ALTERNATE MINIMIZATION AND DOUBLY STOCHASTIC MATRICES

Melvyn B. Nathanson
Department of Mathematics, Lehman College (CUNY), Bronx, New York
melvyn.nathanson@lehman.cuny.edu

Received: 1/4/19, Accepted: 9/28/19, Published: 4/27/20

Abstract

Sinkhorn's alternative minimization algorithm applied to a positive $n \times n$ matrix converges to a doubly stochastic matrix. If the algorithm, applied to a 2×2 matrix, converges in a finite number of iterations, then it converges in at most two iterations, and the structure of such matrices is determined.

1. The Alternate Minimization Algorithm

A *positive matrix* is a matrix with positive coordinates. Let $\mathrm{diag}(x_1, \ldots, x_n)$ denote the $n \times n$ diagonal matrix with coordinates $x_1, \ldots, x_n$ on the main diagonal. A *positive diagonal matrix* is a diagonal matrix whose diagonal coordinates are positive. If A is an $m \times n$ positive matrix, X is an $m \times m$ positive diagonal matrix, and Y is an $n \times n$ positive diagonal matrix, then XA and AY are $m \times n$ positive matrices.

Let $A = (a_{i,j})$ be an $n \times n$ matrix. The ith *row sum* of A is

$$\mathrm{row}_i(A) = \sum_{j=1}^{n} a_{i,j}.$$

The jth *column sum* of A is

$$\mathrm{col}_j(A) = \sum_{i=1}^{n} a_{i,j}.$$

The matrix A is *row stochastic* if $\mathrm{row}_i(A) = 1$ for all $i \in \{1, \ldots, n\}$. The matrix A is *column stochastic* if $\mathrm{col}_j(A) = 1$ for all $j \in \{1, \ldots, n\}$. The matrix A is *doubly stochastic* if it is both row stochastic and column stochastic.

For example, a positive 2×2 matrix A is doubly stochastic if and only if there exist $\alpha, \beta \in (0,1)$ such that $\alpha + \beta = 1$ and

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}.$$

If $\alpha, \beta, \gamma \in (0,1)$ satisfy $\alpha + 2\beta = \beta + 2\gamma = 1$, then the 3×3 symmetric matrix

$$\begin{pmatrix} \alpha & \beta & \beta \\ \beta & \gamma & \gamma \\ \beta & \gamma & \gamma \end{pmatrix}$$

is doubly stochastic.

Let $A = (a_{i,j})$ be a positive $n \times n$ matrix. We have $\operatorname{row}_i(A) > 0$ and $\operatorname{col}_j(A) > 0$ for all $i, j \in \{1, \ldots, n\}$. Define the $n \times n$ positive diagonal matrix

$$X(A) = \operatorname{diag}\left(\frac{1}{\operatorname{row}_1(A)}, \frac{1}{\operatorname{row}_2(A)}, \ldots, \frac{1}{\operatorname{row}_n(A)} \right).$$

Multiplying A on the left by $X(A)$ multiplies each coordinate in the ith row of A by $1/\operatorname{row}_i(A)$, and so

$$\operatorname{row}_i(X(A)A) = \sum_{j=1}^{n} (X(A)A)_{i,j} = \sum_{j=1}^{n} \frac{a_{i,j}}{\operatorname{row}_i(A)} = \frac{\operatorname{row}_i(A)}{\operatorname{row}_i(A)} = 1$$

for all $i \in \{1, 2, \ldots, n\}$. The process of multiplying A on the left by $X(A)$ to obtain the row stochastic matrix $X(A)A$ is called *row scaling* or *row normalization*. We have $X(A)A = A$ if and only if A is row stochastic if and only if $X(A) = I$. Note that the row stochastic matrix $X(A)A$ is not necessarily column stochastic.

Similarly, we define the $n \times n$ positive diagonal matrix

$$Y(A) = \operatorname{diag}\left(\frac{1}{\operatorname{col}_1(A)}, \frac{1}{\operatorname{col}_2(A)}, \ldots, \frac{1}{\operatorname{col}_n(A)} \right).$$

Multiplying A on the right by $Y(A)$ multiplies each coordinate in the jth column of A by $1/\operatorname{col}_j(A)$, and so

$$\operatorname{col}_j(AY(A)) = \sum_{i=1}^{n} (AY(A))_{i,j} = \sum_{i=1}^{n} \frac{a_{i,j}}{\operatorname{col}_j(A)} = \frac{\operatorname{col}_j(A)}{\operatorname{col}_j(A)} = 1$$

for all $j \in \{1, 2, \ldots, n\}$. The process of multiplying A on the right by $Y(A)$ to obtain a column stochastic matrix $AY(A)$ is called *column scaling* or *column normalization*. We have $AY(A) = A$ if and only if $Y(A) = I$ if and only if A is column stochastic. The column stochastic matrix $AY(A)$ is not necessarily row stochastic.

