Ann. Funct. Anal. 7 (2016), no. 3, 386–393
http://dx.doi.org/10.1215/20088752-3605195
ISSN: 2008-8752 (electronic)
http://projecteuclid.org/afa

POLYTOPES OF STOCHASTIC TENSORS

HAIXIA CHANG,[1] VEHBI E. PAKSOY,[2] and FUZHEN ZHANG[2*]

Communicated by Q.-W. Wang

Abstract. Considering $n \times n \times n$ stochastic tensors (a_{ijk}) (i.e., nonnegative hypermatrices in which every sum over one index i, j, or k, is 1), we study the polytope (Ω_n) of all these tensors, the convex set (L_n) of all tensors in Ω_n with some positive diagonals, and the polytope (Δ_n) generated by the permutation tensors. We show that L_n is almost the same as Ω_n except for some boundary points. We also present an upper bound for the number of vertices of Ω_n.

1. Introduction

A square matrix is doubly stochastic if its entries are all nonnegative and each row and column sum is 1. A celebrated result known as *Birkhoff's theorem* about doubly stochastic matrices (see, e.g., [8, p. 549]) states that an $n \times n$ matrix is doubly stochastic if and only if it is a convex combination of some $n \times n$ permutation matrices. Considered as elements in $\mathbb{R}^{n^2}$, the $n \times n$ doubly stochastic matrices form a polytope (ω_n). The Birkhoff's theorem says that the polytope ω_n is the same as the polytope (δ_n) generated by the permutation matrices. A traditional proof of this result is to make use of a lemma which ensures that every doubly stochastic matrix has a positive diagonal (see, e.g., [8, Lemma 8.7.1, p. 548]). By a "diagonal of an n-square matrix" we mean a set of n entries taken from different rows and columns. The n-square doubly stochastic matrices having a positive diagonal form a polytope (l_n) too. Apparently, $\delta_n \subseteq l_n \subseteq \omega_n$. Birkhoff's theorem asserts that the three polytopes ω_n, l_n, and δ_n coincide.

Received Sep. 18, 2015; Accepted Nov. 26, 2015.
*Corresponding author.
2010 *Mathematics Subject Classification.* Primary 15B51; Secondary 52B11.
Keywords. doubly stochastic matrix, extreme point, polytope, stochastic semi-magic cube, stochastic tensor.

In this paper, we consider the counterpart of the Birkhoff theorem for higher dimensions. A multidimensional array of numerical values is referred to as a *tensor* (see, e.g., [9]). It is also known as a hypermatrix [10]. Let $A = (a_{ijk})$ be an $n\times n\times n$ tensor (or an n-tensor cube). We call A a *stochastic tensor* (see [6]), or a *stochastic semi-magic cube* [1], or simply a *stochastic cube* if all $a_{ijk} \geq 0$ and

$$\sum_{i=1}^{n} a_{ijk} = 1, \quad \forall j, k, \tag{1.1}$$

$$\sum_{j=1}^{n} a_{ijk} = 1, \quad \forall i, k, \tag{1.2}$$

$$\sum_{k=1}^{n} a_{ijk} = 1, \quad \forall i, j. \tag{1.3}$$

An n-tensor cube may be interpreted in terms of its slices (see [9]). By a "*slice* of a tensor A," we mean a 2-dimensional section of tensor A obtained by fixing any one of the three indices. For a 3-tensor cube $A = (a_{ijk})$, there are nine slices, each of which is a square matrix. An intersection of any two nonparallel slices is called a *line* (also known as *fiber* or *tube*). That is, a line is a 1-dimensional section of a tensor; it is obtained by fixing all but one of the indices. A *diagonal* of an $n\times n\times n$ tensor cube is a collection of n^2 elements such that no two lie on the same line. A nonnegative tensor is said to have a *positive diagonal* if all the elements of a diagonal are positive. We say such a tensor has the *positive diagonal property.*

The Birkhoff theorem is about the matrices that are 2-way stochastic, while stochastic cubes are 3-way stochastic. Related works on partial or multiple stochasticity such as line or face stochasticity are [3], [5], [7], including the recent ones [9] (on tensor computation), [10] (a survey chapter on tensors and hypermatrices), [6] (on extreme points of tensors), and [4] (a survey on the spectral theory of nonnegative tensors).

