Rank properties for centred three-way arrays

Casper Albers, John Gower, and Henk Kiers

Accepted for publication in:
F. Mola, C. Conversano, M. Vichi (eds),
Classification, (Big) Data Analysis and Statistical Learning,
Studies in Classification, Data Analysis, and Knowledge Organization, Springer, 2017

Abstract When analysing three-way arrays, it is common practice to centre the arrays. Depending on the context, centring is performed over one, two or three modes. In this paper, we outline how centring affects the rank of the array; both in terms of maximum rank and typical rank.

Key words: three-way analysis, multiway analysis, maximum rank, typical rank, Candecomp/Parafac

1 Introduction

Let $\underline{\mathbf{X}}$, of dimension $I \times J \times K$, be a three-way array (also termed a tensor) with entries x_{ijk}. For sake of simplicity we assume that $I \leq J \leq K$ (whenever this is not the case we can make this the case without loss of generality by simply permuting the labels of the array).

In the analysis of arrays, the concept of rank is of importance, for the same reasons why it is important in the analysis of a two-way data matrix. The rank of a matrix is the dimension of the vector space spanned by its columns, i.e. the maximum number of distinct components the array can be decomposed into. For arrays,

Casper Albers
Department of Psychometrics & Statistics, University of Groningen, Groningen, The Netherlands, e-mail: `c.j.albers@rug.nl`, `www.casperalbers.nl`

John Gower
Department of Mathematics & Statistics, The Open University, Milton Keynes, United Kingdom e-mail: `john.c.gower@open.ac.uk`

Henk Kiers
Department of Psychometrics & Statistics, University of Groningen, Groningen, The Netherlands, e-mail: `h.a.l.kiers@rug.nl`

the concept of rank is similar, but now for three dimensions rather than two. (See Section 2 for details.)

In this paper, we study the consequences of centring, over either one, two or three modes, on the rank of the array. Centring three-way arrays is common practice in data analysis; similar to the centring of data matrices prior to performing a principal components analysis.

One should distinguish different types of pre-scaling data. One purpose of pre-scaling is (i) to reduce the effects of incommensurabilities in different parts of the data, or transformations to more acceptable measures such as logs or square-roots, but another is (ii) to isolate different substantive components which deserve separate examination. Normalisation in Principal Component Analysis is an example of (i), while removing the mean is an example of (ii). In this paper we are concerned with (ii) and note that the separate components of analysis not only enhance interpretation but may also reduce rank. Thus, although centring is usually performed solely to improve model fit, e.g. of a Candecomp/Parafac or Tucker3 decomposition, it is important to realise that centring can have a substantive effect. In the analysis of additive models, especially when studying interactions [1, 2], it is common to partition $\underline{\mathbf{X}}$ into parts for the overall mean, main effects, biadditive effects and triadditive effects:

$$x_{ijk} = m + \{a_i + b_j + c_k\} + \{d_{jk} + e_{ik} + f_{ij}\} + g_{ijk} \tag{1}$$

where the terms with a single suffix represent main effects, those with double suffices two factor interactions and g_{ijk} represents contributions from three factor interactions. Some components of the interactions may be regarded as "error". The defining equations are subsumed in the identity:

$$\begin{aligned} \hat{x}_{ijk} = x_{...} &+ \{(x_{i..} - x_{...}) + (x_{.j.} - x_{...}) + (x_{..k} - x_{...})\} \\ &+ \{(x_{.jk} - x_{.j.} - x_{..k} + x_{...}) + (x_{i.k} - x_{i..} - x_{..k} + x_{...}) \\ &+ (x_{ij.} - x_{i..} - x_{.j.} + x_{...})\} \\ &+ (x_{ijk} - x_{.jk} + x_{i.k} + x_{ij.} + x_{i..} + x_{.j.} + x_{..k} - x_{...}) \end{aligned} \tag{2}$$

where the expressions in parentheses in (2) estimate the corresponding parameters in (1).

