The Parafac Model and its Variants

Richard A. Harshman

Dept. of Psychology, University of Western Ontario, London, Ontario, Canada

Presented at the *Workshop on Tensor Decompositions and Applications*

C.I.R.M., Luminy, France, Aug-Sept, 2005; Sponsored by IEEE, siam, CNRS

The Parafac Model and Decomposition, and their Variants

Richard A. Harshman

Dept. of Psychology, University of Western Ontario, London, Ontario, Canada

Presented at the *Workshop on Tensor Decompositions and Applications*
C.I.R.M., Luminy, France, Aug-Sept, 2005; Sponsored by IEEE, siam, CNRS

Acknowledgements

I would like to gratefully acknowledge both the immediate help on organizing this talk, and the long term contributions to the research described in this talk, made by my research associate and colleague Margaret Lundy.

In this same spirit I also would like to gratefully acknowledge the importance of the accumulated contributions to this research work that have been made by other of my colleagues and friends over the years. Thank you.

This research was supported in part by grants from the Natural Science and Engineering Research Council of Canada.

Warning, despite every effort to eliminate them, these slides may still contain a few spelling errors.

Overview of Tutorial:

1. What is Parafac (and why)?

2. Intro to Parafac's Variants, Relatives

3. Brief introduction to (issues in) Parafac data analysis

1. What is Parafac (and why)?

[(a) A note on “*Models*” vs. “*Decompositions*”]

(b) “Parallel Proportional Profiles”– the initial idea;

(c) The basic model and decomposition – a different rationale for each

(d) Outer products (and tensors) as widely useful models of empirical interactions.

(e) The extreme simplicity of Parafac

A note on the distinction between "*Model*" and "*Decomposition*"

Until the first Tensor Decomposition Workshop (Palo Alto, 2004), I often **used the term "model" indiscriminately** -- for either (i) a scientific/statistical model of the structure in an array, or (ii) a decomposition of an array according to mathematical rules. (And some of my colleagues used it to mean simply (iii) an "approximation" of an array.)

Now that I have acquired further colleagues and friends through these workshops, some of whom are concerned with decompositions *per se*, I can see why it is both important and illuminating to more carefully make this distinction. Parts of this tutorial illustrate this. First, though, let's clarify the difference in meaning.

Parafac is sometimes used as a **model** and sometimes as a **decomposition**. It can be helpful to understand the difference.

A (scientific) model differs from a (mathematical) decomposition in:

purpose

mathematical form

interpretation

1. They differ in purpose of representation:

A good *model* is a structural description that is *scientifically informative* (reveals unknown empirically generated patterns and/or measures known patterns) and might be *useful for empirical assessment* of the pattern-generating processes.

A good *decomposition* is a structural description that is *mathematically informative* (reveals important mathematical properties and/or re-expresses a complex object as a product of simpler ones) and might be *useful for mathematical manipulation* of the object.

2. They differ in structural form:

A model usually consists of two parts, one that is "structural" -- an algebraic expression describing the modeled pattern(s) – and one that is "stochastic" -- usually a single symbol representing the residuals or unmodeled variation. Their sum equals the array.

A decomposition usually consists of a single algebraic expression (e.g., a matrix product or sum of several products of the same form) that exactly equals the array.

3. They differ in interpretation:

A model is often interpreted by identifying the causal or logical sources of the patterns reflected in the structural part while the high rank residual matrix is interpreted as reflecting the contributions of “noise” or other errors of fit -- or as not-yet-modeled variation.

A decomposition is interpreted in terms of the mathematical properties of its parts and their relations, and what they reveal about the array.

There is a scientific basis for the mathematical structure of a model:

- Scientific (causal) model of a source process --> its expected kind of patterned influences on measurements

- Expected kinds of patterns --> a *pattern filter* or “structural model” that can detect, extract, and measure such patterns

- Applied to specific data --> reveals particular patterns

- Interpretation of the particular patterns --> particular kinds or amounts of causal source activity that produced them

- Observation of this source activity in this measurement context --> broader theoretical/practical implications

Example of a model: chemical-specific spectral patterns

In scientific factor model the structural part represents the sum of a few patterns, each contributed by an individual physical/empirical source. Each of these pattern is assumed to have a particular mathematical form, specified by the model family. For factor-analysis models, each pattern usually has rank-1 outer-product form.

Example of a decomposition: Singular Value Decomposition

A svd *decomposition* exactly represents the entire matrix by a single matrix expression ($\mathbf{X} = \mathbf{SUV}'$), various aspects of this expression reveal things about the decomposed matrix, such as rank, relative sizes of components etc.

It facilitates manipulation of the represented object by replacing it with orthogonal or diagonal matrixes, which are often easier to work with.