Let Ω_n be the set of all $n \times n \times n$ stochastic tensors (i.e., semimagic cubes). It is evident that Ω_n, regarded as a subset of $\mathbb{R}^{n^3}$, is convex and compact (since it is an intersection of finitely many closed half-spaces and it is bounded). It is not difficult to show that every permutation tensor (i.e., $(0, 1)$ stochastic tensor) is a vertex of Ω_n. Let Δ_n be the polytope generated (i.e., convex combinations of) by the $n \times n \times n$ permutation tensors and let L_n be the set of all $n \times n \times n$ stochastic tensors with the positive diagonal property. Obviously,

$$\Delta_n \subseteq L_n \subseteq \Omega_n.$$

If $n = 2$, a straightforward computation yields that $\Delta_n = L_n = \Omega_n$. For $n \geq 3$, it is known that each of the above inclusions is proper (see Section 3). Furthermore, the number of vertices of Δ_n is equal to the number of Latin squares of order n (see [6, Proposition 2.6]).

We assume that $n \geq 2$ throughout this article. We show a close relation between L_n and Ω_n, and especially the closure of L_n is Ω_n. Moreover, a lower bound for the number of the vertices of Ω_n is available in [1, p. 34]. In the present article, we give an upper bound for the number of vertices of Ω_n.

2. Vectorizing a cube

For an $m \times n$ matrix A with rows $r_1, \ldots, r_m$ and columns $c_1, \ldots, c_n$, let

$$\mathrm{vec}_r(A) = \begin{pmatrix} r_1^T \\ \vdots \\ r_m^T \end{pmatrix}, \qquad \mathrm{vec}_c(A) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

Vectorizing, or "vecing" for short, a matrix (with respect to rows or columns) is a basic method in solving matrix equations. It also plays an important role in computation. We present in this section how to "vec" a cube. This may be useful in tensor computations, which is a popular field currentty.

For $n \times n$ doubly stochastic matrices, we have the following fact by a direct verification. Let $e_n = (1, \ldots, 1) \in \mathbb{R}^n$. An $n \times n$ nonnegative matrix $S = (s_{ij})$ is doubly stochastic if and only if S is a nonnegative matrix satisfying

$$(I_n \otimes e_n)\, \mathrm{vec}_r(S) = e_n^T \qquad \text{and} \qquad (I_n \otimes e_n)\, \mathrm{vec}_c(S) = e_n^T.$$

When a tensor cube is interpreted in terms of slices (see [9, p. 458]), we see that each slice of a stochastic tensor is a doubly stochastic square matrix. An intersection of any two nonparallel slices is a *line* (fiber). An $n \times n \times n$ tensor cube has $3n^2$ lines. Considering each line as a column vector of n components, we stack all the lines in the order of i, j, and k directions (or modes), respectively, to make a column vector of $3n^3$ components. We call such a vector the "line vec" of the cube and denote it by $\mathrm{vec}_\ell(\cdot)$. Note that when vecing a 3rd-order n-dimensional tensor, every entry of the tensor is used three times.

For two cubes A and B of the same size and for any scalar α, we have

$$\mathrm{vec}_\ell(\alpha A + B) = \alpha\, \mathrm{vec}_\ell(A) + \mathrm{vec}_\ell(B)$$

and

$$\langle A, B \rangle = \frac{1}{3} \big\langle \mathrm{vec}_\ell(A), \mathrm{vec}_\ell(B) \big\rangle,$$

where the left inner product is for tensors while the right one is for vectors.

What follows is a characterization of a stochastic tensor through "vecing."