The following elementary identity shows that column scaling can be replaced by row scaling, and conversely. .

Lemma 1. *Let A^t denote the transpose of the $n \times n$ positive matrix $A = (a_{i,j})$. Row and column scaling satisfy the following transpose symmetries:*

$$AY(A) = \left(X(A^t)\ A^t \right)^t$$

and

$$X(A)A = \left(A^t\ Y(A^t) \right)^t.$$

Proof. Let $A^t = (a^t_{i,j})$, where $a^t_{i,j} = a_{j,i}$. We have

$$\mathrm{row}_i(A^t) = \sum_{j=1}^{n} a^t_{i,j} = \sum_{j=1}^{n} a_{j,i} = \mathrm{col}_i(A)$$

and so

$$\begin{aligned} X(A^t) &= \mathrm{diag}\left(\frac{1}{\mathrm{row}_1(A^t)}, \ldots, \frac{1}{\mathrm{row}_n(A^t)}\right) \\ &= \mathrm{diag}\left(\frac{1}{\mathrm{col}_1(A)}, \ldots, \frac{1}{\mathrm{col}_n(A)}\right) \\ &= Y(A). \end{aligned}$$

Because the transpose of a diagonal matrix D is $D^t = D$, we obtain

$$\left(X(A^t)\ A^t\right)^t = \left(\ A^t\right)^t \left(X(A^t)\right)^t = A\ X(A^t) = A\ Y(A).$$

The proof of the identity $X(A)A = \left(A^t\ Y(A^t)\right)^t$ is similar. □

For example, if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $A^t = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ and

$$X(A^t) = \begin{pmatrix} 1/(a+c) & 0 \\ 0 & 1/(b+d) \end{pmatrix} = Y(A).$$

We have

$$\begin{aligned} \left(X(A^t)\ A^t\right)^t &= \begin{pmatrix} a/(a+c) & c/(a+c) \\ b/(b+d) & d/(b+d) \end{pmatrix}^t \\ &= \begin{pmatrix} a/(a+c) & b/(b+d) \\ c/(a+c) & d/(b+d) \end{pmatrix} \\ &= A\ Y(A). \end{aligned}$$

Sinkhorn [4] proved that row and column scaling satisfy the following uniqueness theorem.

Theorem 1. *Let A be a positive matrix, and let X_1, X_2, Y_1, and Y_2 be positive diagonal matrices. If*

$$S_1 = X_1 A Y_1 \qquad \text{and} \qquad S_2 = X_2 A Y_2$$

are doubly stochastic matrices, then

$$S_1 = S_2$$

and there exists $\lambda > 0$ such that

$$X_2 = \lambda X_1 \qquad \text{and} \qquad Y_2 = \lambda^{-1} Y_1.$$

If the positive matrix A is symmetric, then there is a unique positive diagonal matrix D such that $S = DAD$ is doubly stochastic.

The following algorithm is called "alternate minimization" (perhaps, more appropriately called "alternate scaling" or "alternate normalization"). The proof of the convergence of the algorithm is due to Sinkhorn [4] and Sinkhorn and Knopp [5].

Theorem 2. *Let $A = (a_{i,j})$ be a positive $n \times n$ matrix. Construct inductively an infinite sequence of positive $n \times n$ matrices by alternate operations of column scaling and row scaling:*

$$\begin{aligned}
A^{(0)} &= A \\
A^{(1)} &= A^{(0)}\ Y\left(A^{(0)}\right) \\
A^{(2)} &= X\left(A^{(1)}\right)\ A^{(1)} \\
A^{(3)} &= A^{(2)}\ Y\left(A^{(2)}\right) \\
A^{(4)} &= X\left(A^{(3)}\right)\ A^{(3)} \\
A^{(5)} &= A^{(4)}\ Y\left(A^{(4)}\right) \\
&\ \vdots\ .
\end{aligned}$$

The sequence of matrices $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ converges to a doubly stochastic matrix $S(A)$, and there exist positive diagonal matrices X and Y such that

$$S(A) = XAY.$$

The sequence of matrices $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ is called the *alternate minimization sequence* associated with A, and the matrix

$$S(A) = \lim_{\ell \to \infty} A^{(\ell)}$$

is the *alternate minimization limit* (also called the *Sinkhorn limit*) of A.