In MDS (Multidimensional Scaling) the Three-way Decomposition and Three-way Model are distinct:

Carroll and Chang (1970)

CANDECOMP =

A trilinear Canonical Decomposition equivalent to the Parafac Decomposition

INDSCAL =

An Individual Differences Scaling model for judgments of similarity/dissimilarity among a set of stimuli, with the following form:

INDSCAL model

d_{ijk} = subjective distance between stimulus i and j for person k

$$d_{ijk} = \sqrt{w_{k1} d_{ij1}^2 + w_{k2} d_{ij2}^2 + \cdots}$$

$$d_{ijk} = \sqrt{w_{k1}\left(x_{i1} - x_{j1}\right)^2 + w_{k2}\left(x_{i2} - x_{j2}\right)^2 + \cdots}$$

Parafac is both the name of a decomposition and the name of a model:

1. Parafac as a model

Parafac was developed as a generalization of the factor analysis *model* for two-way data, where each factor represents a pattern due to a distinct influence on the data. The model describes contributions of R factors plus error.

The model's purpose is to provide valid approximations of the original source patterns that generated the systematic part of the data, whether or not they are mathematically elegant .

Parafac as both a decomposition and model:

2. Parafac as a decomposition

The (exact, full) Parafac decomposition of an array has many of the same properties as the SVD decomposition of a matrix. It decomposes a rank-R object into R elementary rank-1 parts.

In some numerical experiments (e.g., reported in the 1970 monograph) Parafac was studied as a decomposition of arrays that had less than maximum rank. For example, 12x10x8 arrays with rank 3. This was done to simulate decomposition of the systematic (latent structure) part of real data, which usually has low rank compared to the maximum rank possible given the size of the dataset.

1. What is Parafac (and why)?

[(a) A note on "*Models*" vs. "*Decompositions*"]

(b) "Parallel Proportional Profiles"– the initial idea;

(c) The basic model and decomposition – a different rationale for each

(d) Outer products (and tensors) as widely useful models of emperical interactions.

(e) The extreme simplicity of Parafac

Raymond B. Cattell and Parallel Profiles article

PSYCHOMETRIKA—VOL. 9, NO. 4
DECEMBER, 1944

"PARALLEL PROPORTIONAL PROFILES" AND OTHER PRINCIPLES FOR DETERMINING THE CHOICE OF FACTORS BY ROTATION

RAYMOND B. CATTELL
DUKE UNIVERSITY

The choosing of a set of factors likely to correspond to the real psychological unitary traits in a situation usually reduces to finding a satisfactory rotation in a Thurstone centroid analysis. Seven principles, three of which are new, are described whereby rotation may be determined and/or judged. It is argued that the most fundamental is the principle of "parallel proportional profiles" or "simultaneous simple structure." A mathematical proof of the uniqueness of determination by this means is attempted and equations are suggested for discovering the unique position.

Source Traits or Mathematical Artifacts?

If factor analysis is used merely as a tool to obtain mathematical factors, a relatively small number of which will act as efficient predictors with respect to a relatively large number of individual variables, the problems discussed in this article do not arise. Any one set of mathematical "artifacts" is practically as good as another for prediction from any one test battery.

Ledyard R Tucker and Three-Mode article

PSYCHOMETRIKA—VOL. 31, NO. 3
SEPTEMBER, 1966

SOME MATHEMATICAL NOTES ON THREE-MODE FACTOR ANALYSIS*

LEDYARD R TUCKER

UNIVERSITY OF ILLINOIS

The model for three-mode factor analysis is discussed in terms of newer applications of mathematical processes including a type of matrix process termed the Kronecker product and the definition of combination variables. Three methods of analysis to a type of extension of principal components analysis are discussed. Methods II and III are applicable to analysis of data collected for a large sample of individuals. An extension of the model is described in which allowance is made for unique variance for each combination variable when the data are collected for a large sample of individuals.

Extension of the two-mode factor analytic model to three or more modes of data classification has been suggested by Tucker. Initial discussions of this development appear in the monographs: *Problems in Measuring Change* [8] and *Contributions to Mathematical Psychology* [9]. The latter of these two monographs gives the basic mathematical structure of the proposed model. A further discussion of the mathematical structure was given by Levin in his PhD dissertation *Three-mode factor analysis* [4]. Results of experimental trials of the method were reviewed by Tucker in a paper read at the 1964 Invitational Conference on Testing Problems [10]. Since the Tucker and Levin

Raymond B. Cattell and Parallel Profiles article

PSYCHOMETRIKA—VOL. 9, NO. 4
DECEMBER, 1944

"PARALLEL PROPORTIONAL PROFILES" AND OTHER PRINCIPLES FOR DETERMINING THE CHOICE OF FACTORS BY ROTATION

RAYMOND B. CATTELL
DUKE UNIVERSITY

The choosing of a set of factors likely to correspond to the real psychological unitary traits in a situation usually reduces to finding a satisfactory rotation in a Thurstone centroid analysis. Seven principles, three of which are new, are described whereby rotation may be determined and/or judged. It is argued that the most fundamental is the principle of "parallel proportional profiles" or "simultaneous simple structure." A mathematical proof of the uniqueness of determination by this means is attempted and equations are suggested for discovering the unique position.

