A concise proof of Kruskal's theorem on tensor decomposition

John A. Rhodes[1]

Department of Mathematics and Statistics
University of Alaska Fairbanks
PO Box 756660
Fairbanks, AK 99775

Abstract

A theorem of J. Kruskal from 1977, motivated by a latent-class statistical model, established that under certain explicit conditions the expression of a third-order tensor as the sum of rank-1 tensors is essentially unique. We give a new proof of this fundamental result, which is substantially shorter than both the original one and recent versions along the original lines.

Key words: tensor rank, candecomp, parafac
2008 MSC: Primary 15A69, Secondary 15A72, 15A18

1. Introduction

In [11], J. Kruskal proved that, under certain explicit conditions, the expression of a third-order tensor (*i.e.*, a 3-way array) of rank r as a sum of r tensors of rank 1 is unique, up to permutation of the summands. (See also [9, 10].) This result contrasts sharply with the well-known non-uniqueness of expressions of matrices of rank at least 2 as sums of rank-1 matrices. The uniqueness of this tensor decomposition is moreover of fundamental interest for a number of applications, ranging from Kruskal's original motivation by latent-class models used in psychometrics, to chemistry and signal processing, as mentioned in [13] and its references. In these fields, the expression of a tensor as a sum of rank-1 tensors is often referred to as the Candecomp or Parafac decomposition. Recently, Kruskal's theorem has been used as a general tool for investigating the identifiability of a wide variety of statistical models with hidden variables [1, 2].

As noted in [13], Kruskal's original proof was "rather inaccessible," leading a number of authors to work toward a shorter and more intuitive presentation.

Email address: `j.rhodes@uaf.edu` (John A. Rhodes)

[1]The author thanks the National Science Foundation, for support from NSF grant DMS 0714830, and the Statistical and Applied Mathematical Sciences Institute, for support during a residency in which this work had its origins.

This thread, which continued to follow the basic outline of Kruskal's approach in which his 'Permutation Lemma' plays a key role, culminated in the proof given in [13]. Kruskal's original proof has undergone further streamlining in the manuscript [12]. In this paper, we present a new concise proof of Kruskal's theorem, Theorem 3 below, that follows a different approach. While the resulting theorem is identical, the alternative argument given here offers a new perspective on the role of Kruskal's explicit condition ensuring uniqueness.

While Kruskal's theorem gives a sufficient condition for uniqueness of a decomposition, the condition is in general not necessary. Of particular note are recent independent works of De Lathauwer [6] and Jiang and Sidiropoulos [8], which give a different criterion that can ensure uniqueness. These results require a stronger hypothesis than Kruskal's on one of the three sets of vectors used in the rank-1 tensors, but allow weaker assumptions on the others. See also [14] for the connection between these works.

It would, of course, be highly desirable to obtain conditions — more involved than Kruskal's — that would ensure the essential uniqueness of the expression of a rank r tensor as a sum of rank-1 tensors under a wider range of assumptions on the size and rank of the tensor. Note that both Kruskal's condition and that of [6, 8] can be phrased algebraically, in terms of the non-vanishing of certain polynomials in the variables of a natural parameterization of rank r tensors. This algebraic formulation allows one to conclude that *generic* rank r tensors of certain sizes have unique decompositions. (See [5] for more on generic aspects of tensor rank.) Having explicit understanding of these polynomial conditions is essential for certain applications, such as in [1]. The general problem of determining for which sizes and ranks of generic tensors the decomposition is essentially unique, and what explicit algebraic conditions can ensure uniqueness, remains open.

2. Notation

Throughout, we work over an arbitrary field.

For a matrix such as M_k, we use $\mathbf{m}_j^k$ to denote the jth column, $\bar{\mathbf{m}}_i^k$ to denote the ith row, and m_{ij}^k the (i,j)th entry. We use $\langle S \rangle$ to denote the span of a set of vectors S. With $[r] = \{1, 2, 3, \ldots, r\}$, we denote by $\mathfrak{S}_r$ the symmetric group on $[r]$.

