MULTIWAY DATA ANALYSIS
R. Coppi and S. Bolasco (Editors)

HOW 3-MFA DATA CAN CAUSE DEGENERATE PARAFAC SOLUTIONS, AMONG OTHER RELATIONSHIPS

J. B. Kruskal
AT&T Bell Laboratories
Murray Hill, NJ 07974-2070

R. A. Harshman
University of Western Ontario
London, Ontario, Canada N6A 5C2

M. E. Lundy
University of Western Ontario
London, Ontario, Canada N6A 5C2

This paper discusses relationships among three models: (1) 3-mode factor analysis, (2) PARAFAC-CANDECOMP, and (3) CANDELINC. The most interesting relationship is that data satisfying model (1) can cause degenerate solutions when analyzed with model (2), as described by Theorem 1 and its corollary. Another interesting relationship connecting all three models at once is described by Theorem 2 and its corollaries.

1. INTRODUCTION

In this paper, we discuss relationships between three different models for 3-way arrays. The first is the model used by 3-mode factor analysis, which we abbreviate as 3-MFA. The second is the model used by PARAFAC-CANDECOMP, which we often abbreviate as PARAFAC. The third is the model used by CANDELINC. One of the most interesting relationships is that when data satisfying the 3-MFA model is fitted by PARAFAC, it can cause the PARAFAC solution to have a curious pattern of properties known as degeneracy. This relationship has been clearly demonstrated by a theorem, which is presented below, as well as in real and simulated data. Some basic initial references to 3-MFA are Tucker (1963, 1964, 1966); the original references to PARAFAC and CANDECOMP are Harshman (1970) and Carroll (1970) respectively; and the original reference to CANDELINC is Carroll et al. (1980).

2. PRELIMINARIES

Notation: To state the equations for these models, we introduce notation for the product of a matrix and a 3-array (i.e., a 3-way array). If a is an $I \times R$ matrix and G is an $R \times S \times T$ array, then $a \circledast_1 G$ means the $I \times S \times T$ 3-array whose (i, s, t) element is given by $\sum_r a_{ir} G_{rst}$. The subscript 1 on $\circledast$ indicates that the first subscript of G is involved in the summation. Operations $\circledast_2$ and $\circledast_3$ are defined in a similar manner. Also, if a, b, c and G have sizes that are consonant, then $(a, b, c)\circledast G$ means $a\circledast_1(\circledast_2(\circledast_3 G))$. It is not hard to prove that the order of the operations here makes no difference: this fact is essentially the same as the associativity of matrix multiplication. We also introduce the *identity 3-array of size R*, $I\!I_{3;R}$, which is sometimes abbreviated to $I\!I_3$ or $I\!I$. It is $R \times R \times R$, and its element $I\!I_{r_1 r_2 r_3}$ is 1 if $r_1 = r_2 = r_3$ and 0 otherwise.

Now it is easy to state the equations for the three models. In every case X is an $I \times J \times K$ array of data, possibly after some preprocessing. The 3-MFA model is

$$X \cong (a, b, c)\circledast G \qquad \text{or} \qquad X_{ijk} \cong \sum_{rst} a_{ir}\, b_{js}\, c_{kt}\, G_{rst} ,$$

where the *loading matrices* a, b, c and and the *core* G are to be determined by fitting. The loading matrices are $I \times R$, $J \times S$, $K \times T$, and the core is $R \times S \times T$. The PARAFAC-CANDECOMP model is

$$X \cong (a, b, c)\circledast I\!I_3 \qquad \text{or} \qquad X_{ijk} \cong \sum_r a_{ir}\, b_{jr}\, c_{kr} ,$$

where the *loading matrices* a, b, c are to be determined by fitting. This is sometimes written as $X = [a, b, c]$. The loading matrices have sizes $I \times R$, $J \times R$, $K \times R$, and $I\!I_3$ has size R. The name CANDELINC refers to "CANDEcomp solutions with LINear Constraints". In words, the model seeks the best CANDECOMP (= PARAFAC) *loading matrices* a, b, c subject to the constraints

$$a = \bar{a}\, t_a, \quad b = \bar{b}\, t_b, \quad c = \bar{c}\, t_c,$$

where the *constraint matrices* $\bar{a}$, $\bar{b}$, $\bar{c}$ are given. Stated symbolically, the model is

$$X \cong (\bar{a}\, t_a, \bar{b}\, t_b, \bar{c}\, t_c)\circledast I\!I_3 ,$$

where the constraint matrices are given while the *transformation matrices* t_a, t_b, t_c are to be determined by fitting the data.