Source Traits or Mathematical Artifacts?

If factor analysis is used merely as a tool to obtain mathematical factors, a relatively small number of which will act as efficient predictors with respect to a relatively large number of individual variables, the problems discussed in this article do not arise. Any one set of mathematical "artifacts" is practically as good as another for prediction from any one test battery.

Cattell’s Proposal: Find the rotation that reveals “Parallel Proportional Profiles” for a factor’s loadings in 2 datasets

How to solve the “rotation problem”: obtain two datasets with the same underlying factors but differing in their relative impact of variance contributed.

Rotate factors extracted from both of them until you find, for each dataset, a rotational position in which each factor shows up in both but with different but proportional loadings.

However, there were serious problems:

(1) The method won’t work for correlation matrices. They destroy the proportional effect of factor-size changes because they *rescale* each variable in each of the two matrices to unit variance.

(2) While proportional changes are preserved if the two factor analyses are based on uniformly scaled covariances, it still won’t work. Meredith (1968?) showed that the PP solution for two covariance matrices would actually <u>not</u> be unique after all!

Is Cattell's "Principle of Proportional Profiles" fatally flawed?

It seemed so.

But Meredith's mathematical formulation of PP only required proportional changes of factor loadings in one mode: variables. By allowing the factor axis angles to vary across solutions, it in effect allowed nonproportional variation in the other mode (factor scores).

Later work (on Parafac2 and Paratuck) showed that requiring the same angles between factors (in *three* solutions) could make the covariance factor axes unique.

An alternative approach:

My relatively modest contribution was to modify the application of Cattell's basic idea in two ways:

(1) From comparing two factor analyses to doing a single factor analysis of three-way data.

(2) From analyzing correlations/covariances to direct analysis of the observations themselves.

It was first tested with synthetic data, and it worked!

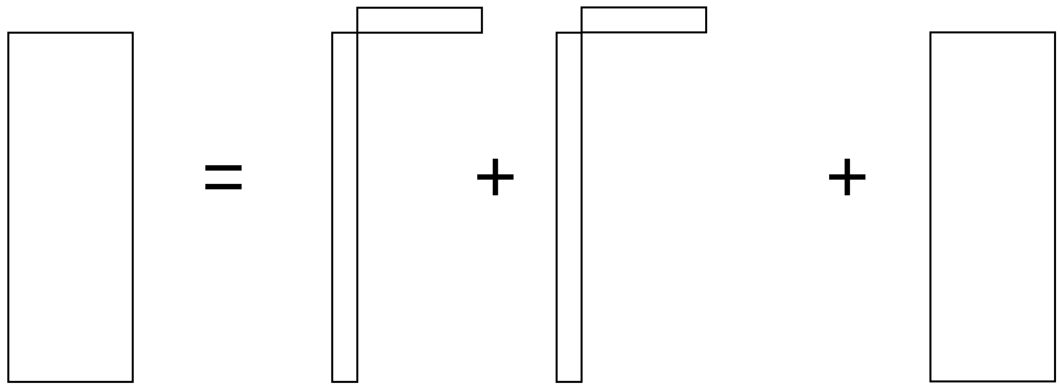
I was lucky that it also worked with the first real datasets on which it was tried. (But only after some basic questions in data preprocessing were solved.)

The utter simplicity of turning a standard two-way factor model (as stated for ‘raw data’) into a Proportional Profiles three-way model:

