PSYCHOMETRIKA—VOL. 53, NO. 4, 579–584
DECEMBER 1988
NOTES AND COMMENTS

EXPLICIT CANDECOMP/PARAFAC SOLUTIONS FOR A CONTRIVED 2 × 2 × 2 ARRAY OF RANK THREE

JOS M.F. TEN BERGE AND HENK A. L. KIERS

UNIVERSITY OF GRONINGEN

JAN DE LEEUW

UNIVERSITY OF LEIDEN

Kruskal, Harshman and Lundy have contrived a special 2 × 2 × 2 array to examine formal properties of degenerate Candecomp/Parafac solutions. It is shown that for this array the Candecomp/Parafac loss has an infimum of 1. In addition, the array will be used to challenge the tradition of fitting Indscal and related models by means of the Candecomp/Parafac process.

Key words: Degenerate Parafac Solutions, trilinear models, component analysis of n-way arrays.

Carroll and Chang (1970) and Harshman (1970) have independently suggested the same method of analyzing three-way arrays, and christened these methods "Candecomp" and "Parafac", respectively. Specifically, let Z denote a $p \times q \times m$ array containing m frontal slabs Z_i, $i = 1, \ldots, m$. Then the Candecomp/Parafac process (CP-process) seeks to minimize the function

$$f(X, Y, C) = \sum_{i=1}^{m} \| Z_i - XC_i Y' \|^2, \tag{1}$$

where X is a $p \times r$ matrix, Y is a $q \times r$ matrix and C_i is a diagonal $r \times r$ matrix, with diagonal elements equal to the elements of the i-th row of an $m \times r$ matrix C.

Kruskal (1977) has generalized the concept of matrix rank to n-way arrays. For $n = 3$ the rank of the array Z, as defined by Kruskal, is the smallest value of r for which $f(X, Y, C)$ can attain its lower bound zero.

Kruskal, Harshman and Lundy (1983, 1985) have analyzed a particular 2 × 2 × 2 array at great length. The array consists of two frontal slabs

$$Z_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \text{ and } Z_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2}$$

which will be referred to as the "KHL data" henceforth. The KHL data array has Rank 3. It has proven to be highly instructive for the study of so-called "degenerate solutions" that can be obtained with the CP-process (Kruskal, Harshman, & Lundy, 1983, 1985; and Harshman & Lundy, 1984, p. 280). Applying the CP-process to the KHL data with $r = 2$ produces a degenerate solution. That is, X, Y and C approach certain matrices of Rank 1. In addition, Kruskal, Harshman and Lundy (1983) have claimed that $f(X, Y, C)$ does not have a minimum but has an infimum, in this case.

The present paper is focussed on the latter claim, for which no formal proof has been

Jan de Leeuw is now at the University of California at Los Angeles.
Requests for reprints should be sent to Jos M.F. ten Berge, Vakgroep Psychologie, RU Groningen, Grote Markt 32, 9712 HV Groningen, THE NETHERLANDS.

0033-3123/88/1200-9072$00.75/0

published so far. Specifically, it is the main purpose of the present paper to show that $f(X, Y, C)$ has an infimum of 1 for the KHL data, when $r = 2$.

In addition, some attention will be paid to the "symmetry claim". That is, Carroll and Chang (1970) have claimed that, when applied to symmetric data the CP-process will, after convergence, produce a solution where X and Y are equal. This claim serves as a rationale for fitting Indscal and related models by the CP-process. The KHL data will be used to refute this rationale.

Solving for C in Terms of X and Y

Before adressing the KHL data, it seems convenient to consider the minimization of $f(X, Y, C)$ in general. Because C_i can be optimized independently of C_j ($j \neq i$) we may first consider the problem of minimizing, for given X and Y,

$$f_i(C_i) = \| Z_i - XC_iY' \|^2 = \text{tr } Z_i'Z_i - 2 \text{ tr } X'Z_iYC_i + \text{tr } X'XC_iY'YC_i, \tag{3}$$

for arbitrary $p \times q$ matrices Z_i, $i = 1, \ldots, m$. Let w_i and c_i be the vectors containing the r diagonal elements of $X'Z_iY$ and C_i, respectively, and define

$$B \equiv (X'X \times Y'Y), \tag{4}$$

where $\times$ stands for the element-wise (Hadamard) product of matrices. It can be verified that the elements of the vector Bc_i are the diagonal elements of $(X'XC_iY'Y)$, and that B is a Gramian matrix (Schur, 1911, p. 14). Therefore, assuming that B is nonsingular, we can express $f_i(C_i)$ equivalently as

$$f_i(C_i) = \text{tr } Z_i'Z_i - 2w_i'c_i + c_i'Bc_i = \text{tr } Z_i'Z_i - w_i'B^{-1}w_i + \| B^{-1/2}w_i - B^{1/2}c_i \|^2. \tag{5}$$

Clearly, $f_i(C_i)$ has a minimum value

$$\min f_i(C_i) = \text{tr } Z_i'Z_i - w_i'B^{-1}w_i, \tag{6}$$

which will be attained if and only if

$$c_i = B^{-1}w_i, \tag{7}$$

$i = 1, \ldots, m$, assuming that B is nonsingular.

