

## An annotated Bibliography of Articles on Degenerate Solutions or Decompositions

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 1 of 10

-- Articles are listed/discussed in approximate chronological and logical order.


-- Links are sometimes provided to information on Kroonenberg's Three-Mode website or my University website. **In the pdf version of this file, most of the links to Kroonenberg's site do not work, but many to my site do. However, the Kroonenberg site is the richest source of information; if you are interested in those links, use the .doc version of this file instead of the .pdf, or go directly to <http://three-mode.leidenuniv.nl/>.**

1. [Harshman](#), R. A., & Lundy, M. E. (1984b).

Data preprocessing and the extended PARAFAC model. In H. G. Law, C. W. Snyder Jr, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216-284). New York: Praeger. [Abstract](#) , [Full text](#)

The first published discussion of degenerate Parafac solutions is contained in a section of this chapter, where the phenomenon is described and named and its cause is investigated and qualitatively identified. A distinction is first made between “soft” degeneracies, which can be eliminated by preprocessing, and “hard-core” degeneracies, which cannot. However, the “hard-core” degeneracies *can* be prevented by applying either a zero-correlation or orthogonality constraint to any one mode; this lowers fit by a few percentage points, but experience with both real and simulated data indicate that the constrained nondegenerate solution approximately recovers the true factors. Some experiments are described that test theories about the causes of degeneracy, and these lead to the conclusion that it is “Tucker variation”—systematic data variation that cannot be represented by a Parafac model at the dimensionality being fit, but that *can* be represented at that dimensionality by a Tucker T3 or T2 model (pp. 278-279). Two key findings essentially prove this connection: (a) when real (i.e., empirical) datasets that produce degenerate Parafac solutions at a particular dimensionality are “filtered” of all but T3-T2 variation (by orthogonal projection onto a T2 subspace), they still produce the *same* degenerate solutions; (b) when artificial (i.e., computer generated) datasets that are constructed to include substantial “Tucker variation” are analyzed by Parafac, they produce a degenerate solution, as predicted, whereas other artificial datasets constructed without “Tucker variation” do not, also as predicted. Although these experiments are reported in abbreviated form, their qualitative findings (about when degeneracy does and does not occur) are easily to verify and are sufficient, in themselves, to establish the authors' conclusion that “Tucker variation” causes degeneracy. However, they shed little light on the mathematical mechanism by which this occurs.

2. Harshman, R. A., & DeSarbo, W. S. (1984).

An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In H. G. Law, C. W. Snyder Jr, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for ultimode data analysis* (pp. 602-642). New York: Praeger. [Abstract](#), [Full text](#) 

## **An annotated Bibliography of Articles on Degenerate Solutions or Decompositions**

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004 p. 2 of 10

This article appears as an appendix to (the volume that contains) [1]. Its importance here is that it supplements the discussion of degeneracy in [1] by presenting actual degenerate solutions encountered during analysis of a particular dataset and illustrating how preprocessing and zero-correlation constraints overcome them. Its broader role is to serve as a step-by-step demonstration and discussion of various aspects of a Parafac analysis.

3. Lundy, M. E., Harshman, R. A., & Kruskal, J. B. (1989).

A two-stage procedure incorporating good features of both trilinear and quadrilinear models. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 123-130).

Amsterdam: Elsevier. [Abstract](#) [Full Text](#)