$$x_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr} + e_{ij}$$

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk}$$

The standard two-way model in graphical form

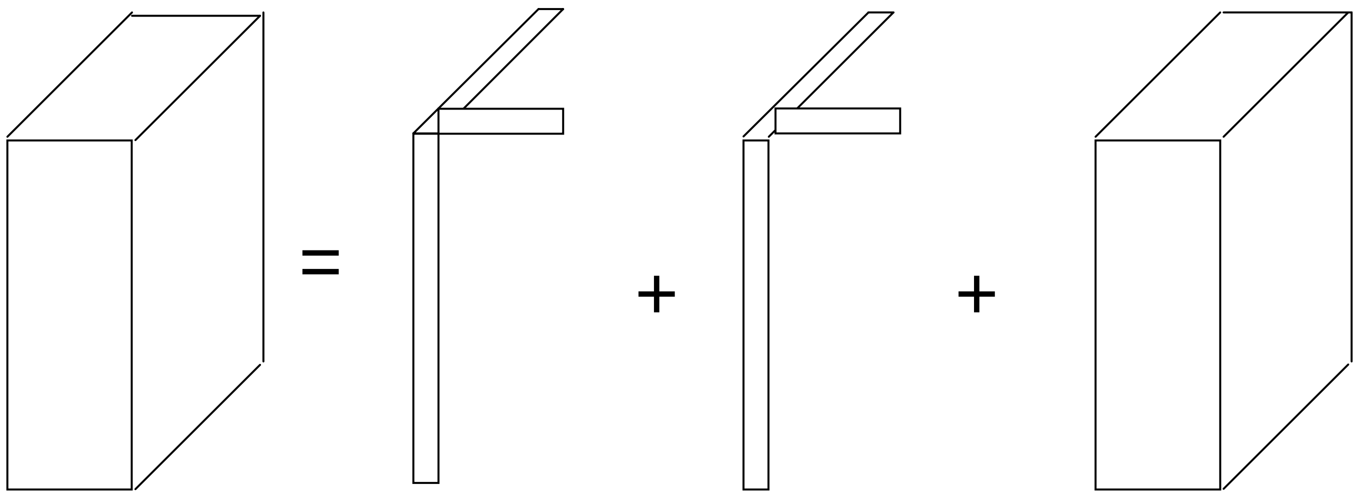


$$\mathbf{X} = \mathbf{a}_1 \mathbf{b}_1' + \mathbf{a}_2 \mathbf{b}_2' + \mathbf{E}$$

rank-1
outer product
matrix

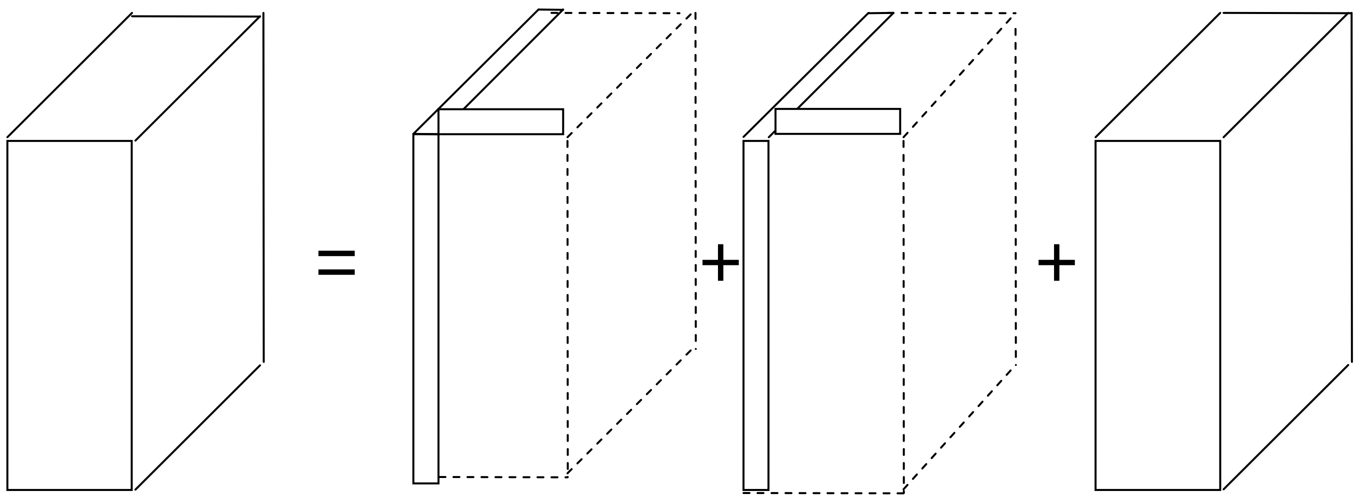
rank-1
outer product
matrix

The three-way Parallel Factor model in graphical form



$$\underline{\mathbf{X}} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2 + \underline{\mathbf{E}}$$

Data array = rank-1 factor contributions plus error



$$\underline{\mathbf{X}} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2 + \underline{\mathbf{E}}$$