It is easy to see that each of these models is a special case of the preceding model. Therefore an array X that can be fit perfectly by CANDELINC using R dimensions (= factors) and some given $\bar{a}$, $\bar{b}$, $\bar{c}$ can also be fit perfectly by PARAFAC-CANDECOMP using R dimensions. Similarly, an array X that can be fit perfectly by PARAFAC using R dimensions can also be fit perfectly by 3-MFA using an $R \times R \times R$ core G. Incidentally, the fact that one model is a special case of another model does not mean that the more special model is uninteresting. The extra assumptions can greatly improve the results of the analysis if they are approximately correct. This principle is heavily used in analysis of variance, time series analysis, and many other areas.

It is less obvious but still elementary that if X is fitted perfectly by 3-MFA with an $R \times S \times T$ core, then it can be fitted perfectly by PARAFAC using some number $\bar{R}$ of dimensions. A weak upper bound for the value of $\bar{R}$ is $\bar{R} \le \min(RS, RT, ST)$. Stronger bounds are known only in a few cases. For example, if $R = S = T = 3$, then a sharp upper bound for $\bar{R}$ is 5 from Kruskal (future). If $R = 2$ and $S = T \ge 2$, then a sharp upper bound for $\bar{R}$ is $\lfloor 3J/2 \rfloor$ from Ja' Ja' (1979) and Kruskal (future), where $\lfloor x \rfloor$ means the largest integer $\le x$.

It might seem that there is not much more to say. However, the relationship between the models is in fact remarkably rich, as we illustrate below. It is possible that a similar richness exists between other models for 3-arrays and their special case models.

3. PARAFAC DEGENERACIES: A CONNECTION WITH 3-MFA

One relationship among the models is closely related to "degenerate" PARAFAC solutions, and probably explains many cases of this curious phenomenon. PARAFAC solutions sometimes display very large positive or negative correlations between some columns of the loading matrices. By "large" we mean magnitude at least 0.9, and often much closer to 1. The factors involved may or may not be substantively interpretable, but the correlation itself is not interpretable. Such solutions are referred to as degenerate. The majority of cases fall into the following pattern, which is referred to here as *two-factor degeneracy*:

(i) Exactly two factors are involved.

(ii) The two factors are highly correlated in all three modes.

(iii) The signs of the three correlations are either all negative, or else consist of one negative and two positive values.

(iv) The magnitudes of the two factors are very large and almost equal.

(v) The triad contributed by one of the two factors approximately cancels the triad contributed by the other factor, so their net contribution has size comparable to that of the normal factors.

(vi) If the PARAFAC stopping criterion is made more stringent, then the magnitudes of the correlations and of the two factors simply get larger, without any qualitative change in the solution.

(vii) The two factors involved in the degeneracy are usually the first two factors, and always become the first two factors if the stopping criterion is made stringent enough.

In several cases of two-factor degeneracy, when the two factors were plotted together, the tiny deviations from perfect correlation appeared to display interpretable factors. In a couple of cases, this appearance was investigated further, and a 2×2 matrix was found that would map the two factors jointly into fully interpretable form. The 2×2 matrix involved was highly-sheared, i.e., its determinant was very small compared to the magnitude of the elements.

The following theorem presents one reason why two-factor degeneracies can occur. The theorem uses the concept of array rank. We refer the reader to the paper by Kruskal in this volume for a thorough discussion of array rank. We note here that this concept has a simple definition, applies to N-way arrays for all N, and coincides with the classical meaning of matrix rank when applied to 2-arrays. We also note that for 3-arrays, rank(X) = the lowest dimensionality of a perfectly fitting PARAFAC solution of X.