The discovery in [1] that “Tucker variation” is the cause of degenerate Parafac solutions affords the possibility of “decoding” the phenomenon--identifying and then interpreting the *specific* data patterns responsible for making a particular analysis become degenerate. Article [3] shows how this can be done using a two-stage analysis procedure the authors call PFCORE. Stage 1 employs the standard response to hard-core degeneracy--a factor-independence constraint that blocks it and allows an interpretable Parafac solution to emerge. Stage 2 uses the loadings from Stage 1 to estimate a T3-type core array that shows the *Tucker-type interactions of the Parafac factors*. Large elements in this array reveal how particular inter-factor angles have varied and covaried across modes as well as the direction and relative size of these variations, all of which may provide useful scientific insights. The advantage of PFCORE over a direct T3 analysis is its potential for a unique factor “rotation”. The method is demonstrated by applying it to ratings of 15 television shows (rated 1 to 7) on 16 evaluative or descriptive scales by each of 40 raters. The unconstrained 2D Parafac solution is nondegenerate; its dimensions are Program Humor and Program Sensitivity. However, the unconstrained 3D solution is degenerate, prompting a Parafac reanalysis in which the “TV Program” mode is constrained to have uncorrelated/orthogonal loadings. The Humor and Sensitivity dimensions are again recovered, plus a third dimension, interpreted as Program Violence. This solution is used to obtain a 3x3x3 core array of factor “interactions” which shows that the problematic (non-Parafac) variation is mainly due to differences across individuals in angle between their humor and violence dimensions; for some raters, these are positively related, for others, just the opposite. The article also discusses more general models, and demonstrates a Monte Carlo method of assessing the significance of improvements in fit. In the example application of this method (a dataset of metaphor ratings), strong T1 structure is indicated.

4. [Kruskal](#), J.B. (1989).

Rank, decomposition, and uniqueness for 3-way and  $N$ -way arrays. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 7-18). Amsterdam: Elsevier. (see also and

5. Kruskal, J. B., Harshman, R. A., & Lundy, M. E. (1989).

How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 115-122). Amsterdam:

Elsevier. [Abstract](#) 

## **An annotated Bibliography of Articles on Degenerate Solutions or Decompositions**

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 3 of 10

These two articles together provide a basic mathematical framework for understanding Parafac degeneracy that is used in almost all subsequent work on the topic. Article [4] is more general in that it explains essential ideas about rank and decomposition of arrays<sup>i</sup>, while [5] uses these ideas to explain degeneracy<sup>ii</sup>. It begins with a description of seven observed properties of actual degenerate solutions, stated as they apply in the simplest case, which is called a “*two-factor degeneracy*”. (When more factors are involved, most properties are the same but correlations and cancellations are more complicated.) For this simple case, (i) two factors are involved, (ii) their loadings are highly correlated in all three modes, (iii) one or all three correlations are negative, (iv) the factors have large loadings but (v) a “normal sized” net contribution because their combined contributions almost cancel out (due to the negative correlations), (vi) their loadings keep growing in size and degree of correlation as further iterations are performed, and thus (vii) they eventually become larger than any other factors.

Article [5] then mathematically defines and geometrically describes a process with quite similar properties to those in the above list, and states a theorem about it that reveals important consequences of trying to fit a rank-3  $2 \times 2 \times 2$  array using a rank-2 Parafac model. The description constitutes a novel and important method of representing and visualizing the degeneracy phenomenon in a geometric framework, one that is built around Kruskal’s surprising discovery about multiple “typical ranks” of three-way arrays (first reported in [4]; see, e.g., [10] for more recent discussion).

Key features of the framework are: (a) Use of the space of possible data arrays (in this case, an 8-dimensional space or 8-space, defined by the possible values in a  $2 \times 2 \times 2$  array) rather than the 12-dimensional space of parameter arrays (defined by the 12 loadings). (b) Critical use of Kruskal’s discovery that the 8-space is divided into two regions of positive volume, one of rank-2 (where the Parafac approximations are located) and one of rank-3 (where the array being fit is located) [for a more precise statement see the end of this bibliography]. (c) Description of the Parafac fitting process as (i) a sequence of points in 8-space, each an approximation, that can be thought of as forming a path through the rank-2 region toward the array being fitted, but (ii) the path cannot reach the actual data array because it is in the rank-3 region, while the approximations, which are generated by a two-factor model, necessarily have rank-2; so (iii) the path converges toward that point on the boundary closest to the data point in 8-space<sup>iii</sup>. (d) Direct linkage of the degeneracy to the fact that the approximation is approaching a point (on the boundary) which it can never reach.