high-rank data

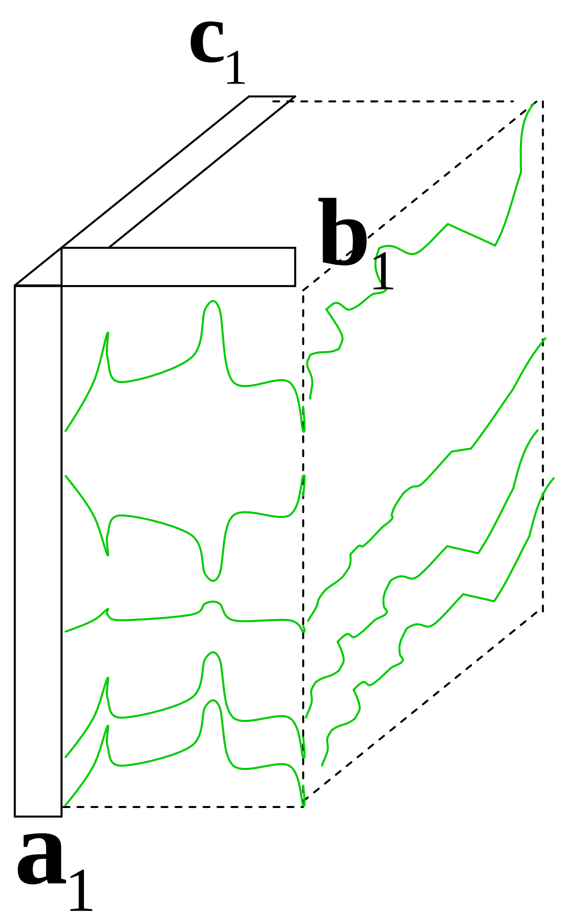
rank-1 outer product

rank-1 outer product

high-rank random values

A single Parafac component is a rank-1 array. This is a logical choice for the basic unit of a decomposition: it has the simplest possible pattern.

In a rank-1 (or outer product) array, the same single pattern is repeated over and over for all fibers of a given orientation; it is simply stepped up or down in size (or reversed in sign) from one fiber to the next



$$\mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1$$

The rank-1 array plays a natural role in both the model and the decomposition

For the model, it is an empirically natural form of variation in the influence of a single cause across levels of the array. A single influence will produce an outer product pattern of variation whenever the impact of its influence varies proportionally across levels of each mode, so that the influence on a single cell is the product of its influence on that level of Mode A, times its influence on that level of Mode B etc. ***(This point will arise in each of my subsequent talks and has great significance because it is a key reason that our tensor applications can be so scientifically useful.)***

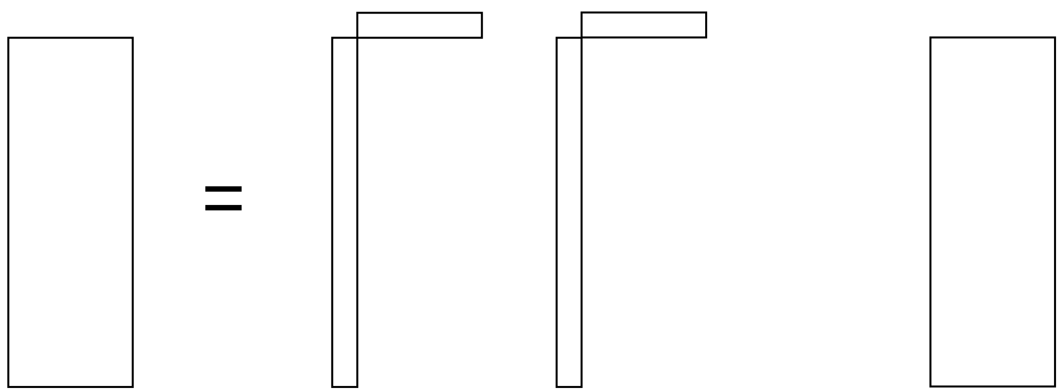
For a decomposition, it is mathematically natural

Four ways to write it

First, four ways to write the standard factor analysis (bilinear) model

Then, for each, how to modify it to represent the Parafac (trilinear) model

Again, to guide our intuition, here is the standard two-way model first in graphical form



$$\mathbf{X} \quad = \quad \mathbf{a}_1\, \mathbf{b}_1' \; + \; \mathbf{a}_2\, \mathbf{b}_2' \quad + \quad \mathbf{E}$$

The standard *bilinear* model (Factor Analysis and PCA)

