Some Numerical Results on the Rank of Generic Three-Way Arrays over $\mathbb{R}$

Choulakian, V.
Université de Moncton, Moncton, N.B., E1A 3E9, Canada.
email: vartan.choulakian@umoncton.ca

1 June 2009

Abstract. The aim of this paper is the introduction of a new method for the numerical computation of the rank of a three-way array, $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$, over $\mathbb{R}$. We show that the rank of a three-way array over $\mathbb{R}$ is intimately related to the real solution set of a system of polynomial equations. Using this, we present some numerical results based on the computation of Gröbner bases.

Key words: Tensors; three-way arrays; Candecomp/Parafac; Indscal; generic rank; typical rank; Veronese variety; Segre variety; Gröbner bases.

AMS classification: Primary 15A69; Secondary 15A72, 15A18.

1. Introduction

Let $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$ be a tensor of order 3, sometimes named a three-way array or a three-mode data set. A rank 1 or a decomposed tensor is

$$\mathbf{D} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}, \tag{1}$$

where $\mathbf{a} \in \mathbb{R}^I$, $\mathbf{b} \in \mathbb{R}^J$ and $\mathbf{c} \in \mathbb{R}^K$, and $\otimes$ is the tensor product, sometimes named also outer product. $\mathbf{X}$ can be expressed as a sum of decomposed tensors given in (1),

$$\mathbf{X} = \sum_{\alpha=1}^{r} \mathbf{D}_\alpha. \tag{2}$$

The rank of $\mathbf{X}$ is defined to be the minimal integer r, see for instance Kruskal (1977, 1989). In data analysis, this implies that the rank of a three-way array is the smallest number of components that provide a perfect fit in Candecomp/Parafac (CP), see for instance, (Carroll and Chang, 1970, and Harshman, 1970). In statistics CP is considered a natural extension of singular value decomposition or principal components analysis to three-way data.

There is quite a literature concerning the value of maximal rank, generic rank or typical rank of three-way arrays in the area of statistics, algebraic complexity theory and algebraic geometry. Some references, among others, are: Ja' Ja' (1979), Kruskal (1977, 1983, 1989), Strassen (1983), Ten Berge (1991, 2000, 2004a, 2004b), Ten Berge and Kiers (1999), Ten Berge and Stegeman (2006), Comon and Ten Berge (2008), Bürgisser et al (1997), Catalisano et al (2002), Friedland (2008) and Abo et al. (2006). Friedland (2008) provides an up todate survey with some new results on the generic rank of three-way arrays.

First, we give the following

Definition 1: A dataset is called *generic* if its elements are randomly generated from a continuous distribution.

The generic and typical ranks are defined in the following way by Comon and Ten Berge (2008): Given that the rank of $I \times J \times K$ arrays is bounded, one can partition the arrays according to the rank values. Generic rank is defined to be true almost everywhere; while typical ranks are associated with the rank values that occur with positive probability. So, if there is a single typical rank, then it may be called generic rank; that is, a generic rank is typical, but the converse is not true.

Ten Berge (2000) classified three-way arrays into three classes: very tall, tall and compact. Let $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$ be a tensor of order 3 with $I \geq J \geq K$. The array $\mathbf{X}$ is called very tall when $I \geq KJ$; $\mathbf{X}$ is tall when $KJ - J < I \leq KJ - 1$; $\mathbf{X}$ is compact when $I \leq KJ - J$. The generic rank of the very tall arrays is very well known and easiest to prove: it is KJ. Ten Berge (2000) showed that all tall three-way arrays have generic rank I; and the tallest among the compact arrays, that is when $I = KJ - J$, have typical rank $\{I, I+1\}$. Ten Berge and Stegeman (2006) provided some further results on the compact case. Friedland (2008) showed that: typical rank($12 \times 4 \times 4$) $\geq$12, typical rank($11 \times 4 \times 4$) $\geq$11, and typical rank($I \times J \times K$) $\geq I$ for $(I, J, K) = ((J-1)^2+1, J, J)$ when $J \geq 2$. These results are all based on mathematical proofs. However, the rank computation problem has also been approached from a numerical point of view: Comon and ten Berge (2008) and Friedland (2008) applied Terracini's lemma, based on the numerical calculation of the maximal rank of the Jacobian matrix of (2), to obtain numerically the generic rank of some three-way arrays. The numerical method based on Terracini's lemma, when used to evaluate rank over $\mathbb{R}$, gives the generic rank when the typical rank is single-valued, and the smallest typical rank value otherwise.

Two well known facts are: a) There is no known method to calculate the rank of a given three-way dataset, Martin (2004, AIM tensor workshop); b) A three-way array over $\mathbb{R}$ may have a different rank than the same array considered over $\mathbb{C}$, (Kruskal, 1989).

We shall be concerned by the numerical computation of the rank of a three-way array over $\mathbb{R}$ only.

Computationally, the most primitive approach to the numerical evaluation of the typical rank of three-way arrays is based on the alternating least square (ALS) minimization algorithm: It is to run ALS many times to convergence on many generic three-way arrays of a given format, and to check whether or not the fit is perfect for a given number of components. But as a referee remarked, this approach has 2 problems: First, we do not know how many three-way arrays of a given format to examine before a valid inference can be drawn. For instance, when 100000 arrays have been examined and all seem to have the same rank α, it does not follow that α is indeed the generic rank for that array format. After all, a different rank may occur with an extremely small yet positive probability. Second, the decision of when to terminate ALS is hazardous, because even if the residual sum of squares is, say, $exp(-32)$, this does not prove that it is zero; in fact, it may have zero as infimum. The present paper relieves us from both above mentioned problems: It offers a straightforward method of determining the rank of any given array over $\mathbb{R}$, based on inspection of the number of real roots of a system of certain polynomial equations.

The real solution set of a system of polynomial equations is called semi-algebraic set in real algebraic geometry, see Basu, Pollack and Roy (2006) or Friedland (2008). Semi-algebraic sets are open sets and are composed of a finite union of connected components, where each component is called a basic semi-algebraic set. The main problem can be reformulated as: For a given tensor $\mathbf{X}$ over $\mathbb{R}$ calculate the number of connected components where each component is characterized by a unique real rank value. Our numerical results will shed some light on this. The numerical results on simulated datasets will be obtained by computing the Gröbner bases using Maple 12 of the system of polynomial equations characterizing the dataset. We note that generic datasets and random numbers are generated from integers between -99 and 99.

The paper is organized as follows. In section 2 we present the main lemma which provides a necessary and sufficient condition that a three-way array can be expressed as a sum of a fixed number of decomposed tensors. All results in this paper will be based on this lemma. In section 3, we show how the lemma can be applied to compute the rank of a generic tensor over $\mathbb{R}$ numerically for some cases. In section 4, we show another application of the lemma for the computation of rank for nongeneric particular datasets. In section 5, we show how the lemma can be applied to compute the rank of generic I symmetric $J \times J$ arrays, named INDSCAL arrays, over $\mathbb{R}$. And finally in section 6 we conclude.

