JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 275–284

Transforming three-way arrays to multiple orthonormality

Jos M. F. ten Berge[1*†], Henk A. L. Kiers[1], Takashi Murakami[2] and René van der Heijden[1]

[1]*University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands*
[2]*Nagoya University, Nagoya, Japan*

SUMMARY

This paper is concerned with the question to what extent the concept of rowwise or columnwise orthonormality can be generalized to three-way arrays. Whereas transforming a three-way array to multiple orthogonality is immediate, transforming it to multiple orthonormality is far from straightforward. The present paper offers an iterative algorithm for such transformations, and gives a proof of monotonical convergence when only two modes are orthonormalized. Also, it is shown that a variety of three-way arrays do not permit double orthonormalization. This is due to the order of the arrays, and holds regardless of the particular elements of the array. Studying three-way orthonormality has proven useful in exploring the possibilities for simplifying the core, to guide the search for equivalent direct transformations to simplicity; see Murakami *et al*. (*Psychometrika* 1998; **63**: 255–261) as an example. Also, it appears in various contexts of the mathematical study of three-way analysis. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: simplicity; orthonormality; three-way component analysis

INTRODUCTION

Transforming a matrix to rowwise or columnwise orthonormality is an elementary topic in matrix analysis. Extensions to three-way arrays, where three-way orthonormality is defined in terms of orthonormality of the matrices containing all horizontal slices and/or all lateral slices and/or all vertical slices of the array, do not seem to have been considered from a systematic point of view. However, the concept of three-way orthonormality does occasionally arise in different areas of three-way analysis. First of all, a context in which multiple orthonormality has appeared is that of computational efficiency. Kiers [1] has shown how to implement a CANDECOMP/PARAFAC analysis [2,3] for large data sets with multicollinearity. One particular step in his method relies on approximate three-way orthonormality of a three-way array.

Another example can be encountered in Tucker's three-way principal component analysis (3PCA) [4]. Application of this method has always been hampered by the lack of simplicity of the so-called core array, which attributes weights to the joint impact of any triple of components from three different modes. Recently, advances have been made in efforts to transform the core array in 3PCA to have a large number of zero weights, thus attaining a solution for 3PCA that possesses largely the

* Correspondence to: J. M. F. ten Berge, Heijmans Institute, University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands.
† E-mail: j.m.f.ten.berge@ppsw.rug.nl

same type of simplicity that is characteristic of CANDECOMP/PARAFAC. Murakami *et al*. [5] and Ten Berge and Kiers [6] have offered explicit transformations to simplicity for core arrays of order $P \times Q \times R$ when $P = QR - 1$ and when $P > Q$ and $R = 2$ respectively. These patterns of simplicity were first encountered by an iterative procedure for oblique simplicity rotation of core matrices [7], and, notably, by iterative orthonormalization of the core in three directions. For instance, when a $4 \times 3 \times 2$ array is iteratively orthonormalized in two directions, and when, upon convergence, one additional transformation is carried out to the effect that the first frontal slab is diagonalized (and hence fully simplified) by its singular vectors, we typically end up with a transformed array with 18 of the 24 elements zero. Thus transformation to simplicity is another context in which the concept of multiple orthonormality has appeared.

Unfortunately, little is known of iterative orthonormalization of three-way arrays. The present paper attempts to shed some light on conceptual and computational aspects of transforming three-way arrays to multiple orthonormality. In particular, monotonical convergence is proven for iterative orthonormalization in two directions. In addition, it is shown that, for arrays of a certain order, double orthonormality is not possible. The behavior of the procedure of iterative orthonormalization for such cases is examined.

BASIC DEFINITIONS

A $P \times Q$ matrix $\mathbf{X}$ is said to be columnwise orthonormal when $\mathbf{X}'\mathbf{X}$ is the identity matrix $\mathbf{I}_Q$, and rowwise orthonormal when $\mathbf{X}\mathbf{X}'$ is $\mathbf{I}_P$. It is obvious that, when $P \neq Q$, these two forms of orthonormality are incompatible, because $\mathrm{tr}(\mathbf{X}\mathbf{X}') = \mathrm{tr}(\mathbf{X}'\mathbf{X})$. This trace argument no longer applies when we relax the definition of rowwise orthonormality to 'proportionality of $\mathbf{X}\mathbf{X}'$ to $\mathbf{I}_P$'. However, even under the relaxed definition, double orthonormality for a matrix is not possible when $P \neq Q$, because it would imply that $\mathbf{X}$ has both rank P and rank Q.

