PSYCHOMETRIKA—VOL. 59, NO. 3, 377–380
SEPTEMBER 1994

A SIMPLIFICATION OF A RESULT BY ZELLINI ON THE MAXIMAL RANK OF SYMMETRIC THREE-WAY ARRAYS

ROBERTO ROCCI

UNIVERSITY OF ROME "LA SAPIENZA"

JOS M. F. TEN BERGE

UNIVERSITY OF GRONINGEN

Zellini (1979, Theorem 3.1) has shown how to decompose an arbitrary symmetric matrix of order $n \times n$ as a linear combination of $\frac{1}{2}n(n+1)$ fixed rank one matrices, thus constructing an explicit tensor basis for the set of symmetric $n \times n$ matrices. Zellini's decomposition is based on properties of persymmetric matrices. In the present paper, a simplified tensor basis is given, by showing that a symmetric matrix can also be decomposed in terms of $\frac{1}{2}n(n+1)$ fixed binary matrices of rank one. The decomposition implies that an $n \times n \times p$ array consisting of p symmetric $n \times n$ slabs has maximal rank $\frac{1}{2}n(n+1)$. Likewise, an unconstrained INDSCAL (symmetric CANDECOMP/PARAFAC) decomposition of such an array will yield a perfect fit in $\frac{1}{2}n(n+1)$ dimensions. When the fitting only pertains to the off-diagonal elements of the symmetric matrices, as is the case in a version of PARAFAC where communalities are involved, the maximal number of dimensions can be further reduced to $\frac{1}{2}n(n-1)$. However, when the saliences in INDSCAL are constrained to be nonnegative, the tensor basis result does not apply. In fact, it is shown that in this case the number of dimensions needed can be as large as p, the number of matrices analyzed.

Key words: INDSCAL, CANDECOMP, PARAFAC, three-way rank, tensor rank.

Zellini (1979, Theorem 3.1) has proven that the set of real symmetric $n \times n$ matrices has tensor rank $\frac{1}{2}n(n+1)$ in the real field. Zellini's proof is based on the decomposition of symmetric matrices in terms of a fixed basis of eigenvectors of persymmetric matrices. The same result (for real symmetric matrices) will now be obtained without resorting to persymmetry. Instead, a simple basis of binary rank one matrices will be constructed. In addition, it will be shown that Zellini's result has direct implications for the number of dimensions needed to perfectly fit the scalar products version of the INDSCAL model (Carroll & Chang, 1970), and for the problem of maximal rank of symmetric three-way arrays, more generally.

A Binary Basis for the Set of Symmetric $n \times n$ Matrices

Let S be an arbitrary symmetric $n \times n$ matrix with elements $s_{jk} = s_{kj}$, $j, k = 1, \ldots, n$. Let I_n be the $n \times n$ identity matrix, with columns $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Construct t binary symmetric matrices, with $t = \frac{1}{2}n(n+1)$, as follows. For $j = 1, \ldots, n$, let

$$T_{jj} = \mathbf{e}_j\mathbf{e}_j', \tag{1}$$

of rank one, and let, for $j, k = 1, \ldots, n$, $(j < k)$,

$$T_{jk} = (\mathbf{e}_j\mathbf{e}_k' + \mathbf{e}_k\mathbf{e}_j'), \tag{2}$$

of rank two. It is trivially verified that S can be expressed as the linear combination

Requests for reprints should be sent to Roberto Rocci, Department of Statistics, University of Rome "La Sapienza", Piazzale Aldo Moro, 5—00185 Rome, Italy.

0033-3123/94/0900-92074$00.75/0

$$S = \sum_{j \le k}^{n} s_{jk} T_{jk}. \tag{3}$$

The linear span of $T_1, \ldots, T_t$ does not change if we replace every rank two matrix T_{jk}, for $j < k$, by $U_{jk} = T_{jk} + T_{jj} + T_{kk}$. Note that $U_{jk} = (\mathbf{e}_j + \mathbf{e}_k)(\mathbf{e}_j + \mathbf{e}_k)'$, a matrix of rank one. Clearly, all resulting t matrices are now of rank 1, and they are linearly independent. Writing the n rank one matrices T_{jj}, and the $\frac{1}{2}n(n - 1)$ rank one matrices U_{jk}, $j < k$, as $\mathbf{a}_i\mathbf{a}_i'$, for certain vectors $\mathbf{a}_i$, and collecting these vectors in an $n \times t$ matrix, we obtain a matrix A such that, regardless of the elements of S, there always exists a diagonal matrix D for which $S = ADA'$.

It should be noted that the matrix A above has a very natural form: Its columns consist of all possible n-vectors such that all but one or all but two elements are zero, the nonzero elements being 1. An alternative way of saying is that A contains the n columns of I_n, and all pairwise sums of these columns. The elements of the diagonal matrix D can be expressed explicitly in terms of elements of S or they can be computed by solving an equation in Vec-notation, to be discussed below, see (6).