Given matrices M_l of size $s_l \times r$, the matrix triple product $[M_1, M_2, M_3]$ is an $s_1 \times s_2 \times s_3$ tensor defined as a sum of r rank-1 tensors by

$$[M_1, M_2, M_3] = \sum_{i=1}^{r} \mathbf{m}_i^1 \otimes \mathbf{m}_i^2 \otimes \mathbf{m}_i^3,$$

so

$$[M_1, M_2, M_3](j, k, l) = \sum_{i=1}^{r} m_{ji}^1 m_{ki}^2 m_{li}^3.$$

A matrix A of size $t \times s_l$ acts on an $s_1 \times s_2 \times s_3$ tensor T 'in the lth coordinate.' For example, with $l = 1$

$$(A *_1 T)(i, j, k) = \sum_{n=1}^{s_1} a_{in} T(n, j, k),$$

so that $A *_1 T$ is of size $t \times s_2 \times s_3$. One then easily checks that

$$A *_1 [M_1, M_2, M_3] = [AM_1, M_2, M_3],$$

with similar formulas applying for actions in other coordinates.

Definition. The *Kruskal rank*, or *K-rank*, of a matrix is the largest number j such that *every* set of j columns is independent.

Definition. We say a triple of matrices (M_1, M_2, M_3) is *of type* $(r; a_1, a_2, a_3)$ if each M_i has r columns and the K-rank of M_i is at least $r - a_i$.

In a slight abuse of notation, we will say a product $[M_1, M_2, M_3]$ is of type $(r; a_1, a_2, a_3)$ when the triple (M_1, M_2, M_3) is of that type.

Note that with this definition, type $(r; a_1, a_2, a_3)$ implies type $(r; b_1, b_2, b_3)$ as long as $a_i \leq b_i$ for each i. Thus a_i is a bound on the gap between the K-rank of the matrix M_i and the number r of its columns. Intuitively, when the a_i are small it should be easier to identify the M_i from the product $[M_1, M_2, M_3]$.

3. The proof

We begin by establishing a lemma that generalizes a basic insight that has been rediscovered many times over the last half century, in which matrix diagonalizations arising from matrix slices of a third-order tensor are used to understand the tensor decomposition. A few such instances of the appearance of this idea include [3, 4], and other such references are mentioned in [7] where the idea is exploited for computational purposes. Note that [3] attributes an earlier occurrence to unpublished notes of P.F. Lazarsfeld.

Lemma 1. *Suppose* (M_1, M_2, M_3) *is of type* $(r; 0, 0, r-1)$*;* N_1, N_2, N_3 *are matrices with* r *columns; and* $[M_1, M_2, M_3] = [N_1, N_2, N_3]$*. Then there is some permutation* $\sigma \in \mathfrak{S}_r$ *such that the following holds:*

Let $\mathcal{I} \subseteq [r]$ *be any maximal subset (with respect to inclusion) of indices with the property that* $\langle \{\mathbf{m}_i^3\}_{i \in \mathcal{I}} \rangle$ *is 1-dimensional. Then*

1. $\langle \{\mathbf{m}_i^j\}_{i \in \mathcal{I}} \rangle = \langle \{\mathbf{n}_{\sigma(i)}^j\}_{i \in \mathcal{I}} \rangle$*, for* $j = 1, 2, 3$ *and*
2. $\mathcal{I}$ *is also maximal for the property that* $\langle \{\mathbf{n}_{\sigma(i)}^3\}_{i \in \mathcal{I}} \rangle$ *is 1-dimensional.*

Proof. That (M_1, M_2, M_3) is of type $(r; 0, 0, r-1)$ means M_1, M_2 have full column rank, and M_3 has no zero columns.

Choose some vector $\mathbf{c}$ that is not orthogonal to any of the columns of M_3, so that $\mathbf{c}^T M_3$ has no zero entries. (Such a vector may not exist with entries in a fixed finite field, but always does if we allow entries of $\mathbf{c}$ to be in the algebraic closure, for instance.) Then

$$A = \mathbf{c}^T *_3 [M_1, M_2, M_3] = [M_1, M_2, \mathbf{c}^T M_3] = M_1 \operatorname{diag}(\mathbf{c}^T M_3) M_2^T$$

is a matrix of rank r. Since

$$A = \mathbf{c}^T *_3 [N_1, N_2, N_3] = [N_1, N_2, \mathbf{c}^T N_3] = N_1 \operatorname{diag}(\mathbf{c}^T N_3) N_2^T,$$

N_1 and N_2 must also have rank r, and $\mathbf{c}^T N_3$ has no zero entries. These two expressions for A also show that the span of the columns of M_j is the same as that of the columns of N_j for $j = 1, 2$. Expressing the columns of M_j and N_j in terms of a basis given by the columns of M_j, we may henceforth assume $M_1 = M_2 = I_r$, the $r \times r$ identity, and N_1, N_2 are invertible. Thus $A = \operatorname{diag}(\mathbf{c}^T M_3)$.