For three-way arrays, (relaxed) orthonormality in more than one direction is less constrained. Define the $P \times QR$ matrix $\mathbf{X}_\mathrm{a} = [\mathbf{X}_1|\ldots|\mathbf{X}_R]$ containing R frontal $P \times Q$ slices of the $P \times Q \times R$ three-way array $\underline{\mathbf{X}}$ next to each other. Define the $Q \times PR$ matrix $\mathbf{X}_\mathrm{b} = [\mathbf{X}_1'|\ldots|\mathbf{X}_R']$ containing the *transposed* frontal slices of $\underline{\mathbf{X}}$ next to each other; this matrix version of $\underline{\mathbf{X}}$ is identical to the B-mode matricized version of $\underline{\mathbf{X}}$ [8] up to a permutation of the columns. Let $\mathbf{X}_\mathrm{c}$ be the C-mode matricized version of $\underline{\mathbf{X}}$, defined as the $R \times PQ$ matrix with row r obtained by stringing elements of $\mathbf{X}_r$ columnwise out into a row vector, $r = 1,\ldots,R$.

It is not difficult to see that arrays can be orthonormal in three directions, in the sense that the inner product matrices $\mathbf{X}_\mathrm{a}\mathbf{X}_\mathrm{a}'$, $\mathbf{X}_\mathrm{b}\mathbf{X}_\mathrm{b}'$ and $\mathbf{X}_\mathrm{c}\mathbf{X}_\mathrm{c}'$ are proportional to an identity matrix. For instance, when $\underline{\mathbf{X}}$ is the $3 \times 2 \times 2$ array with frontal slices

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 0 \\ 0 & a \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} 0 & 0 \\ a & 0 \\ 0 & 1 \end{bmatrix} \tag{1}$$

with $a = 0{\cdot}707$, then $\mathbf{X}$ has strict A-mode orthonormality because $[\mathbf{X}_1|\mathbf{X}_2]$ is rowwise orthonormal; it has relaxed B-mode orthonormality because $\mathbf{X}_\mathrm{b}\mathbf{X}_\mathrm{b}'$ is proportional to $\mathbf{I}_2$, and it also has relaxed C-mode orthonormality because the sums of squares of $\mathbf{X}_1$ and $\mathbf{X}_2$ are equal and $\mathrm{tr}(\mathbf{X}_1'\mathbf{X}_2) = 0$. From now on, however, we shall ignore the distinction between strict and relaxed orthonormality and simply say that an array has orthonormality in any mode when the inner product matrix for the elements of that mode is proportional to an identity matrix.

We shall consider three-way arrays of order $P \times Q \times R$, usually with $P \geq Q \geq R > 1$. It is easy to verify that an orthogonal rotation in one mode affects the inner products for that mode, but leaves the inner products for the other modes unaffected. For instance, when $\mathbf{X}_1,\ldots,\mathbf{X}_R$ are postmultiplied by an orthogonal rotation matrix $\mathbf{T}'$, we have a B-mode rotation. This rotation amounts to postmultiplying $\mathbf{X}_a$ by $(\mathbf{I}_R \otimes \mathbf{T}')$, premultiplying $\mathbf{X}_b$ by $\mathbf{T}$ and postmultiplying $\mathbf{X}_c$ by $(\mathbf{T}' \otimes \mathbf{I}_P)$ [8]. It is obvious that $\mathbf{T}'\mathbf{X}_b\mathbf{X}_b'\mathbf{T}$ may differ from $\mathbf{X}_b\mathbf{X}_b'$, but the B-mode rotation will affect neither $\mathbf{X}_a\mathbf{X}_a'$ (the A-mode inner product matrix) nor $\mathbf{X}_c\mathbf{X}_c'$ (the C-mode inner product matrix). As a result, triple orthogonality, in the sense that the inner product matrix of a mode is a diagonal matrix, can be obtained at once by just three orthogonalizing rotations, e.g. by eigenvectors of the inner product matrices for each mode.

The straightforward transformations to multiple orthogonality do not generalize to multiple ortho*normality*. This means that, in general, iterative procedures are needed to attain multiple orthonormality, if it can be attained at all. We shall now describe a method of iteratively transforming an array to A- and B-mode orthonormality, when that is possible.

MONOTONICITY OF ITERATIVE DOUBLE ORTHONORMALIZATION

Transformations to double orthonormality cannot be expressed in closed form. However, an iterative procedure of alternately orthonormalizing the A-mode and the B-mode appears to converge. In this section the method is described and a convergence proof is developed.

Consider the matrix $\mathbf{X}_{a0} = [\mathbf{X}_1|\ldots|\mathbf{X}_R]$. Define the $Q \times PR$ matrix $\mathbf{X}_{b0}$ with the transposed versions of the frontal slices of $\mathbf{X}_{a0}$ next to each other. Iterate as follows.

Step 1. Orthonormalize the A-mode by taking $\mathbf{X}_{a1} = (\mathbf{X}_{a0}\mathbf{X}_{a0}')^{-1/2}\mathbf{X}_{a0}$. Construct $\mathbf{X}_{b1}$ by transposing all frontal slices of $\mathbf{X}_{a1}$.

Step 2. Orthonormalize the B-mode by taking $\mathbf{X}_{b2} = (\mathbf{X}_{b1}\mathbf{X}_{b1}')^{-1/2}\mathbf{X}_{b1}$, and construct the associated $\mathbf{X}_{a2}$ by transposing all slices of $\mathbf{X}_{b2}$. Note that we now have B-mode orthonormality, but not A-mode orthonormality.