Scalar form

$$x_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr} + e_{ij}$$

Matrix form

$$\mathbf{X} = \mathbf{AB'} + \mathbf{E}$$

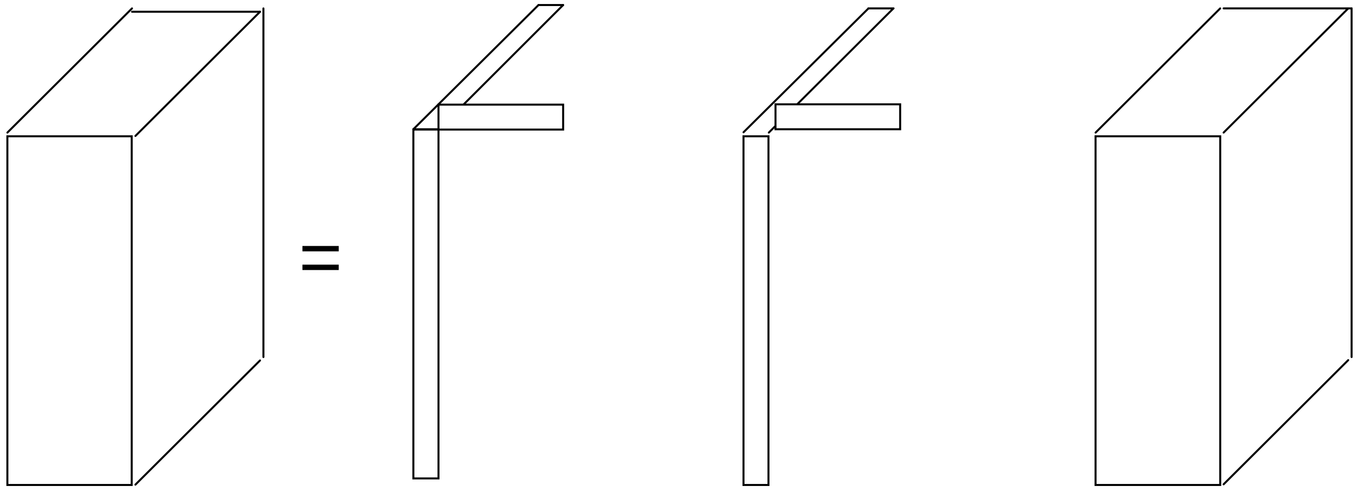
Tensor product form

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r + \mathbf{E}$$

Array Index Notation

$$\mathbf{X}_{\mathrm{IJ}} = \mathbf{A}_{\mathrm{IR}} \mathbf{B}_{\mathrm{JR}} + \mathbf{E}_{\mathrm{IJ}}$$

The trilinear Parallel Factor model in graphical form



$$\mathbf{X} \quad = \quad \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 \; + \; \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2 \; + \; \mathbf{E}$$

The *trilinear* or three-way Parallel Factor generalization

Scalar form $$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk}$$

Matrix form $$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}' + \mathbf{E}_k$$

Tensor product form $$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r + \underline{\mathbf{E}}$$

Array Index Notation $$\mathbf{X}_{\mathrm{IJK}} = \mathbf{A}_{\mathrm{IR}}\,\mathbf{B}_{\mathrm{JR}}\,\mathbf{C}_{\mathrm{KR}}$$

The trilinear or three-way Parallel Factor generalization

Scalar form

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk}$$

Matrix form

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}' + \mathbf{E}_k$$

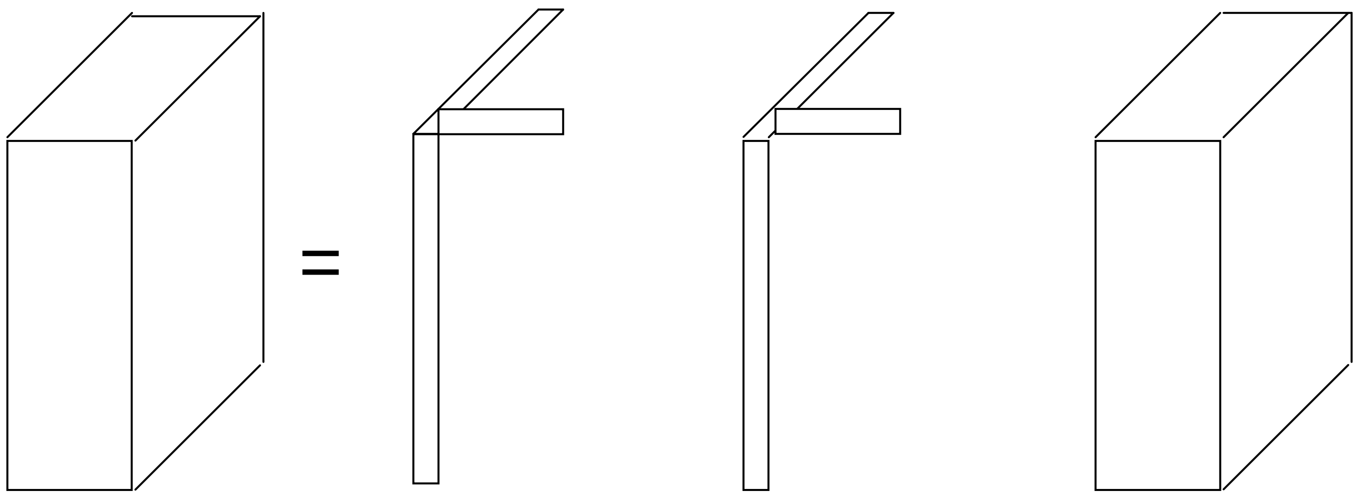
Tensor product form

$$\underline{\underline{\mathbf{X}}} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r + \underline{\underline{\mathbf{E}}}$$