Finally, summing over i yields

$$\min f(X, Y, C \mid C) = \sum_{i=1}^{m} \text{tr } Z_i'Z_i - \text{tr } B^{-1}WW', \tag{8}$$

where W is the $r \times m$ matrix containing the vectors w_i, $i = 1, \ldots, m$, column-wise.

The problem of minimizing $f(X, Y, C \mid C)$ over X and Y can thus be converted into the problem of maximizing

$$g_r(X, Y) \equiv \text{tr } B^{-1}WW' = \sum \text{tr } Z_i'Z_i - \min f(X, Y, C \mid C) \tag{9}$$

provided that B is nonsingular. As a matter of convenience, the identification constraints

$$\text{Diag } (X'X) = \text{Diag } (Y'Y) = I_r \tag{10}$$

will be adopted throughout this paper.

Additional Theory for $2 \times 2 \times 2$ Arrays

So far our treatment of minimizing $f(X, Y, C)$ has been completely general, apart from the requirement that B must be non-singular. In this section the special case of $2 \times 2 \times 2$ arrays will be elaborated, that is, the case $p = q = m = 2$.

First, suppose that we set $r = 1$. Then B is the 1×1 identity matrix, see (4) and (10), and W reduces to the 1×2 'matrix'

$$W = [x'Z_1 y, x'Z_2 y]. \tag{11}$$

From (9) and (10) it is clear that maximizing $g_1(X, Y)$ amounts to maximizing

$$\text{tr } WW' = (x'Z_1 y)^2 + (x'Z_2 y)^2. \tag{12}$$

Next, consider the case $r = 2$. Using (10) we can define

$$X = \begin{pmatrix} \sin\alpha & \sin\beta \\ \cos\alpha & \cos\beta \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} \sin\gamma & \sin\delta \\ \cos\gamma & \cos\delta \end{pmatrix}, \tag{13}$$

for certain α, β, γ and δ. For future reference we also define

$$\lambda \equiv \sin(\alpha - \beta)\sin(\gamma - \delta) = |X| \cdot |Y|, \tag{14}$$

and

$$\mu \equiv \cos(\alpha - \beta)\cos(\gamma - \delta). \tag{15}$$

Solutions for the KHL Data With $r = 2$

For the KHL data the above definitions imply that, for $r = 2$, we have

$$B = \begin{pmatrix} 1 & \mu \\ \mu & 1 \end{pmatrix}, \quad \text{and} \quad WW' = \begin{pmatrix} 1 & \mu - \lambda \\ \mu - \lambda & 1 \end{pmatrix}, \tag{16}$$

thus yielding, for nonsingular B or, equivalently, for $\mu^2 < 1$,

$$g_2(X, Y) = 2 + 2\mu\lambda(1 - \mu^2)^{-1} \equiv h(\alpha, \beta, \gamma, \delta). \tag{17}$$

Noting that $\text{tr} \sum Z_i' Z_i = 4$ for the KHL data we arrive at

$$\inf f(X, Y, C) = 4 - \sup g_2(X, Y) = 4 - \sup h(\alpha, \beta, \gamma, \delta) \tag{18}$$

if B is nonsingular.

Using the expressions obtained above, we can readily derive the following results for the KHL data.

Lemma 1. If both X and Y are singular, then, for $r = 2$, $\min f(X, Y, C) = 3$; otherwise $\inf f(X, Y, C)$ is given by (18).

Proof. First, note that, for $r = 2$, B is singular if and only if $\mu^2 = 1$, or, equivalently, if and only if both X and Y are singular. It follows that (18) applies when either X or Y or both are nonsingular. Next, consider the case where X and Y are indeed singular. Then $XC_i Y'$, the fitted part of Z_i, can be written as $c_i^* xy'$ where x and y are the first columns of X and Y, respectively, and c_i^* is a scalar. It follows that, without loss of fit, we may consider the case $r = 1$ instead of $r = 2$. Applying (12) to the special case of the KHL data yields tr $WW' = 1$ and therefore $\min f(X, Y, C) = f(X, Y, C) = 3$.

Lemma 2. For $r = 2$ we have $1 < f(X, Y, C) \leq 3$.