Step 3. Orthonormalize the A-mode by taking $\mathbf{X}_{a3} = (\mathbf{X}_{a2}\mathbf{X}_{a2}')^{-1/2}\mathbf{X}_{a2}$. Compute $\mathbf{X}_{b3}$ by transposing all frontal slices of $\mathbf{X}_{a3}$.

Step 4. Orthonormalize the B-mode by taking $\mathbf{X}_{b4} = (\mathbf{X}_{b3}\mathbf{X}_{b3}')^{-1/2}\mathbf{X}_{b3}$, and construct the associated $\mathbf{X}_{a4}$ by transposing all frontal slices of $\mathbf{X}_{b4}$. Note that we now have B-mode orthonormality, but not A-mode orthonormality.

Etc.

It will now be shown that the procedure monotonically increases a certain function. First we need the well-known result that, for any given $n \times k$ matrix $\mathbf{Y}$ of rank $n \le k$, the rowwise orthonormal $n \times k$ matrix $\mathbf{Z}$ that gives the best least squares approximation to $\mathbf{Y}$ is obtained as $\mathbf{Z} = (\mathbf{YY}')^{-1/2}\mathbf{Y}$. Thus $\mathbf{Z}$ minimizes $\|\mathbf{Z}-\mathbf{Y}\|^2$ subject to the constraint $\mathbf{ZZ}' = \mathbf{I}_n$. Hence $\mathbf{Z}$ maximizes $\mathrm{tr}(\mathbf{Z}'\mathbf{Y})$ subject to $\mathbf{ZZ}' = \mathbf{I}_n$. This means that $\mathbf{X}_{a1}$ maximizes $\mathrm{tr}(\mathbf{X}_{a1}'\mathbf{X}_{a0})$ subject to $\mathbf{X}_{a1}\mathbf{X}_{a1}' = \mathbf{I}_P$. Likewise, $\mathbf{X}_{b2}$ maximizes $\mathrm{tr}(\mathbf{X}_{b2}'\mathbf{X}_{b1})$ subject to $\mathbf{X}_{b2}\mathbf{X}_{b2}' = \mathbf{I}_Q$; $\mathbf{X}_{a3}$ maximizes $\mathrm{tr}(\mathbf{X}_{a3}'\mathbf{X}_{a2})$ subject to $\mathbf{X}_{a3}\mathbf{X}_{a3}' = \mathbf{I}_P$; $\mathbf{X}_{b4}$ maximizes $\mathrm{tr}(\mathbf{X}_{b4}'\mathbf{X}_{b3})$ subject to $\mathbf{X}_{b4}\mathbf{X}_{b4}' = \mathbf{I}_Q$; and so on.

A next step in proving monotonical convergence of the iterative process consists of expressing the optimality of the trace functions involved in terms of trace inequalities. We start with the optimality of Step 3. In Step 3 we maximize $f(\mathbf{X}_{a3}) = \mathrm{tr}(\mathbf{X}_{a3}'\mathbf{X}_{a2})$ subject to rowwise orthonormality for $\mathbf{X}_{a3}$. Because $\mathbf{X}_{a1}$ also has rowwise orthonormality, it cannot outperform $\mathbf{X}_{a3}$, hence $\mathrm{tr}(\mathbf{X}_{a3}'\mathbf{X}_{a2}) \ge \mathrm{tr}(\mathbf{X}_{a1}'\mathbf{X}_{a2})$. In Step 4 we maximize $g(\mathbf{X}_{b4}) = \mathrm{tr}(\mathbf{X}_{b4}'\mathbf{X}_{b3})$ subject to rowwise orthonormality for $\mathbf{X}_{b4}$. Because $\mathbf{X}_{b2}$ also has rowwise orthonormality, it cannot outperform $\mathbf{X}_{b4}$, hence $\mathrm{tr}(\mathbf{X}_{b4}'\mathbf{X}_{b3}) \ge \mathrm{tr}(\mathbf{X}_{b2}'\mathbf{X}_{b3})$. Likewise, Step 5 yields $\mathrm{tr}(\mathbf{X}_{a5}'\mathbf{X}_{a4}) \ge \mathrm{tr}(\mathbf{X}_{a3}'\mathbf{X}_{a4})$, Step 6 yields $\mathrm{tr}(\mathbf{X}_{b6}'\mathbf{X}_{b5}) \ge \mathrm{tr}(\mathbf{X}_{b4}'\mathbf{X}_{b5})$, and so on.