This framework is supported and clarified by Theorem 1 in [5], which states that in the situation described above: (a) the sum of squared residual errors has an infimum, but no minimum; (b) the loadings for the factors diverge; and (c) their loadings approach proportionality (and do so in such a way that they almost “cancel” and hence the fitted array converges). A corollary of the theorem extends it to higher rank arrays: for any integer  $R$  there exist rank- $R$  arrays for which the lower-rank Parafac approximation<sup>iv</sup> has an infimum rather than a minimum value. In a two-factor approximation, the loadings behave as described in Theorem 1, while in a higher-rank approximation, at least the

## An annotated Bibliography of Articles on Degenerate Solutions or Decompositions

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 4 of 10

largest two factors behave in this way. The article also discusses other relationships between models, but these are less related to degeneracy.


*Apparent and real limitations.* Some of the seeming limitations of generality in this method of representing degeneracies, and in the associated theorems, are more apparent than real. As the authors point out, the results apply not only to arrays meeting the stated conditions, but also to larger arrays that have *cores* meeting the specified conditions (when exactly decomposed via T3). And it also seems likely that more complicated degeneracies will have a related explanation, since they are observed to have the same empirical properties. Three-factor degeneracies, for example, might be explained "...in connection with 3x3x3 cores of rank 4 or 5" ([5] p. 120). The most serious limitation of this article is that its authors do not give proofs of the theorems that are stated. Fortunately, as indicated below, subsequent authors *have* been able to provide proofs for some of the most critical propositions and even add to them.

6. [Ten Berge](#), J.M.F., Kiers, H.A.L., & De Leeuw, J. (1988).

Explicit candecomp/parafac solutions for a contrived 2x2x2 array of rank three. *Psychometrika*, 53, 579-584.

This article makes an important contribution toward evaluating and proving the Theorems that were stated in [5] without proof. First, it proves part "(a)" of Theorem 1 which states that, in the conditions specified, the Parafac sum of squared residuals approaches an infimum rather than minimum, and hence the fitting process does not converge. This is arguably the most critical (and surprising) part of the theorem. With this statement proved, the divergence and increasing proportionality of loadings (statements "(b)" and "(c)" of Theorem 1) seem plausible as additional characteristics of degenerate solutions that also move toward a discontinuity at the boundary between the rank-2 and rank-3 regions of the 8 dimensional space. Second, it takes "(a)" further than [5] by showing that the value of the infimum = 1 if the data array contains the values given in the rank-3 example on p. 9 of [4] (mistakenly called the KHL array). Finally, it points out that the common belief that "symmetric data implies a symmetric solution" does not necessarily hold when the solution is degenerate.

7. [Mitchell](#), B.C., & Burdick, D.S. (1994).

Slowly converging Parafac sequences: Swamps and two-factor degeneracies. *Journal of Chemometrics*, 8, 155-168. 

A new and unexpected aspect of degenerate behavior is revealed in this article. The continuously improving array approximations may *temporarily* display Theorem 1 properties (b) and (c) (see [5]) –diverging size and increasingly higher proportionality– and along with this the fit may improve in smaller and smaller increments, suggesting property (a), but then, after many iterations, the approximations lose these characteristics and make more rapid progress toward a nondegenerate converged solution. The authors refer to this as Parafac passing through a "swamp", and report that it was observed in roughly 1/3 to 1/4 of sixty-four simulated 4D chemical datasets that they studied. Though unexpected, the article points out that such behavior is not inconsistent with Theorem 1 in [5]: "perhaps, as the PARAFAC sequence progresses toward [its] target, it comes close to

## **An annotated Bibliography of Articles on Degenerate Solutions or Decompositions**

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 5 of 10

a region of higher rank ...” (p.165) slowing it down in just the way Theorem 1 would describe. If the data array is not in the higher rank region, the sequence could gradually work its way around the ‘swamp; and resume fast progress toward its target. The authors acknowledge that this is “purely conjecture”, but they are subsequently shown correct by mathematical work in [10]. They also point out that the concern of some users that Parafac encounters local optima may be based on misinterpretation of swamps. By “waiting out” swamps these authors encountered no local optima in their synthetic datasets. [Note: When uniqueness holds, local optima can arise legitimately in various ways, such as when subsets of R factors are extracted from a space that has a higher true dimensionality]

8. [Rayens](#), W.S., & Mitchell, B.C. (1997).