Now let S_i denote the slice of $[M_1, M_2, M_3] = [N_1, N_2, N_3]$ with fixed third coordinate i, so S_i is an $r \times r$ matrix. Recalling that $\bar{\mathbf{m}}_i^j$ and $\bar{\mathbf{n}}_i^j$ denote the ith rows of M_j and N_j, we have

$$S_i = \operatorname{diag}(\bar{\mathbf{m}}_i^3) = N_1 \operatorname{diag}(\bar{\mathbf{n}}_i^3) N_2^T.$$

Note the matrices

$$S_i A^{-1} = \operatorname{diag}(\bar{\mathbf{m}}_i^3) \operatorname{diag}(\mathbf{c}^T M_3)^{-1} = N_1 \operatorname{diag}(\bar{\mathbf{n}}_i^3) \operatorname{diag}(\mathbf{c}^T N_3)^{-1} N_1^{-1},$$

for various choices of i, commute. Thus their (right) simultaneous eigenspaces are determined. But from the two expressions for $S_i A^{-1}$ we see its α-eigenspace is spanned by the set

$$\{\mathbf{e}_j = \mathbf{m}_j^1 \mid m_{i,j}^3 / (\mathbf{c}^T \mathbf{m}_j^3) = \alpha\},$$

and also by the set

$$\{\mathbf{n}_j^1 \mid n_{i,j}^3 / (\mathbf{c}^T \mathbf{n}_j^3) = \alpha\}.$$

A simultaneous eigenspace for the $S_i A^{-1}$ is thus spanned by the set $\{\mathbf{e}_j\}_{j \in \mathcal{I}}$ where $\mathcal{I}$ is a maximal set of indices with the property that if $j, k \in \mathcal{I}$, then

$$m_{i,j}^3 / (\mathbf{c}^T \mathbf{m}_j^3) = m_{i,k}^3 / (\mathbf{c}^T \mathbf{m}_k^3), \text{ for all } i.$$

This condition is equivalent to $\mathbf{m}_j^3$ and $\mathbf{m}_k^3$ being scalar multiples of one another. Such a set $\mathcal{I}$ is therefore exactly of the sort described in the statement of the lemma. As the simultaneous eigenspaces are also spanned by similar sets defined in terms of the columns of N_1, one may choose a permutation σ so that claim 2 holds, as well as claim 1 for $j = 1$.

The case $j = 2$ of claim 1 is similarly proved using the transposes of A and the S_i. As the needed permutation of the columns of the N_j in the two cases of $j = 1, 2$ is dependent only on the maximal sets $\mathcal{I}$, a common σ may be chosen. Finally, the case $j = 3$ follows from equating eigenvalues in the two expressions giving diagonalizations for $S_i A^{-1}$, to see that for all i

$$m_{i,j}^3 / \mathbf{c}^T \mathbf{m}_j^3 = n_{i,\sigma(j)}^3 / \mathbf{c}^T \mathbf{n}_{\sigma(j)}^3,$$

so $\mathbf{m}_j^3$ and $\mathbf{n}_{\sigma(j)}^3$ are scalar multiples of one another. □

Since the diagonalizations used in this argument were obtained using only matrix inversion and multiplication, we emphasize that no assumption that the field be algebraically closed is needed.

This lemma quickly yields a special case of Kruskal's theorem, when two of the matrices in the product are assumed to have full column rank.

Corollary 2. *Suppose* (M_1, M_2, M_3) *is of type* $(r; 0, 0, r-2)$*;* N_1, N_2, N_3 *are matrices with* r *columns; and* $[M_1, M_2, M_3] = [N_1, N_2, N_3]$*. Then there exists some permutation matrix* P *and invertible diagonal matrices* D_i *with* $D_1D_2D_3 = I_r$ *such that* $N_i = M_iD_iP$*.*

Proof. Since (M_1, M_2, M_3) is also of type $(r; 0, 0, r-1)$, we may apply Lemma 1. As in the proof of that lemma, we may also assume $M_1 = M_2 = I_r$. But M_3 has K-rank at least 2, so every pair of columns is independent. Therefore the maximal sets of indices in Lemma 1 are all singletons. Thus with P acting to permute columns by σ, the one-dimensionality of all eigenspaces shows there is a permutation P and invertible diagonal matrices D_1, D_2 with $N_i = M_iD_iP = D_iP$ for $j = 1, 2$.