The final ingredient of the proof is the observation that Step j of the process, $j = 1,2,\ldots$, generates matrices $\mathbf{X}_{aj}$ and $\mathbf{X}_{bj}$ which have exactly the same elements, differently arranged. Therefore

$\mathrm{tr}(\mathbf{X}'_{\mathrm{a}j}\mathbf{X}_{\mathrm{a}(j+1)}) = \mathrm{tr}(\mathbf{X}'_{\mathrm{b}j}\mathbf{X}_{\mathrm{b}(j+1)})$ for every j. With this equality we can write the optimum trace values associated with the successive steps as

$$\begin{aligned} \mathrm{tr}(\mathbf{X}'_{\mathrm{b}2}\mathbf{X}_{\mathrm{b}1}) &= \mathrm{tr}(\mathbf{X}'_{\mathrm{a}2}\mathbf{X}_{\mathrm{a}1}) \quad \text{after Step 2} \\ \mathrm{tr}(\mathbf{X}'_{\mathrm{a}3}\mathbf{X}_{\mathrm{a}2}) &= \mathrm{tr}(\mathbf{X}'_{\mathrm{b}3}\mathbf{X}_{\mathrm{b}2}) \quad \text{after Step 3} \\ \mathrm{tr}(\mathbf{X}'_{\mathrm{b}4}\mathbf{X}_{\mathrm{b}3}) &= \mathrm{tr}(\mathbf{X}'_{\mathrm{a}4}\mathbf{X}_{\mathrm{a}3}) \quad \text{after Step 4} \\ \mathrm{tr}(\mathbf{X}'_{\mathrm{a}5}\mathbf{X}_{\mathrm{a}4}) &= \mathrm{tr}(\mathbf{X}'_{\mathrm{b}5}\mathbf{X}_{\mathrm{b}4}) \quad \text{after Step 5} \end{aligned}$$

and so on. The trace value after Step 3 is at least as high as $\mathrm{tr}(\mathbf{X}'_{\mathrm{a}2}\mathbf{X}_{\mathrm{a}1})$, which is the optimum trace value obtained by Step 2; the trace value after Step 4 is at least as high as $\mathrm{tr}(\mathbf{X}'_{\mathrm{b}3}\mathbf{X}_{\mathrm{b}2})$, which is the optimum trace value after Step 3; and so on. This shows that the iterative procedure of A- and B-mode orthonormalization indeed generates a monotonically increasing series of trace values.

It is important to see that the series of trace values is bounded above. Specifically, when j is odd, the sum of squared elements of $\mathbf{X}_{\mathrm{a}j}$ (rowwise orthonormal) is P, and the sum of squares of $\mathbf{X}_{\mathrm{a}(j+1)}$ is Q, because that matrix has the same elements as the columnwise orthonormal matrix $\mathbf{X}_{\mathrm{b}(j+1)}$. Therefore, by the Schwarz inequality, we obtain the upper bound $\mathrm{tr}(\mathbf{X}'_{\mathrm{a}j}\mathbf{X}_{\mathrm{a}(j+1)}) \leq (PQ)^{1/2}$. The existence of an upper bound implies that the iterations must converge to a stable value of the optimal trace. The upper bound will be attained if and only if $\mathbf{X}_{\mathrm{a}j}$ and $\mathbf{X}_{\mathrm{a}(j+1)}$ are proportional, which means that A- and B-mode orthonormality would have been attained. As we shall see below, this is not always possible.

The iterative process of orthonormalizing the A-mode and the B-mode has been described in terms of a specific, uniquely defined transformation. For instance, B-mode orthonormality is implemented by premultiplying $\mathbf{X}_{\mathrm{b}j}$ by $(\mathbf{X}_{\mathrm{b}j}\mathbf{X}'_{\mathrm{b}j})^{-1/2}$. However, any other way of orthonormalizing $\mathbf{X}_{\mathrm{b}j}$ would also be permitted. Technically, this means that any further premultiplication of $(\mathbf{X}_{\mathrm{b}j}\mathbf{X}'_{\mathrm{b}j})^{-1/2}$ by an orthonormal matrix would also be allowed, because, as we have seen earlier, orthogonal rotations in a mode do not affect the rowwise inner products of any other mode, and they do not affect the inner product matrix of the same mode when that mode is orthonormal. If we allow such a rotation, however, we destroy the optimal fit and optimal trace properties which we have used above to demonstrate convergence. Therefore it is convenient to consider an alternative function, also optimized by the orthonormalizing transformation, but independent of any further rotation. Such a function will now be defined in terms of the singular values of the matrix to be orthonormalized.

