Answers to Some Questions about Explicit Sinkhorn Limits posed by Mel Nathanson

Shalosh B. EKHAD and Doron ZEILBERGER

Preface

At the Jan. 2018 Joint Mathematics Meetings, Avi Wigderson gave a series of three fascinating lectures [W], whose starting point was the *Sinkhorn algorithm*. One of the people in the audience was Mel Nathanson, and this lead him to write two papers [N1][N2] inspired by this algorithm.

A square matrix is *row-stochastic* if all its rows add-up to 1 and is *column-stochastic* if all its columns add-up to 1. It is *doubly-stochastic* if it is both row- and column- stochastic.

Consider the following question.

"Given a square matrix, A, with positive entries, can you find diagonal matrices X and Y, and a doubly-stochastic matrix, S, such $S = XAY$?"

The Sinkhorn algorithm gives, very fast, an *approximate* answer, as follows. Let $R(A)$ be the operation that inputs a matrix A with positive entries and outputs the row-stochastic matrix obtained by normalizing each row, i.e. dividing each row by its sum. Analogously, let $C(A)$, be the operation that inputs a matrix A with positive entries and outputs the column-stochastic matrix obtained by normalizing each column, i.e. dividing each column by its sum.

The Sinkhorn algorithm proceeds by *alternating* these two 'correction' operations. Surprisingly, after few iterations you get something that is *approximately* doubly-stochastic.

If A is also symmetric, then one can take $X = Y$, and then one is looking for the unique diagonal matrix, X, and for the unique symmetric doubly-stochastic matrix, S, such that $S = XAX$.

Mel Nathanson's questions

Nathanson wondered if one can find explicit expressions, in terms of the entries of A, for the entries of the Sinkhorn limit, S. He also commented that one should be able to do it using *Gröbner bases*. He also wondered whether there exist matrices for which the Sinkhorn algorithm terminates in a *finite* number of steps, and settled the question for the 2×2 case.

The Maple package SINKHORN.txt

One of us (DZ) wrote a Maple package, `SINKHORN.txt`, available from the front of the present article

`http://sites.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/sinkhorn.html` ,

that lead to the solutions, by the other author (SBE), of some of Nathanson's questions. Following Nathanson's advice we used Gröbner bases (the Buchberger algorithm).

Answers to some of Nathanson's questions

The following theorem completely answers the *central problem* (problem 1, p. 26) in Nathanson's article [N2].

Theorem 1. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

be the *generic*, ('symbolic'), 3×3 **symmetric** matrix, with **positive coefficients**. Its Sinkhorn limit, let's call it S, is a certain symmetric 3×3 **doubly-stochastic** matrix whose $(1,1)$ entry, s_{11}, is given by