The triadditive model for given choices of $P \le I, Q \le J, R \le K$ and S is given by

$$x_{ijk} = m + a_i + b_j + c_k + \sum_{p=1}^{P} d_{jp}\tilde{d}_{kp} + \sum_{q=1}^{Q} e_{iq}\tilde{e}_{kq} + \sum_{r=1}^{R} f_{ir}\tilde{f}_{jr} + \sum_{s=1}^{S} g_{is}\tilde{g}_{js}\tilde{\tilde{g}}_{ks} + \varepsilon_{ijk} \tag{3}$$

(By taking $S = 0$, one obtains the biadditive model.) To make this model identifiable, zero-sum identification constraints are required. Without such constraints, exactly the same fit would be obtained if, e.g., a non-zero value ε was added to all a_i and subtracted from all b_j. Requiring zero-sums is in line with the concept of marginality [12], i.e. first fitting an overall effect, then main effects on the residuals, then biadditive effects on the residiuals, and so on. In biadditive models, zero-sum constraints are straightforward, but this is not the case in triadditive models since for triadditive models, some forms of centring change the form of the model. One consequence is that the least-squares estimates of the triadditive interaction parameters depend on how exactly, i.e. by how many components, each of the biadditive terms is modelled [2, 7]. To bypass these issues, one may fit the triadditive part conditional on the main effects and the saturated biadditive components of the model. That is, we fit the triadditive part of the model to the biadditive residual table:

$$z_{ijk} = x_{ijk} - x_{\cdot jk} - x_{i\cdot k} - x_{ij\cdot} + x_{i\cdot\cdot} + x_{\cdot j\cdot} + x_{\cdot\cdot k} - x_{\cdot\cdot\cdot}. \tag{4}$$

Triadditive interactions in (3) may be modelled using a truly triadic model such as the Candecomp algorithm [6], minimising

$$\sum_{i,j,k,r} (z_{ijk} - a_{ir} b_{jr} c_{kr})^2 \tag{5}$$

(see next Section).

Thus, centring over one or two modes, can be seen as taking out main effects or two-way interactions, respectively, and analyse them separately. It is important to wonder whether it is sensible for the problem at hand to perform the chosen type of centring. In the words of [11]: ‘It is important that the final model or models should make sense physically: at a minimum, this usually means that interactions should not be included without main effects nor higher-degree polynomial terms without their lower-degree relatives.’

In this paper, we study the effect of various types of centring on the rank of three way arrays. This paper is organised as follows. In Section 2 we establish notation and recall relevant definitions from literature. Section 3 hosts the main theorem on the rank properties of centred arrays. We conclude with a series of examples in Section 4.

2 Notation and known results

We adhere to the standardised notation and terminology as proposed by [8]. The mode A matricised version of $\underline{\mathbf{X}}$ is given by the $I \times JK$ matrix $\mathbf{X}_a$ with all vertical fibers of a three-way array collected next to each other. Mode B and Mode C matricised versions are defined in analogous ways. The vectorisation operator vec implies column-wise vectorisation and $\otimes$ is used for the Kronecker product. Furthermore, array $\underline{\mathbf{G}}$ is the so-called superidentity core array with elements $g_{pqr} = 1$ if $p = q = r$ and $g_{pqr} = 0$ otherwise. Finally, $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ and $\mathbf{1}$ are column vectors with all values either 0 or 1, respectively, all of accommodating size.

There is a considerable literature on the ranks of general three-way arrays, summarised by [4], [13, Section 2.6] and [9, Section 8.4]. There are two types of rank to be considered: maximum rank and typicial rank.

Definition 1. The *maximum rank* of three-way array $\mathbf{X}$, with dimension $I \times J \times K$, is defined as the smallest value of R that can give exact fit for

$$\sum_{i,j,k=1}^{I,J,K} \sum_{r=1}^{R} \left(z_{ijk} - a_{ir} b_{jr} c_{kr}\right)^2 . \tag{6}$$

Definition 2. The *typical rank* is defined by [4, p.3] as follows: "The typical rank of a three-way array is the smallest number of rank-one arrays that have the array as their sum, when the array is generated by random sampling from a continuous distribution."

An earlier definition of typical rank by [10] is given in a more complicated way [4], but on [10, p.96] (bottom paragraph) seems to converge to Ten Berge's definition. So we follow the latter one. Since typical rank can be smaller than maximal rank (see [4] for examples), it will be of more practical usefulness than maximal rank, as this already provides a practical upper-bound to the number of components one wants to decompose the array in.

When J is small (close to I), the rank of $\underline{\mathbf{X}}$ is less than the upper bound K but it seems to coincide with the upper bound when $K \geq IJ$. These results are less simple than those for matrices, but have in common more concern with good low-rank approximations to (6) rather than with the rank itself. The three-way interaction in (4) is free both of main effects and of two-way interactions, and so all its margins are null. Thus, the three-way table $\mathbf{Z} = \{z_{ijk}\}$ is a special form of a triadditive table and it may be expected to have special properties. In particular, we may expect it to have lower triadditive rank than for unconstrained triadditivity. Also, when only some of the modes are centred, the rank is expected to be reduced. A formal result that establishes this expectation, is given in the following section.

