Matrix Scaling Limits in Finitely Many Iterations

Melvyn B. Nathanson

Abstract The alternate row and column scaling algorithm applied to a positive $n \times n$ matrix A converges to a doubly stochastic matrix $S(A)$, sometimes called the *Sinkhorn limit* of A. For every positive integer n, a two parameter family of row but not column stochastic $n \times n$ positive matrices is constructed that become doubly stochastic after exactly one column scaling.

2010 Mathematics Subject Classification 11C20 · 11B75 · 11J68 · 11J70

1 The Alternate Scaling Algorithm

A *positive matrix* is a matrix with positive coordinates. A *nonnegative matrix* is a matrix with nonnegative coordinates. Let $D = \operatorname{diag}(x_1, \ldots, x_n)$ denote the $n \times n$ diagonal matrix with coordinates $x_1, \ldots, x_n$ on the main diagonal. The diagonal matrix D is *positive* if its coordinates $x_1, \ldots, x_n$ are positive. If $A = (a_{i,j})$ is an $m \times n$ positive matrix, if $X = \operatorname{diag}(x_1, \ldots, x_m)$ is an $m \times m$ positive diagonal matrix, and if $Y = \operatorname{diag}(y_1, \ldots, y_n)$ is an $n \times n$ positive diagonal matrix, then $XA = (x_i a_{i,j})$, $AY = (a_{i,j} y_j)$, $XAY = (x_i a_{i,j} y_j)$ are $m \times n$ positive matrices.

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

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

The jth *column sum* of A is

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

M. B. Nathanson (✉)
Lehman College (CUNY), Bronx, NY 10468, USA
e-mail: melvyn.nathanson@lehman.cuny.edu

M. B. Nathanson (ed.), *Combinatorial and Additive Number Theory III*,
Springer Proceedings in Mathematics & Statistics 297,
https://doi.org/10.1007/978-3-030-31106-3_12

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

Let $A = (a_{i,j})$ be a nonnegative $n \times n$ matrix such that $\text{rowsum}_i(A) > 0$ and $\text{colsum}_j(A) > 0$ for all $i, j \in \{1, \ldots, n\}$. Define the $n \times n$ positive diagonal matrix

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

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

$$(X(A)A)_{i,j} = \frac{a_{i,j}}{\text{rowsum}_i(A)}$$

and

$$\begin{aligned}\text{rowsum}_i\,(X(A)A) &= \sum_{j=1}^{n}(X(A)A)_{i,j} = \sum_{j=1}^{n}\frac{a_{i,j}}{\text{rowsum}_i(A)}\\ &= \frac{\text{rowsum}_i(A)}{\text{rowsum}_i(A)} = 1\end{aligned}$$

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*. 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

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

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

$$(AY(A))_{i,j} = \frac{a_{i,j}}{\text{colsum}_j(A)}$$

and

$$\begin{aligned}\text{colsum}_j(AY(A)) &= \sum_{i=1}^{n}(AY(A))_{i,j} = \sum_{i=1}^{n}\frac{a_{i,j}}{\text{colsum}_j(A)}\\ &= \frac{\text{colsum}_j(A)}{\text{colsum}_j(A)} = 1\end{aligned}$$

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*. 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.

Let A be a positive $n \times n$ matrix. Alternately row scaling and column scaling the matrix A produces an infinite sequence of matrices that converges to a doubly stochastic matrix This result (due to Brualdi, Parter, and Schnieder [1], Letac [3], Menon [4], Sinkhorn [7], Sinkhorn–Knopp [8], Tverberg [9], and others) is classical.