Suppose that we orthonormalize a matrix $\mathbf{Y}$ rowwise by defining $\mathbf{Z} = (\mathbf{YY}')^{-1/2}\mathbf{Y}$. Then the optimal trace value involved is $\mathrm{tr}(\mathbf{Z}'\mathbf{Y}) = \mathrm{tr}(\mathbf{YY}')^{1/2} = \mathrm{tr}(\mathbf{D})$, where $\mathbf{Y} = \mathbf{PDQ}'$ is the singular value decomposition of $\mathbf{Y}$. Clearly, premultiplying the transformation matrix $(\mathbf{YY}')^{-1/2}$ by an orthonormal matrix $\mathbf{T}$ will give a trace value $\mathrm{tr}(\mathbf{TD})$ that differs from $\mathrm{tr}(\mathbf{D})$. However, the sum of singular values of $\mathbf{Y}$ is fixed and will not be affected by $\mathbf{T}$. Therefore the sum of singular values of the matrix to be orthonormalized can be evaluated as a monotonically increasing function of the iterative process of orthonormalizing the A-mode and the B-mode. Specifically:

- after Step 1, the function value is the sum of singular values of $\mathbf{X}_{\mathrm{a}0}$;
- after Step 2, the function value is the sum of singular values of $\mathbf{X}_{\mathrm{b}1}$;
- after Step 3, the function value is the sum of singular values of $\mathbf{X}_{\mathrm{a}2}$;

and so on. The series of function values thus obtained is identical to the optimal trace values obtained by the specific orthonormalization of the type $\mathbf{Z} = (\mathbf{YY}')^{-1/2}\mathbf{Y}$, but it remains unaltered when any other type of orthonormalizing transformation is used.

The convergence proof above shows that the algorithm must converge to a stable value for the sum of singular values. This means that at some point the function value can be regarded as having reached convergence. When the function value has attained its upper bound $(PQ)^{1/2}$, double orthonormality has indeed been attained and additional iterations would not affect the elements of the array. However, there is no guarantee that the upper bound will be attained. Situations where this happens inevitably will be discussed in the next section.

NECESSARY CONDITIONS FOR A- AND B-MODE ORTHONORMALITY

Suppose that the iterative method of double orthonormalization is applied to a $5 \times 3 \times 2$ array. Then the iterations do converge in terms of the sum of singular values, but they do not achieve double orthonormality. In fact, the sum of singular values converges to 3·864, which falls short of the upper bound $\sqrt{PQ} = \sqrt{15} = 3{\cdot}873$. In the present section, such discrepancies will be explained.

Consider an array $\underline{\mathbf{X}}$ of order $P \times Q \times R$, with $P \geq Q \geq R > 1$. We shall only concern ourselves with cases where $P < QR$, because A- and B-mode orthonormality is never problematic when $P = QR$, and it is impossible when $P > QR$.

Suppose that the array has double orthonormality. That is, suppose we have $\mathbf{X}_{\mathrm{a}}\mathbf{X}_{\mathrm{a}}' = \lambda\mathbf{I}_P$ and $\mathbf{X}_{\mathrm{b}}\mathbf{X}_{\mathrm{b}}' = \mu\mathbf{I}_Q$ for certain scalars λ and μ. Without loss of generality we set $\lambda = 1$, so $\mu = P/Q$. Because $\mathbf{X}_{\mathrm{a}}$ is rowwise orthonormal, it can be completed to a square orthonormal matrix. Let $\mathbf{Y}_{\mathrm{a}}$ be such a completing matrix (complement) containing R submatrices $\mathbf{Y}_1,\ldots,\mathbf{Y}_R$ of order $(QR - P) \times Q$. Then the matrix

$$\begin{bmatrix} \mathbf{X}_{\mathrm{a}} \\ \mathbf{Y}_{\mathrm{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \,|\, \ldots \,|\, \mathbf{X}_R \\ \mathbf{Y}_1 \,|\, \ldots \,|\, \mathbf{Y}_R \end{bmatrix}$$

is square and orthonormal. This implies that $\mathbf{X}_r'\mathbf{X}_r + \mathbf{Y}_r'\mathbf{Y}_r = \mathbf{I}_Q, r = 1, \ldots, R$. Summing over r yields

$$\sum_r (\mathbf{X}_r'\mathbf{X}_r + \mathbf{Y}_r'\mathbf{Y}_r) = R\mathbf{I}_Q \tag{2}$$

whence, using $\sum_r (\mathbf{X}_r'\mathbf{X}_r) = \mathbf{X}_{\mathrm{b}}\mathbf{X}_{\mathrm{b}}' = P/Q\mathbf{I}_Q$, we have

$$\sum_r (\mathbf{Y}_r'\mathbf{Y}_r) = R\mathbf{I}_Q - (P/Q)\mathbf{I}_Q = (R - P/Q)\mathbf{I}_Q \tag{3}$$

From (3) it is evident that $\sum_r (\mathbf{Y}_r'\mathbf{Y}_r)$ is proportional to the identity matrix $\mathbf{I}_Q$. This means that the $Q \times R(QR-P)$ supermatrix $[\mathbf{Y}_1' | \ldots | \mathbf{Y}_R']$ has rank Q. Hence it must have at least Q columns. From these considerations we have the following necessary condition for double orthonormality.