3 Main result

Theorem 1. *Let the class of real-valued three-way arrays $I \times J \times K$ have at most maximum rank $f(I,J,K)$, where $f(I,J,K)$ denotes a particular function of I, J, and K. Then, a three-way array obtained by centring an array from this class of arrays will have rank at most equal to $f(I^*,J^*,K^*)$, where the starred versions denote $(I-1)$ or I, $(J-1)$ or J, $(K-1)$ or K, respectively, depending on whether or not the array has been centred across the first, second and/or third mode respectively.*

Before we prove Theorem 1, we make three remarks.

Remark 1. It should be mentioned that [5, p. 375] already mentioned that double centring symmetric matrices "has a rank-reducing impact on the symmetric array" and they give a concise proof for that. The above Theorem follows the same reasoning as [5] but gives a more general result.

Remark 2. We conjecture that the analogous theorem where "maximal rank" is replaced by "typical rank" also holds. For several classes of arrays of size $I \times J \times K$, the typical rank has been given as a function $f(I,J,K)$ of I, J and K, and our conjecture is that, like for the maximal rank, upon centring the array across the first, second and/or third mode, the typical rank should be given by $f(I^*,J^*,K^*)$, where the starred versions denote $(I-1)$ or I, $(J-1)$ or J, and $(K-1)$ or K, respectively, depending on whether or not the array has been centred across the first, second and/or third mode respectively. In fact, [5] apply this reasoning. This may very well be correct, but we do not know whether we can still consider a class of random arrays which (all in the same way) have been double centred and from which two slices have been chopped off as "generated by random sampling from a continuous distribution"[1]

Remark 3. We have no knowledge of any encompassing function $f(I,J,K)$ describing the maximal rank of $I \times J \times K$ arrays, but there are results for some general classes of $I \times J \times K$ arrays for the maximal or typical rank (see below), for example, $f(I,J,K) = I$ for all arrays for which $JK - J < I < JK$, and f now denotes typical rank [3]. However, in many cases no results are less general, and the function f in fact refers to a partially known mapping of the set $\{I,J,K\}$ on the real field $\mathbb{R}$. The mapping can be deduced from the literature, the latest summary of which (to our knowledge) has been given by [4].

Proof. (of Theorem 1). Recall that the maximum rank of a three-way array $\underline{\mathbf{X}}$ is given by the smallest number R for which for all i,j,k it holds that $x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}$. In matrix notation, this is

$$\mathbf{X}_a = \mathbf{A}\mathbf{G}_a(\mathbf{C} \otimes \mathbf{B})', \tag{7}$$

where $\mathbf{X}_a$ and $\mathbf{G}_a$ denote the A-mode matricised versions of $\mathbf{X}$ and $\mathbf{G}$, respectively and $\mathbf{A}(I \times R)$, $\mathbf{B}(J \times R)$ and $\mathbf{C}(K \times R)$ denote the component matrices for the three modes. The following equivalent expressions can be given upon B- or C-mode matricisation:

$$\mathbf{X}_b = \mathbf{B}\mathbf{G}_b(\mathbf{A} \otimes \mathbf{C})', \tag{8}$$

and

$$\mathbf{X}_c = \mathbf{C}\mathbf{G}_c(\mathbf{B} \otimes \mathbf{A})'. \tag{9}$$

Obviously,

$$\mathbf{X}_a = \mathbf{A}\mathbf{G}_a(\mathbf{C} \otimes \mathbf{B})' \quad \text{iff} \quad \mathbf{S}\mathbf{X}_a(\mathbf{U} \otimes \mathbf{T})' = \mathbf{S}\mathbf{A}\mathbf{G}_a(\mathbf{U}\mathbf{C} \otimes \mathbf{T}\mathbf{B})', \tag{10}$$

for any nonsingular square matrices $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{U}$. Now suppose that $\underline{\mathbf{X}}$ is centred across mode A, then for the vector $\mathbf{u} = (1,1,\ldots,1)'$ it holds that

$$\mathbf{u}'\mathbf{A}\mathbf{G}_a(\mathbf{C} \otimes \mathbf{B})' = \mathbf{0}'. \tag{11}$$

[1] Technically, this is a matter of assessing the class' Lebesgue measure, to which we have no clue. To give an example that generally performed transformations may alter 'randomness' properties, consider for instance squaring all values, which clearly affects the Lebesgue measure of subclasses of the class of such arrays. However, because [5]'s transformations, as our own, are rank preserving, we expect that the results that are only proven for the maximal rank, also hold for the typical rank of classes of arrays.