Applications to the Scalar Products Version of the INDSCAL Model

The INDSCAL model is a highly popular tool for analyzing symmetric matrices of proximities or dissimilarities. In the (quasi) scalar product version of INDSCAL, which is also known as symmetric CANDECOMP/PARAFAC, the model reads

$$S_i = AD_iA' + E_i \tag{4}$$

$i = 1, \ldots, p$, where S_i is the i-th observed symmetric matrix of order $n \times n$, typically pertaining to judge i, A is an $n \times r$ group space matrix, D_i is a diagonal matrix of idiosyncratic weights (*saliences*) for judge i, and E_i is a matrix of residuals for judge i (Carroll & Chang, 1970; Harshman, 1970). In practice, the number r of dimensions is chosen as the smallest value that still permits an adequate least squares fit. That is, it should be possible to obtain a low value for the sum of squared elements of $E_1, \ldots, E_p$.

It is well-known that it usually takes more than n dimensions to attain a perfect fit, with $E_1, \ldots, E_p = 0$. It will now be shown that $\frac{1}{2}n(n + 1)$ is an upper bound to this number of dimensions. In fact, this is immediate from the previous section. Specifically, for $i = 1, \ldots, p$, S_i can be decomposed as $S_i = AD_iA'$, where A is either based on Zellini (1979) or it is the binary n by $\frac{1}{2}n(n + 1)$ matrix derived in the previous section, with columns $\mathbf{e}_1, \ldots, \mathbf{e}_n, (\mathbf{e}_1 + \mathbf{e}_2), \ldots, (\mathbf{e}_{n-1} + \mathbf{e}_n)$. To find D_i, it is convenient to note that $S_i = AD_iA'$ if and only if

$$\text{Vec}(S_i) = (A \times A)\mathbf{d}_i, \tag{5}$$

where the vector $\mathbf{d}_i$ contains the diagonal elements of D_i and $\times$ represents the columnwise Kronecker product. That is, the j-th column of $(A \times A)$ is the Kronecker product of the j-th column of A and itself. The coefficient matrix $(A \times A)$ is of full column rank, which means that $\mathbf{d}_i$ can be obtained at once as

$$[(A \times A)'(A \times A)]^{-1}(A \times A)'\text{Vec}(S_i). \tag{6}$$

It has thus been shown how to compute an explicit INDSCAL solution in $\frac{1}{2}n(n + 1)$ dimensions, and that $\frac{1}{2}n(n + 1)$ is an upper bound to the INDSCAL rank.

Admittedly, there is no guarantee that D_i will be nonnegative, for i =

$1, \ldots, p$. The present results have no bearing on INDSCAL subject to a constraint of nonnegative weights, to be treated below.

Although $\frac{1}{2}n(n + 1)$ is an upper bound to the number of dimensions in (unconstrained) INDSCAL, it will overestimate this number when p, the number of symmetric matrices analyzed, is less than the bound. However, when $p \geq \frac{1}{2}n(n + 1)$, the bound is sharp. This will be shown in the next section.

It was pointed out by J. Douglas Carroll (personal communication, 1993) that the upper bound $\frac{1}{2}n(n + 1)$ to the number of dimensions in INDSCAL can also be obtained by means of singular value decompositions of $S_1, \ldots, S_p$, followed by a selection of $\frac{1}{2}n(n + 1)$ linearly independent vectors from the much larger set of vectors involved. In this way it can be shown that a basis of $\frac{1}{2}n(n + 1)$ rank one matrices can always be constructed for any given set of symmetric matrices $S_1, \ldots, S_p$. This differs from the approach adopted in this paper, where the basis is known in advance. The resulting upper bound, however, is the same in either approach.

From Upper Bound to Maximal Rank

The number of dimensions r needed in INDSCAL is closely related to the rank of the three-way array S of order $n \times n \times p$, consisting of symmetric slabs $S_1, \ldots, S_p$. It is well-known (e.g., Kruskal, 1989) that the latter rank is the smallest number of dimensions needed to decompose S_i, $i = 1, \ldots, p$, as

$$S_i = AD_iB', \tag{7}$$

where D_i is a diagonal matrix, and, contrary to (5), there is no constraint that A and B are equal. It will now be shown that, when $p \geq \frac{1}{2}n(n + 1)$, the maximal number of dimensions in INDSCAL and the maximal three-way rank coincide and are equal to $\frac{1}{2}n(n + 1)$. To verify this, write (7) equivalently in Vec-notation as

$$\text{Vec}(S_i) = (A \times B)\mathbf{d}_i \tag{8}$$

for $i = 1, \ldots, p$, where $\times$ again refers to the column-wise Kronecker product. Consider a set of $\frac{1}{2}n(n + 1)$ matrices $S_1, \ldots, S_p$ such that their Vec's form a linearly independent set (the matrices T_{jj} and T_{jk} of (1) and (2) are a case in point). If there is a solution for (8), the matrix $(A \times B)$ must at least have $\frac{1}{2}n(n + 1)$ columns. It follows that $\frac{1}{2}n(n + 1)$ is not merely an upper bound but is in fact the maximal rank of a symmetric $n \times n \times p$ array with $p \geq \frac{1}{2}n(n + 1)$.