Nathanson [5, 6] proved that if A is a 2×2 positive matrix that is not doubly stochastic but becomes doubly stochastic after a finite number L of scalings, then L is at most 2, and the 2×2 row stochastic matrices that become doubly stochastic after exactly one column scaling were computed explicitly. An open question was to describe $n \times n$ matrices with $n \geq 3$ that are not doubly stochastic but become doubly stochastic after finitely many scalings. Ekhad and Zeilberger [2] discovered the following row-stochastic but not column stochastic 3×3 matrix, which requires exactly one column scaling to become doubly stochastic:

$$A = \begin{pmatrix} 1/5 & 1/5 & 3/5 \\ 2/5 & 1/5 & 2/5 \\ 3/5 & 1/5 & 1/5 \end{pmatrix}. \tag{1}$$

Column scaling A produces the doubly stochastic matrix

$$AY(A) = \begin{pmatrix} 1/6 & 1/3 & 3/6 \\ 2/6 & 1/3 & 2/6 \\ 3/6 & 1/3 & 1/6 \end{pmatrix}.$$

The following construction generalizes this example. For every $n \geq 3$, there is a two parameter family of row-stochastic $n \times n$ matrices that require exactly one column scaling to become doubly stochastic

Let $A = \left(a_{i,j}\right)$ be an $m \times n$ matrix. For $i = 1, \ldots, m$, we denote the ith row of A by

$$\text{row}_i(A) = \left(a_{i,1}, a_{i,2}, \ldots, a_{i,n}\right).$$

Theorem 1 *Let k and ℓ be positive integers, and let $n > \max(2k, 2\ell)$. Let x and z be positive real numbers such that*

$$0 < x + z < \frac{1}{k} \quad \text{and} \quad x + z \neq \frac{2}{n} \tag{2}$$

and let

$$y = \frac{x+z}{2} \quad \text{and} \quad w = \frac{1 - k(x+z)}{n - 2k}. \tag{3}$$

The $n \times n$ matrix A such that

$$row_i(A) = \begin{cases} (\underbrace{x, x, \ldots, x}_{k}\, \underbrace{w, w, \ldots, w}_{n-2k}\, \underbrace{z, z, \ldots, z}_{k} & \text{if } i \in \{1, 2, \ldots, \ell\} \\ (\underbrace{y, y, \ldots, y}_{k}\, \underbrace{w, w, \ldots, w}_{n-2k}\, \underbrace{y, y, \ldots, y}_{k} & \text{if } i \in \{\ell+1, \ell+2, \ldots, n-\ell\} \\ (\underbrace{z, z, \ldots, z}_{k}\, \underbrace{w, w, \ldots, w}_{n-2k}\, \underbrace{x, x, \ldots, x}_{k} & \text{if } i \in \{n-\ell+1, n-\ell+2, \ldots, n\} \end{cases}$$

is row stochastic but not column stochastic. The matrix obtained from A after one column scaling is doubly stochastic.

Proof If

$$i \in \{1, 2, \ldots, \ell\} \cup \{n-\ell+1, n-\ell+2, \ldots, n\}$$

then

$$\text{rowsum}_i(A) = k(x+z) + (n-2k)w = 1.$$

If

$$i \in \{\ell+1, \ell+2, \ldots, n-\ell\}$$

then

$$\text{rowsum}_i(A) = 2ky + (n-2k)w = 1.$$

Thus, the matrix A is row stochastic.

If

$$j \in \{1, 2, \ldots, k\} \cup \{n-k+1, n-k+2, \ldots, n\}$$

then

$$\text{colsum}_j(A) = \ell x + (n-2\ell)y + \ell z = ny = \frac{n}{2}(x+z) \neq 1.$$

If

$$j \in \{k+1, k+2, \ldots, n-k\}$$

then

$$\text{colsum}_j(A) = nw \neq 1.$$

Thus, matrix A is not column stochastic.