Consider an infinite sequence (a_n,b_n,c_n) of PARAFAC solutions obtained by the standard iterative process when fitting a $2\times2\times2$ array X of rank 3. Let the fitted arrays be $X^{(n)} = X = [a_n, b_n, c_n] \equiv (a_n,b_n,c_n) \circledast I\!I_3$, and let $X^{(n)}$ approach some limit $X^{(0)}$, as shown in Figure 1.

This figure displays the 8-dimensional space of all $2\times2\times2$ arrays. To understand this figure, you need a few facts. The maximum rank of any such array is 3. The set A_3 of arrays of rank 3, and the set A_2 of arrays of rank 2, each fill non-zero volumes of 8-dimensional space. Define the *boundary* to consist of all arrays in the mathematical boundary of either A3 or A2.

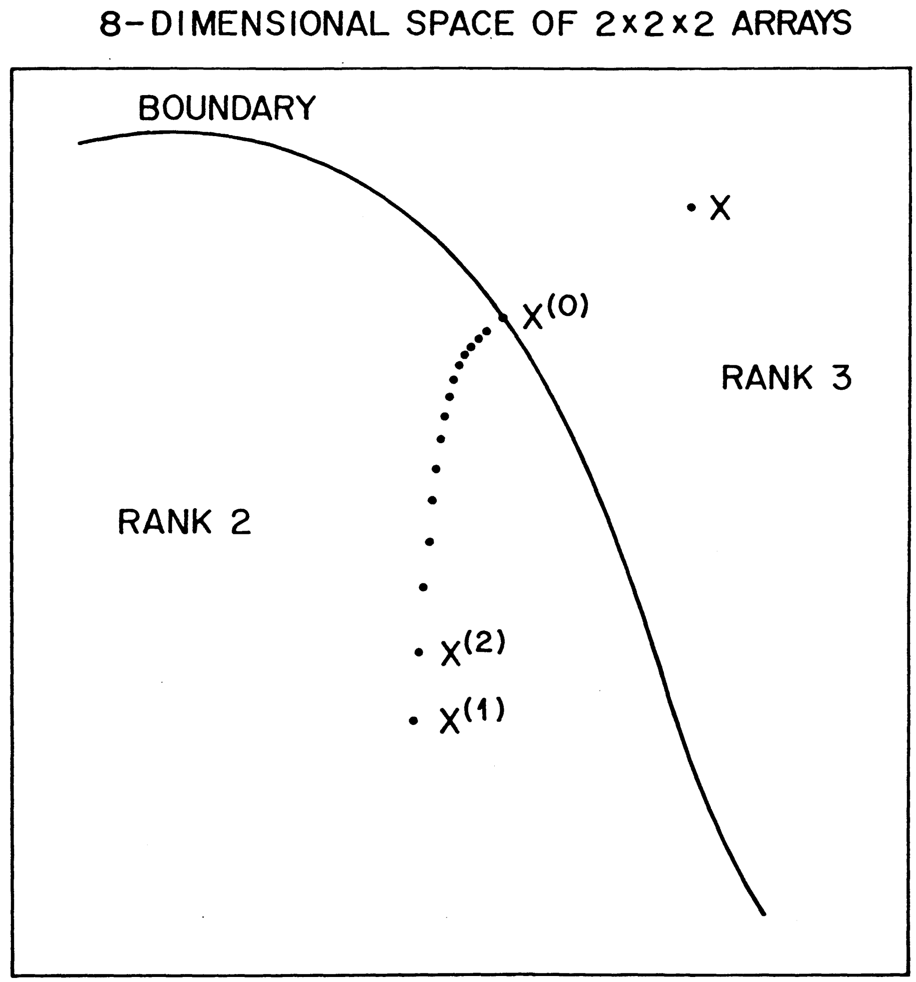
8-DIMENSIONAL SPACE OF 2×2×2 ARRAYS
BOUNDARY
X
X(0)
RANK 3
RANK 2
X(2)
X(1)

FIGURE 1

(Note: arrays in the mathematical boundary of a set may or may not belong to the set.) Many and perhaps all boundary arrays are in the boundary of both A3 and A2. The boundary is a 7-dimensional manifold, and has 8-dimensional volume 0.