For example, if

$$A = A^{(0)} = \begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix}$$

then the next three matrices in the sequence $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ are

$$\begin{aligned} A^{(1)} &= A^{(0)}\, Y\left(A^{(0)}\right) = \begin{pmatrix} 1/4 & 3/7 \\ 3/4 & 4/7 \end{pmatrix} \\ A^{(2)} &= X\left(A^{(1)}\right)\, A^{(1)} = \begin{pmatrix} 7/19 & 12/19 \\ 21/37 & 16/37 \end{pmatrix} \\ A^{(3)} &= A^{(2)}\, Y\left(A^{(2)}\right) = \begin{pmatrix} 37/94 & 111/187 \\ 57/94 & 76/187 \end{pmatrix}. \end{aligned}$$

Let P and Q be positive diagonal $n \times n$ matrices. It follows from Theorem 1 that the alternate minimization limit of the positive $n \times n$ matrix A is equal to the alternate minimization limit of the matrix PAQ. In particular, the matrices A and $X(A)A$ have the same limits, and so it makes no difference if we start the alternate minimization sequence by column scaling or by row scaling.

Let A be a positive matrix, and let $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ be the alternate minimization sequence of matrices constructed in Theorem 2. If $A^{(L)}$ is doubly stochastic for some L, then $A^{(\ell)} = A^{(L)}$ for all $\ell \geq L$, and so the sequence of matrices $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ is eventually constant. In this presumably exceptional case, we say that the alternate minimization algorithm terminates in at most L steps. Note that, if the $n \times n$ matrix A has positive rational coordinates, then the matrix $A^{(\ell)}$ has positive rational coordinates for all $\ell \geq 1$. It follows that, if the Sinkhorn limit has irrational coordinates, then the alternate minimization algorithm cannot terminate in a finite number of steps. In Section 3, we prove that, for 2×2 matrices, if the algorithm terminates in a finite number of steps, then the algorithm terminates in at most two steps.

There is a vast literature on alternate minimization algorithms and Sinkhorn limits. For a recent survey, see Idel [2]. In complexity theory, it is the asymptotics of the approximating sequence $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ that is important (for example, Allen-Zhu, Li, Oliveira, and Wigderson [1]). This paper is concerned with number theoretic aspects of the algorithm, and with the classification of matrices for which the alternate minimization algorithm terminates in a finite number of steps. It is also of interest to consider the application of the algorithm to simultaneous approximation of irrational numbers by rational numbers.

2. Alternate Minimization Limits for 2×2 matrices

Theorem 3. *Let*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a positive 2×2 *matrix. Define the positive diagonal matrices*

$$X = \begin{pmatrix} \sqrt{cd} & 0 \\ 0 & \sqrt{ab} \end{pmatrix}$$

and

$$Y = \begin{pmatrix} \left(a\sqrt{cd} + c\sqrt{ab}\right)^{-1} & 0 \\ 0 & \left(b\sqrt{cd} + d\sqrt{ab}\right)^{-1} \end{pmatrix}$$

The limit of the alternate minimization sequence $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ *is the doubly stochastic matrix*

$$S(A) = XAY = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \tag{1}$$

with

$$\alpha = \frac{\sqrt{ad}}{\sqrt{ad} + \sqrt{bc}} \qquad \text{and} \qquad \beta = \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}. \tag{2}$$

Proof. Simply compute the product XAY. That the matrix XAY is the alternate minimization limit follows from uniqueness (Theorem 1). □

Corollary 1. *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2^+(\mathbf{Q})$. *The alternate minimization limit of the matrix* A *has rational coordinates if and only if* ad/bc *is the square of a rational number.*

For example, if $A_1 = \begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix}$, then

$$\begin{aligned} S(A_1) &= \begin{pmatrix} 2\sqrt{3} & 0 \\ 0 & \sqrt{3} \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} \left(5\sqrt{3}\right)^{-1} & 0 \\ 0 & \left(10\sqrt{3}\right)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1/5 & 0 \\ 0 & 1/10 \end{pmatrix} \\ &= \begin{pmatrix} 2/5 & 3/5 \\ 3/5 & 2/5 \end{pmatrix}. \end{aligned}$$

If $A_2 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, then

$$\begin{aligned} S(A_2) &= \begin{pmatrix} 2\sqrt{3} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} \left(2\sqrt{3} + 3\sqrt{2}\right)^{-1} & 0 \\ 0 & \left(4\sqrt{3} + 4\sqrt{2}\right)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{6} - 2 & 3 - \sqrt{6} \\ 3 - \sqrt{6} & \sqrt{6} - 2 \end{pmatrix}. \end{aligned}$$

Because A_2 has rational coefficients and $S(A_2)$ has irrational coefficients, the alternate minimization algorithm for A_2 must have infinite length, that is, does not terminate in a finite number of steps.