Thus $[M_1, M_2, M_3] = [N_1, N_2, N_3]$ implies

$$[I_r, I_r, M_3] = [D_1P, D_2P, N_3] = [D_1, D_2, N_3P^T] = [I_r, I_r, N_3P^TD_1D_2],$$

which shows $M_3 = N_3P^TD_1D_2$. Setting $D_3 = (D_1D_2)^{-1}$, we find $N_3 = M_3D_3P$. ◻

We now use the lemma to give a new proof of Kruskal's Theorem 4a of [11] in its full generality. Note that the condition on the a_i stated in the following theorem is equivalent to Kruskal's condition that $(r-a_1)+(r-a_2)+(r-a_3) \geq 2r+2$. Kruskal's work also presents several variants of the Theorem that are slightly stronger but with more complicated assumptions, which are not considered here.

Theorem 3 (Kruskal, [11])**.** *Suppose* (M_1, M_2, M_3) *is of type* $(r; a_1, a_2, a_3)$ *with* $a_1+a_2+a_3 \leq r-2$*;* N_1, N_2, N_3 *are matrices with* r *columns, and* $[M_1, M_2, M_3] = [N_1, N_2, N_3]$*. Then there exists some permutation matrix* P *and invertible diagonal matrices* D_i *with* $D_1D_2D_3 = I_r$ *such that* $N_i = M_iD_iP$*.*

Proof. We need only consider $a_1 + a_2 + a_3 = r - 2$. We proceed by induction on r, with the case $r = 2$ (and 3) already established by Corollary 2. We may also assume $a_1 \leq a_2 \leq a_3$, We may furthermore restrict to $a_2 \geq 1$, since the case $a_1 = a_2 = 0$ is established by Corollary 2.

We first claim that it will be enough to show that, for some $1 \leq i \leq 3$, there is some set of indices $\mathcal{J} \subset [r]$, $1 \leq |\mathcal{J}| \leq r - a_i - 2$, and a permutation $\sigma \in \mathfrak{S}_r$ such that

$$\langle \{\mathbf{m}_j^i\}_{j\in\mathcal{J}} \rangle = \langle \{\mathbf{n}_{\sigma(j)}^i\}_{j\in\mathcal{J}} \rangle. \tag{1}$$

To see this, if there is such a set $\mathcal{J}$, assume for convenience $i = 1$ (the cases $i = 2, 3$ are similar), and the columns of M_i, N_i have been reordered so that

$\sigma = id$ and $\mathcal{J} = [s]$. Let Π be a matrix with nullspace the span described in equation (1). Then

$$[\Pi M_1, M_2, M_3] = \Pi *_1 [M_1, M_2, M_3] = \Pi *_1 [N_1, N_2, N_3] = [\Pi N_1, N_2, N_3].$$

But since the first s columns of ΠM_1 and ΠN_1 are zero, these triple products can be expressed as triple products of matrices with only $r - s$ columns. That is, using the symbol '$\widetilde{\ }$' to denote deletion of the first s columns,

$$[\Pi \widetilde{M}_1, \widetilde{M}_2, \widetilde{M}_3] = [\Pi \widetilde{N}_1, \widetilde{N}_2, \widetilde{N}_3].$$

For $i = 2, 3$, since M_i has K-rank $\geq r - a_i$, the matrix $\widetilde{M}_i$ has K-rank $\geq \min(r - a_i, r - s)$. Since the nullspace of Π is spanned by the first s columns of M_1, and M_1 has K-rank $\geq r - a_1$, one sees that $\Pi \widetilde{M}_1$ has K-rank $\geq r - s - a_1$, as follows: For any set of $r - s - a_1$ columns of $\Pi \widetilde{M}_1$, consider the corresponding columns of M_1, together with the first s columns. This set of $r - a_1$ columns of M_1 is therefore independent, so the span of its image under Π is of dimension $r - s - a_1$. This span must then have as a basis the chosen set of $r - s - a_1$ columns of $\Pi \widetilde{M}_1$, which are therefore independent. Thus $[\Pi \widetilde{M}_1, \widetilde{M}_2, \widetilde{M}_3]$ is of type $(r - s; a_1, b_2, b_3)$, where $b_i = \max(0, a_i - s)$ for $i = 2, 3$. Note also that $s \leq r - a_1 - 2$ implies $a_1 + b_2 + b_3 \leq r - s - 2$.