The column scaling matrix for A is the positive diagonal matrix

$$Y(A) = \text{diag}\left(\underbrace{\frac{1}{ny}, \ldots, \frac{1}{ny}}_{k}, \underbrace{\frac{1}{nw}, \ldots, \frac{1}{nw}}_{n-2k}, \underbrace{\frac{1}{ny}, \ldots, \frac{1}{ny}}_{k}\right).$$

For the column scaled matrix $AY(A)$, we have the following row sums. If

$$i \in \{1, 2, \ldots, \ell\} \cup \{n-\ell+1, n-\ell+2, \ldots, n\}$$

then

$$\text{rowsum}_i(AY(A)) = \frac{kx}{ny} + \frac{(n-2k)w}{nw} + \frac{kz}{ny} = \frac{k(x+z)}{ny} + 1 - \frac{2k}{n} = 1.$$

If

$$i \in \{\ell+1, \ell+2, \ldots, n-\ell\}$$

then

$$\text{rowsum}_i(A) = \frac{2ky}{ny} + \frac{(n-2k)w}{nw} = \frac{2k}{n} + 1 - \frac{2k}{n} = 1.$$

Thus, the matrix $AY(A)$ is row stochastic. This completes the proof. □

For example, let $k = \ell = 1$ and $n = 3$, and let w, x, y, z be positive real numbers such that

$$0 < x + z < 1, \qquad x + z \neq \frac{2}{3}$$

$$y = \frac{x+z}{2} \qquad \text{and} \qquad w = 1 - x - z.$$

The matrix

$$A = \begin{pmatrix} x & w & z \\ y & w & y \\ z & w & x \end{pmatrix}, \tag{4}$$

is row stochastic but not column stochastic. By Theorem 1, column scaling A produces a doubly stochastic matrix. Choosing $x = 1/5$ and $z = 3/5$, we obtain the matrix (1).

Here is another example. Let $k = 2$, $\ell = 3$, and $n = 7$. Choosing

$$x = \frac{1}{4}, \quad y = \frac{3}{16}, \quad z = \frac{1}{8}, \quad w = \frac{1}{12}$$

we obtain the row but not column stochastic matrix

$$A = \begin{pmatrix} 1/4 & 1/4 & 1/12 & 1/12 & 1/12 & 1/8 & 1/8 \\ 1/4 & 1/4 & 1/12 & 1/12 & 1/12 & 1/8 & 1/8 \\ 1/4 & 1/4 & 1/12 & 1/12 & 1/12 & 1/8 & 1/8 \\ 3/16 & 3/16 & 1/12 & 1/12 & 1/12 & 3/16 & 3/16 \\ 1/8 & 1/8 & 1/12 & 1/12 & 1/12 & 1/4 & 1/4 \\ 1/8 & 1/8 & 1/12 & 1/12 & 1/12 & 1/4 & 1/4 \\ 1/8 & 1/8 & 1/12 & 1/12 & 1/12 & 1/4 & 1/4 \end{pmatrix}.$$

Column scaling produces the doubly stochastic matrix

$$AY(A) = \begin{pmatrix} 4/21 & 4/21 & 1/7 & 1/7 & 1/7 & 2/21 & 2/21 \\ 4/21 & 4/21 & 1/7 & 1/7 & 1/7 & 2/21 & 2/21 \\ 4/21 & 4/21 & 1/7 & 1/7 & 1/7 & 2/21 & 2/21 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 2/21 & 2/21 & 1/7 & 1/7 & 1/7 & 4/21 & 4/21 \\ 2/21 & 2/21 & 1/7 & 1/7 & 1/7 & 4/21 & 4/21 \\ 2/21 & 2/21 & 1/7 & 1/7 & 1/7 & 4/21 & 4/21 \end{pmatrix}.$$

Theorem 2 *Every $n \times n$ matrix A constructed in Theorem 1 satisfies* $\det(A) = 0$.

Proof There are three cases.

If $k > 1$ or $n - 2k > 1$, then A has two equal columns and $\det(A) = 0$.

If $\ell > 1$ or $n - 2\ell > 1$, then A has two equal rows and $\det(A) = 0$.