Theorem 4. *Consider the positive symmetric matrix*

$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}.$$

Let

$$\lambda = \left(abd + b^2\sqrt{ad}\right)^{-1/2}$$

and

$$D = \begin{pmatrix} \lambda\sqrt{bd} & 0 \\ 0 & \lambda\sqrt{ab} \end{pmatrix}.$$

The Sinkhorn limit of A is the doubly stochastic matrix

$$S(A) = DAD = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

with

$$\alpha = \frac{\sqrt{ad}}{\sqrt{ad}+b} \qquad \text{and} \qquad \beta = \frac{b}{\sqrt{ad}+b}.$$

Proof. The row scaling matrix

$$X(A) = \begin{pmatrix} \sqrt{bd} & 0 \\ 0 & \sqrt{ab} \end{pmatrix}$$

and the column scaling

$$Y(A) = \begin{pmatrix} \left(a\sqrt{bd} + b\sqrt{ab}\right)^{-1} & 0 \\ 0 & \left(b\sqrt{bd} + d\sqrt{ab}\right)^{-1} \end{pmatrix}$$

satisfy

$$D = \lambda X(A) = \lambda^{-1} Y(A).$$

By Theorem 3, the matrix

$$XAY = (\lambda X)A\left(\lambda^{-1}Y\right) = DAD = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

is doubly stochastic with $\alpha = \sqrt{ad}/(\sqrt{ad}+b)$. This completes the proof. $\square$

For example, if $A_3 = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, then

$$D = \begin{pmatrix} \sqrt{2}/2 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix}$$

and

$$\begin{aligned} DA_3D &= \begin{pmatrix} \sqrt{2}/2 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & 0 \\ 0 & \sqrt{2}/4 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \end{aligned}$$

If $A_4 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, then

$$D = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix}$$

and

$$\begin{aligned} DA_4D &= \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \end{aligned}$$

Note that the matrices $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ have the same Sinkhorn limits.

3. Limits for 2 × 2 Matrices in Finitely Many Steps

Theorem 5. *Let A be a positive 2×2 matrix that is not doubly stochastic. If the column scaled matrix $AY(A)$ is doubly stochastic, then A is a matrix of the form*

$$A = \begin{pmatrix} a & ct \\ c & at \end{pmatrix} \tag{3}$$

and

$$S(A) = AY(A) = \begin{pmatrix} a/(a+c) & c/(a+c) \\ c/(a+c) & a/(a+c) \end{pmatrix}.$$

If the row scaled matrix $X(A)A$ is doubly stochastic, then A is a matrix of the form

$$A = \begin{pmatrix} a & b \\ bt & at \end{pmatrix} \tag{4}$$

and

$$S(A) = X(A)A = \begin{pmatrix} a/(a+b) & b/(a+b) \\ b/(a+b) & a/(a+b) \end{pmatrix}.$$

For example, column scaling the matrix $\begin{pmatrix} 1 & 12 \\ 3 & 4 \end{pmatrix}$ and row scaling the matrix $\begin{pmatrix} 1 & 3 \\ 12 & 4 \end{pmatrix}$ both produce the doubly stochastic matrix $\begin{pmatrix} 1/4 & 3/4 \\ 3/4 & 1/4 \end{pmatrix}$.

Proof. Column scaling a matrix of the form (3) and row scaling a matrix of the form (4) both produce doubly stochastic matrices.

Conversely, let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The column scaled matrix

$$AY(A) = \begin{pmatrix} a/(a+c) & b/(b+d) \\ c/(a+c) & d/(b+d) \end{pmatrix}$$

is doubly stochastic if and only if

$$\frac{a}{a+c} + \frac{b}{b+d} = \frac{c}{a+c} + \frac{d}{b+d} = 1$$

if and only if

$$ab = cd.$$

Defining $t = b/c = d/a$, we obtain

$$A = \begin{pmatrix} a & ct \\ c & at \end{pmatrix} \qquad \text{and} \qquad S(A) = AY(A) = \begin{pmatrix} a/(a+c) & c/(a+c) \\ c/(a+c) & a/(a+c) \end{pmatrix}.$$

Similarly, the row scaled matrix

$$X(A)A = \begin{pmatrix} a/(a+b) & b/(a+b) \\ c/(c+d) & d/(c+d) \end{pmatrix}$$

is doubly stochastic if and only if

$$\frac{a}{a+b} + \frac{c}{c+d} = \frac{b}{a+b} + \frac{d}{c+d} = 1$$

if and only if

$$ac = bd.$$

Defining $t = c/b = d/a$, we obtain

$$A = \begin{pmatrix} a & b \\ bt & at \end{pmatrix} \qquad \text{and} \qquad S(A) = X(A)A = \begin{pmatrix} a/(a+b) & b/(a+b) \\ b/(a+b) & a/(a+b) \end{pmatrix}.$$