Choosing $\mathbf{S}$ as a non-singular matrix the first $I-1$ rows of which are not centred (e.g. by taking these equal to the first $I-1$ rows of the $I \times I$ identity matrix) and the last row is the vector $\mathbf{u}'$. Then, the last row of $\mathbf{SA}$ and hence of

$$\mathbf{SX}_a = \mathbf{SAG}_a(\mathbf{C} \otimes \mathbf{B})' \tag{12}$$

equals $\mathbf{0}'$. Thus, the matricised array $\mathbf{SX}_a$ can be written as $\begin{pmatrix} \mathbf{Y}_a \\ \mathbf{0} \end{pmatrix}$, in other words, as the concatenation of the $(I-1) \times J \times K$ array $\mathbf{Y}_a$ containing the first $I-1$ rows of $\mathbf{SX}_a$ and the vector $\mathbf{0}$. For array $\underline{\mathbf{Y}}$, written in matricised form $\mathbf{Y}_a$, it holds that it has rank at most equal to $f(I-1,J,K)$. Hence, it has a decomposition as in (7) for $R = f(I-1,J,K)$. As a consequence, $\mathbf{SX}_a$ can be written as

$$\begin{pmatrix} \mathbf{Y}_a \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^*\mathbf{G}_a(\mathbf{C} \otimes \mathbf{B})' \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^* \\ \mathbf{0} \end{pmatrix} \mathbf{G}_a(\mathbf{C} \otimes \mathbf{B})',$$

where $\mathbf{A}^* = \mathbf{SA}$ and thus $\mathbf{SX}_a$ has a decomposition in $R = f(I-1,J,K)$ components. As a consequence, because of (10), also $\mathbf{X}_a$ has a decomposition in $R = f(I-1,J,K)$ components, from which it follows immediately that $\mathbf{X}_a$ has at most rank $f(I-1,J,K)$.

This concludes the proof of the theorem for centring across mode A. Centring across mode B or C can be proven completely analogously, using matricised forms (8) and (9). □

4 Examples

In this section, we give a few examples.

Example 1. $100 \times 3 \times 2$ arrays.
The Theorem could be seen as stating that centring across one mode will always reduce the maximal rank of a class of arrays by a factor $(G-1)/G$ where G denotes I, J, or K depending on the mode across which we centre. This, however, need not be true, as is obvious in the case where $I >> JK$. Suppose we deal with the class of $100x3x2$ arrays. Then the typical rank will at most be 6 [4]. In this case, the rank does not depend on I at all (since $I > JK$). Hence, centring across mode A, will lead to $R = f(I-1,J,K)$, which also equals 6 [4]. However, centring across mode B and C, does have an effect on the maximal rank. Provided that this is $JK = 6$, centring only across mode B reduces it to $(J-1)K = 2 \times 2 = 4$, centring only across mode C reduces it to $J(K-1) = 3 \times 1 = 3$ and centring across both modes reduces it to $(J-1)(K-1) = 2 \times 1 = 2$, a threefold reduction compared to the original typical rank.

Example 2. $10 \times 4 \times 3$ arrays.
Following [4], for the class of arrays of size $10 \times 4 \times 3$, the typical rank is 10. Table 1 gives the typical rank for all combinations of centring of such arrays.

Clearly, in this case, the effect of single centring depends on the mode that is centred (see rows 2–4 in the table). This is even more so for the effect of double centring (rows 5–7).

Mode A	Mode B	Mode C	$I^* \times J^* \times K^*$	Typical rank
N	N	N	$10 \times 4 \times 3$	10
C	N	N	$9 \times 4 \times 3$	9
N	C	N	$10 \times 3 \times 3$	9
N	N	C	$10 \times 4 \times 2$	8
C	C	N	$9 \times 3 \times 3$	9
C	N	C	$9 \times 4 \times 2$	8
N	C	C	$10 \times 3 \times 2$	6
N	N	N	$9 \times 3 \times 2$	6

Table 1 Example of effects of (combinations of) centring of modes of $10 \times 4 \times 3$ arrays. In the table C means centring across that mode, and N means not centring across that mode. Results are derived from Table 1 from [4]. The lines separate no centring, single centring, double centring and triple centring.