If $k = \ell = 1$ and $n = 3$, then

$$A = \begin{pmatrix} x & w & z \\ y & w & y \\ z & w & x \end{pmatrix}$$

and

$$\det(A) = w(x - z)(x + z - 2y) = 0.$$

This completes the proof. □

Theorem 2 is of interest for the following reason. Let $A = \left(a_{i,j}\right)$ be an $n \times n$ matrix. If $\det(A) \neq 0$, then the system of linear equations

$$\begin{aligned} a_{1,1}t_1 + a_{2,1}t_2 + \cdots + a_{n,1}t_n &= 1 \\ a_{1,2}t_1 + a_{2,2}t_2 + \cdots + a_{n,2}t_n &= 1 \\ &\vdots \\ a_{1,n}t_1 + a_{2,n}t_2 + \cdots + a_{n,n}t_n &= 1 \end{aligned}$$

has a unique solution. Equivalently, if $\det(A) \neq 0$, then there exists a unique $n \times n$ diagonal matrix $T = \operatorname{diag}(t_1, \ldots, t_n)$ such that the matrix $B = TA$ is column stochastic.

Suppose that the matrix A is positive and row stochastic. If $t_i > 0$ for all $i \in \{1, \ldots, n\}$, then T is invertible and $B = TA$ is a positive column stochastic matrix. Setting $X = T^{-1}$, we have $XB = A$. Moreover, X is the row scaling matrix associated to B. Thus, if A is a row stochastic matrix such that column scaling A produces a doubly stochastic matrix, then we have pulled A back to a column stochastic matrix B, and we have increased by 1 the number of scalings needed to get a doubly stochastic matrix.

Unfortunately, the matrices constructed in Theorem 1 have determinant 0.

2 Open Problems

1. Does there exist a positive 3×3 row stochastic but not column stochastic matrix A with nonzero determinant such that A becomes doubly stochastic after one column scaling?
2. Let A be a positive 3×3 row stochastic but not column stochastic matrix that becomes doubly stochastic after one column scaling. Does $\det(A) = 0$ imply that A has the shape of matrix (4)?
3. Here is the inverse problem: Let A be an $n \times n$ row-stochastic matrix. Does there exist a column stochastic matrix B such that row scaling B produces A (equivalently, such that $X(B)B = A$)? Compute B.
4. Modify the above problems so that the matrices are required to have rational coordinates.
5. Determine if, for positive integers $L \geq 3$ and $n \geq 3$, there exists a positive $n \times n$ matrix that requires exactly L scalings to reach a doubly stochastic matrix.
6. Classify all matrices for which the alternate scaling algorithm terminates in finitely many steps.

Acknowledgements Supported in part by a grant from the PSC-CUNY Research Award Program.

References

1. R. A. Brualdi, S. V. Parter, and H. Schneider, *The diagonal equivalence of a nonnegative matrix to a stochastic matrix*, J. Math. Anal. Appl. 16 (1966), 31–50.
2. S. B. Ekhad and D. Zeilberger, *Answers to some questions about explicit Sinkhorn limits posed by Mel Nathanson*, arXiv:1902.10783, 2019.
3. G. Letac, *A unified treatment of some theorems on positive matrices*, Proc. Amer. Math. Soc. **43** (1974), 11–17.
4. M. V. Menon, *Reduction of a matrix with positive elements to a doubly stochastic matrix*, Proc. Amer. Math. Soc. 18 (1967), 244–247.
5. M. B. Nathanson, *Alternate minimization and doubly stochastic matrices*, arXiv:1812.11935, 2018.
6. M. B. Nathanson, *Matrix scaling and explicit doubly stochastic limits*, Linear Algebra and its Applications 578 (2019), 111–132; https://doi.org/10.1016/j.laa.2019.05.004
7. R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist. **35** (1964), 876–879.
8. R. Sinkhorn and P. Knopp, *Concerning nonnegative matrices and doubly stochastic matrices*, Pacific J. Math. **21** (1967), 343–348.
9. H. Tverberg, *On Sinkhorn's representation of nonnegative matrices*, J. Math. Anal. Appl. 54 (1976), no. 3, 674–677.