Two-factor degeneracies and a stabilization of Parafac. *Chemometrics and Intelligent Laboratory Systems*, 38, 173-181. [Abstract](#) ✓

The discovery (in [7]) that periods of *temporary* degenerate behavior can arise and then go away as an analysis passes through a ‘swamp’ stimulated the authors of this article to develop a modified analysis method that tries to take a path around the periphery of these ‘swamps’, thereby greatly reducing the iterations required to reach the solution. They accomplish this by replacing the standard OLS multiple regression in the Alternating Least Squares algorithm with *ridge regression*, a well known technique used in statistics to ‘regularize’ (improve stability of) predictor weights in multiple regression when some of the predictors are highly collinear. In effect, the method penalizes the goodness-of-fit when there are large (positive or negative) regression weights. Since large loadings and highly collinear factors characterize both degenerate solutions and ‘swamps’, this regularization is well suited to “change the error gradient” near swamps so that the sequence of improved solutions avoids ‘swampy’ regions of the space (but see [10] re ‘shadows’ cast by swamps). The authors point out that ridge regression somewhat biases the resulting parameter estimates, and discuss ways to select the optimum size of the ridge weights. Applying their methods, they demonstrate that the number of iterations needed to achieve a converged solution in the presence of swamps is substantially reduced, often by a factor of 10 or even more.

[[Note: Since the authors’ simulation was constructed to study swamps, all their analyses were at the correct dimensionality (i.e., used the same number of factors that was used to generate the data, in this case, 4). However, when the solution has lower rank than the systematic variation in the data (before error), which is a key condition assumed in Theorem 1, there are particular circumstances where unavoidable degenerate solutions will occur. One such circumstance might be when the data rank is higher than the size of any mode of the data array (as the remark on p.120 in [5] suggests). In cases of true degeneracy rather than a swamp, one must impose a constraint such as a factor independence or positivity to overcome the degeneracy. The constraint forces the analysis to find the position in the lower-rank array space that is as close to the target in the higher-rank region as is feasible without having partially canceling factors; thus it stops short of that region near the boundary where severe distortions and swamp behavior emerge.]]

## An annotated Bibliography of Articles on Degenerate Solutions or Decompositions

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 6 of 10

9. [Kiers](#) H.A.L., Ten Berge, J.M.F., & Rocci, R. (1997).

Uniqueness of three-mode factor analysis models with sparse cores: The  $3 \times 3 \times 3$  case. *Psychometrika*, 62, 349-374.

Three aspects of this article are most relevant to degeneracy (it also makes other important contributions not noted here): (a) it provides a constrained T3 model that, in effect, extends Parafac to include “Tucker-variation” directly into the model and hence represent it in the basic (Stage 1) analysis, *without losing uniqueness* (a formal proof of this is given); (b) alternatively, the model can be used as Stage 3 of an “Extended PFCORE” analysis, in which the Stage 2 core array is itself modeled, and then abstracted into a few key parameters that represent particular three-way interactions of factors; (c) the article provides, as a demonstration, an extension of the PFCORE analysis of TV ratings data carried out in [3], thus adding to our information about that specific instance of Tucker-variation causing degeneracy; and, finally, (d) the full (9-parameter) version of their Extended Parafac model can represent *all* the T3 variation among the extracted dimensions, and this makes an empirical result reported in [11] particularly significant, possibly requiring a refinement of the “Tucker-variation” theory of degeneracy.

10. [Paatero](#), P. (2000).

Construction and analysis of degenerate PARAFAC models. *Journal of Chemometrics*, 14, 285-299.