2. Main Lemma

Let $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$ be a three-way dataset and $2 \leq K \leq J \leq I$. The lemma provides a necessary and sufficent condition that the tensor $\mathbf{X}$ can be expressed as a sum of I

decomposed tensors; that is

$$\begin{aligned} \mathbf{X} &= \sum_{\alpha=1}^{I} \mathbf{D}_\alpha, \\ &= \sum_{\alpha=1}^{I} \mathbf{a}_\alpha \otimes \mathbf{b}_\alpha \otimes \mathbf{c}_\alpha, \end{aligned} \tag{3}$$

where $\{\mathbf{a}_\alpha |\ \alpha = 1, ..., I\}$ is a basis for $\mathbb{R}^I$, $\mathbf{c}_\alpha \in \mathbb{R}^K$ and $\mathbf{b}_\alpha \in \mathbb{R}^J$. Note that if (3) is true, then $rank(\mathbf{X}) \leq I$. We denote by $\mathbf{X}_k \in \mathbb{R}^{I\times J}$ the kth slice in $\mathbf{X}$ for $k = 1, ..., K$.We note that (3) can be written as

$$\begin{aligned} \mathbf{X}_k &= \sum_{\alpha=1}^{I} c_{k\alpha} \mathbf{a}_\alpha \otimes \mathbf{b}_\alpha \quad \text{for} \quad k = 1, ..., K, \\ &= \mathbf{A}\mathbf{D}(\mathbf{c}_k)\mathbf{B}' \quad \text{for} \quad k = 1, ..., K, \end{aligned} \tag{4}$$

where $\mathbf{A} = (\mathbf{a}_1\ \mathbf{a}_2 ... \mathbf{a}_I) \in \mathbb{R}^{I\times I}$, $\mathbf{B} = (\mathbf{b}_1\ \mathbf{b}_2 ... \mathbf{b}_I) \in \mathbb{R}^{J\times I}$, $\mathbf{C} = (c_{k\alpha}) \in \mathbb{R}^{K\times I}$ and $\mathbf{D}(\mathbf{c}_k) = \mathbf{Diag}(\mathbf{c}_k) \in \mathbb{R}^{I\times I}$ is a diagonal matrix with diagonal elements $c_{k\alpha}$. Note that the vector $\mathbf{c}_k \in \mathbb{R}^I$ represents the kth row of $\mathbf{C}$.

We consider the system of polynomial equations

$$\mathbf{s}'_\alpha \mathbf{X}_k = c_{k\alpha} \mathbf{b}'_\alpha \quad \text{for} \ \ k = 1, ..., K \text{ and } \alpha = 1, ..., I, \tag{5}$$

where $\{\mathbf{s}_\alpha |\ \alpha = 1, ..., I\}$ is a basis for $\mathbb{R}^I$ and $\mathbf{c}_\alpha \in \mathbb{R}^K$, and $\mathbf{b}_\alpha \in \mathbb{R}^J$. We note that (5) can be written as

$$\mathbf{S}'\mathbf{X}_k = \mathbf{D}(\mathbf{c}_k)\mathbf{B}' \text{ for } k = 1, ..., K, \tag{6}$$

where $\mathbf{S}$ has columns $\mathbf{s}_\alpha$.

Lemma 1: (6) is a necessary and sufficient condition for (4).
Proof: Let $\mathbf{I} = \mathbf{A}\mathbf{S}'$, then (4) is true if and only if (6) is true.

Remark 1: a) To see if (5) is true, we solve the system of polynomial equations

$$\mathbf{s}'\mathbf{X}_k = c_k \mathbf{b}' \quad \text{for} \ \ k = 1, ..., K, \tag{7}$$

for $\mathbf{s} \in \mathbb{R}^I$, $\mathbf{b} \in \mathbb{R}^J$ and $\mathbf{c} \in \mathbb{R}^K$.

b) We note that (7) has two indeterminacies: It can be rewritten as $\mathbf{s}'_*\mathbf{X}_k = c_{k*}\mathbf{b}'_*$ for $k = 1, ..., K,$, where for instance, $\mathbf{s}_* = \lambda\mathbf{s}$ for any scalar $\lambda \neq 0$, $c_{k*} = \mu c_k$ for any scalar $\mu \neq 0$, and $\mathbf{b}_* = \lambda\mathbf{b}/\mu$. To eliminate these indeterminacies, hereafter, we fix

$$c_1 = 1 \quad \text{and} \quad s_I = 1. \tag{8}$$

c) Theorem 2.4 of Friedland (2008) provides another characterization for (3) or (4): It states that each slice $\mathbf{X}_k \in span(\mathbf{a}_1 \otimes \mathbf{b}_1, ..., \mathbf{a}_I \otimes \mathbf{b}_I)$ and the rank($\mathbf{X}$) equals the minimal dimension of the $span(\mathbf{a}_1 \otimes \mathbf{b}_1, ..., \mathbf{a}_I \otimes \mathbf{b}_I)$.

d) The necessary condition, when rank($\mathbf{X}$) $=I$, which was shown to be also sufficient afterwards, was used many times by Ten Berge and his coworkers, Ten Berge (2000), Ten Berge (2004a), and Ten Berge, Sidiropoulos and Rocci (2004).

3. Rank computation

We shall suppose in the sequel that $\mathbf{X} \in \mathbb{R}^{I\times J\times K}$ is a generic three-way array and $2 \leq K \leq J \leq I \leq KJ$. Then we have the following well known inequality: rank($\mathbf{X}$) $\geq I$. We will check if $\mathbf{X}$ has rank I. By the Main Lemma , the tensor $\mathbf{X}$ has rank I, if for parameter vectors $\mathbf{s} \in \mathbb{R}^I$, $\mathbf{b} \in \mathbb{R}^J$ and $\mathbf{c} \in \mathbb{R}^K$ the system of polynomial equations (7) subject to (8) have I real solutions $(c_{k\alpha}, \mathbf{b}_\alpha, \mathbf{s}_\alpha)$ for $\alpha = 1, ..., I$, such that the elements of the set $\{\mathbf{s}_\alpha |\ \alpha = 1, ..., I\}$ is a basis for $\mathbb{R}^I$; that is, (7) with (8) has I real isolated solutions. Let us see how can we know if this is true. The system of polynomial equations (7) with (8) is equivalent to

$$\mathbf{s}'(\mathbf{X}_k - c_k\mathbf{X}_1) = \mathbf{0}' \text{ for } \ k = 2, ..., K. \tag{9}$$

So the number of equations, neq, in (9) is

$$neq = (K-1)J, \tag{10}$$

and the number of degrees of freedom or the number of free variables , df, is

$$df = (I-1) + (K-1), \tag{11}$$

because of (8) there are $(K-1)$ free c_k 's and $(I-1)$ free s_i 's.

We are interested in the study of the number of solutions of (9) over $\mathbb{R}$ for *generic* data. We distinguish three cases named, minimal when $neq = df$, overdetermined when $df < neq$, and, underdetermined when $df > neq$. We note that Abo et al. (2006) also distinguished three cases that they named subabundant, superabundant and equiabundant: these were used for induction purposes.

3.1. Case 1: Minimal System($\mathbf{neq = df}$). When $I = (K-1)(J-1)+1$, $neq = df$, and the system (9) is called minimal. The number of real solutions is *bounded*; an upper bound is provided by Khovanskii's theorem, see Sturmfels (2002),

Theorem 1 (Khovanskii): Consider n polynomials in n variables involving m distinct monomials in total. The number of isolated roots in the positive orthant $(R_+)^n$ of any such system is at most $2^{\binom{m}{2}}(n+1)^m$.

In our case $n = neq = df = (K-1)J$ and $m = I-1+(K-1)(I-1) = K(I-1)$. The number of isolated roots in the positive orthant $(R_+)^{df}$ of any such system is at most $2^{\binom{m}{2}}(df+1)^m$ and m is the number of distinct monomials in the system (9). So (9) may or may not have I real isolated solutions. In case (9) has I real isolated solutions, then rank($\mathbf{X}$) $= I$; otherwise we embed it, which is discussed later on.