A few additional facts will be of interest to some readers. The boundary includes arrays of ranks 3, 2, 1, 0. All arrays of rank 1 or 0 belong to it, but almost all arrays in the boundary have rank 3, i.e., the set of boundary arrays having ranks 2, 1, or 0 has 7-dimensional volume 0 within the boundary.

Theorem 1: Suppose an $I \times J \times K$ array X with rank(X) ≥ 3 has a perfect 3-MFA solution with a $2 \times 2 \times 2$ core G having rank 3. (These conditions imply rank(X) = 3.) Suppose we fit a 2-factor PARAFAC solution (a, b, c) to X. Let $Q = Q(a, b, c)$ be the sum-of-squared-residual-errors, i.e., the sum of squares of $X - (a, b, c) \circledast \mathbb{I}_3$.

(a) Mathematically, Q has an infimum but *no* minimum! Every PARAFAC solution (a, b, c) can be improved.

(b) Every infinite sequence of approximate PARAFAC solutions (a_n, b_n, c_n) whose Q values Q_n approach the infimum *must* fail to converge. In fact, the sequence must diverge to infinity, i.e., the largest absolute value among the elements of the 3 matrices approaches infinity. Nevertheless the associated fitted arrays $X^{(n)} = (a_n, b_n, c_n) \circledast \mathbb{I}_3$ may and typically do converge.

(c) Every such sequence diverges in a special manner. The three infinite sequences a_n, b_n, c_n each satisfy a property like the following, which is described only for a_n. Let a_n' be a_n with its two columns rescaled to have length 1. Then a_n' converges to a matrix (of rank 1) whose two columns are either equal to one another or negatives of one another. If we set $\lambda_a = 1$ or -1 accordingly, then $\lambda_a \lambda_b \lambda_c = -1$.

By comparing the properties in this theorem with the properties of two-factor degeneracies described above, we see a very close agreement. Furthermore, the phenomenon described by this theorem is not restricted to data arrays which have a perfect 3-MFA solution with a $2 \times 2 \times 2$ core, nor is it restricted to two-factor solutions, as we see in the following corollary. By the way, the arrays described in the following corollary are very common.

Corollary: Given integers R and S with $R \overset{>}{\geq} S \geq 2$ and $R \geq 3$, there exist $I \times J \times K$ arrays X with the following properties.

(i) X has a perfect 3-MFA solution with an $R \times R \times R$ core of rank R.

(ii) For the S-factor PARAFAC solutions, Q has an infimum but no minimum, and every such solution can be improved.

(iii) If $S = 2$, conditions (b) and (c) hold for the S-factor solutions, i.e., every infinite sequence of solutions whose Q values approach the infimum must diverge to infinity, and after rescaling must converge as described in (c).

(iv) For $S > 2$, conditions (b) and (c) hold, roughly speaking, for the first two factors. More precisely, in every infinite sequence of solutions whose Q values approach the infimum, the first two factors diverge to infinity, and after rescaling converge as described in (c).

Notice one important difference between the theorem and the corollary. The theorem states that for every array satisfying certain conditions the solutions must behave as described, while the corollary merely states the existence of data arrays whose solutions behave as described.

We do not give a proof of the corollary, but only briefly indicate one way to construct arrays of the kind described. Start with a $2 \times 2 \times 2$ core of rank 3, and form the "direct sum" of it with a "superdiagonal" $(R-2) \times (R-2) \times (R-2)$ core. Multiply the result by a trio of loading matrices with columns of length 1 to form X. The value of S depends on the relative size of the factors contributed by the $2 \times 2 \times 2$ part of the core and the factors contributed by the remainder of the core. If the former factors are small compared with T of the remaining factors and large compared with the rest, $S = T+2$.

The phenomenon of two-factor degeneracy can also be demonstrated with synthetic data. Error-free data was generated in accordance with the conditions in Theorem 1 and PARAFAC used to seek a solution. The results of this simulation were in agreement with the corollary. Furthermore, the phenomenon is not restricted to error-free data. When a small or moderate level of error was added to the data before the PARAFAC analysis, a generally similar result was obtained. Because this experiment was done long ago, before the theorem was discovered, there is no record of the precise ways in which the error changed the appearance of the degeneracy.