We may thus apply the inductive hypothesis to $[\Pi \widetilde{M}_1, \widetilde{M}_2, \widetilde{M}_3] = [\Pi \widetilde{N}_1, \widetilde{N}_2, \widetilde{N}_3]$, and, after an allowed permutation and scalar multiplication of the columns of the N_i, conclude that $\widetilde{M}_i = \widetilde{N}_i$ for $i = 2, 3$. But this means we can now take the set $\mathcal{J}$ described in equation (1) to be a singleton set $\{j\}$, with $j > s$, and $i = 2$. Again applying the argument developed thus far implies that, allowing for a possible permutation and rescaling, all but the jth columns of M_3 and N_3 are identical. As $\mathbf{m}_j^3 = \mathbf{n}_j^3$, this shows $M_3 = N_3$. Applying this argument yet again, with $i = 3$, and varying choices of j, then shows $M_1 = N_1$ and $M_2 = N_2$, up to the allowed permutation and rescaling. The claim is thus established.

We next argue that some set of columns of some M_i, N_i meets the hypotheses of the above claim.

Let Π_3 be any matrix with nullspace $\langle \{\mathbf{n}_i^3\}_{1 \leq i \leq a_1 + a_2} \rangle$, spanned by the first $a_1 + a_2$ columns of N_3. Let $\mathcal{Z}$ be the set of indices of all zero columns of $\Pi_3 M_3$. Since every set of $r - a_3 = a_1 + a_2 + 2$ columns of M_3 is independent, $|\mathcal{Z}| \leq a_1 + a_2$. Note also that at least 2 columns of $\Pi_3 M_3$ are independent, since the span of any $a_1 + a_2 + 2$ columns of $\Pi_3 M_3$ is at least 2 dimensional.

Let $\mathcal{S}_1, \mathcal{S}_2$ be any disjoint subsets of $[r]$ such that $|\mathcal{S}_1| = a_2$, $|\mathcal{S}_2| = a_1$, $\mathcal{Z} \subseteq \mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}$, and $\mathcal{S}$ excludes at least two indices of independent columns of $\Pi_3 M_3$. Let $\Pi_1 = \Pi_1(\mathcal{S}_1)$ be any matrix with nullspace $\langle \{\mathbf{m}_i^1\}_{i \in \mathcal{S}_1} \rangle$, and let $\Pi_2 = \Pi_2(\mathcal{S}_2)$ be any matrix with nullspace $\langle \{\mathbf{m}_i^2\}_{i \in \mathcal{S}_2} \rangle$.

Now consider

$$\begin{aligned}[\Pi_1 M_1, \Pi_2 M_2, \Pi_3 M_3] &= \Pi_3 *_3 (\Pi_2 *_2 (\Pi_1 *_1 [M_1, M_2, M_3])) \\ &= \Pi_3 *_3 (\Pi_2 *_2 (\Pi_1 *_1 [N_1, N_2, N_3]))) = [\Pi_1 N_1, \Pi_2 N_2, \Pi_3 N_3].\end{aligned}$$

By the specification of the nullspace of Π_3, the columns of all N_i with indices in $[a_1+a_2]$ can be deleted in this last product. In the first product, one can similarly delete the columns of the M_i with indices in $\mathcal{S}$, due to the specifications of the nullspaces of Π_1 and Π_2. Using '$\widetilde{\ }$' to denote the deletion of these columns, we have

$$[\Pi_1\widetilde{M_1}, \Pi_2\widetilde{M_2}, \Pi_3\widetilde{M_3}] = [\Pi_1\widetilde{N}_1, \Pi_2\widetilde{N}_2, \Pi_3\widetilde{N}_3], \tag{2}$$

where these products involve matrix factors with $r - a_1 - a_2 = a_3 + 2$ columns.