Example 1: $I \times I \times 2$ **arrays:** $neq = df = I$

This class of arrays is discussed in detail by Ten Berge (1991), who showed that the typical rank of such arrays is $\{I, I+1\}$. To check if the rank of a generic $I \times I \times 2$ array is I, it suffices to solve (9), which reduces to finding the real roots of the determinantal equation $det(\mathbf{X}_2 - c_2\mathbf{X}_1) = 0$. If $det(\mathbf{X}_2 - c_2\mathbf{X}_1) = 0$ has I real roots, then $rank(\mathbf{X}) = I$, otherwise $rank(\mathbf{X}) = I + 1$. Simulation results for 5000 generic $3 \times 3 \times 2$ arrays produced one real root 51.76% and 3 real roots 48.24% of the time. So we can deduce that Pr(rank ($3 \times 3 \times 2$ array) = 3) ≈ 48.24% and Pr(rank ($3 \times 3 \times 2$ array) = 4) ≈ 51.76%.

Example 2: $I \times J \times 3$ **arrays with** $I = 2J - 1$**:** $neq = df = 2J$

a) $5 \times 3 \times 3$ **arrays:** $neq = df = 6$. This class of arrays is also discussed in Ten Berge (2004a), where Ten Berge showed that generic $5 \times 3 \times 3$ arrays have either rank 5 or rank 6 with positive probability. Further, he showed that a closed form solution for the case when the array has rank 5 corresponds to finding the number of real roots of a sixth degree polynomial equation: if there are 6 real roots, then the array has rank 5, otherwise its rank is 6. Table 1 displays the number of real roots obtained by solving the system (9) for 1000 simulated generic arrays. First, we note that the solution set of (9) always admitted 6 roots, as expected according to Ten Berge (2004a); further, the number of real solutions is an even number or zero. Second, Pr(rank ($5 \times 3 \times 3$ array) = 5) ≈ 6.8% and Pr(rank ($5 \times 3 \times 3$ array) = 6) ≈ 93.2%.

Table 1: Simulation results for 1000 generic $5 \times 3 \times 3$ **arrays**.

real *roots*	0	2	4	6
counts	47	501	384	68

b) $7 \times 4 \times 3$ **arrays:** $neq = df = 8$. Table 2 displays the number of real roots obtained by solving the system (9) for 1000 simulated generic arrays. First, we note that the solution set of (9) always admitted 10 roots and the number of real solutions is an even number or zero. Second, Pr(rank ($7 \times 4 \times 3$ array) = 7) ≈ 4.2%.

Table 2: Simulation results for 1000 generic $7 \times 4 \times 3$ **arrays**.

real *roots*	0	2	4	6	8	10
counts	16	268	456	218	40	2

c) $9 \times 5 \times 3$ **arrays:** $neq = df = 10$. Table 3 displays the number of real roots obtained by solving the system (9) for 1000 simulated generic arrays. First, we note that the solution set of (9) always admitted 15 roots and the number of real solutions is an odd number. Second, Pr(rank $(9\times5\times3$ array$) = 9) \approx 6\%$ and Pr(rank $(9\times5\times3$ array$) = 10) \approx 94\%$. This latter result follows from Ten Berge (2000, Result 5) or see Example 5 describing tallest compact arrays, by embedding $9 \times 5 \times 3$ arrays into $10 \times 5 \times 3$ arrays.

Table 3: Simulation results for 1000 generic $9 \times 5 \times 3$ arrays.

real *roots*	1	3	5	7	9	11	13
counts	34	290	404	212	51	8	1

Example 3: $I \times J \times 4$ **arrays with** $I = 3J - 2$**:** $neq = df = 3J$

Numerical computations showed that $\neq$ (*roots* of $10\times4\times4$ arrays) = 20; $\neq$ (*roots* of $13 \times 5 \times 4$ arrays) = 35 and $\neq$ (*roots* of $16 \times 6 \times 4$ arrays) = 56. Table 4 shows that Pr(rank $(10 \times 4 \times 4$ array$) = 10) \approx 7.8\%$.

Table 4: Simulation results for 1000 generic $10 \times 4 \times 4$ arrays.

real *roots*	0	2	4	6	8	10	12	14
counts	2	78	284	342	216	58	14	6

Remark 2: a) To calculate a Gröbner basis for (9) in Example 2 for $I\times J\times 3$ arrays with $I = 2J - 1$, we used pure lexicographic order given by the following sequence $(s_1, ..., s_{I-1}, c_3, c_2)$ of the free variables. In all cases the Gröbner basis, denoted by G_β, consisted of $(K-1)J$ polynomials having the following form: $G_1(c_2) = 0$, $G_2(c_2, c_3) = \text{poly}_2(c_2) + c_3 = 0$, $G_{3,\alpha}(c_2, s_\alpha) = \text{poly}_\alpha(c_2) + s_\alpha = 0$ for $\alpha = 1, ..., I - 1$. It is important to note that this particular form of the Gröbner basis polynomials, G_β, shows that the degree of the polynomial $G_1(c_2) = 0$, denoted by $degG_1(c_2)$, represents the number of roots of the system (9). An introduction to Gröbner basis can be found in, among others, Cox et al. (2007). Example 6 show quite in detail the Gröbner basis application to a generic array.

b) The Maple 12 commands to do the computations in Example 2 are shown in Appendix 1.

c) For $I\times I\times 2$ arrays and $I \geq 2$, $det(\mathbf{X}_2 - c_2\mathbf{X}_1) = G_1(c_2) = 0$, where $G_1(c_2) = 0$ is the first element of the Gröbner basis. This phenomenon will be also seen for tallest compact arrays, see Examples 4, 5 and 6.

A reviewer noted that the right hand side of (7) is a Segre variety, which is the image of the Segre map, $\Sigma_{(K-1),(J-1)}$. The Segre map sends an element of the projective space $P^{(K-1)} \times P^{(J-1)}$ into P^{KJ-1}. While the left hand side of (7) is a linear space of projective dimension $I-1 = (K-1)(J-1)$. So, (7), will represent the intersection of the linear space with the Segre map, and the number of intersections is the degree of the Segre variety given in (12), see for instance Harris (1992, p. 233). This result is summarized in the following

Theorem 2: Let $I = (K-1)(J-1)+1$ and $2 \le K \le J \le I$, then for generic data the number of roots (real or complex) of the polynomial system (9) is

$$degG_1(c_2) = \binom{K-1+J-1}{K-1}. \tag{12}$$

Corollary 1: For minimal systems and $3 \le K \le J \le I$, $I < degG_1(c_2)$.

Proof: Let $n = J-1$ and $m = K-1$. We have to show that

$$mn + 1 \le \frac{(m+n)!}{n!m!} \quad \text{for} \quad 2 \le m \le n.$$

It is true for $m = 2$. For $m \ge 3$, we have

$$\begin{aligned} \frac{(m+n)!}{n!m!} &= \left[\frac{(n+1)}{m}\frac{(n+2)}{m-1}\cdots\frac{(n+m-2)}{3}\right]\left[\frac{(n+m-1)}{2}\frac{(n+m)}{1}\right] \\ &\ge \left[\frac{(n+m-1)}{2}\frac{(n+m)}{1}\right]. \end{aligned}$$

So, it is sufficient to show that $(n+m)(n+m-1) \ge 2(mn+1)$, which is easily seen to be true.

Corollary 2: The typical rank of arrays with a minimal system have more than one rank value and the minimum attained value is I.

Proof: The rank of a generic array with a minimal system is I, if the number of real roots of $G_1(c_2)$ is greater than or equal to I; otherwise its rank is greater than I.

We note that Corollary 2 generalizes Friedland (2008), who showed that: typical rank$(I \times J \times K) \ge I$ for $(I, J, K) = ((J-1)^2+1, J, J)$ when $J \ge 2$.