Theorem 2.1. *Let e_k be the all 1 row vector of k components, where k is a positive integer. An $n \times n \times n$ nonnegative cube $C = (c_{ijk})$ is stochastic if and only if*

$$(I_m \otimes e_n)\, \mathrm{vec}_\ell(C) = e_m^T \quad \textit{where } m = 3n^2.$$

Proof. The proof comes by a direct verification. □

3. The convex set L_n

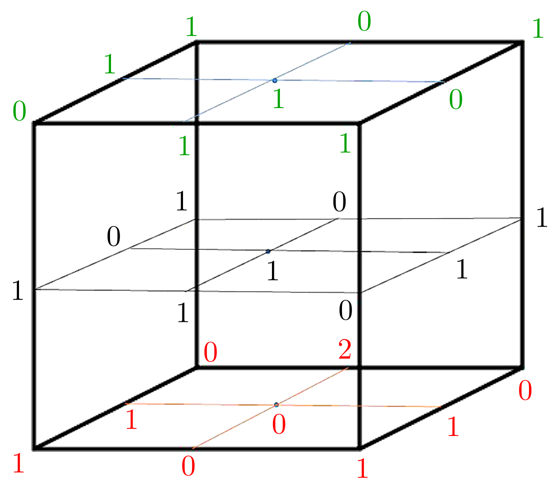
It is known that if $A = (a_{ij})$ is an $n \times n$ doubly stochastic matrix, then A has a positive diagonal; that is, there exist n positive entries of A such that no two of these entries are on the same row and same column (the positive diagonal property). Does the polytope of stochastic tensor cubes have the positive diagonal property? Let $A = (a_{ijk})$ be an $n \times n \times n$ stochastic cube. Is it true that there always exist n^2 positive entries of A such that no two of these entries lie on the same line? In short, does a stochastic tensor have the positive diagonal property?

It is easy to see that every permutation tensor is an extreme point of Ω_n. Apparently, the set of nonnegative tensors of the same size forms a cone; that is, if A and B (of the same size) have the positive diagonal property, then so are aA and $A + bB$ for any positive scalars a, b. Obviously, every permutation tensor possesses the positive diagonal property, and so does any convex combination of finitely many permutation tensors. However, some stochastic tensor cubes fail to have the positive diagonal property as the following example shows.

Note that Δ_n and Ω_n are convex and compact (in $\mathbb{R}^{n^3}$), while L_n is convex but not compact. In what follows, Example 3.1 shows that a stochastic tensor cube need not have a positive diagonal (unlike the case of doubly stochastic matrices); Example 3.2 shows a stochastic tensor cube where the positive diagonal property need not be generated by permutation tensors.

Example 3.1. Let E be the $3 \times 3 \times 3$ stochastic tensor cube:

1/2



which can be "flattened" to be a 3×9 matrix

$$\frac{1}{2}\begin{bmatrix} 0 & 1 & 1 & \vdots & 1 & 1 & 0 & \vdots & 1 & 0 & 1 \\ 1 & 1 & 0 & \vdots & 0 & 1 & 1 & \vdots & 1 & 0 & 1 \\ 1 & 0 & 1 & \vdots & 1 & 0 & 1 & \vdots & 0 & 2 & 0 \end{bmatrix}.$$

One may verify (by starting with the entry 2 at the position (3, 2, 3)) that E has no positive diagonal and E is not a convex combination of the permutation tensors. So $L_3 \subset \Omega_3$. (In fact, E is an extreme point of Ω_3; see, e.g., [1].)

Example 3.2. Taking the stochastic tensor cube F with the flattened matrix

$$\begin{bmatrix} 0 & 0.6 & 0.4 & \vdots & 1 & 0 & 0 & \vdots & 0 & 0.4 & 0.6 \\ 0.6 & 0 & 0.4 & \vdots & 0 & 0.4 & 0.6 & \vdots & 0.4 & 0.6 & 0 \\ 0.4 & 0.4 & 0.2 & \vdots & 0 & 0.6 & 0.4 & \vdots & 0.6 & 0 & 0.4 \end{bmatrix},$$

we see that $F \in L_3$ by choosing the positive elements $\times$ as follows:

$$\left[\begin{array}{ccccccccccc} & \times & & \vdots & \times & & & \vdots & & & \times \\ \times & & & \vdots & & & \times & \vdots & & \times & \\ & & \times & \vdots & & \times & & \vdots & \times & & \end{array}\right].$$