$$s_{11} = a_{11} z \quad ,$$

where z is the positive root of the **quartic equation**

$$\begin{gathered}
({a_{1,1}}^4{a_{2,2}}^2{a_{3,3}}^2 - {a_{1,1}}^4 a_{2,2}{a_{2,3}}^2 a_{3,3} - 2\,{a_{1,1}}^3{a_{1,2}}^2 a_{2,2}{a_{3,3}}^2 \\
+{a_{1,1}}^3{a_{1,2}}^2{a_{2,3}}^2 a_{3,3} + 2\,{a_{1,1}}^3 a_{1,2}a_{1,3}a_{2,2}a_{2,3}a_{3,3} - 2\,{a_{1,1}}^3{a_{1,3}}^2{a_{2,2}}^2 a_{3,3} + {a_{1,1}}^3{a_{1,3}}^2 a_{2,2}{a_{2,3}}^2 \\
+{a_{1,1}}^2{a_{1,2}}^4{a_{3,3}}^2 - 2\,{a_{1,1}}^2{a_{1,2}}^3 a_{1,3}a_{2,3}a_{3,3} + 3\,{a_{1,1}}^2{a_{1,2}}^2{a_{1,3}}^2 a_{2,2}a_{3,3} - {a_{1,1}}^2{a_{1,2}}^2{a_{1,3}}^2{a_{2,3}}^2 \\
-2\,{a_{1,1}}^2 a_{1,2}{a_{1,3}}^3 a_{2,2}a_{2,3} + {a_{1,1}}^2{a_{1,3}}^4{a_{2,2}}^2 - a_{1,1}{a_{1,2}}^4{a_{1,3}}^2 a_{3,3} + 2\,a_{1,1}{a_{1,2}}^3{a_{1,3}}^3 a_{2,3} - a_{1,1}{a_{1,2}}^2{a_{1,3}}^4 a_{2,2})\,z^4 \\
+(-4\,{a_{1,1}}^3{a_{2,2}}^2{a_{3,3}}^2 + 4\,{a_{1,1}}^3 a_{2,2}{a_{2,3}}^2 a_{3,3} + 4\,{a_{1,1}}^2{a_{1,2}}^2 a_{2,2}{a_{3,3}}^2 - 3\,{a_{1,1}}^2{a_{1,2}}^2{a_{2,3}}^2 a_{3,3} - 2\,{a_{1,1}}^2 a_{1,2}a_{1,3}a_{2,2}a_{2,3}a_{3,3} \\
+4\,{a_{1,1}}^2{a_{1,3}}^2{a_{2,2}}^2 a_{3,3} - 3\,{a_{1,1}}^2{a_{1,3}}^2 a_{2,2}{a_{2,3}}^2 - 2\,a_{1,1}{a_{1,2}}^2{a_{1,3}}^2 a_{2,2}a_{3,3} + 2\,a_{1,1}{a_{1,2}}^2{a_{1,3}}^2{a_{2,3}}^2 - {a_{1,2}}^4{a_{1,3}}^2 a_{3,3} \\
+2\,{a_{1,2}}^3{a_{1,3}}^3 a_{2,3} - {a_{1,2}}^2{a_{1,3}}^4 a_{2,2})\,z^3 \\
+(6\,{a_{1,1}}^2{a_{2,2}}^2{a_{3,3}}^2 - 6\,{a_{1,1}}^2 a_{2,2}{a_{2,3}}^2 a_{3,3} - 2\,a_{1,1}{a_{1,2}}^2 a_{2,2}{a_{3,3}}^2 + 3\,a_{1,1}{a_{1,2}}^2{a_{2,3}}^2 a_{3,3} \\
-2\,a_{1,1}a_{1,2}a_{1,3}a_{2,2}a_{2,3}a_{3,3} - 2\,a_{1,1}{a_{1,3}}^2{a_{2,2}}^2 a_{3,3} + 3\,a_{1,1}{a_{1,3}}^2 a_{2,2}{a_{2,3}}^2 + 2\,{a_{1,2}}^3 a_{1,3}a_{2,3}a_{3,3} \\
-3\,{a_{1,2}}^2{a_{1,3}}^2 a_{2,2}a_{3,3} - {a_{1,2}}^2{a_{1,3}}^2{a_{2,3}}^2 + 2\,a_{1,2}{a_{1,3}}^3 a_{2,2}a_{2,3})\,z^2 \\
+(-4\,a_{1,1}{a_{2,2}}^2{a_{3,3}}^2 + 4\,a_{1,1}a_{2,2}{a_{2,3}}^2 a_{3,3} - {a_{1,2}}^2{a_{2,3}}^2 a_{3,3} + 2\,a_{1,2}a_{1,3}a_{2,2}a_{2,3}a_{3,3} - {a_{1,3}}^2 a_{2,2}{a_{2,3}}^2)\,z \\
+{a_{2,2}}^2{a_{3,3}}^2 - a_{2,2}{a_{2,3}}^2 a_{3,3} \;=\; 0 \quad .
\end{gathered}$$

Furthermore the diagonal matrix X, such that $S = XAX$ has its $(1,1)$ entry, x_{11}, given explicitly by

$$x_{11} = \sqrt{z} \quad .$$

The other five entries of the 3×3 symmetric matrix S, and the other two non-zero entries of the 3×3 diagonal matrix X are too long to be presented here, but are readily available (free of charge, and no advertisements!) from the following url `http://sites.math.rutgers.edu/~zeilberg/tokhniot/oSINKHORN3.txt` .