Array Index Notation

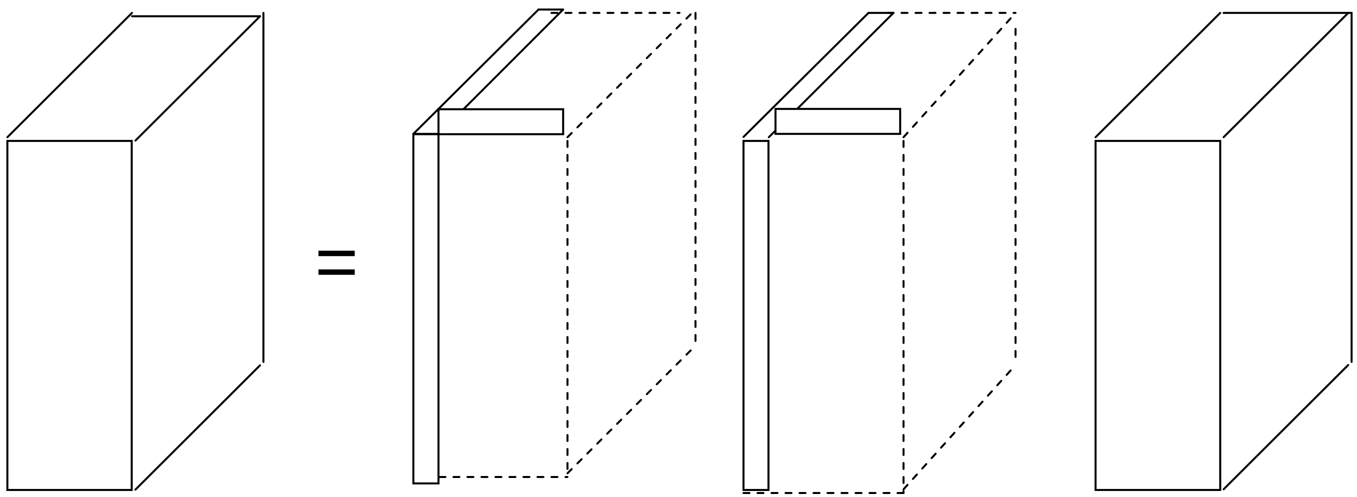
$$\mathbf{X}_{\mathrm{IJK}} = \mathbf{A}_{\mathrm{IR}}\, \mathbf{B}_{\mathrm{JR}}\, \mathbf{C}_{\mathrm{KR}} + \mathbf{E}_{\mathrm{IJK}}$$

The trilinear Parallel Factor model in graphical form



$$\mathbf{X} \quad = \quad \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 \quad + \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2 \quad + \quad \mathbf{E}$$

The trilinear Parallel Factor model in graphical form



$$\mathbf{X} \quad = \quad \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 \quad + \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \otimes \mathbf{c}_2 \quad + \quad \mathbf{E}$$

rank-1
outer product

rank-1
outer product

1. What is Parafac (and why)? (cont.)

(e) Uniqueness --

(a) Why and when (necessary conditions for uniqueness)

(b) Partial uniqueness and alternative unique sols (e.g., if too few factors extracted)

(c) Surface vs. deep uniqueness -- don’t confuse them!

(d) Artificial uniqueness due to interaction of true structure with proportional error

1. What is Parafac (and why)? (cont.)

(f) Limitations and Drawbacks of the strong model:

reduced generality (compared to Tucker)

specific factor variation requirements

degenerate solutions

2. Selected Data Analysis Issues with Parafac: Before-analysis

(a) Appropriateness of model (e.g., is outer-product structure plausible, is there system vs. object variation, likelihood of substantial Tucker variation – factor axis angle changes.)

(b) What kind of preprocessing (centering, standardization) is needed; *tensor models require ”ratio scale” data*

(c) How will you estimate the number of factors?

(d) How will you assess convergence (of fit vs. of “rotation”)

(e) Will conditions for deep uniqueness be fulfilled?

(f) ... Other issues, see e.g. Bro book (downloadable), Smilde, Bro and Giladi book (buyable on line)

2. Selected Data Analysis Issues with Parafac: After-analysis

Quality assessment and analysis problem diagnostics –

Can do Reliability assessment by split half, jack-knifing, leave-one-out, etc.

Can compute confidence bounds around loadings and do significance tests by modern “computer intensive” methods such as randomization and permutation tests, etc.

Is there statistically significant and analutically adequate system variation? Do a randomization test!

Checking error type (uniform variance vs proportional variance vs log-normal or other exponential variance, etc.)

If nonuniform, check influence of outliers, consider using weighted least squares estimation (available in “n–way toolbox” version of Parafac).

2. Selected Data Analysis Issues with Parafac: further questions include

Is amount of Tucker variation small enough? (pros and cons of Corcondia), alternative tests under development

Is there reason to suspect weak degeneracy – shearing of solution space because a solution is mildly contaminated by Tucker variation -- in a mild swamp... (e.g., Stanley analysis of “emotion space” showed a mild distortion with some choices of centering or standardization)

Broader questions of validity: Scientific judgment of meaningfulness of factors; making testible predictions that depend on solutions results; looking for convergent replication in different circumstances.

2. Selected Data Analysis Issues with Parafac: Degenerate Solutions

There are three kinds of degeneracy:

Temporary (passing through a swamp),

Bounded but permanent (solution is located in a swamp)

Unbounded (solution is divergent because no optimum solution exists, the fit of every solution can be improved by increasing shear and making it more degenerate)

The cause of degeneracy is Tucker variation.