This article makes several major contributions to the study of degenerate Parafac solutions. (a) It provides (i) exact descriptions – one-parameter families of mathematical models – showing how factor loadings could change as a solution goes through the process of degenerate collapse<sup>v</sup>; two main families are described, both consistent with [5]; (ii) these models also imply the form of the approximating array at each point in the process; (iii) the limit points for the sequences of rank-2 approximating arrays are what Paatero calls “degenerate arrays” -- rank-3 arrays that can be approximated to arbitrary precision by a rank-2 array; arrays of this form would be found in boundary of the rank-3 region and the degenerate process would approach the one that is least distant from the target (in the 8-space); (iv) the two main model families take different paths toward the boundary in the 8-dimensional array space yet approach the same rank-3 limit array. (b) It provides the first detailed description and algebraic models of “higher order” degeneracies, ones involving more than two factors (for example, it provides a model of a three factor solution in which all three factors diverge toward infinity yet the estimated array converges). (c) It answers one of the basic questions posed by Kruskal a decade earlier: what is the *shape* of the boundary between the rank-2 and rank-3 regions in the 8 dimensional array space. It turns out to be parabolic if the space is sliced in certain directions, hyperbolic in others, with its “top” corresponding to a saddle-point on the 7-dimensional hypersurface. (e) It confirms the conjecture made in [7] that ‘swamps’ are *temporary* occurrences of the phenomenon described in [5], and shows how to construct analysis problems that will exhibit predictable swamp behavior<sup>vi</sup>; (e) It demonstrates a graphical method of displaying the path through array space generated by the sequence of



## An annotated Bibliography of Articles on Degenerate Solutions or Decompositions

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 7 of 10

gradually improved Parafac approximations *and its relation to the rank-2 boundary*, which appears as a parabola below which lies the region of higher-rank. (f) It uses these array-space plots to present some typical results of numerical experiments demonstrating (i) constructed ‘swamp’ behavior<sup>vii</sup>, (ii) a fully degenerate suboptimal solution where the sequence becomes “trapped” and is unable to surmount the parabolic obstacle, instead falling “forever” into a swamp—even though there is a valid solution further off in the rank-2 region. (iii) differences in the behavior of Parafac algorithms, specifically standard ALS, Paatero’s PMF ([13]), Ryan and Mitchell’s regularization ([8]), and Bro’s line search ([14]), in the presence of swamps and on approach to the true optimum.

11. [Zijlstra](#), B. J. H., & Kiers, H. A. L. (2002).

Degenerate solutions obtained from several variants of factor analysis. *Journal of Chemometrics*, 16, 596-605.

This article significantly enlarges our perspective on degenerate solutions (of the kind considered by this bibliography) by investigating how widely they are found and looking for universals that describe and govern their behavior. The authors first review the literature to construct a definition—ending up with loading profiles that become increasingly proportional<sup>viii</sup> but negatively related, while also growing (diverging) in size until they become excessively large (in the subset of loadings not size-standardized). They then conduct a series of simulation studies, looking for this loading behavior in several “variants of factor analysis”, all of which have at least some elements that are uniquely determined up to scaling and permutations. These consisted of the “shifted multiplicative model” (SHMM) and a component model for multi-trait multi-method matrices (MTMM), (both two-way models) and Constrained Tucker3, Parafac, and Parafac-2 (which are three-way models). After multiple attempts with each, they can identify two or more degenerate solutions with *all* models (though indirect-fit Parafac2 only as local optima). They conclude that degenerate solutions are characteristic, not just of Parafac, but more broadly of many (possibly all) factor-type models that have unique solutions! Furthermore, “degenerate solutions obtained with these models all share the same features”. In contrast, they also show logically why models without unique solutions can *not* show degenerate solutions. “Rotational uniqueness” is a prerequisite for unique solutions, but the precise structure of the unique aspects of the model can vary considerably.

[[There is a particularly puzzling and potentially significant aspect of this paper in relation to the theory of degeneracy presented in my talk at this Tensor Decomposition Workshop. In [11] they find that the full 9 parameter form of the constrained Tucker3 (which is called C3TMFA-EP in [9]) *can* exhibit degenerate solutions. If all Tucker variation among the extracted dimensions is fit by the 9-parameter model, then why would it ever show degenerate solutions? However, it turns out that the 9-parameter model may *not* always be adequate to obtain the same fit as a full Tucker T3 model!]]