This completes the proof. $\square$

Theorem 6. *Let A be a positive 2×2 row stochastic matrix that is not column stochastic. If column scaling A produces a doubly stochastic matrix $S(A) = AY(A)$, then $A = \begin{pmatrix} a & 1-a \\ a & 1-a \end{pmatrix}$ with $a \neq 1/2$, and $S(A) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$.*

Let A be a positive 2×2 column stochastic matrix that is not row stochastic. If row scaling A produces a doubly stochastic matrix $S(A) = X(A)A$, then $A = \begin{pmatrix} a & a \\ 1-a & 1-a \end{pmatrix}$ with $a \neq 1/2$, and $S(A) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$.

Proof. Let A be a positive 2×2 matrix that is not doubly stochastic. By Theorem 5, if column scaling A produces a doubly stochastic matrix, then $A = \begin{pmatrix} a & ct \\ c & at \end{pmatrix}$ for some $t > 0$. If A is also row stochastic, then

$$a + ct = c + at = 1$$

and so

$$(a - c)(1 - t) = 0.$$

If $t = 1$, then $a + c = 1$ and A is doubly stochastic, which is absurd. Therefore, $t \neq 1$ and $a = c$. It follows that $A = \begin{pmatrix} a & at \\ a & at \end{pmatrix} = \begin{pmatrix} a & 1-a \\ a & 1-a \end{pmatrix}$ with $a \neq 1/2$ and $AY(A) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$.

Similarly, if row scaling A produces a doubly stochastic matrix, then Theorem 5 implies that $A = \begin{pmatrix} a & b \\ bt & at \end{pmatrix}$ for some $t > 0$. If A is also column stochastic, then

$$a + bt = b + at = 1$$

and so

$$(a - b)(1 - t) = 0.$$

If $t = 1$, then $a + b = 1$ and A is doubly stochastic, which is absurd. Therefore, $t \neq 1$ and $a = b$. It follows that $A = \begin{pmatrix} a & a \\ at & at \end{pmatrix} = \begin{pmatrix} a & a \\ 1-a & 1-a \end{pmatrix}$ with $a \neq 1/2$ and $X(A)A = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$. This completes the proof. □

Theorem 7. *Let A be a positive 2×2 matrix that is not doubly stochastic. If the alternate minimization algorithm produces a doubly stochastic matrix $S(A)$ in a finite number of steps, then the algorithm terminates in at most two steps.*

Suppose that the algorithm terminates in exactly two steps. The matrix $S(A) = A^{(2)}$ is obtained from $A^{(1)}$ by column scaling if and only if there exist positive real numbers p, r, and t with $t \neq 1$ such that

$$A = \begin{pmatrix} p & pt \\ r & rt \end{pmatrix}.$$

The matrix $S(A) = A^{(2)}$ is obtained from $A^{(1)}$ by row scaling if and only if there exist positive real numbers p, q, and t with $t \neq 1$ such that

$$A = \begin{pmatrix} p & q \\ pt & qt \end{pmatrix}.$$

In both cases, the alternate minimization limit is $S(A) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$.

Note that we obtain the limit matrix $S(A)$ either by first column scaling and then row scaling, or by first row scaling and then column scaling.

Proof. Let L be a positive integer such that the alternate minimization algorithm for A terminates in exactly L steps. There is a sequence of matrices $\left(A^{(\ell)}\right)_{\ell=0}^{L}$ with $A^{(0)} = A$ and $A^{(L)} = S(A)$ such that, for $\ell = 1, \ldots, L$, the matrix $A^{(\ell)}$ is obtained from $A^{(\ell-1)}$ by alternate column and row scalings.

Suppose that $L \geq 3$. There are two cases. Either $A^{(L)}$ is obtained from $A^{(L-1)}$ by column scaling, or $A^{(L)}$ is obtained from $A^{(L-1)}$ by row scaling,

If $A^{(L)}$ is obtained from $A^{(L-1)}$ by column scaling, then $A^{(L-1)}$ is obtained from $A^{(L-2)}$ by row scaling, and $A^{(L-2)}$ is obtained from $A^{(L-3)}$ by column scaling. We have the diagram

$$A = A^{(0)} \longrightarrow \ldots \longrightarrow A^{(L-3)} \xrightarrow{\text{col}} A^{(L-2)} \xrightarrow{\text{row}} A^{(L-1)} \xrightarrow{\text{col}} A^{(L)} = S(A).$$