On the other hand, if F is written as $x_1P_1+\cdots+x_kP_k$, where all x_i are positive with sum 1, and each P_i is a permutation tensor of the same size, then each P_i has 0 as its entry at position (j_1, j_2, j_3) where $F_{j_1j_2j_3} = 0$; that is, every P_i takes the form

$$P_i = \left[\begin{array}{ccccccccccc} 0 & * & * & \vdots & 1 & 0 & 0 & \vdots & 0 & * & * \\ * & 0 & * & \vdots & 0 & \star & * & \vdots & * & * & 0 \\ * & * & * & \vdots & 0 & * & * & \vdots & * & 0 & * \end{array}\right].$$

There exists only one such permutation tensor with 0 in the $(2, 2)$ position $(\star)$ in the second slice. Therefore, $F \notin \Delta_3$. So the inclusions $\Delta_3 \subset L_3 \subset \Omega_3$ are proper.

Next we show that the closure of L_n is Ω_n and also that every interior point of Ω_n belongs to L_n. So L_n is "close" to Ω_n except for some points of the boundary.

Theorem 3.3. *The closure of the set of all $n \times n \times n$ stochastic tensors having a positive diagonal is the set of all $n \times n \times n$ stochastic tensors. In symbols,*

$$\mathrm{cl}(L_n) = \Omega_n.$$

Moreover, every interior point (tensor) of Ω_n has a positive diagonal. Consequently, a stochastic tensor that does not have the positive diagonal property belongs to the boundary $\partial\Omega_n$ of the polytope Ω_n.

Proof. Since $L_n \subseteq \Omega_n$, we have $\mathrm{cl}(L_n) \subseteq \Omega_n$. For the other way around, observe that every permutation tensor is in L_n. If $P, Q \in \Omega_n$, where P is a permutation tensor, then $tP+(1-t)Q$ belongs to L_n for any $0 < t \leq 1$. Thus, for any $Q \in \Omega_n$, if we set $t = \frac{1}{m}$, we get $\lim_{m\to\infty}(\frac{1}{m}P+(1-\frac{1}{m})Q) = Q$. This says that $\Omega_n \subseteq \mathrm{cl}(L_n)$. It follows that $\Omega_n = \mathrm{cl}(L_n)$.

We now show that every interior point of Ω_n lies in L_n. Let B be an interior point of Ω_n. Then there is an open ball, denoted by $\mathcal{B}(B)$, centered at B, inside Ω_n. Take a permutation tensor A, say, in Δ_n. Then $tA + (1-t)B \in L_n$ for any $0 < t \leq 1$. Suppose that the intersection point of the sphere $\partial\,\mathrm{cl}(\mathcal{B}(B))$ with the line $tA + (1-t)B$ is at C. Let C' be the corresponding point of C under the antipodal mapping with respect to the center B. Then B is between A and C', so B can be written as $sA + (1-s)C'$ for some $0 < s < 1$. By the above discussion, $B = sA + (1-s)C'$ is in L_n. That is, every interior point of Ω_n lies in L_n. □

4. An upper bound for the number of vertices

The Birkhoff polytope (i.e., the set of doubly stochastic matrices) is the convex hull of its extreme points—the permutation matrices. The Krein–Milman theorem (see, e.g., [11, p. 96]) states that every compact convex polytope is the convex hull of its vertices. A fundamental question of polytope theory is that of an upper (or lower) bound for the number of vertices (or even faces). Determining the number of vertices (and faces) of a given polytope is a computationally difficult problem in general (see, e.g., the texts on polytopes [2] and [11]).

Ahmed et al. [1, p. 34] gave a lower bound $(n!)^{2n}/n^{n^2}$ for the number of vertices (extreme points) of Ω_n through an algebraic combinatorial approach. We present an upper bound and our approach is analytic.