3.2. Case 2: Underdetermined System(df > neq).

When $(K-1)(J-1)+2 \le I \le IJ$, $df > neq$, and the system (9) is called underdetermined. The upper bound for the number of isolated roots of (9) is *infinity; so* (9) may or may not have I real isolated solutions: So the attained minimum bound for the rank of a generic three-way array is, $b_{\min} = I$. Before discussing two general classes studied in detail by Ten Berge (2000), we introduce some notation.

The system (9) can be written as

$$\mathbf{s}'\mathbf{\Gamma} = \mathbf{s}'\left[(\mathbf{X}_2 - c_2\mathbf{X}_1), (\mathbf{X}_3 - c_3\mathbf{X}_1), ..., \mathbf{X}_K - c_K\mathbf{X}_1\right] = \mathbf{0}', \tag{13}$$

where the number of columns of the matrix $\mathbf{\Gamma}$ is

$$n\,\mathrm{col}\,\mathbf{\Gamma} = (K-1)J, = neq; \tag{14}$$

and the number of rows of $\mathbf{\Gamma}$ is

$$n\,\mathrm{row}\,\mathbf{\Gamma}{=}I, \tag{15}$$

and $\mathbf{\Gamma}$ is a matrix function of the parameters $c_2, ..., c_K$. We also define

$$nbil = K - 1, \tag{16}$$

which represents the minimal number of c_k parameters that can be specialized to make the system of polynomial equations (13) linear. In algebraic geometry, the replacement of variables by specific values is called specialization.

Example 4: **Tall arrays:** $df - neq \geq nbil$

These are arrays when $(K-1)J < I \leq KJ$ and $I \geq J \geq K$, whose generic rank is I, as shown by Ten Berge (2000, Result 2). This implies that (15) $>$ (14), that is $I > (K-1)J$, or, $df - neq \geq nbil = K - 1$, where $nbil$ is given in (16). By assigning random values to the $(K-1)c_k$'s in (13), we reduce (13) to a system of linear equations, which will have a solution for any generic data; so (13) will admit I real and isolated solutions; from which we deduce that the generic rank of tall arrays is I.

Example 5: **Tallest compact arrays:** $n\,\mathrm{col}\,\mathbf{\Gamma} = n\,\mathrm{row}\,\mathbf{\Gamma}$ and $K \geq 3$

These are arrays when $I = J(K-1)$, $I \geq J \geq K$ and $K \geq 3$. Note that we exclude $I \times I \times 2$ arrays for $I \geq 2$ discussed in Example 1. Ten Berge (2000, Results 3, 4 and 5) discussed this case.

When $I = (K-1)J$ and $K \geq 3$, it implies that (14) = (15), that is, $\mathbf{\Gamma}$ is a square matrix. Solving (13) for c_k 's for $k = 2, ..., K$ is equivalent to solving $det(\mathbf{\Gamma})$ =0. The leading monomial in $det(\mathbf{\Gamma})$ =0 is $\prod_{k=2}^{K} c_k^J$. If J is an odd integer, then (13) will have infinite number of real solutions: Assign random continuous numbers to c_k 's for $k = 3, ..., K$, and solve for c_2. This corresponds to Result 5 in Ten Berge (2000), which states: When $I = J(K-1)$ and $I \geq J \geq K$ and $K \geq 3$ and J is odd, then the typical rank is I. If J is an even integer, then (13) may have infinite number of real solutions or finite number of real solutions or 0 real solution: For instance for $J = 4$ and $K = 3$, the polynomial $f(c_2, c_3) = 3c_2^4 c_3^4 + 1$ has 0 real solution, the polynomial $f(c_2, c_3) = 3c_2^4 c_3^4 - 1$ has infinite number of real and distinct solutions, and the polynomial $f(c_2, c_3) = 3c_2^4(c_3^4 - 1)$ has a finite number of real solutions. Ten Berge (2000) specifically discussed the case of $8 \times 4 \times 3$ arrays, where he stated that typical rank of such arrays is $\{8, 9\}$ and for randomly sampled data the rank of 9 is extremely rare. Similarly, Friedland(2008, Th.7.2) showed that typical rank of $12 \times 4 \times 4$ arrays has more than one value. We conducted a limited simulation study on generic $8 \times 4 \times 3$ and $12 \times 4 \times 4$ arrays; and each time we got I real isolated solutions. The simulation study was done in the following way: For a generic dataset let $f(c_2, c_3..., c_K) = det(\mathbf{\Gamma})$ =0; assign random values to the parameters $c_3..., c_K$, then solve for c_2. This shows that for generic data, when $I = J(K-1)$ and $K \geq 3$ the rank is I with very high probability. Also, see example 6.

3.3. Example 6. We consider a simulated generic dataset of size $7 \times 4 \times 3$ having the following three slices

$$\mathbf{X}'_1 := \begin{pmatrix} [-50, -38, -98, -93, -32, 8, 44] \\ [-22, -18, -77, -76, -74, 69, 92] \\ [45, 87, 57, -72, -4, 99, -31] \\ [-81, 33, 27, -2, 27, 29, 67] \end{pmatrix}$$

$$\mathbf{X}'_2 := \begin{pmatrix} [99, -25, 24, -61, 31, 25, 50] \\ [60, 51, 65, -48, -50, 94, 10] \\ [-95, 76, 86, 77, -80, 12, -16] \\ [-20, -44, 20, 9, 43, -2, -9] \end{pmatrix}$$

$$\mathbf{X}'_3 := \begin{pmatrix} [90, -82, 29, 52, 42, -62, 22] \\ [80, -70, 70, -13, 18, -33, 14] \\ [19, 41, -32, 82, -59, -68, 16] \\ [88, 91, -1, 72, 12, -67, 9] \end{pmatrix}$$

Our aim is to find the rank of $\mathbf{X}$, by representing it as in (6). This dataset has a minimal system of polynomial equations. We solve equation (9) via Gröbner basis using the lexicograhic order $(s_1, s_2, s_3, s_4, s_5, s_6, c_3, c_2)$. The first two polynomials of the Gröbner basis are

$G_1(c_2) = 0 =$
$-25879797508399905866360381811483872416558357311425629 4$
$-1987946767932180125365724555441379125561037244553323732{*}c_2$
$-7447583055793225423658520296174635567495579387052082486{*}c_2^2$
$+18477292423934054741969006645285810935999768448664319668{*}c_2^3$
$+16286857667624818445824550468864864953766140760544740734 4{*}c_2^4$
$+22324671325209561198922665813216562379249229294549244662{*}c_2^5$
$-93044594774454916354246852601811640731920664188422515202{*}c_2^6$
$+103499036526817564025434273115607968972407114567429495674 6{*}c_2^7$
$-139921510983826984867148291317620071655282519570059139007 5{*}c_2^8$
$+155346700794650490501115016130585172314320583287574900147{*}c_2^9$