The theorem and the simulation show that hypothetical and simulated data satisfying a certain kind of 3-MFA model can cause a two-factor degeneracy. To show that no other aspect of the data set is involved in the two-factor degeneracy, an experiment was done. A set of real data displaying two-factor degeneracy was "filtered" so that the resulting modified data set fits the 3-MFA model perfectly. When the filtered data set was analyzed by PARAFAC, it led to a degeneracy that looked just like the one from the unfiltered data set. (To filter the data, 3-MFA was applied to the data to find an $R \times S \times T$ core and three loading matrices that are $I \times R$, $J \times S$, and $K \times T$. For each loading matrix, the perpendicular projection matrix onto its column space was formed. Filtering consisted of multiplying the data array by the three perpendicular projection matrices.)

Though three-factor degeneracies have not yet been explained, we believe that ultimately they will be, perhaps in connection with $3 \times 3 \times 3$ cores of rank 4 or 5.

4. SOME CONNECTIONS WITH OTHER METHODS

The following theorem is very elementary, and its chief significance lies in its two corollaries. Corollary 2a is the basis for a method called PFCORE, which is described in Lundy, Harshman, and Kruskal (this volume). Corollary 2b makes a three-way connection between CANDELINC, 3-MFA, and PARAFAC. This connection provides a viewpoint for a method described in Carroll (1980) by which CANDELINC can be used to greatly speed up the computation of a PARAFAC-CANDECOMP solution.

> *Theorem 2:* Suppose X is fitted perfectly by 3-MFA with loading matrices $\bar{a}$, $\bar{b}$, $\bar{c}$ and $R \times S \times T$ core G. Suppose (a, b, c) is a solution obtained by the PARAFAC program with $\bar{R}$ factors. Regardless of how well or poorly this solution fits X, even if PARAFAC convergence is incomplete, the column space of a is contained in that of $\bar{a}$ so there is a matrix t_a for which $a = \bar{a}\, t_a$; and similarly for b and c. If in addition $\text{rank}(a) = \text{rank}(\bar{a})$,

then the column space of a is the same as the column space of $\bar{a}$, so t_a is nonsingular and $\bar{a} = a\, t_a^{-1}$; and likewise for b and c.

Corollary 2a: Suppose $R = S = T = \bar{R}$. Then $\bar{a} = a\, t_a^{-1}$, $\bar{b} = b\, t_b^{-1}$, $\bar{c} = c\, t_c^{-1}$ for some nonsingular matrices t_a, t_b, t_c.

Corollary 2b: Suppose $R = S = T = \bar{R}$ and a CANDELINC analysis of X is done using $\bar{a}$, $\bar{b}$, $\bar{c}$ as the CANDELINC constraint matrices. Then the least-squares CANDELINC loading matrices are also least-squares PARAFAC loading matrices.

Proof of Theorem 2: The *fibers* of X in direction v are all vectors which can be formed by letting the v-th subscript of X vary and fixing all the other subscripts. The following argument will be phrased using the fibers in direction 1 and the first loading matrix a, but applies also to fibers in directions 2 or 3 and the second or third loading matrices b or c respectively.

On the first substep of each iteration of PARAFAC, the columns of a, are determined by least-squares regression to a set of equations of the form

fiber of X in direction 1 $\;\cong\;$ linear combination of columns of a.

By the nature of regression, it follows that each column of a determined in this way must be some linear combination of the fibers in direction 1. However, since X is fit perfectly by 3-MFA with loading matrices $\bar{a}$, $\bar{b}$, $\bar{c}$, each fiber in direction 1 of X is a linear combination of the columns of $\bar{a}$. Thus each column of a is a linear combination of the columns of $\bar{a}$, and the theorem is proved.

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., & Kruskal, J. B. (1980). CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, *45*, 3-24.

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

Ja' Ja', J. (1979). Optimal Evaluation of Pairs of Bilinear Forms. *SIAM Journal on Computing*, *8*(3), 443-461.

Kruskal, J. B. (future). The Maximum Rank of Arrays.

Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), *Problems in Measuring Change*. Madison, Wisconsin: University of Wisconsin Press.

Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Fredericksen & H. Gulliksen (Eds.), *Contributions to Mathematical Psychology*. New York: Holt, Rinhart, and Winston.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279-311.