Nonnegative Saliences

So far we have only considered the unconstrained scalar products version of INDSCAL (symmetric CANDECOMP/PARAFAC) model. In practice, it is typically desired to have $D_1, \ldots, D_p$, the diagonal matrices of saliences, nonnegative. In that case the tensor basis result described above is of no avail. In fact, if the constraint of nonnegative saliences is imposed, the number of dimensions needed to decompose a symmetric $n \times n \times p$ array may require as many as p dimensions, *regardless how large p may be*. To see this, consider the case of an array consisting of p nonproportional psd rank one arrays $S_1, \ldots, S_p$. Decomposing any slab S_i as AD_iA', when S_i has rank 1, and D_i is constrained to be nonnegative definite, implies that $AD_i^{1/2}$ must have rank 1, hence AD_i must be of rank one. Writing S_i as $\mathbf{u}_i\mathbf{u}_i'$, $i = 1, \ldots, p$, it follows that $\mathbf{u}_i$ must be proportional to a column of A. Hence, when none of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$ are proportional, we need p dimensions to perfectly fit the INDSCAL model with nonnegative saliences. In this case, the matrices $S_1, \ldots, S_p$ are each decomposed independently. Interestingly, this is a case where the INDSCAL solution is unique in the sense that only permutations and reflections are allowed (Harshman, 1972), yet the uniqueness conditions given by Harshman (1972) and Kruskal (1989) are not met. Computational methods for INDSCAL subject to the constraint of nonnegative saliences have been examined by ten Berge, Kiers and Krijnen (1993).

Discussion

Zellini's result implies that taking $r = \frac{1}{2}n(n + 1)$ dimensions in INDSCAL guarantees the existence of a perfectly fitting solution, regardless of the number p of symmetric matrices S_i involved. Also, it has been shown that a $n \times n \times p$ array has maximal rank $\frac{1}{2}n(n + 1)$ when the array consists of $p \geq \frac{1}{2}n(n + 1)$ symmetric slabs. Comparing this to what is known in general about the maximum rank of a $n \times n \times p$ array, it can be seen that symmetry entails a lower rank, as is to be expected. For instance, a $3 \times 3 \times 6$ array may have rank 7 (Atkinson & Stevens, 1979; also see Franc, 1992, pp. 214–215) but it can have rank 6 at most in case of symmetry. Similarly, the $4 \times 4 \times 12$ array may have rank 14 (same references), but it has rank 10 at most in case of symmetry.

Although the value of $r = \frac{1}{2}n(n + 1)$ is sharp when $p \geq r$, it may overestimate the number of INDSCAL dimensions needed when $p < r$. For instance, when $n = 3$ and $p = 2$, the maximum rank is 4; when $n = 3$ and $p = 3$, the maximum rank is 5 (Kruskal, 1989). In such cases, the rank one matrices used in the decomposition of $S_1, \ldots, S_p$ are not known in advance but depend on the elements of $S_1, \ldots, S_p$.

Harshman's PARAFAC procedure for symmetric CANDECOMP/PARAFAC has an option for "communality estimation" as in ordinary factor analysis. That is, it allows the free determination of the diagonal entries of $S_1, \ldots, S_p$. It is obvious from the binary basis derived above that in this case the maximal number of dimensions needed to obtain a perfect fit is $\frac{1}{2}n(n - 1)$, because only the off-diagonal elements need to be fitted.

References

Atkinson, M. D., & Stevens, N. M. (1979). On the multiplicative complexity of a family of bilinear forms. *Linear Algebra and its Applications*, *27*, 1–8.

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

Franc, A. (1992). *Étude algébrique des Multitableaux: Apports de l'Algèbre Tensorielle* [Algebraic study of multiway tables: Contributions from tensor algebra]. Unpublished doctoral dissertation, University of Montpellier II.

Harshman, R. A. (1970). *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis* (UCLA Working Papers in Phonetics, 16:1–84). Los Angeles: University of California at Los Angeles.

Harshman, R. A. (1972). *Determination and proof of minimum uniqueness conditions for PARAFAC1* (UCLA Working Papers in Phonetics, 22:111–117). Los Angeles: University of California at Los Angeles.

Kruskal, J. B. (1989). Rank, decomposition, and uniqueness for 3-way and *N*-way arrays. In R. Coppi & S. Bolasco, (Eds.); *Multiway data analysis*. Amsterdam: North-Holland.

ten Berge, J. M. F., Kiers, H. A. L. & Krijnen, W. P. (1993). Computational solutions for the problem of negative saliences and nonsymmetry in INDSCAL. *Journal of Classification*, *10*, 115–124.

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

Manuscript received 11/16/92
Final version received 9/10/93