$+6450726303789537576787170010007192178213154527772261680268{*}c_2^{10}$
$G_2(c_2, c_3) = 0 =$
23547338655791229204617338928506186026357940595072461145844743562575
17224332611787380814907080082015469486345955534220731803730371551646
24134092641462097641695580678924043455393253397537452244857196895154
46461325491822959772758622053185481177195343628678109510431558799243
57732629782484420615672656492158360322133109432630494581048689547
$-3097250590883846738286889878288351798304609282451856596410303546337$
36319437905844941144694584532179492363486873159383686238249006746894
40160319769886619561748249254067684300497233927857246816557510045452
29549030244795643979218209546091280718239589534262958680509337880753
$427578723122948642870383378005580770133313107270509303255691909320{*}c_2$
$-1903803609627374035621320978361604020106824484275906499213001947264$
95672033886734748658480731859038594412417938078895776898706276221196
35127294306367325389388675757732402447319340359820738619607197620136
52879058280068724250611547838816805478213287696041214145895488077794
$7039714261670007939310542108233323980869417122270525843479004885596{*}c_2^2$
$+8450512390839763624967161124974667624579807298666582130379782638299$
901903452364927022770162589839439882861673620370340791082608420840404
311331236303835885178287071688967159857905276468573727498262705288023
535530120799333316114432548548504549670733473680409194666521803651767
$1193123383965058588912264378749348068202790932902763246688455431{*}c_2^3$
$+4336608402971539347025237252135145455841035812345961936163213074515$
668913809930686458789021032092183097107068630380980654723975384806834
399547265746928595076966467194028602153464501847919685450547334048541
838603267464810745721708090718604776111130817877753185112995125825196
$9900571656166285598237774354646551889675112530575080545846831937{*}c_2^4$
$-29667122481091026364901639098151640003965888696689225008930393424968$
602525252447427436681754274345497424043473832230915504701237012993339
710812856574652111731271369111860472809751620103390836422947014561918
596742170800703187841128631115509897386033559201960217645649903785096
$9895485234230652379590298859552730627182446420246826454883883689{*}c_2^5$
$+8713668020397159975202209402993411331481527445489159653742816589285$
782666925647284363167235229251382288691898280884602035659684679625961
618445548964513132925721174946043787600439882519732665839121162038859
547938786254023880112832091100880202921766725381340769795219852772659
$31253523019517845556699435901635488455864886970958223801767254199{*}c_2^6$
$-78305675171958244526488263512192141094477411091651712476670327217351$
782883871707407108769516955504305407699176179301277874612313054094467

$$
\begin{aligned}
&597035177196760703985825236814433985312481352680953166404723733165962\\
&317279471444717834466650359718509004877505793494350037939473327528086\\
&086797222271563841516298832735291488845800360601997816074967 9848{*}c_2^7\\
&-5052259920358362162559563711521744883860734306629909564563700911 7677\\
&845425111468649945267084093503724529301981623817384478696550110017337\\
&119404751862738670370910467517021683484411445974804570961028420156557\\
&202505669968683614281388897157201250167813030135004835015786338672614\\
&037510989474695713072179459238022542102052991592155908191171478{*}c_2^8\\
&+3531570778002236274768723374593078025694670204694037564345870988381 5\\
&717091334531092687692414120601353572599659606631610854904141003914876\\
&375870980951760199224617620310327264953286419665073175139662877717976\\
&464011962351866981273603060052461996153251279426523988643704170224549\\
&350504160917950639530767935075745711987600896457583856972369 6328{*}c_2^9\\
&+2619606453792314898725984402386883999147230568960575124238685834365 4\\
&158387703055190547211883116255260309881480267447955878169277786134938\\
&237906112297170134620139058793181439502365320984021683968720825019581\\
&715365315571505452391632141389861154616280828170929049200565639622657\\
&407952496761241790311916420891554034603210659065541299430393 59{*}c_3
\end{aligned}
$$

The polynomials $G_{3,6}(s_6, c_2) = 0, G_{3,5}(s_5, c_2) = 0, ..., G_{3,1}(s_1, c_2) = 0$ have the same form as $G_2(c_2, c_3) = 0$ given above.

The polynomial $G_1(c_2) = 0$ has only four real roots, which are: -1.871987136, -0.3332612900, -0.2556946431, 0.2733107997; so the rank of the dataset is greater than 7. We embed it by joining the following vectors to the three slices: $\mathbf{v}_1' = (1 \;\; 0 \;\; 0 \;\; 0)$, $\mathbf{v}_2 = \mathbf{v}_3 = \mathbf{0}$. The embedded dataset is $\mathbf{X}_1^e = (\mathbf{X}_1' \;\; \mathbf{v}_1')'$, $\mathbf{X}_2^e = (\mathbf{X}_2' \;\; \mathbf{v}_2')'$ and $\mathbf{X}_3^e = (\mathbf{X}_3' \;\; \mathbf{v}_3')'$. The rank of the embedded dataset will be calculated by two distinct methods.

First, for the embedded dataset we see that $n \,\mathrm{col}\, \mathbf{\Gamma} = n \,\mathrm{row}\, \mathbf{\Gamma} = 8$, so we can calculate the determinant of $\mathbf{\Gamma}$ as in Example 5, which is:

$$
\begin{aligned}
&det(\mathbf{\Gamma}) = 0 = 111296195967997{*}c_2^4 - 163212875913821{*}c_2^3 - 288078435761246{*}c_2^3{*}c_3\\
&+188384423078426{*}c_2^2 + 139757151961919{*}c_2^2{*}c_3 - 123835533958927{*}c_2^2{*}c_3^2\\
&+3188520736473{*}c_2 + 1745777654358{*}c_2{*}c_3 + 145702375007129{*}c_2{*}c_3^2\\
&+154156258186696{*}c_2{*}c_3^3 - 30068441704134{*}c_3 - 78231890782721{*}c_3^2\\
&-9292669314727{*}c_3^3 + 24148992371016{*}c_3^4
\end{aligned}
$$

Following the argument in Example 5, we note that there is a slight possibility that there will not be eight distinct values of $(\widetilde{c_2}, \widetilde{c_3})$ such that the $det(\mathbf{\Gamma}) = 0$, because it is of degree 4. So, in general, following this approach of computing we can not assert that typical rank of generic $7 \times 4 \times 3$ arrays is $\{7, 7+1\}$. However, let us continue our computation as in Example 5. We obtain the $\mathbf{C}$ matrix of Lemma 1,rounded to 2 decimal digits,

$$\mathbf{C} := \begin{pmatrix} [1,1,1,1,1,1,1,1] \\ [5.68, 63.84, 0.6, 44.31, 3.49, 9.33, 3.93, 21.29] \\ [-38, 91, -1, 63, -23, -63, -26, 30] \end{pmatrix}$$

where the third row of $\mathbf{C}$ represents the randomly generated eight $\widetilde{c_3}$ values, and the second row represents the corresponding eight $\widetilde{c_2}$ values obtained by solving the $det(\mathbf{\Gamma})$ =0 after plugging the $\widetilde{c_3}$ values in it.

Now, we can obtain the $\mathbf{S}$ matrix of Lemma 1 by solving

$$\mathbf{s}'(\mathbf{X}_k^e - \widetilde{c_k}\mathbf{X}_1^e) = \mathbf{0}' \text{ for } \; k = 2, 3 \tag{17}$$

eight times: The ith column of $\mathbf{S}$ corresponds to the eigenvector of the matrix $\mathbf{\Gamma}(\widetilde{c_{2i}}, \widetilde{c_{3i}})$ associated with the unique null eigenvalue

$$\mathbf{S} := 10^{-3}\times \begin{pmatrix} [-6.85,\ -6.54,\ -4.89,\ -6.51,\ 7.02,\ 6.74,\ 6.97,\ 6.44] \\ [-4.02,\ 5.45,\ 8.94,\ 5.44,\ 3.85,\ 4.11,\ 3.90,\ 5.40] \\ [9.29,\ -5.99,\ 14.8,\ -5.97,\ -9.49,\ -9.16,\ -9.43,\ -5.91] \\ [-16.4,\ -6.87,\ -34.4,\ -6.83,\ 16.7,\ 16.2,\ 16.6,\ -6.71] \\ [-4.31,\ 0.995,\ -6.31,\ 1.01,\ 4.15,\ 4.40,\ 4.20,\ 1.08] \\ [-11.5,\ -5.37,\ -43.7,\ -5.36,\ 11.8,\ 11.4,\ 11.7,\ -5.32] \\ [-3.22,\ -6.50,\ -6.24,\ -6.50,\ 3.20,\ 3.24,\ 3.20,\ -6.52] \\ [-1000,\ -1000,\ -1000,\ -1000,\ -1000,\ -1000,\ -1000,\ -1000] \end{pmatrix}$$