The matrix $A^{(L-1)} = X(A^{(L-2)})A^{(L-2)}$ is row stochastic but not column stochastic. By Theorem 6, $A^{(L-1)} = \begin{pmatrix} a & 1-a \\ a & 1-a \end{pmatrix}$ with $a \neq 1/2$. If $L \geq 3$, then the matrix $A^{(L-2)}$ is column stochastic, and $A^{(L-1)} = X(A^{(L-2)})A^{(L-2)}$. We have

$$A^{(L-2)} = \begin{pmatrix} u & v \\ 1-u & 1-v \end{pmatrix}$$

for some $u, v \in (0,1)$, and

$$\begin{aligned} \begin{pmatrix} a & 1-a \\ a & 1-a \end{pmatrix} &= A^{(L-1)} = X(A^{(L-2)})A^{(L-2)} \\ &= \begin{pmatrix} u/(u+v) & v/(u+v) \\ (1-u)/(2-u-v) & (1-v)/(2-u-v) \end{pmatrix}. \end{aligned}$$

Therefore,

$$\frac{u}{u+v} = a = \frac{1-u}{2-u-v}.$$

Equivalently,

$$2u - u^2 - uv = u(2-u-v) = (1-u)(u+v) = u + v - u^2 - uv$$

and so $u = v$ and

$$A^{(L-2)} = \begin{pmatrix} u & u \\ 1-u & 1-u \end{pmatrix}.$$

Thus, the matrix

$$\begin{aligned} A^{(L-1)} = X(A^{(L-2)})A^{(L-2)} &= \begin{pmatrix} 1/(2u) & 0 \\ 0 & 1/(2-2u) \end{pmatrix} \begin{pmatrix} u & u \\ 1-u & 1-u \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \end{aligned}$$

is doubly stochastic, which is absurd. Therefore, $A^{(L-2)}$ is not column stochastic, and so $L \leq 2$.

Suppose that $L = 2$. Let $A = A^{(0)} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$. Because $A^{(1)}$ is row stochastic but not column stochastic and $A^{(2)}$ is doubly stochastic, there exists $a \in (0,1)$, $a \neq 1/2$, such that

$$\begin{aligned} \begin{pmatrix} a & 1-a \\ a & 1-a \end{pmatrix} &= A^{(1)} = X\left(A^{(0)}\right)A^{(0)} \\ &= \begin{pmatrix} p/(p+q) & q/(p+q) \\ r/(r+s) & s/(r+s) \end{pmatrix} \end{aligned}$$

and so

$$\frac{p}{p+q} = \frac{r}{r+s}.$$

Equivalently, $ps = qr$ and $s = qr/p$. Thus, with $t = q/p$, we obtain

$$A = A^{(0)} = \begin{pmatrix} p & q \\ r & qr/p \end{pmatrix} = \begin{pmatrix} p & pt \\ r & rt \end{pmatrix}.$$

If $t = 1$, then $A^{(1)}$ is doubly stochastic, which is absurd. Therefore, $t \neq 1$. Thus, if $L = 2$, then the alternate minimization sequence is

$$A = \begin{pmatrix} p & pt \\ r & rt \end{pmatrix} \to \begin{pmatrix} 1/(1+t) & t/(1+t) \\ 1/(1+t) & t/(1+t) \end{pmatrix} \to \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

A similar argument works in the second case, where the matrix $A^{(L)}$ is obtained from $A^{(L-1)}$ by row scaling. This completes the proof. $\square$

4. An Alternate Minimization Limit for an $n \times n$ Matrix

There are no formulae analogous to (1) and (2) for the alternate minimization limit of a positive 3×3 matrix. Nathanson [3] has explicitly computed the alternate minimization limits of some classes of symmetric positive 3×3 matrices. Here is a simple example of an explicit calculation.

Let $n \geq 3$ and $K > 0$. We consider the positive symmetric $n \times n$ matrix

$$A = \begin{pmatrix} K & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \\ \vdots & & & & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}.$$

By Theorems 1 and 2, there exists a unique positive diagonal matrix

$$D = \operatorname{diag}(x_1, x_2, \ldots, x_n) = \begin{pmatrix} x_1 & 0 & 0 & 0 & 0 \\ 0 & x_2 & 0 & 0 & 0 \\ 0 & 0 & x_3 & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & 0 & x_n \end{pmatrix}$$

such that the matrix

$$S(A) = DAD = \begin{pmatrix} Kx_1^2 & x_1x_2 & x_1x_3 & \cdots & x_1x_n \\ x_2x_1 & x_2^2 & x_2x_3 & \cdots & x_2x_n \\ x_3x_1 & x_3x_2 & x_3^2 & \cdots & x_3x_n \\ \vdots & & & & \vdots \\ x_nx_1 & x_nx_2 & x_nx_3 & \cdots & x_n^2 \end{pmatrix}$$