Comment: Of course 'explicit' is in the eyes of the beholder, and some people may argue that Sinkhorn's algorithm that produces (extremely fast!) the desired doubly-stochastic matrix S to any desired accuracy is explicit enough. But to pure mathematicians it only gives 'approximations'. Our solution is as explicit as it can get, even if you insist that the entries are 'solvable by radicals', since z satisfies a certain explicit *quartic* equation, with coefficients that are polynomials in the six entries of A.

Since the general case is so complicated, Nathanson [N2] (problem 2, p. 26) also asked for the Sinkhorn matrices of two special cases. The next theorem answers the first part of problem 2.

Theorem 2. Let K and L be arbitrary positive numbers, and let

$$A = \begin{pmatrix} K & 1 & 1 \\ 1 & L & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad .$$

Its Sinkhorn limit, let's call it S, is a certain symmetric 3×3 **doubly-stochastic** matrix whose $(1,1)$ entry, s_{11}, is given by

$$s_{11} = Kz \quad ,$$

where z is the positive root of the quartic equation

$$L+(-4\,LK+1)\,z+\left(6\,LK^2-2\,LK-3\,K-1\right)z^2-(K-1)\left(4\,LK^2-3\,K-1\right)z^3+K\,(K-1)^2\,(LK-1)\,z^4=0 \quad .$$

Furthermore the diagonal matrix X, such that $S = XAX$ has its $(1,1)$ entry, x_{11}, given explicitly by

$$x_{11} = \sqrt{z} \quad .$$

The other five entries of the 3×3 symmetric matrix S, and the other two non-zero entries of the 3×3 diagonal matrix X are available here:

`http://sites.math.rutgers.edu/~ zeilberg/tokhniot/oSINKHORN4.txt` .

The next theorem answers the second part of problem 2 of [N2].

Theorem 3. Let K, L and M be arbitrary positive numbers, and let

$$A = \begin{pmatrix} K & 1 & 1 \\ 1 & L & 1 \\ 1 & 1 & M \end{pmatrix}$$

Its Sinkhorn limit, let's call it S, is a certain symmetric 3×3 **doubly-stochastic** matrix whose $(1,1)$ entry, s_{11}, is given by

$$s_{11} = Kz \quad ,$$

where z is the positive root of the quartic equation

$$LM(LM-1)+(-4\,KL^2M^2+4\,KML+2\,LM-L-M)z$$

$$+(6\,K^2L^2M^2-6\,K^2LM-2\,KML^2-2\,KLM^2-2\,KML+3\,LK+3\,KM-3\,LM+2\,L+2\,M-1)z^2+$$

$$(-4\,K^3L^2M^2+4\,K^3LM+4\,K^2L^2M+4\,K^2LM^2-2\,K^2LM-3\,K^2L$$

$$-3\,K^2M-2\,KML+2\,K-L-M+2)z^3$$

$$+(KM-1)K(LK-1)(KML-K-L-M+2)z^4=0 \quad .$$

Furthermore the diagonal matrix X, such that $S=XAX$ has its $(1,1)$ entry, x_{11}, given explicitly by

$$x_{11}=\sqrt{z} \quad .$$

The other five entries of the 3×3 symmetric matrix S, and the other two entries of the diagonal matrix X are available here:

`http://sites.math.rutgers.edu/~ zeilberg/tokhniot/oSINKHORN5.txt` .

The next fact answers, in the **affirmative**, problem 5 (p. 27) in [N2].

Fact 4: The matrix

$$A=\begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{3}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$$

is row-stochastic (check!), but not column-stochastic (check!), but applying column-scaling to it yields the matrix

$$\begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$$

that **is** doubly-stochastic (check!).