The matrix $\mathbf{B}' = \mathbf{S}'\mathbf{X}_1 = \mathbf{A}^{-1}\mathbf{X}_1$ of Lemma 1 is

$$\mathbf{B} := 10^{-2}\times \begin{pmatrix} [1.06,\ -1.47,\ 22.7,\ -2.11,\ -1.84,\ -0.625,\ -1.61,\ -4.37] \\ [-2.25,\ -1.24,\ -171,\ -1.79,\ 3.73,\ 1.35,\ 3.30,\ -3.72] \\ [2.49,\ -0.077,\ -22.7,\ -0.111,\ -4.22,\ -1.48,\ -3.70,\ -0.228] \\ [3.85,\ -0.280,\ -69.4,\ -4.00,\ -6.39,\ -2.31,\ -5.65,\ -0.806] \end{pmatrix}$$

And $\mathbf{A}' = \mathbf{S}^{-1}$; finally, we obtain $\widetilde{\mathbf{X}}_k = \mathbf{A}\mathbf{D}(\mathbf{c}_k)\mathbf{B}' = \mathbf{X}_k$. This was numerically verified.

A second approach to compute the matrices $\mathbf{S}$, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ is via the Gröbner basis for the embedded system (17) using the lexicographic order $(s_1, s_2, s_3, s_4, s_5, s_6, c_3)$; note that c_2 is a free variable. The first Gröbner basis polynomial is

$G_1(c_2, c_3) = 0 = 111296195967997{*}c_2^4{-}163212875913821{*}c_2^3{-}288078435761246{*}c_2^3{*}c_3$
$+188384423078426{*}c_2^2{+}139757151961919{*}c_2^2{*}c_3{-}123835533958927{*}c_2^2{*}c_3^2$
$+3188520736473{*}c_2{+}1745777654358{*}c_2{*}c_3{+}145702375007129{*}c_2{*}c_3^2$
$+154156258186696{*}c_2{*}c_3^3{-}30068441704134{*}c_3{-}78231890782721{*}c_3^2$
$-9292669314727{*}c_3^3{+}24148992371016{*}c_3^4,$

which equals $det(\mathbf{\Gamma})$. This shows that both approaches are identical for this particular problem.

4. Another Application of The Main Lemma

Consider **nongeneric** dataset of size $4 \times 4 \times 3$

$$\mathbf{X}_1 := \begin{pmatrix} [-872410, 509152, -155756, 301976] \\ [-669515, 355308, -105576, 215236] \\ [349983, -898362, 265770, -79182] \\ [3285, -185950, 180998, 97398] \end{pmatrix}$$
$$\mathbf{X}_2 := \begin{pmatrix} [-403995, 481229, 24054, 201485] \\ [-243133, 337616, -4344, 94484] \\ [317091, -174294, -2454, -206076] \\ [-317457, 112640, 183938, 289254] \end{pmatrix}$$
$$\mathbf{X}_3 := \begin{pmatrix} [-274447, 214327, -280750, 108851] \\ [-252456, 116912, -145020, 92016] \\ [-127464, -713802, 599526, 54318] \\ [-38790, -204608, 236662, 21168] \end{pmatrix}$$

To see if the rank of $\mathbf{X}$ is 4, we solve the system (9) composed of 8 polynomial equations in 5 variables via Gröbner basis using the pure lexicograhic order $(s_1, s_2, s_3, c_3, c_2)$. The elements of the Gröbner basis are

$G_1(c_2) = 0 = -266104 + 1131869c_2 + 1855673c_2^2 - 10091484c_2^3 + 3934656c_2^4$;

$G_2(c_2, c_3) = 0 = 70150154675210213 - 61657878275323159c_2 - 70078073768841556 8c_2^2 + 308891236767911424c_2^3 + 628616789525725c_3$;

$G_{3,3}(c_2, s_3) = 0 = -790119586832666088459321810985 57 + 377170833747374430067039542008 86c_2 + 901077269210427745705210304730192c_2^2 - 390331538460948950190867958454016c_2^3 + 10995656448714576020137029824 55s_3$;

$G_{3,2}(c_2, s_2) = 0 = 3935310306421428023441642821994 9 - 190977009897456095062042069799807c_2 + 209064381129999539517569775784236c_2^2 - 57940861139941085694575004742848c_2^3 + 549782822435728801006851491227 5s_2$;

$G_{3,1}(c_2, s_1) = 0 = -292815752925406189572563201863 16 - 395996753150175561163175671614 7c_2 + 390527303469098244882389504161956c_2^2 - 165798278803217428934162052760128c_2^3 + 10995656448714576020137029824 55s_1$.

The polynomial $G_1(c_2) = 0$ is of degree 4 and it has four real roots, which are: $-.3369565217, .2929292929, .2962962963, 2.312500000$. So the rank of the dataset is 4 by the main Lemma. Such datasets have been characterized by their defining equations in Landsberg and Manivel (2006).

5. INDSCAL arrays

Let $\mathbf{X} \in \mathbb{R}^{I \times J \times J}$ be a tensor of order 3,where the ith slice $\mathbf{X}_i \in \mathbb{R}^{J \times J}$ for $i = 1, ..., I$ is symmetric. INDSCAL, proposed by Carroll and Chang (1970), is a statistical method used in psychometrics to analyse such arrays. For this reason, we shall name such an array an INDSCAL array to distinguish it from a general three-way array $\mathbf{Y} \in \mathbb{R}^{I \times J \times K}$ discussed above, where such a decomposition is usually named

PARAFAC, see Harshman (1970). A rank 1 INDSCAL array or a decomposed tensor is

$$\mathbf{D} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{b}, \tag{18}$$

where $\mathbf{a} \in \mathbb{R}^I$ and $\mathbf{b} \in \mathbb{R}^J$.

The following theoretical results are known for generic INDSCAL data $\mathbf{X} \in \mathbb{R}^{I\times J\times J}$: a) By Zellini (1979), see also Rocci and Ten Berge (1994), if $I \geq J(J+1)/2$, then $\text{rank}(X) = J(J+1)/2$. b) $I \times 2 \times 2$ and $I \times 3 \times 3$ arrays are studied by Ten Berge, Sidiropoulos and Rocci (2004). The rank computation problem has also been approached from a numerical point of view by Comon and ten Berge (2008), who applied applied Terracini's lemma, based on the numerical calculation of the maximal rank of the Jacobian matrix of (2), to obtain numerically the generic rank of some INDSCAL three-way arrays. The numerical method based on Terracini's lemma, when used to evaluate rank over $\mathbb{R}$, gives the generic rank when the typical rank is single-valued, and the smallest typical rank value otherwise.

For INDSCAL data (7) becomes

$$\mathbf{s}'\mathbf{X}_k = b_k\mathbf{b}' \quad \text{for} \quad k = 1, ..., J, \tag{19}$$

for $\mathbf{X}_k \in \mathbb{R}^{I\times J}, \mathbf{s} \in \mathbb{R}^I$ and $\mathbf{b} \in \mathbb{R}^J$.