Result 1

For arrays with $P < QR$, orthonormality in A- and B-mode simultaneously is not feasible when

$$R(QR - P) < Q \tag{4}$$

At this point it will be instructive to revisit the $5 \times 3 \times 2$ case. The matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are of order 1×3, yet $[\mathbf{Y}_1' | \mathbf{Y}_2']$ must have rank 3. This cannot be, because it has only two columns. Therefore a $5 \times 3 \times 2$ array cannot have orthonormality in A- and B-mode simultaneously. Incidentally, B- and C-mode orthonormality for this array is possible, but that is of no concern here. The sheer fact that (4) is violated means that A-mode orthonormality is incompatible with B-mode orthonormality for a

$5 \times 3 \times 2$ array. In the sequel we shall say that a $5 \times 3 \times 2$ array is AB-unfeasible. The reason is that its complement, the $1 \times 3 \times 2$ array, is AB-unfeasible.

The unfeasibility Result 1 can be extended by considering the $Q \times PR$ matrix $\mathbf{X}_\mathrm{b} = [\mathbf{X}'_1| \ldots |\mathbf{X}'_R]$. It is readily verified that this matrix resists AB-orthonormality if and only if the original matrix $\mathbf{X}_\mathrm{a}$ resists AB-orthonormality. Therefore a $P \times Q \times R$ array is AB-unfeasible if and only if the complementary $(PR - Q) \times P \times R$ array is AB-unfeasible. This yields the following.

Result 2

When, for certain values of P, Q and R, with $P \geq Q$, a $P \times Q \times R$ array is AB-unfeasible because it violates (4), then the complementary array, which is of order $(PR - Q) \times P \times R$, is also AB-unfeasible, and *vice versa*.

Again, an example will be instructive. The $3 \times 1 \times 2$ array is AB-unfeasible. Therefore, upon transposing the frontal slices, it is clear that the $1 \times 3 \times 2$ array is AB-unfeasible. Therefore its complement, the $5 \times 3 \times 2$ array, is AB-unfeasible; so its complement, the $7 \times 5 \times 2$ array, is also AB-unfeasible; so its complement, the $9 \times 7 \times 2$ array, is also AB-unfeasible; and so on. It should be noted that the latter two arrays do satisfy (4). Still, arrays of this size are AB-unfeasible because they are generated, by iterative completion to orthonormality, from a smaller AB-unfeasible array.

The $5 \times 3 \times 2$ array is a special case of the arrays with $P = QR - 1$, extensively studied in Reference [5]. All arrays of this class violate (4) if and only if $Q > R$, because when (4) is violated we have $R(QR - P) = R(QR - QR + 1) = R < Q$.

In order to show that an array of a specific order is AB-feasible, we can adopt two approaches: either we run the iterative method for a random array of the desired order to convergence to see whether or not double orthonormality results, or we work our way back to smaller arrays by using Result 2 backwards. For instance, when we need to determine whether or not a $9 \times 6 \times 2$ array is AB-feasible, we notice that it is complementary to the $6 \times 3 \times 2$ array, which clearly is AB-feasible. Therefore the $9 \times 6 \times 2$ array is also AB-feasible. A few examples with $R = 2$, where (4) reduces to $2P \geq 3Q$, are given in Table I.

For arrays with $P \geq Q \geq 3$, AB-feasibility requires that $P/Q \leq 8/3$; see (4). When $Q = 3$, there is no violation of (4) because $P < QR$ implies that $P < 9$. When $Q = 4$ and $R = 3$, the smallest value of P which violates (4) is $P = 11$. All cases with smaller P appear to be AB-feasible. In general,

Table I.

P	*Q*	*R*	
4	3	2	AB-feasible
5	3	2	violates (4)
6	3	2	AB-feasible (trivial)
6	4	2	AB-feasible
7	4	2	violates (4)
7	5	2	satisfies (4), but is complementary to a $5 \times 3 \times 2$ array
8	5	2	violates (4)
9	5	2	violates (4)
9	6	2	AB-feasible (complementary to a $6 \times 3 \times 2$ array)
9	7	2	satisfies (4), but is complementary to a $5 \times 7 \times 2$ array
10	7	2	satisfies (4), but is complementary to a $4 \times 7 \times 2$ array
11	7	2	violates (4)

AB-unfeasibility is to be expected when the ratio of the highest dimension of the array (P) to the smallest dimension (R) is high. This explains why most instances reported above involve arrays with $R = 2$.

So far, we have exclusively been concerned with AB-feasibility in terms of the dimensions P, Q and R of the arrays. However, when the arrays are feasible in that sense, there still may be internal barriers against double orthonormality. For instance, when $\mathbf{X}_a$ is rank-deficient, having a rank less than P, transformations will not be able to restore the rank, so A-mode orthonormality will not be possible. A more subtle counter-example is the following. When

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & a & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{5}$$

with $a = \sqrt{0{\cdot}5}$, then the array has no exterior barriers against AB-feasibility, nor are there problems of rank deficiency. Still, this array resists a transformation to double orthonormality. Clearly, the array has A-mode orthonormality, and orthonormalizing the B-mode would rescale the two unit elements to a. The next A-mode orthonormalization would reset these values at unity, and we will be back where we started. It is tempting to believe that, for this array, the iterative procedure is trapped at a local maximum. However, this explanation is invalid, because changing the array by arbitrary initial transformations is of no avail. We are facing an array which cannot be transformed to AB-orthonormality even though its order is compatible with AB-orthonormality. Fortunately, it can be shown that such arrays arise in practice with probability zero. This is because arrays of order $4 \times 3 \times 2$ can almost surely be transformed to a simple form which does allow AB-orthonormality [6]. For all practical purposes we can therefore say that a $4 \times 3 \times 2$ array is AB-feasible. In the next section we examine the iterative double orthonormalization procedure when applied to AB-unfeasible arrays.