Proof. The case where B is singular follows from Lemma 1. In the nonsingular case we note that $(\mu + \lambda)$ and $(\mu - \lambda)$ are cosines hence $(\mu + \lambda)^2 \leq 1$ and $(\mu - \lambda)^2 \leq 1$. It follows that $|2\mu\lambda| \leq 1 - \mu^2 - \lambda^2 < 1 - \mu^2$ if $\lambda \neq 0$. If $\lambda = 0$ then $h(\alpha, \beta, \gamma, \delta) = 2$; if $\lambda \neq 0$ then we have from $\mu^2 < 1$ that $-1 < 2\mu\lambda(1 - \mu^2)^{-1} < 1$ hence $1 < h(\alpha, \beta, \gamma, \delta) < 3$. It follows that $1 < f(X, Y, C) < 3$ in the nonsingular case. Therefore, $1 < f(X, Y, C) \leq 3$ for $r = 2$.

Theorem 1. We have $f(X, Y, C) > 1$ for all X, Y, and C and $\inf f(X, Y, C) = 1$.

Proof. The first part is immediate from Lemma 2. To prove the second part, take $\beta = \delta = 0$ and $\gamma = \alpha$. Then $\lambda = 1 - \mu$ and $h(\alpha, 0, \alpha, 0) = 2 + 2\mu(1 + \mu)^{-1}$. If we let α tend to 0 then $\mu = \cos^2 \alpha$ tends to 1, $h(\alpha, 0, \alpha, 0)$ tends to 3, and $f(X, Y, C)$ tends to 1.

The technical interpretation of Theorem 1 is that $f(X, Y, C)$ is discontinuous at $\alpha = \beta = \gamma = \delta = 0$, where $f(X, Y, C) = 3$ (cf. Lemma 1), and that $f(X, Y, C)$ does not have a minimum, just as has been claimed by Kruskal, Harshman and Lundy (1983). An immediate consequence for the CP-process is that it cannot converge to a solution with $f(X, Y, C) = 1$ because such a solution does not exist.

Symmetry of the Solutions

Carroll and Chang (1970, p. 287) have claimed that "the basic symmetry of the data", reflected in $Z_i = Z_i'$, $i = 1, \ldots, m$, guarantees that, when the CP-process finally converges, X and Y will have proportional columns. This claim of "symmetric solutions" for symmetric data has been repeated, among others, by Harshman and Lundy (1984, p. 135) and Carroll and Pruzansky (1984, p. 382), as a rationale for using the CP-process to fit Indscal and related models.

In order to relate this rationale to the above development, we may revisit the proof of Theorem 1. Clearly, this proof implies that $f(X, X, C) > 1$ for all X and C while inf $f(X, X, C) = 1$. Thus it is seen that requiring symmetry ($X = Y$) does not involve additional loss. On the other hand, truly asymmetric solutions can also be constructed. If we let $\alpha - \beta = \gamma - \delta$ then $h(\alpha, \beta, \gamma, \delta) = 2 + 2\mu(1 + \mu)^{-1}$ with $\mu^2 = \cos^2 (\alpha - \beta)$. Therefore, letting $\beta = \alpha - \varepsilon$ and $\delta = \gamma - \varepsilon$, with ε tending to zero, already gives $h(\alpha, \beta, \gamma, \delta)$ tending to 3, hence $f(X, Y, C)$ tending to 1. Because α and γ need not be the same, this implies that X and Y can be very different. In fact, applying the CP-process to the KHL data reveils that asymmetric accumulation points can occur.

It should be noted that we have not disproved the claim that the CP-process, when convergent, converges to symmetric solutions for symmetric data. In the example we have examined, the CP-process is nonconvergent, and the accumulation points in the sequences generated by the CP-process can be either symmetric or asymmetric. However, we have shown that using the CP-process to fit Indscal and related models is not generally justified, because the alledgedly symmetric solutions are not guaranteed.

References

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.

Carroll, J. D., & Pruzansky, S. (1984). The CANDECOMP-CANDELINC family of Models and Methods for Multidimensional Data Analysis. In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, & R. P. McDonald (Eds), *Research methods for multimode data analysis* (pp. 372–402). New York: Praeger.

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

Harshman, R. A., & Lundy, M. E. (1984). The PARAFAC model. In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122–215). New York: Praeger.

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

Kruskal, J. B., Harshman, R. A., & M. E. Lundy (1983). *Some relationships between Tucker's three-mode factor analysis and PARAFAC/CANDECOMP*. Paper presented at the annual meeting of the Psychometric Society, Los Angeles.

Kruskal, J. B., Harshman, R. A., & Lundy, M. E. (1985). *Several mathematical relationships between PARAFAC-CANDECOMP and three-mode factor analysis.* Paper presented at the annual meeting of the Classification Society, St. John's, Newfoundland.

Schur, J. (1911). Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen [Comments on the theory of bounded bilinear forms with infinite numbers of variables]. *Journal für die reine und angewandte Mathematik, 140,* 1–28.

Manuscript received 2/5/87
Final manuscript 7/22/87