Theorem 4.1. *Let $v(\Omega_n)$ be the number of vertices of the polytope Ω_n. Then*

$$v(\Omega_n) \leq \frac{1}{n^3} \cdot \binom{p(n)}{n^3 - 1} \quad \text{where } p(n) = n^3 + 6n^2 - 6n + 2.$$

Proof. Considering Ω_n as defined by (1.1)–(1.3), we want to know the number of independent equations (lines) that describe Ω_n. For each horizontal slice (an $n \times n$ doubly stochastic matrix), $2n - 1$ independent lines are needed and sufficient. So there are $n(2n - 1)$ independent horizontal lines from n horizontal slices. Now consider the vertical lines: there are n^2 vertical lines. However, $(2n - 1)$ of them, say, on the most right and back, have been determined by the horizontal lines (as each line sum is 1). Thus, $n^2 - (2n - 1) = (n - 1)^2$ independent vertical lines are needed. So there are $n(2n-1)+(n-1)^2 = 3n^2-3n+1$ independent lines in total to define the tensor cube. It follows that we can view Ω_n defined by (1.1)–(1.3) as the set of all vectors $x = (x_{ijk}) \in \mathbb{R}^{n^3}$ satisfying

$$\sum_{i=1}^{n} x_{ijk} = 1, \quad 1 \leq j \leq n, 1 \leq k \leq n, \tag{4.1}$$

$$\sum_{j=1}^{n} x_{ijk} = 1, \quad 1 \leq i \leq n, 1 \leq k \leq n - 1, \tag{4.2}$$

$$\sum_{k=1}^{n} x_{ijk} = 1, \quad 1 \leq i \leq n - 1, 1 \leq j \leq n - 1, \tag{4.3}$$

$$x_{ijk} \geq 0, \quad 1 \leq i, j, k \leq n. \tag{4.4}$$

We may rewrite (4.1)–(4.3) and (4.4), respectively, as

$$Ax = u, \qquad Bx \geq 0,$$

where A is a $(3n^2 - 3n + 1) \times n^3$ $(0, 1)$ matrix, u is the all 1 column vector in $\mathbb{R}^{3n^2-3n+1}$, and B is an $n^3 \times n^3$ $(0, 1)$ matrix. Let $m = n^3$.

A subset of $\mathbb{R}^m$ is a convex hull of a finite set if and only if it is a bounded intersection of closed half-spaces (see [11, p. 29]). The polytope Ω_n is generated by the $p = n^3+6n^2-6n+2$ half-spaces defined by the linear inequalities $Ax \geq u$, $Ax \leq u$, and $Bx \geq 0$, $x \in \mathbb{R}^m$. Let e be a vertex of Ω_n. We claim that at least m equalities $he = 1$ or 0 hold, where h is a row of A or row of B, that is, $Ce = w$, where C is a $k \times n^3$ $(k \geq m)$ matrix consisting some rows of A and some rows of B, and w is a $(0, 1)$ column vector. If, otherwise, k equalities hold for $k < m$, let $K = \{x \in \mathbb{R}^m \mid Cx = w\}$. K is an affine space and $e \in K$. Since C is a $k \times m$ matrix, the affine space K has dimension at least $m - k \geq 1$. Let $O = \{x \in \mathbb{R}^m \mid B'x > 0\} \cap K$, where B' is a submatrix of B for which $B'e > 0$;

O is open in K. Since $e \in O$, e is an interior point of O, and thus it cannot be an extreme point of Ω_n. We then have that every extreme point e lies on at least m supporting hyperplanes $h(x) := hx = w$ in (4.1)–(4.4) that define Ω_n.

To show the upper bound, we use induction on n by reducing the problem to a polytope (a supporting hyperplane of Ω_n) of lower dimensions. Let V_m be the maximum value of the vertices of polytopes formed by any p supporting hyperplanes in $\mathbb{R}^m$. We show

$$V_m \leq \frac{1}{m}\binom{p}{m-1} = \frac{1}{n^3}\binom{n^3+6n^2-6n+2}{n^3-1}.$$