THE ITERATIVE PROCEDURE APPLIED TO UNFEASIBLE ARRAYS

It has been shown above that double orthonormality will not arise when the arrays are AB-unfeasible. In that case the upper bound $(PQ)^{1/2}$ to the sum of singular values of the final array will not be sharp. For instance, for $5 \times 3 \times 2$ arrays the procedure will converge to an array with frontal slices

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & a \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ a & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

with $a = \sqrt{0{\cdot}5}$. This array has A-mode orthonormality, but the B-mode inner product matrix is merely diagonal, with diagonal elements 1·5, 2 and 1·5. Thus the sum of singular values of $\mathbf{X}_b$ is $2\sqrt{1{\cdot}5} + \sqrt{2} = 3{\cdot}864$, which falls short of the upper bound $\sqrt{(3 \times 5)} = 3{\cdot}873$. The discrepancy of these values will now be explained in general terms.

Specifically, we shall now describe a sharper upper bound than $(PQ)^{1/2}$ for AB-unfeasible cases which satisfy (4). Let

$$Q > R(QR - P) \tag{7}$$

Suppose that $\mathbf{X}_\mathrm{a} = [\mathbf{X}_1|\ldots|\mathbf{X}_R]$ has orthonormality for the A-mode, but merely orthogonality for the B-mode, which means that $\sum_r \mathbf{X}_r'\mathbf{X}_r$ is a diagonal matrix Λ not proportional to $\mathbf{I}_Q$. Construct the $(QR-P) \times QR$ matrix $\mathbf{Y}_\mathrm{a} = [\mathbf{Y}_1|\ldots|\mathbf{Y}_R]$ as the one that completes $\mathbf{X}_\mathrm{a}$ to a square orthonormal matrix. Then the $Q \times R(QR-P)$ matrix $\mathbf{Y}_\mathrm{b} = [\mathbf{Y}_1'|\ldots|\mathbf{Y}_R']$ has (almost surely) rank $R(QR-P)<Q$ and satisfies $\mathbf{Y}_\mathrm{b}\mathbf{Y}_\mathrm{b}' = R\mathbf{I}_Q - \Lambda$. It follows that $R\mathbf{I}_Q - \Lambda$ has $Q-R(QR-P)$ diagonal elements zero, hence Λ has $Q-R(QR-P)$ diagonal elements equal to R. Because the sum of all diagonal elements of Λ is P, the other $QR^2-PR = R(QR-P)$ diagonal elements sum to $P-R(Q-QR^2+PR) = (R^2-1)(QR-P)$. Because the square roots of the diagonal elements are the singular values that are monotonically increased, their sum has to be at a maximum when the iterative procedure has attained its maximum value. It is well known from the Schwarz inequality that the sum of a set of non-negative numbers with a fixed sum of squares is at a maximum when these numbers are all equal, which means that they are all $[(R^2-1)/R]^{1/2}$. Therefore the sum of singular values of $\mathbf{X}_\mathrm{b}$ (the trace of $\Lambda^{1/2}$) is bounded above by

$$[Q - R(QR - P)]R^{1/2} + (QR^2 - PR)[(R^2 - 1)/R]^{1/2} \tag{8}$$

This upper bound can be used to verify convergence in terms of the function value when double orthonormalization is applied to arrays that violate (4).

It should be noted that the upper bound (8) equals $\sqrt{PQ}$ when $Q = QR^2-PR$. For the $5 \times 3 \times 2$ array, which satisfies (7), we find the upper bound $\sqrt{2} + 2\sqrt{1\cdot 5} = 3\cdot 864$, which is the very value reported above. This shows that the iterative process has indeed attained the maximum value. Likewise, for the $7 \times 4 \times 2$ array the upper bound $\sqrt{PQ} = \sqrt{28} = 5\cdot 292$ cannot be attained, but (8) yields $2\sqrt{2} + 2\sqrt{1\cdot 5} = 5\cdot 278$, which is indeed the value attained in practice.

It is important to note that the maximum reported in (8) exclusively applies to cases which are AB-unfeasible on account of Result 1. Finding the maximum for complementary cases, which owe their AB-unfeasibility to Result 2, is far more complicated.

The monotonicity of the iterative procedure, in combination with the boundedness of the sum of singular values, still does not imply that the procedure must always converge to its upper bound. However, practical experience invariably reveals that the upper bound is indeed attained.