We note that (19) has two indeterminacies, scale and sign: It can be rewritten as $\widetilde{\mathbf{s}}'\mathbf{X}_k = \widetilde{b}_k\widetilde{\mathbf{b}}'$ for $k = 1, ..., J,$, where for instance, $\widetilde{\mathbf{s}} = \lambda\mathbf{s}$ for any scalar $\lambda > 0$ and $\widetilde{\mathbf{b}} = \lambda^{1/2}\mathbf{b}$. The second indeterminacy is the sign indeterminacy of $\mathbf{b}$: replacing $\mathbf{b}$ by $-\mathbf{b}$ in (19) does not change the equality in (19).To eliminate both indeterminacies, hereafter, we fix

$$b_1 = 1. \tag{20}$$

We will represent the set of solutions of (19) subject to (20) by V (Veronese variety).

We are interested in the study of the number of solutions of (19) subject to (20) over $\mathbb{R}$ for *generic* INDSCAL data for $2 \leq J, I \leq J(J+1)/2$. We distinguish three cases named, minimal when $I = 1 + J(J-1)/2$, overdetermined when $I > 1 + J(J-1)/2$, and, underdetermined when $I < 1 + J(J-1)/2$. The overdetermined systems is similar to the one discussed above.

Theorem 3 (minimal system=Veronese variety): Let $I = 1 + J(J-1)/2$ and $2 \leq J \leq I$, then for generic INDSCAL data the number of roots (real or complex) of the polynomial system (19) is

$$degV = 2^{J-1}. \tag{21}$$

Proof: Let $[b_1, ..., b_J]$ be an element of the projective space $P^{(J-1)}$. We note that the right hand side of (19) is a Veronese variety of degree $d = 2$, which is the image of the Veronese map, ν_2, defined by

$$\nu_2 : P^{(J-1)} \to P^N,$$

by sending

$$[b_1, ..., b_J] \to [b_1^2, b_1 b_2, ..., b_J b_{J-1}, b_J^2],$$

where the image has $N + 1 = \binom{J-1+2}{2}$ elements composed of binomials in $b_1, ..., b_J$. While the left hand side of (19) is a general linear space of projective dimension $I - 1$. The number of intersections of the general linear space with the Veronese varity is finite, when $I - 1 = N - (J - 1)$; that is

$$I = 1 + J(J - 1)/2. \tag{22}$$

When (22) is true, the finite number of intersections is the degree of the Veronese variety given in (21), see for instance Harris (1992, p. 231).

Corollary 1: The typical rank of INDSCAL arrays with a minimal system have more than one rank value and the minimum attained value is I.

Proof: For minimal systems and $2 \leq J \leq I$, $I \leq degV$. The rank of a generic INDSCAL array with a minimal system is I, if the number of real roots of V is greater than or equal to I; otherwise its rank is greater than I.

5.1. Example 7. We consider a simulated generic dataset of size $4 \times 3 \times 3$ having the following 4 slices

$$\mathbf{X}_1 := \begin{pmatrix} [54, 107, 161] \\ [107, 58, 13] \\ [161, 13, 134] \end{pmatrix} \mathbf{X}_2 := \begin{pmatrix} [114, -49, -125] \\ [-49, -144, -76] \\ [-125, -76, -8] \end{pmatrix}$$

$$\mathbf{X}_3 := \begin{pmatrix} [-44, 7, -48] \\ [7, -36, -11] \\ [-48, -11, -154] \end{pmatrix} \mathbf{X}_4 := \begin{pmatrix} [50, 92, -4] \\ [92, 100, 1] \\ [-4, 1, -100]] \end{pmatrix}$$

INDSCAL $4 \times 3 \times 3$ arrays have been studied in detail by Ten Berge, Sidiropoulos, and Rocci (2004), where it is shown that if a certain polynomial of degree 4 has 4 real roots, then $rank(\mathbf{X}) = 4$, otherwise the rank is 5.

The Gröbner basis with pure lexicographic order given by the following sequence $(b_1, b_2, s_1, s_2, s_3, s_4)$ of the free variables is formed of 6 polynomials listed below. The first polynomial $G_4(s_4) = 0$ is of degree 4, as shown by ten Berge, Sidiropoulos and Rocci (2004) and Theorem 3, and it has 2 real roots -0.1881015674e-2, 0.7632125093e-1, so the rank of the dataset is greater than 4.

$G_4(s_4)$ =73376693603417736544444527-42937278193692708586613457 68*s_4
-2118632967757969942338642095 76*s_4^2-148692057913121404650687471 4272*s_4^3
+6572869203364733498016667374 8496*s_4^4,
$G_3(s_3,s_4)$ =1756097391380257390467471580317590062736668311394805 4931
-16122325737716095255145266890166937889196573572800540 0638*s_4
-58376960723805941081594102401861541155954316153630799 26796*s_4^2
+13183979237017459316244442359314659795257569739741011 83796472*s_4^3
+52388060785251915671654412347200942895791534199521523 73008*s_3,
$G_2(s_2,s_4)$ =96282398253033709933602079934711919657694784300073 85299
+11068252823433558277754489464864186098568340630108319 931766*s_4
+70599500802293105120029293272762762401144855060418803 1890116*s_4^2
-91182081299629927366092221532631872387972172839379925 54624760*s_4^3
+20955224314100766268661764938880377158316613679808609 492032*s_2,
$G_1(s_1,s_4)$ =-9384923940854434492010502183000235854601288296806481877
-19648227009019956225893987147505154519035126270041471 23640*s_4
+19141224660419798871046145092036847545567187777703796 83036*s_4^2
+68954053727665256715278346278701263570661246053781939 2520176*s_4^3
+26194030392625957835827206173600471447895767099760761 86504*s_1,
$G_{2,4}(b_2,s_4)$ =-155028823048701914654384720617223421617571775035134363431
-66211886029016478638439782073183667186014332650277499 221046*s_4
-33249808278806029901377730579858953393318745063721372 13739628*s_4^2
+44167907819442655873419676087304190072683931413772512 430408456*s_4^3
+13097015196312978917913603086800235723947883549880380 93252*b_2,
$G_1(b_1,s_4)$ =-151481990958486610814337256773617960880927702617415439 6973
-34540431627437701624090279595654955772656754239689727 3802586*s_4
-76966498746097488396981151215439423133970629245095948 73506300*s_4^2
+16134289335517885269973132508442992832462033369653936 5418650824*s_4^3
+19050203921909787516965240853527615598469648799826008 62912*b_1

5.2. Example 8. We consider a simulated generic INDSCAL dataset of size $7 \times 4 \times 4$ having the following 7 slices

$$\mathbf{X}_1 := \begin{pmatrix} [140, 86, -110, -4] \\ [86, -182, 70, 36] \\ [-110, 70, 104, 183] \\ [-4, 36, 183, 148] \end{pmatrix} \mathbf{X}_2 := \begin{pmatrix} [-20, 100, 173, -56] \\ [100, 128, 101, 75] \\ [173, 101, 124, 65] \\ [-56, 75, 65, -158] \end{pmatrix}$$

$$\mathbf{X}_3 := \begin{pmatrix} [178, 15, -186, 52] \\ [15, 196, 119, -148] \\ [-186, 119, -138, 43] \\ [52, -148, 43, -110] \end{pmatrix} \mathbf{X}_4 := \begin{pmatrix} [-8, -137, 21, 20] \\ [-137, -60, 64, 5] \\ [21, 64, -24, -14] \\ [20, 5, -14, -128] \end{pmatrix}$$

$$\mathbf{X}_5 := \begin{pmatrix} [194, -164, -36, -6] \\ [-164, 2, -114, -110] \\ [-36, -114, 64, -127] \\ [-6, -110, -127, -18] \end{pmatrix} \mathbf{X}_6 := \begin{pmatrix} [-22, 74, 85, -40] \\ [74, -198, 23, -53] \\ [85, 23, -152, 18] \\ [-40, -53, 18, -48] \end{pmatrix}$$

$$\mathbf{X}_7 := \begin{pmatrix} [-94, 109, -16, 90] \\ [109, 124, 164, -93] \\ [-16, 164, 98, -134] \\ [90, -93, -134, 184] \end{pmatrix}$$

The Gröbner basis with pure lexicographic order given by the following sequence $(s_1, ..., s_7, b_1, b_2, b_3)$ of the free variables is formed of 10 polynomials, but only the first one is shown below. The first polynomial $G_3(b_3) = 0$ is of degree 8, as shown in Theorem 3, and it has 2 real roots -4.615952848, 1.035693119, so the rank of the dataset is greater than 7.