The matrix $\Pi_1\widetilde{M_1}$ in fact has full column rank. To see this, note that it can also be obtained from M_1 by (a) first deleting columns with indices in $\mathcal{S}_2$, then (b) multiplying on the left by Π_1, and finally (c) deleting the columns arising from those in M_1 with indices in $\mathcal{S}_1$. Since M_1 has K-rank at least $r - a_1$, step (a) produces a matrix with $r - a_1$ columns, and full column rank. Since the nullspace of Π_1 is spanned by certain of the columns of this matrix, step (b) produces a matrix whose non-zero columns are independent. Step (c) then deletes all zero columns to give a matrix of full column rank. Similarly, the matrix $\Pi_2\widetilde{M_2}$ has full column rank.

Noting that $\Pi_3\widetilde{M_3}$ has no zero columns since $\mathcal{Z} \subseteq \mathcal{S}$, we may thus apply Lemma 1 to the products of equation (2). In particular, we find that there is some $\sigma \in \mathfrak{S}_r$ with $\sigma([r] \smallsetminus \mathcal{S}) = [r] \smallsetminus [a_1 + a_2]$ such that if $\mathcal{I}$ is a maximal subset of $[r] \smallsetminus \mathcal{S}$ with respect to the property that $\langle\{\Pi_3\mathbf{m}_i^3\}_{i\in I}\rangle$ is 1-dimensional, then

$$\langle\{\Pi_j\mathbf{m}_i^j\}_{i\in\mathcal{I}}\rangle = \langle\{\Pi_j\mathbf{n}_{\sigma(i)}^j\}_{i\in\mathcal{I}}\rangle \tag{3}$$

for $j = 1, 2, 3$.

Since we chose $\mathcal{S}$ to exclude indices of two independent columns of $\Pi_3 M_3$, there will be such a maximal subset $\mathcal{I}$ of $[r] \smallsetminus \mathcal{S}$ that contains at most half the indices. We thus pick such an $\mathcal{I}$ with $|\mathcal{I}| \le \lfloor (r - a_1 - a_2)/2 \rfloor = \lfloor a_3/2 \rfloor + 1$, and consider two cases:

Case $a_1 = 0$**:** Then $\mathcal{S}_2 = \emptyset$, and Π_2 has trivial nullspace and thus may be taken to be the identity. Since $a_3 \ge a_2 \ge 1$, this implies $|\mathcal{I}| \le a_3 = r - a_2 - 2$. The sets $\{\mathbf{m}_i^2\}_{i\in\mathcal{I}}$ and $\{\mathbf{n}_{\sigma(i)}^2\}_{i\in\mathcal{I}}$ therefore satisfy the hypotheses of the claim.

Case $a_1 \ge 1$**:** Note that $|\mathcal{I}| + a_2 + 1 \le \lfloor a_3/2 \rfloor + a_2 + 2 < a_2 + a_3 + 2 = r - a_1$, so for any index k, the columns of M_1 indexed by $\mathcal{I} \cup \mathcal{S}_1 \cup \{k\}$ are independent. This then implies that for $j = 1$ the spanning set on the left of equation (3) is independent, so the spanning set on the right is as well. Thus the set $\{\mathbf{n}_{\sigma(i)}^1\}_{i\in\mathcal{I}}$ is also independent. Note next that equation (3) implies that, for $i \in \mathcal{I}$, there are scalars b_j^i, c_k^i such that

$$\mathbf{n}_{\sigma(i)}^1 - \sum_{j\in\mathcal{I}} b_j^i \mathbf{m}_j^1 = \sum_{k\in\mathcal{S}_1} c_k^i \mathbf{m}_k^1. \tag{4}$$

Now for any $p \in \mathcal{S}_1$, $q \in \mathcal{S}_2$, let

$$\mathcal{S}_1' = (\mathcal{S}_1 \smallsetminus \{p\}) \cup \{q\}, \quad \mathcal{S}_2' = (\mathcal{S}_2 \smallsetminus \{q\}) \cup \{p\}.$$

Choosing Π_1' and Π_2' to have nullspaces determined as above by the index sets $\mathcal{S}_1'$ and $\mathcal{S}_2'$, and applying Lemma 1 to $[\Pi_1' M_1, \Pi_2' M_2, \Pi_3 M_3] = [\Pi_1' N_1, \Pi_2' N_2, \Pi_3 N_3]$, similarly shows that for some permutation σ' and any $i' \in \mathcal{I}$ there are scalars $d_k^{i'}, f_k^j$ such that

$$\mathbf{n}^1_{\sigma'(i')} - \sum_{j\in\mathcal{I}} d_j^{i'}\mathbf{m}^1_j = \sum_{l\in\mathcal{S}_1'} f_l^{i'}\mathbf{m}^1_l. \tag{5}$$