For $m = 8$ (i.e., $n = 2$), this is easy to check since Ω_2 has only two vertices. Assume that the upper bound inequality holds for the polytopes in the spaces $\mathbb{R}^k$, $k < m$. So Ω_n is formed by p supporting hyperplanes $H_t = \{x \mid h_t(x) = u\}$, $t = 1, \ldots, p$. Since H_t is a face of Ω_n, the vertices of Ω_n lying in H_t are the vertices of H_t. As H_t has smaller dimension than Ω_n (see [2, p. 32]) and it is formed by at most $p-1$ hyperplanes, by the induction hypothesis, H_t has at most V_{m-1} vertices, each of which lies in at least m hyperplanes. We arrive at

$$\begin{aligned} v(\Omega_n) &\leq \frac{1}{m}\sum_{t=1}^{p} v(H_t) \\ &\leq \frac{p}{m} \cdot V_{m-1} \\ &= \frac{p}{m} \cdot \frac{1}{m-1} \cdot \binom{p-1}{m-2} \\ &= \frac{1}{m}\binom{p}{m-1} \\ &= \frac{1}{n^3}\binom{n^3+6n^2-6n+2}{n^3-1}. \end{aligned}$$

□

Acknowledgments. Zhang expresses thanks to Richard Stanley, Zejun Huang, and Rajesh Pereira for drawing his attention to a number of references. All the authors would also like to express their thanks to the referee and to Zhongshan Li for some corrections and discussions. Chang's work was undertaken during the academic year 2014–2015, when she was Visiting Professor at Nova Southeastern University, and she thanks that institution for its hospitality.

Chang's work was partially supported by National Natural Science Foundation of China (NSFC) grant (11501363). Zhang's work was partially supported by NSFC grant (11571220).

References

1. M. Ahmed, J. De Loera, and R. Hemmecke, "Polyhedral cones of magic cubes and squares" in *Discrete and Computational Geometry*, Algorithms Combin. **25**, Springer, Berlin, 2003, 25–41. MR2038468. DOI 10.1007/978-3-642-55566-4_2. 387, 389, 391
2. A. Brondsted, *An Introduction to Convex Polytopes*, Grad. Texts in Math. **90**, Springer, New York, 1983. MR0683612. 391, 392

3. R. A. Brualdi and J. Csima, *Stochastic patterns*, J. Combin. Theory Ser. A **19** (1975) 1–12. Zbl 0304.05011. MR0439661. 387
4. K. Chang, L. Qi, and T. Zhang, *A survey on the spectral theory of nonnegative tensors*, Numer. Linear Algebra Appl. **20** (2013), no. 6, 891–912. Zbl pre06383880. MR3141883. DOI 10.1002/nla.1902. 387
5. J. Csima, *Multidimensional stochastic matrices and patterns*, J. Algebra **14** (1970), 194–202. Zbl 0206.02104. MR0289541. 387
6. L.-B. Cui, W. Li, and M. K. Ng, *Birkhoff–von Neumann theorem for multistochastic tensors*, SIAM. J. Matrix Anal. & Appl. **35** (2014), no. 3, 956–973. Zbl pre06381244. MR3231983. DOI 10.1137/120896499. 387
7. P. Fischer and E. R. Swart, *Three dimensional line stochastic matrices and extreme points*, Linear Algebra Appl. **69** (1985), 179–203. Zbl 0574.15013. MR0798372. DOI 10.1016/0024-3795(85)90075-8. 387
8. R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge Univ. Press, Cambridge, 2013. MR2978290. 386
9. T. G. Kolda and B. W. Bader, *Tensor Decompositions and Applications*, SIAM Rev. **51** (2009), no. 3, 455–500. Zbl 1173.65029. MR2535056. DOI 10.1137/07070111X. 387, 388
10. L.-H. Lim, "Tensors and hypermatrices" in *Handbook of Linear Algebra*, 2nd ed., Discrete Math. Appl. (Boca Raton), Chapman and Hall/CRC, Boca Raton, FL, 2013, Chapter 15. MR3013937. 387
11. G. M. Ziegler, *Lectures on Polytopes*, Grad. Texts in Math. **152**, Springer, New York, 1995. MR1311028. DOI 10.1007/978-1-4613-8431-1. 390, 391

[1]School of Statistics and Mathematics, Shanghai Finance University, Shanghai 201209, P.R. China.

E-mail address: hcychang@163.com

[2]Department of Mathematics, Nova Southeastern University, 3301 College Ave., Fort Lauderdale, FL 33314, USA.

E-mail address: vp80@nova.edu; zhang@nova.edu