## An annotated Bibliography of Articles on Degenerate Solutions or Decompositions

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 8 of 10

### Other secondarily related items (e.g., cited in [1-11]), some with short comments:

12. [Ten Berge](#), J.M.F. (1991).

Kruskal's polynomial for  $2 \times 2 \times 2$  arrays and a generalization to  $2 \times n \times n$  arrays.

*Psychometrika*, 56, 631-636.

This article enriches our understanding of array rank in several ways. First, it clearly explains a polynomial that Kruskal devised for determining the rank of a  $2 \times 2 \times 2$  array and then extends it to  $2 \times n \times n$  arrays<sup>ix</sup>. The methods used for discussing, analyzing and proving these things are educational and informative in themselves. There are also likely to be future applications of the author's technique for evaluating the rank of a  $2 \times n \times n$  array.

13. [Paatero](#), P. (1997).

A weighted non-negative least squares algorithm for three-way "PARAFAC" factor analysis. *Chemometrics and Intelligent Laboratory Systems*, 38, 223-242.

Published Abstract: A time-efficient algorithm PMF3 is presented for solving the three-way PARAFAC (CANDECOMP) factor analytic model. In contrast to the usual alternating least squares, the PMF3 algorithm computes changes to all three modes simultaneously. This typically leads to convergence in 40 .. 100 iteration steps. The equations of the weighted multilinear least squares fit are given. The optional non-negativity is achieved by imposing a logarithmic penalty function. The algorithm contains a possibility for dynamical reweighting of the data during the iteration, allowing a robust analysis of outlier-containing data. The problems typical of PARAFAC models are discussed (but not solved): multiple local solutions, degenerate solutions, non-identifiable solutions. The question of how to verify the solution is discussed at length.

14. C. A. Andersson and R. Bro

The N-way Toolbox for MATLAB. *Chemometrics & Intelligent Laboratory Systems*. 52 (1):1-4, 2000.

Published abstract: This communication describes a free toolbox for MATLABW for analysis of multiway data. The toolbox is called "The N-way Toolbox for MATLAB" and is available on the internet at <http://www.models.kvl.dk/source/nwaytoolbox/>. This communication is by no means an attempt to summarize or review the extensive work done in multiway data analysis but is intended solely for informing the reader of the existence, functionality, and applicability of the N-way Toolbox for MATLAB. <http://www.models.kvl.dk/source/nwaytoolbox/> You can [download a PDF-file of the paper here](#).

15 . Kruskal, J. B. (1977).

Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics ([Corrections](#), 17-1-1984; [fulltext PDF with corrections](#)). *Linear Algebra and Its Applications*, 18, 95-138. [Abstract](#) ✓[(*Corrections also available from The Three-Mode Company*)



## An annotated Bibliography of Articles on Degenerate Solutions or Decompositions

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 9 of 10

Some other Kruskal-related links on Kroononberg's Three-Mode website:

Kruskal, J. B. (1981).

Multilinear models for data analysis. *Behaviormetrika*, 10, 1-20. [Abstract](#) ✓

Kruskal, J. B. (1983a).

Multilinear methods. *Proceedings of Symposia in Applied Mathematics*, 28, 75-104.

[Abstract](#) ✓



Kruskal, J. B. (1983b).

*Statement of some current results about three-way arrays*. Informal notes, Place unknown, Month unknown [Abstract](#) | [fulltext PDF](#)

Kruskal, J. B. (1984).

Multilinear methods. In H. G. Law, C. W. Snyder Jr, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 36-62). New York: Praeger.

[Abstract](#) (not yet available) ✓

Kruskal, J. B. (1985).