Note that since the same Π_3 was used, the set $\mathcal{I}$ is unchanged here, and σ and σ' must have the same image on $\mathcal{I}$. Picking $i' \in \mathcal{I}$ so that $\sigma'(i') = \sigma(i)$, and subtracting equation (4) from (5) shows

$$\sum_{j\in\mathcal{I}} (b_j^i - d_j^{i'})\mathbf{m}^1_j = \sum_{k\in\mathcal{S}_1\smallsetminus\{p\}} (f_k^{i'} - c_k^i)\mathbf{m}^1_k + f_q^{i'}\mathbf{m}^1_q - c_p^i\mathbf{m}^1_p.$$

But since the columns of M_1 appearing in this equation are independent, we see that $f_q^{i'} = c_p^i = 0$. By varying p, we conclude that $\mathbf{n}^1_{\sigma(i)} \in \langle\{\mathbf{m}^1_i\}_{i\in\mathcal{I}}\rangle$. Thus $\langle\{\mathbf{n}^1_{\sigma(i)}\}_{i\in\mathcal{I}}\rangle \subseteq \langle\{\mathbf{m}^1_i\}_{i\in\mathcal{I}}\rangle$. Since both of these spanning sets are independent, and of the same cardinality, their spans must be equal. Since $|\mathcal{I}| \le r - a_1 - 2$, the set $\mathcal{I}$ satisfies the hypotheses of the claim.

$\square$

References

[1] Elizabeth S. Allman, Catherine Matias, and John A. Rhodes, *Identifiability of parameters in latent structure models with many observed variables*, Ann. Statist. **37** (2009), no. 6A, 3099–3132.

[2] Elizabeth S. Allman and John A. Rhodes, *The identifiability of covarion models in phylogenetics*, IEEE/ACM Trans. Comput. Biol. Bioinformatics **6** (2009), 76–88.

[3] T. W. Anderson, *On estimation of parameters in latent structure analysis*, Psychometrika **19** (1954), 1–10. MR MR0075492 (17,756e)

[4] Joseph T. Chang, *Full reconstruction of Markov models on evolutionary trees: identifiability and consistency*, Math. Biosci. **137** (1996), no. 1, 51–73. MR 97k:92011

[5] P. Comon, J. M. F. ten Berge, L. De Lathauwer, and J. Castaing, *Generic and typical ranks of multi-way arrays*, Linear Algebra Appl. **430** (2009), no. 11-12, 2997–3007. MR MR2517853

[6] Lieven De Lathauwer, *A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalizations*, SIAM J. Matrix Anal. Appl. **28** (2006), no. 3, 642–666.

[7] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, *Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition*, SIAM J. Matrix Anal. Appl. **26** (2004), no. 2, 295–327.

[8] Tao Jiang and Nicholas D. Sidiropoulos, *Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints*, IEEE Trans. Signal Processing **52** (2004), no. 9, 2625–2636.

[9] J. B. Kruskal, *Rank, decomposition, and uniqueness for* 3*-way and* N*-way arrays*, Multiway data analysis (Rome, 1988), North-Holland, Amsterdam, 1989, pp. 7–18. MR MR1088949

[10] Joseph B. Kruskal, *More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition and individual differences scaling*, Psychometrika **41** (1976), no. 3, 281–293. MR MR0488592 (58 #8119)

[11] ———, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra and Appl. **18** (1977), no. 2, 95–138. MR MR0444690 (56 #3040)

[12] J.M. Landsberg, *Kruskal's theorem*, (2009), preprint, `arXiv:0902.0543`.

[13] Alwin Stegeman and Nicholas D. Sidiropoulos, *On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition*, Linear Algebra Appl. **420** (2007), 540–552.

[14] Alwin Stegeman, Jos M.F. ten Berge, and Lieven De Lathauwer, *Sufficient conditions for uniqueness in Candecomp/Parafac and Indscal with random component matrices*, Psychometrika **71** (2006), 219–229.