By multiplying the first row of A by 10, the second row by 5 and the third row by 15 we get the matrix

$$M=\begin{pmatrix} 2 & 2 & 6 \\ 2 & 1 & 2 \\ 9 & 3 & 3 \end{pmatrix} \quad ,$$

that achieves its Sinkhorn limit after only **two** steps (or one double step). In other words M is not doubly-stochastic but $C(R(M))$ is.

Using procedure `MelNprob5(T,var)` in the Maple package `SINKHORN.txt`, one can concoct many other such examples.

On the 3×3 matrix of all 1s except for the $(1,1)$ entry

In section 13 of [N2], the following matrix is discussed

$$A(r) := \begin{pmatrix} \frac{r(r+1)}{2} & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad .$$

Its Sinkhorn limit , let's call it $S(r)$ is:

$$S(r) = \begin{pmatrix} \frac{r}{r+2} & \frac{1}{r+2} & \frac{1}{r+2} \\ \frac{1}{r+2} & \frac{r+1}{2(r+2)} & \frac{r+1}{2(r+2)} \\ \frac{1}{r+2} & \frac{r+1}{2(r+2)} & \frac{r+1}{2(r+2)} \end{pmatrix} \quad .$$

The diagonal matrix, $X(r)$, such that $X(r)A(r)X(r) = S(r)$ is

$$X(r) = \sqrt{\frac{2}{(r+1)(r+2)}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{r+1}{2} & 0 \\ 0 & 0 & \frac{r+1}{2} \end{pmatrix} \quad .$$

It is asked in [N2] whether the Sinkhorn algorithm applied to $A(r)$ can terminate after a finite number of steps. This is unlikely for the following reason. We Use procedure `MelNsec13(r,k)` in our Maple package, It inputs a symbol `r`, and a positive integer `k`, and outputs the difference between the sums of the first and second rows when row-scaling followed by column-scaling is applied `k` times. This is a necessary condition for being doubly-stochastic. By trying out `MelNsec13(r,k)` for `k` from 1 to 6 it appears that the numerator is always $3\,((r+2)(r-1))^{2k}$, hence only vanishes when $r = 1$ or $r = -2$ producing the all 1-matrix. The fact that this holds for all k could presumably proved rigorously by mathematical induction.

What about Larger sizes?

Theorem 1 was obtained via procedure `ExacGS` in our Maple package `SINKHORN.txt` . It would be too much for Maple (and probably also for SINGULAR and even for MAGMA) to do the analogous theorem for a *generic*, symbolic symmetric $n \times n$ matrix for $n \geq 4$. But it does a good job, for *numerical* matrices, finding the *exact* Sinkhorn limits in terms of algebraic numbers.

If one had a sufficiently large computer, one would be able to state the analog of Theorem 1 for $n \times n$ matrices, for any *specific* $n \geq 4$, but now the degree of the defining equation for z is 2^{n-1}, rather then $2^2 = 4$, and the coefficients of that defining equation are polynomials in the $n(n+1)/2$ entries of the symmetric $n \times n$ matrix A.

References

[N1] Melvyn B. Nathanson, *Alternate minimization and doubly stochastic matrices*, `https://arxiv.org/abs/1812.11930` .

[N2] Melvyn B. Nathanson, *Matrix scaling, explicit Sinkhorn limits, and Arithmetic*, `https://arxiv.org/abs/1902.04544` .

[W] Avi Wigderson, *Alternate minimization and scaling algorithms: theory, applications, connections*, (The first of a 3-lecture series (the AMS Colloquium lectures) given at the Joint Mathematics Meetings, San Diego, California - January 10, 2018).
Abstract, slides, and video available from `https://www.math.ias.edu/avi/talks` .

Shalosh B. Ekhad and Doron Zeilberger, Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA.
Email: `[DoronZeil, ShaloshBEkhad] at gmail dot com` .

Exclusively published in the Personal Journal of Shalosh B. Ekhad and Doron Zeilberger and arxiv.org .

Written: Feb. 27, 2019.