ORTHONORMALITY IN THREE DIRECTIONS, AND AN EXAMPLE

So far, we have exclusively dealt with the clarification of iterative A- and B-mode orthonormalization. However, it is obvious that the process can be generalized to also include C-mode orthonormalization, by inserting another orthonormalizing step into the process. Unfortunately, this process is much less understood than that of double orthonormalization. There is no monotonical increase of the sum of singular values of the array, nor of any other obvious function that is improved by orthonormalization. Nevertheless, practical experience has consistently revealed that such a process does converge to triple orthonormality when the array is ABC-feasible. In fact, for certain arrays, iterative transformation to AB-orthonormality already tends to give C-mode orthonormality as a bonus. The $8 \times 3 \times 3$ array is a case in point. However, the precise conditions for this to occur remain another matter of further research.

To demonstrate the practical utility of iterative orthonormalization, we present an example from three-way principal component analysis. The data are taken from Reference [9] and consist of a number of pollution measures for the Meaudret river taken at different locations and at different times. A Tucker-3 component analysis, with three components for locations, three for the measures of pollution and two for the time of year, results in the core matrix, consisting of two 3×3 slices (Table II). This array can be transformed to triple orthonormality by iterative transformations. A final

Table II.

2·307	−0·151	−0·031	−0·165	−0·308	−0·207
0·161	0·938	−0·496	0·318	0·421	0·301
0·034	0·376	0·715	−0·047	−0·006	0·047

Table III.

0·817	0·000	0·000	0·000	0·000	0·000
0·000	0·408	0·000	0·000	0·000	−0·707
0·000	0·000	0·408	0·000	0·707	0·000

orthogonal rotation, based on a singular value decomposition of the first slice of the array, generates the transformed core as Table III. It is clear that the array has triple orthogonality in every mode. More importantly, the array displays an amazing degree of simplicity, which greatly reduces the cognitive complexity of the core. This demonstrates the power of iterative orthonormalization in simplifying the core array.

DISCUSSION

In practice, the iterative algorithm presented in this paper invariably seems to converge to its upper bound, which means that double orthonormality is obtained after convergence when the order of the array is feasible. By implication, the maximum of the sum of singular values is related to double orthonormality. A referee raised the intriguing question of how intermediate values of the sum of singular values might be related to measures of (departure from) double orthonormality. The present authors have not found a simple answer. However, it should be obvious that the singular values do depend directly on departure from *single* orthonormality when orthonormality holds for the other mode.

Specifically, suppose we implement Step 1 of the algorithm, which yields $\mathbf{X}_{a1}$ of order $P \times QR$, with sum of squares P. Step 1 also prescribes that we permute the array to get $\mathbf{X}_{b1}$ of order $Q \times PR$, still with sum of squares P. Now the distance between $\mathbf{X}_{b1}$ and its nearest rowwise orthonormal matrix depends on its singular values only. That is, when $\mathbf{X}_{b1} = \mathbf{PDQ}'$ is the SVD of $\mathbf{X}_{b1}$, then its nearest rowwise orthonormal matrix is $\mathbf{PQ}'$, and the sum of squares of $\mathbf{PDQ}' - \mathbf{PQ}'$ is $P + Q - 2\text{tr}(\mathbf{D})$. Because the algorithm monotonically increases the sum of these singular values, it brings the array closer to orthonormality for the mode which is going to be orthonormalized in the next step. In this sense, intermediate values of the sum of singular values are also related to double orthonormality.

It should be clear that this paper has been meant to open the topic of multiple orthonormality. Much work remains to be done before the topic can be closed.

REFERENCES

1. Kiers HAL. A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity. *J. Chemometrics* 1998; **12**: 155–171.
2. Carroll JD, Chang J-J. Analysis of individual differences in multidimensional scaling via an *n*-way generalization of 'Eckart–Young' decomposition. *Psychometrika* 1970; **35**: 283–319.

3. Harshman RA. Foundations of the PARAFAC procedure: models and conditions for PARAFAC1. *UCLA Working Papers Phonet.* 1970; **16**: 1–84.
4. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**: 279–311.
5. Murakami T, Ten Berge JMF, Kiers HAL. A case of extreme simplicity of the core matrix in three-mode principal components analysis. *Psychometrika* 1998; **63**: 255–261.
6. Ten Berge JMF, Kiers HAL. Simplicity of core arrays in three-way principal component analysis and the typical rank of $P \times Q \times 2$ arrays. *Linear Algebra Appl.* 1999; **294**: 169–179.
7. Kiers HAL. Three-way simplimax for oblique rotation of the three-mode factor analysis core to simple structure. *Comput. Statist. Data Anal.* 1998; **28**: 307–324.
8. Kiers HAL. Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics* 2000; **14**: 105–122.
9. Doledec S, Chessel D. Rythmes saisonniers et composantes stationnelles en milieu aquatique. I. Description d'un plan d'observation complet par projection de variables. *Acta Oecol./Oecol. Gen.* 1987; **8**: 403–406.

