The first term of $\vec{x}_4$ in (12) for A_2 is

$$\frac{\left(\dfrac{1}{a^2-bc}\right)^2-\left(\dfrac{1}{b^2-ac}\right)\left(\dfrac{1}{c^2-ab}\right)}{\left(\dfrac{1}{a^2-bc}\right)}=a(3abc-a^3-b^3-c^3).$$

Similarly, the other two terms in $\vec{x}_4$ for A_2 are $b(3abc-a^3-b^3-c^3)$ and $c(3abc-a^3-b^3-c^3)$. Factoring out the common term $3abc-a^3-b^3-c^3$ from this $\vec{x}_4$ results in $\vec{x}_4=(a,b,c)$. Repeating this for $\vec{x}_5$ and $\vec{x}_6$, and forming yet another matrix A_3, results in

$$A_3=\begin{bmatrix} a & c & b \\ b & a & c \\ c & b & a \end{bmatrix},$$

which is of course the original matrix A. This cyclic behavior is quite intriguing and potentially useful. With the aid of Maple, we were able to find analytic solutions to (6) for general 3×3 matrices. Surprisingly, however, regardless of how we manipulated and simplified the equations arising from solving (6), none of our CAS software could find the solutions in the case of a 4×4 circulant matrix, let alone a general 4×4 matrix. If we could find one set of solutions to (6) for a 4×4 circulant matrix, we expect that we could iterate as above to produce a cyclic pattern of matrices whose columns are the solutions, as done in the 3×3 circulant case. We expect that in general for an $n\times n$ circulant matrix, this cycle would require n iterations to return to the original matrix. Indeed, we are currently attempting to use this fact, as well as some of the properties discussion in Section 4, to help us induce solutions for the 4×4 circulant case, from which we then hope to derive a more general pattern of solutions for the $n\times n$ circulant case.

6 Open questions

We have found the general solution to (6) for $n\le3$. We feel that the formula for the solution for $n=3$ is excessively complex and therefore of limited use. We have consequently considered some specific 3×3 cases, most interestingly the circulant case. We have also developed some interesting and useful theory for the general case, including when A is block diagonal, as well as how permuting rows and/or columns of the given matrix A affects the solution $\vec{x}$. There are still a number of open questions that we are currently investigating, including the following.

- We are currently working on developing the solution for the 4×4 circulant case, from which we hope to find a pattern which can be extended to the $n\times n$ circulant case.
- We are trying to find an upper bound on the number of solutions to (6) in the general case. For the non-diagonal $n=1,2$ and 3 cases the maximum number of solutions is 1, 2 and 6, respectively. For the 4×4 case we have numerically found up to 20 solutions. Our numerical results thus far suggest that the maximum number of solutions in the general (non-diagonal or block diagonal) case may be $(2n)!/(n!)^2$, the formula for central binomial coefficients. We are currently investigating if (and why) this would be true, based on the problem itself, rather than simply because 1, 2, 6 and 20 happen to be the first four terms of this series.
- On a related note, it appears that an enormous amount of work could be done regarding numerically finding solutions to (6), especially for larger matrices. We numerically found solutions to (6), including by using built-in *solve*-type functions in Matlab and Maple for modified formulations of the problem. This was useful in the 3×3 case prior to finding analytic formulae for solutions, and thus far has been the only way to find solutions in the general 4×4 (or higher) case. However, there was not a particular approach that was either reliable in finding all solutions or that seemed superior to the other numerical approaches, thus we have not made it a part of our discussion in this paper. We are still trying to develop a more systematic approach to numerically finding solutions in higher dimension cases.

- Finally, we currently trying to develop a better characterization of the non-Sinkhorn-Knopp solutions to (6), that is, the solutions for which $S(\vec{x})$ in (7) has row and column sums of 1 but that are not doubly stochastic. It seems that these other solutions result in a matrix S that is as *non*-doubly stochastic as possible for a given A. That is, they maximize some sort of energy that is related to how "close" to being doubly stochastic the matrix S is, while the single positive solution resulting in the doubly stochastic S minimizes (most likely makes 0) that energy function.

7 Acknowledgements

We would like to thank Charlie Johnson for suggesting this interesting problem to study. Also, we are grateful to the very helpful comments and suggestions from the referees, in particular, regarding a theorem and two proofs in our paper.

References

[1] Michael Bacharach, *Biproportional Matrices and Input-Output Change*, CUP, 1970.

[2] Richard A. Brualdi, Seymour V. Parter and Hans Schneider, *The diagonal equivalence of a nonnegative matrix to a stochastic matrix*, J. of Mathematical Analysis and Applications, 16 (1966), pp. 31 - 50.

[3] Anjela Y. Govan, Amy N. Langville and Carl D. Meyer, *The offense-defense approach to ranking team sports*, J. of Quantitative Analysis in Sports, 5 (2009), Article 4.

[4] P. J. Gray, "Circulant Matrices," John Wiley, New York, 1979.

[5] Philip A. Knight, *The Sinkhorn-Knopp Algorithm: Convergence and Applications*, SIAM J. on Matrix Analysis and Applications, 30 (2008), pp 261-275.

[6] Philip A. Knight and Daniel Ruiz, *A Fast Algorithm for Matrix Balancing*, in "Web Information Retrieval and Linear Algebra Algorithms," edited by Andreas Frommer, Michael W. Mahoney and Daniel B. Szyld, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

[7] B. Lamond and N. F. Stewart, *Bregman's balancing method*, Transportion Research 15B (1981), pp. 239 - 248.

[8] Oren E. Livne and Gene H. Golub, *Scaling by binormalization*, Numerical Algorithms, 35 (2004), pp. 97 - 120.

[9] Marvin Marcus and Morris Newman, *The permanent of a symmetric matrix*, Amer. Math. Soc. Notices, 8 (1961), p. 595.

[10] Richard Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist., 35 (1964), pp. 876 - 879.

[11] Richard Sinkhorn and Paul Knopp, *Concerning nonnegative matrices and doubly stochastic matrices*, Pacific J. Math. 21 (1967), pp. 343 - 348.

About the authors:

Kristen Anderson graduated in mathematics from Pepperdine University in 2010 and is currently a student at the Hult International Business School.

Ashley Burt graduated in mathematics from Pepperdine University in 2009 and is currently a medical student at the University of Washington.

Will Cousins graduated in mathematics from Pepperdine University in 2009 and is currently a Ph.D. student in mathematics at North Carolina State University.

Brent Hancock graduated in mathematics from Pepperdine University in 2009 and is currently a Ph.D. student in mathematics at the University of California Riverside.

David Strong is an Associate Professor of Mathematics at Pepperdine University. He earned a Ph.D. in Applied Mathematics from UCLA in 1997 and works in image processing and numerical linear algebra. He currently serves on the MAA Subcommittee on Undergraduate Research.