is doubly stochastic. Equivalently,

$$Kx_1^2 + x_1 \sum_{j=2}^{n} x_j = 1$$

and

$$x_i \sum_{j=1}^{n} x_j = 1$$

for $i = 2, 3, \ldots, n$. It follows that

$$x_i = \frac{1}{\sum_{j=1}^{n} x_j}$$

for $i = 2, 3, \ldots, n$, and

$$S(A) = \begin{pmatrix} \alpha & \beta & \beta & \cdots & \beta \\ \beta & \gamma & \gamma & \cdots & \gamma \\ \beta & \gamma & \gamma & \cdots & \gamma \\ \vdots & & & & \vdots \\ \beta & \gamma & \gamma & \cdots & \gamma \end{pmatrix}$$

where

$$\begin{aligned}
\alpha &= Kx_1^2 \\
\beta &= x_1x_2 = \frac{1-\alpha}{n-1} \\
\gamma &= x_2^2 = \frac{1-\beta}{n-1} = \frac{n-2+\alpha}{(n-1)^2}.
\end{aligned}$$

We obtain

$$\left(\frac{1-\alpha}{n-1}\right)^2 = \beta^2 = x_1^2x_2^2 = \frac{\alpha}{K}\left(\frac{n-2+\alpha}{(n-1)^2}\right)$$

and so

$$(K-1)\alpha^2 - (2K+n-2)\alpha + K = 0.$$

If $K=1$, then $\alpha = \beta = \gamma = 1/n$.

If $K \neq 1$, then

$$\alpha = \frac{2(K-1)+n \pm \sqrt{4(n-1)K+(n-2)^2}}{2(K-1)}.$$

The inequality $0 < \alpha < 1$ implies that

$$\alpha = \frac{2(K-1)+n - \sqrt{4(n-1)K+(n-2)^2}}{2(K-1)}$$

if $K > 1$ and if $0 < K < 1$.

For example, if $n = 3$, then

$$\alpha = \frac{2K+1-\sqrt{8K+1}}{2(K-1)}.$$

If $n = 3$ and $K = 2$, then

$$\alpha = \frac{5-\sqrt{17}}{2}, \qquad \beta = \frac{-3+\sqrt{17}}{4}, \qquad \gamma = \frac{7-\sqrt{17}}{8},$$

$$x_1 = \sqrt{\frac{5-\sqrt{17}}{4}} \qquad \text{and} \qquad x_2 = \frac{-3+\sqrt{17}}{\sqrt{5-\sqrt{17}}},$$

and

$$S(A) = DAD = \begin{pmatrix} \frac{5-\sqrt{17}}{2} & \frac{-3+\sqrt{17}}{4} & \frac{-3+\sqrt{17}}{4} \\ \frac{-3+\sqrt{17}}{4} & \frac{7-\sqrt{17}}{8} & \frac{7-\sqrt{17}}{8} \\ \frac{-3+\sqrt{17}}{4} & \frac{7-\sqrt{17}}{8} & \frac{7-\sqrt{17}}{8} \end{pmatrix}.$$

If $n = 3$ and $K = 3$, then

$$\alpha = \frac{1}{2}, \qquad \beta = \frac{1}{4}, \qquad \gamma = \frac{3}{8}.$$

For $n = 3$ and integers $K \geq 2$, the doubly stochastic matrix $S(A)$ is rational if and only if K is a triangular number, that is, a number of the form $K = (k^2 + k)/2$ for some positive integer k. In this case, we have

$$\alpha = \frac{k^2 - k}{k^2 + k - 2}$$
$$\beta = \frac{k - 1}{k^2 + k - 2}$$
$$\gamma = \frac{k^2 - 1}{2(k^2 + k - 2)}.$$

If $n = 4$, then

$$\alpha = \frac{K + 1 - \sqrt{3K + 1}}{K - 1}.$$

If $n = 4$ and $K = 2$, then

$$\alpha = 3 - \sqrt{7}, \qquad \beta = \frac{-2 + \sqrt{7}}{3}, \qquad \gamma = \frac{5 - \sqrt{7}}{9},$$

If $n = 4$ and $K = 5$, then

$$\alpha = 1/2, \qquad \beta = 1/6, \qquad \gamma = 5/18.$$

5. Open Problems

Problem 1. Does there exist a positive 3×3 matrix that is row stochastic but not column stochastic, and becomes doubly stochastic after one column scaling? This is equivalent to asking if there is a positive 3×3 matrix that, with respect to the alternate minimization algorithm, has finite length $L \geq 2$.

Problem 2. Let $n \geq 3$. Does there exist an integer $L^*(n)$ such that, if A is a positive $n \times n$ matrix for which the alternate minimization algorithm terminates in a finite number of steps, then the alternate minimization algorithm terminates in at most $L^*(n)$ steps?