*Rank of N-way arrays and the geometry of 2x2x2 arrays*. Technical Memorandum, AT&T Bell Laboratories, Murray Hill NJ. [Abstract](#) | [fulltext PDF](#) ✓

Links to other related material on three-way (and n-way) array decompositions can be found at:

<http://three-mode.leidenuniv.nl/>

endnote: A precise statement of the dual-rank property that Kruskal discovered is as follows: "In the 8-dimensional Euclidean space of all 2x2x2 arrays, there are two different ranks,  $r=2$  and  $r=3$ , such the set of all arrays of rank  $r$  has Lebesgue measure greater than 0"

---

<sup>i</sup> Article [4] is listed here because it is a companion to [5]. Kruskal's work on array rank and uniqueness appears in several places; some is written for mathematicians and some (as here) for scientists and data analysts; see the supplementary publication list and bibliographic link further down.

<sup>ii</sup> Some of the articles in this bibliography, including [6], [11], and to a lesser degree [12], exclusively or primarily cite unpublished sources (specifically, 1983 and 1985 conference talks by the same authors) as the origin of the ideas and results in [5]. However, after consulting with all the authors, it is clear that (a) the text of these talks is not being distributed and (b) all concerned prefer that the printed sources be cited, both because they are publicly available and because they better reflect the original sequence of research thinking and results.

<sup>iii</sup> It is natural to ask whether the path can actually reach the boundary between the rank-2 and rank-3 regions. Practical experience suggests that it cannot, since if it could it should happen at least occasionally in practice. Thus practical experience is consistent with the mathematical framework.

<sup>iv</sup> There is a typo in the first line of the corollary; it should read  $R > S$ , not  $R \geq S$ . The latter would obviously give perfect fit.

<sup>v</sup>The models are very simple; a single parameter controls the weighting of "generating vectors" for each pair of factors in each mode. By gradually increasing the value of this parameter, one can see the full

## An annotated Bibliography of Articles on Degenerate Solutions or Decompositions

Supplementary information for a talk by Richard A. Harshman, presented at the *Workshop On Tensor Decompositions* sponsored by the American Institute of Mathematics and Stanford University, among others, and held in Palo Alto, California, July 2004

p. 10 of 10

---

picture of degenerative change, including (i) a coordinated increase in factor loading correlations, which, being negative, produce (ii) overall cancellation of most of the factor contributions, but (iii) this is counterbalanced by increasing magnitudes of individual factor loadings, which (iv) diverge toward infinity as the parameter approaches a specific value, while (v) all of these loading changes are balanced in such a way that they continually improve the fit of the two-factor approximation to the rank-3 target array. Thus, the elementary algebraic relations in the model reveal the mathematical structure of the process of degeneration it describe. It would seem to me ([r.a.h.]) that the degeneracy parameter is interpretable as the solution's position along a path moving toward the boundary between the rank-2 and rank-3 regions of the 8 dimensional array space described in [5]. I believe that by appropriately relabeling and reinterpreting the "generating vectors" and reexpressing their weightings, at least some of Paatero's "models" can be turned into literal descriptions of the degeneracy phenomenon in particular datasets in terms of the actual factors underlying those datasets.

<sup>vi</sup> This is done by locating the analysis starting point and the target "data" array so that the top of the parabolically shaped boundary between rank-2 and rank-3 regions in the array space is a partial obstacle, forcing the sequence of improving approximations to pass close to it.

<sup>vii</sup> The rate of progress through the space (i.e., relative distance between successive approximations) slows down dramatically when the path goes close to the rank-3 region, then slowly moves around the obstacle; it speeds up again as it moves away from the boundary region and goes directly toward the target.

<sup>viii</sup> Editor's note: Zijlstra and Kiers' use of cosines to measure proportionality has advantages, but is also vulnerable to a confound when there are large shared baselines. Some others use both a product of cosines and a product of correlations to assess profile similarity.

<sup>ix</sup> We thank Lim (p.c., 2005) for noting that this polynomial is also found in the literature on hyperdeterminants, cf. I.M. Gelfand, M.M. Kapranov, A.V. Zelevinsky, "Discriminants, Resultants, and Multidimensional Determinants," Birkhauser, 1994, p. 448.