Example 3. $2 \times J \times K$ arrays.
A third special case is concerned with triadditive interactions arrays, such as $\underline{\mathbf{Z}}$ as given in Equation (4), with $I = 2$ and $J, K > 2$. In this case, the rank is $J - 1$ and there are various ways decomposing the array into three component matrices with perfect fit. A convenient decomposition is the following. As $\underline{\mathbf{Z}}$ has zero-sum marginals, it is clear that $\mathbf{A} \propto (\mathbf{1}, -\mathbf{1})'$ (with dimension $2 \times (J-1)$) and it's convenient to choose $\mathbf{A} \propto (\mathbf{1}, -\mathbf{1})'$. Then, the matrices $\mathbf{B}$ ($J \times (J-1)$) and $\mathbf{C}$ ($K \times (J-1)$) can be obtained from the $J \times K$ matrix $\mathbf{Z}_1 = -\mathbf{Z}_2$ through a singular value decomposition, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ denote the first and second horizontal slices of $\underline{\mathbf{Z}}$.

However, a simpler decomposition emerges upon writing

$$\mathbf{Z}_1 = \begin{pmatrix} \mathbf{Z}_1^* \\ -\mathbf{1}'\mathbf{Z}_1^* \end{pmatrix},$$

where $\mathbf{Z}_1^*$ contains the first $J-1$ rows of $\mathbf{Z}$. Then, obviously, $\mathbf{Z}_1 = \mathbf{B}\mathbf{C}'$, where $\mathbf{B} = (\mathbf{I}, -\mathbf{1})'$, with $\mathbf{I}$ of order $(J-1) \times (J-1)$, and $\mathbf{C}' = \mathbf{Z}_1^*$. As, clearly, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ all have $J-1$ columns, thus constituting a rank $J-1$ decomposition of $\underline{\mathbf{Z}}$. The convenience of this solution lies in that of the three component matrices, only $\mathbf{C}$ contains values that relate to the data itself.

5 Conclusion

To conclude, it has been seen that centring often, but not always reduced the rank of arrays. Sometimes, the reduction is dramatic, and comes close to practical values. For instance, a researcher should not be surprised to find perfect PARAFAC fit already for $R = 2$ when analysing a $100 \times 3 \times 2$ array which has been centred across B- and C-mode.

References

[1] Albers, C.J., Gower, J.C.: A contribution to the visualisation of three-way arrays. Journal of multivariate analysis **132**, 1–8 (2014)

[2] Albers, C.J., Gower, J.C.: Visualising interactions in bi- and triadditive models for three-way tables. Submitted for publication (2015)

[3] ten Berge, J.M.F.: The typical rank of tall three-way arrays. Psychometrika **65**, 525–532 (2000)

[4] ten Berge, J.M.F.: Simplicity and typical rank results for three-way arrays. Psychometrika **76**, 3 – 12 (2011)

[5] ten Berge, J.M.F., Sidiropoulos, N.D., Rocci, R.: Typical rank and indscal dimensionality for symmetric threeway arrays of order $I \times 2 \times 2$ or $I \times 3 \times 3$. Linear Algebra and its Applications **388**, 363 – 377 (2004)

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

[7] Gower, J.C.: The analysis of three-way grids. In: P. Slater (ed.) Dimensions of Intra Personal Space, vol. 2, The Measurement of Intra Personal Space by Grid Technique, pp. 163–173. Chicester: Wiley (1977)

[8] Kiers, H.A.L.: Towards a standardized notation and terminology in multiway analysis. Journal of Chemometrics **14**, 105–122 (2000)

[9] Kroonenberg, P.M.: Applied Multiway Data Analysis. Hoboken, New Jersey: Wiley (2008)

[10] Lickteig, T.: Typical tensorial rank. Linear Algebra and its Applications **69**, 95 – 120 (1985)

[11] McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Boca Raton, Florida: Chapman & Hall/CRC (1989)

[12] Nelder, J.A.: A reformulation of linear models. Journal of the Royal Statistical Society, Series A (General) **140**, 48–77 (1977)

[13] Smilde, A.K., Bro, R., Geladi, P.: Multi-way analysis with applications in the chemical sciences. Hoboken, New Jersey: John Wiley & Sons (2004)