Problem 3. Let K be a subfield of $\mathbf{R}$, and let $M_n^+(K)$ be the set of positive $n \times n$ matrices with coordinates in K. If $A \in M_n^+(K)$, then $A^{(\ell)} \in M_n^+(K)$ for all matrices in the alternate minimization sequence $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$. It follows that if $S(A) \notin M_n^+(K)$, then the alternate minimization algorithm for the matrix A has infinite length. Thus, if $A \in M_n^+(\mathbf{Q})$ and if the doubly stochastic limit $S(A)$ contains an irrational coordinate, then the alternate minimization algorithm has infinite length. In this case, the coordinates in the matrices $\left(A^{(\ell)}\right)_{\ell=0}^{\infty}$ are sequences of rational numbers that simultaneously converge to the coordinates of $S(A)$. It is of interest to understand the rate of convergence.

Problem 4. Let $\mathbf{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix} \in \mathbf{R}^m$ and $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbf{R}^n$ be vectors with positive coordinates such that

$$\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} c_j.$$

Let $A = (a_{i,j})$ be an $m \times n$ matrix. The matrix is A is $\mathbf{r}$-row stochastic if

$$\mathrm{row}_i(A) = \sum_{j=1}^{n} a_{i,j} = r_i$$

for all $i \in \{1, \ldots, m\}$. The matrix is A is $\mathbf{c}$-column stochastic if

$$\mathrm{col}_j(A) = \sum_{i=1}^{m} a_{i,j}$$

for all $j \in \{1, \ldots, n\}$. The matrix is A is $(\mathbf{r}, \mathbf{c})$ stochastic if it is both $\mathbf{r}$-row stochastic and $\mathbf{c}$-column stochastic.

Let A be a positive $m \times n$ matrix, and let

$$X_{\mathbf{r}}(A) = \mathrm{diag}\left(\frac{r_1}{\mathrm{row}_1(A)}, \frac{r_2}{\mathrm{row}_2(A)}, \ldots, \frac{r_m}{\mathrm{row}_m(A)} \right)$$

and

$$Y_{\mathbf{c}}(A) = \mathrm{diag}\left(\frac{c_1}{\mathrm{col}_1(A)}, \frac{c_2}{\mathrm{col}_2(A)}, \ldots, \frac{c_n}{\mathrm{col}_n(A)} \right).$$

The matrix $X_{\mathbf{r}}(A)\ A$ is $\mathbf{r}$-row stochastic, and the matrix $A\ Y_{\mathbf{c}}(A)$ is $\mathbf{r}$-column stochastic. The analogous $(\mathbf{r}, \mathbf{c})$-alternate minimization algorithm applied to a positive $m \times n$ matrix always converges to an $(\mathbf{r}, \mathbf{c})$-stochastic matrix.

Let $m, n \geq 2$. Does there exist an integer $L^*(m,n)$ such that, if A is a positive $m \times n$ matrix for which the $(\mathbf{r}, \mathbf{c})$-alternate minimization algorithm terminates in a finite number of steps, then the $(\mathbf{r}, \mathbf{c})$-alternate minimization algorithm terminates in at most $L^*(m,n)$ steps?

Problem 5. Does there exist a constant C_n with the following property: If A is an positive $n \times n$ matrix such that the alternate minimization algorithm, starting with row scaling, terminates in N_1 steps, and the alternate minimization algorithm, starting with column scaling, terminates in N_2 steps, then $|N_1 - N_2| < C_n$?

Note added in proof. S. B. Ekhad and D. Zeilberger (*Answers to some questions about explicit Sinkhorn limits posed by Mel Nathanson*, arXiv:1902.10783) solved

Problem 1 by constructing a positive 3×3 matrix that is row stochastic but not column stochastic, and becomes doubly stochastic after one column scaling. M. B. Nathanson (*Matrix scaling limits in finitely many iterations*, arXiv:1903.06778) generalized this construction to $n \times n$ matrices.

Alex Cohen (unpublished) solved Problem 2 by proving that $L^*(n) = 2$ for all $n \geq 3$. This also solves Problem 5. Extending Cohen's proof, Nathanson (unpublished) solved Problem 4 by showing that $L^*(m, n) = 2$ for all $m, n \geq 2$.

References

[1] Zeyuan Allen-Zhu, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson, *Much faster algorithms for matrix scaling*, 58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017, IEEE Computer Soc., Los Alamitos, CA, 2017, pp. 890–901.

[2] Martin Idel, *A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps*, arXiv:1609.06349, 2016.

[3] Melvyn B. Nathanson, *Matrix scaling and explicit doubly stochastic limits*, Linear Algebra Appl. 578 (2019), 111–132.

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

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