$G_3(b_3)$ =-267319790697212354162205439965563724346086890209668628287611296
-418573483979735109514695930195818332286961955303805928337210144*b_3
-53224562968122644847846329140305933491773608156555442814188832*b_3^2
+190260260230311283947025128614232688395842607775283676963072480*b_3^3
+172806709139583658797792038234309205181892588602072553465939944*b_3^4
+164461253658569745584828839860332360925297521683057843681458288*b_3^5
+95090716874891491062104108972298040778579595301835996945880112*b_3^6
+24575461542028106507015735163996805821952192308876843008456228*b_3^7
+2240887382441309183839416634048576470976843441637962999441259*b_3^8

6. Conclusion

We introduced a new method to compute ranks of three-way arrays, by showing that it is intimately related to the solution set of a system of polynomial equations, which is a well developed and active area of mathematics known as algebraic geometry. The new method was used to compute numerically the ranks of some sizes of three-way arrays over ℝ via Gröbner basis.

The problem of computing the rank of overdetermined systems by solving embedded polynomial systems is a work in progress.

Acknowledgements

This research was financed by the Natural Sciences and Engineering Research Council of Canada. The author thanks Pr. J. Ten Berge and an unknown reviewer for their constructive comments, and Marc Sormany for help in programming.

Appendix 1

Below the matrix $\mathbf{Y}k = \mathbf{X}_k'$.

> K := 3;

```
> J := 8;
> I := 15;
> with(LinearAlgebra);
> Y1 := RandomMatrix(J, I);
> Y2 := RandomMatrix(J, I);
> Y3 := RandomMatrix(J, I);
> S := Vector(1 .. I, 1);
> for h from 1 to I-1 do
  S[h] := s_h end do;
> M1 := Y2-c_2*Y1;
> P1 := M1.S;
> M2 := Y3-c_3*Y1;
> P2 := M2.S;
> poly := [seq(P1[l], l = 1 .. J), seq(P2[n], n = 1 .. J)];
> with(Groebner);
> liste := seq(s_h, h = 1 .. I-1);
> polyG := Basis(poly, plex(liste, c_3, c_2));
> polyG[1];
> S := [solve(polyG[1])];
> nops(S);
> fsolve(polyG[1]);
> nops([fsolve(polyG[1])]);
> fsolve(polyG[1], a, complex);
> nops([fsolve(polyG[1], c_3, complex)]);
```

References

Abo, H., Ottaviani, G., Peterson, C. (2006). Induction for secant varieties of Segre varieties. arXiv:math.AG/0607191 v3, 2 August.

Basu, S., Pollack, R. and Roy, M-F. (2006). Algorithms in Real Algebraic Geometry. Springer, Berlin, 2nd edition.

Bürgisser, P., Clausen, M., Shokrollahi, M.A. (1997). Algebraic Complexity Theory, with the collaboration of Thomas Lickteig. *Grundlehren der Mathematischen Wissenschaften*, 315. Springer-Verlag, Berlin.

Carroll, J.D. and Chang J.J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, 35, 283-319.

Catalisano, M.V., Geramita, A.V., Gimigliano, A. (2002). Ranks of tensors, secant varieties and fat points. *Linear Algebra Appli.* 355, 263-285. Erratum, *Linear Algebra Appli.* 367 (2003), 347-348.

Comon, P. and ten Berge, J. (2008). Generic and typical ranks of three-way arrays. arXiv:0802.2371v1[cs.OH], 17 February.

Cox, D., Little, J. and O'Shea, D. (2007). Ideals, Varities, and Algorithms. Third edition, Springer, N.Y.

Friedland, S. (2008). On the generic rank of 3-tensors. ArXiv:0805.3777v2[math.AG], 27 May.

Harris, J. (1992). *Algebraic Geometry: A first course.* Springer, N.Y.

Harshman, R. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16, 1-84.

Ja' Ja', J. (1979). Optimal evaluation of pairs of bilinear forms. *SIAM Journal of Computing*, 8, 443-462.

Kruskal, J.B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18, 95-138.

Kruskal, J.B. (1983). Statement of some current results about three-way arrays. Unpublished manuscript, AT&T Bell Laboratories, Murray Hill, N.J, downloadable from http://www.leidenuniv.nl/fsw/three-mode/biblogr/biblio_k.htm

Kruskal, J.B. (1989). Rank, decompostion, and uniqueness for 3-way and N-way arrays. In R. Coppi & S. Bolasco (Eds.), *Multiway data Analysis* (pp. 7-18). Amsterdam: North Holland.

Landsberg, J.M. and Manivel, L. (2006). Generalizations of Strassen's equations for secant varieties of Segre varieties. ArXiv:math/0601097v1[math.AG], 5 January.

Martin,M C. (2004). Tensor Decompositions Workshop Discussion Notes, American Institute of Mathematics (AIM), available online.

Rocci, R. and Ten Berge, J. (1994). A simplification of a result by Zellini on the maximal rank of symmetric three-way arrays. *Psychometrika*, 59, 377-380.

Strassen, V. (1983). Rank and optimal computations of generic tensors. *Linear Algebra Appl.* 52/53, 645-685.

Sturmfels, B. (2002). *Solving Systems of Polynomial Equations.* CBMS Reg. Conf. Ser. in Mathematics, 97, AMS.

Ten Berge, J.M.F. (1991). Kruskal's polynomial for $2 \times 2 \times 2$ arrays and a generalization to $2 \times n \times n$ arrays. *Psychometrika*, 56, 631-636.

Ten Berge, J.M.F. (2000). The typical rank of tall three-way arrays. *Psychometrika*, 65, 525-532.

Ten Berge, J.M.F. (2004a). Partial uniqueness in CANDECOMP\PARAFAC. *Journal of Chemometrics*, 18, 12-16.

Ten Berge, J.M.F. (2004b). Simplicity and typical rank of three-way arrays, with applications to Tucker-3 analysis with simple cores. *Journal of Chemometrics*, 18, 17-21.

Ten Berge, J.M.F. and Kiers, H.A.L. (1999). Simplicity of core arrays in three-way principal component analysis and the typical rank of $P \times Q \times 2$ arrays. *Linear Algebra its Applications*, 294, 169-179.

Ten Berge, J.M.F. and Stegeman, A. (2006). Symmetry transformations for square sliced three-way arrays, with applications to their typical rank. *Linear Algebra its Applications*, 418 (1), 215-224.

Ten Berge, J.M.F., Sidiropoulos, N.D. and Rocci, R. (2004). Typical rank and INDSCAL dimensionality for symmetric three-way arrays of order $I \times 2 \times 2$ or $I \times 3 \times 3$. *Linear Algebra its Applications*, 388, 363-377.

Zellini, P. (1979). On the optimal computation of a set of symmetric and persymmetric bilinear forms. *Linear Algebra its Applications*, 23, 101-119.

/end