Currently, the best “cure” known is to impose constraints on the parameters, typically either a positivity constraint on all loadings, or a factor-independence constraint (zero-correlation among factors) on loadings for one mode

For a subsequent talk...

3. Parafac's Variants and Relatives:

Variants: Indirect-fit Parafac, Parafac2, Paratuck, Paralind,

Relatives: (Tucker T3, T3), DEDICOM, Shifted Factor Analysis

Faux Amis / Nemesis: arrays that are *not* tensors (e.g., simplex structures).

Introduction to Tucker's model

$$x_{ijk} = \sum_r \sum_s \sum_t a_{ir} b_{js} c_{kt} g_{rst}$$

- What kind of variation does this allow?

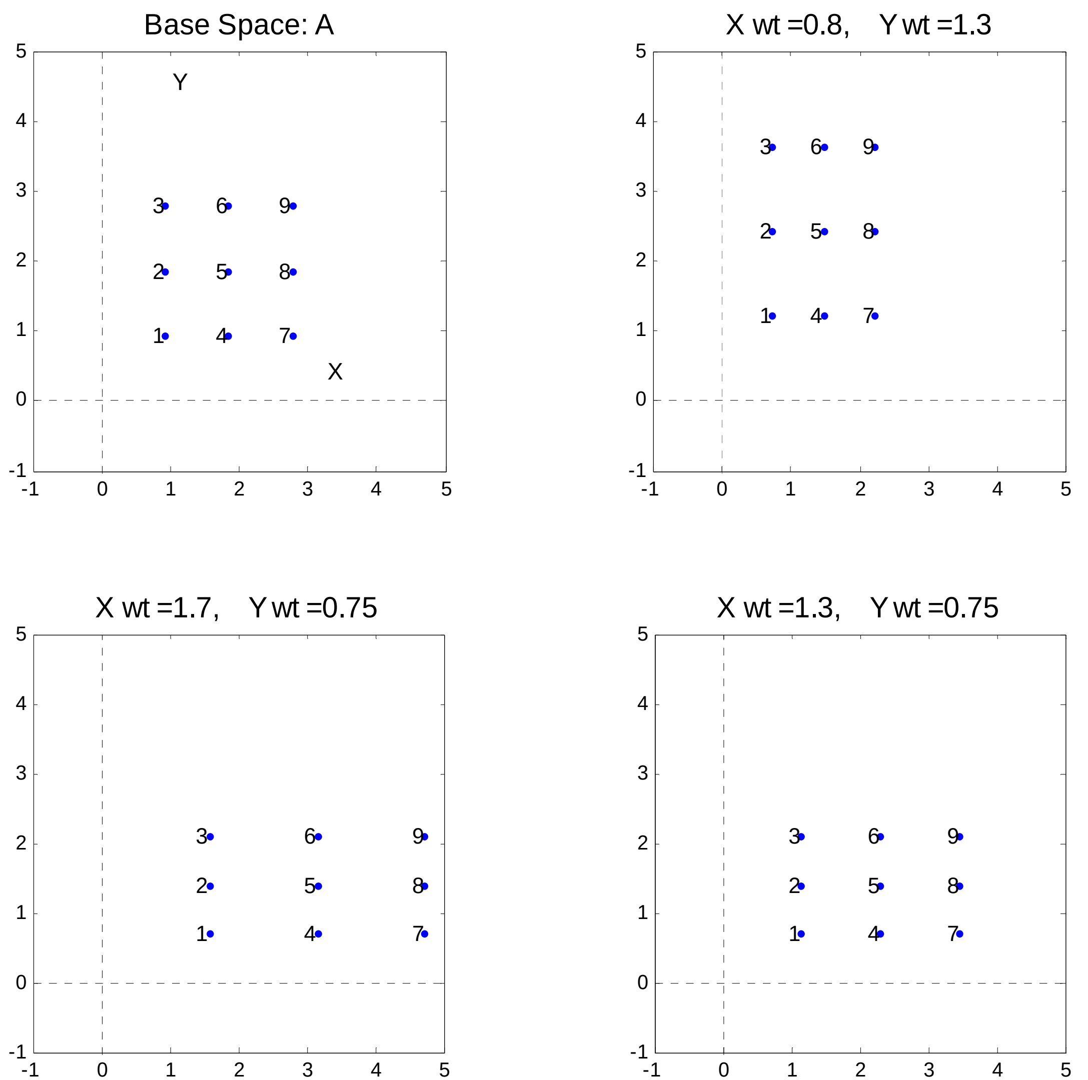
Review:**
Variations across slices of an array

- Two kinds of variation can occur, alone or in combination: a) variations in the length of factor axes or basis vectors (which can be represented by Parafac and the Tucker models), and/or b) variations in their skew or orientation relative to the points (which can be represented by the Tucker models).

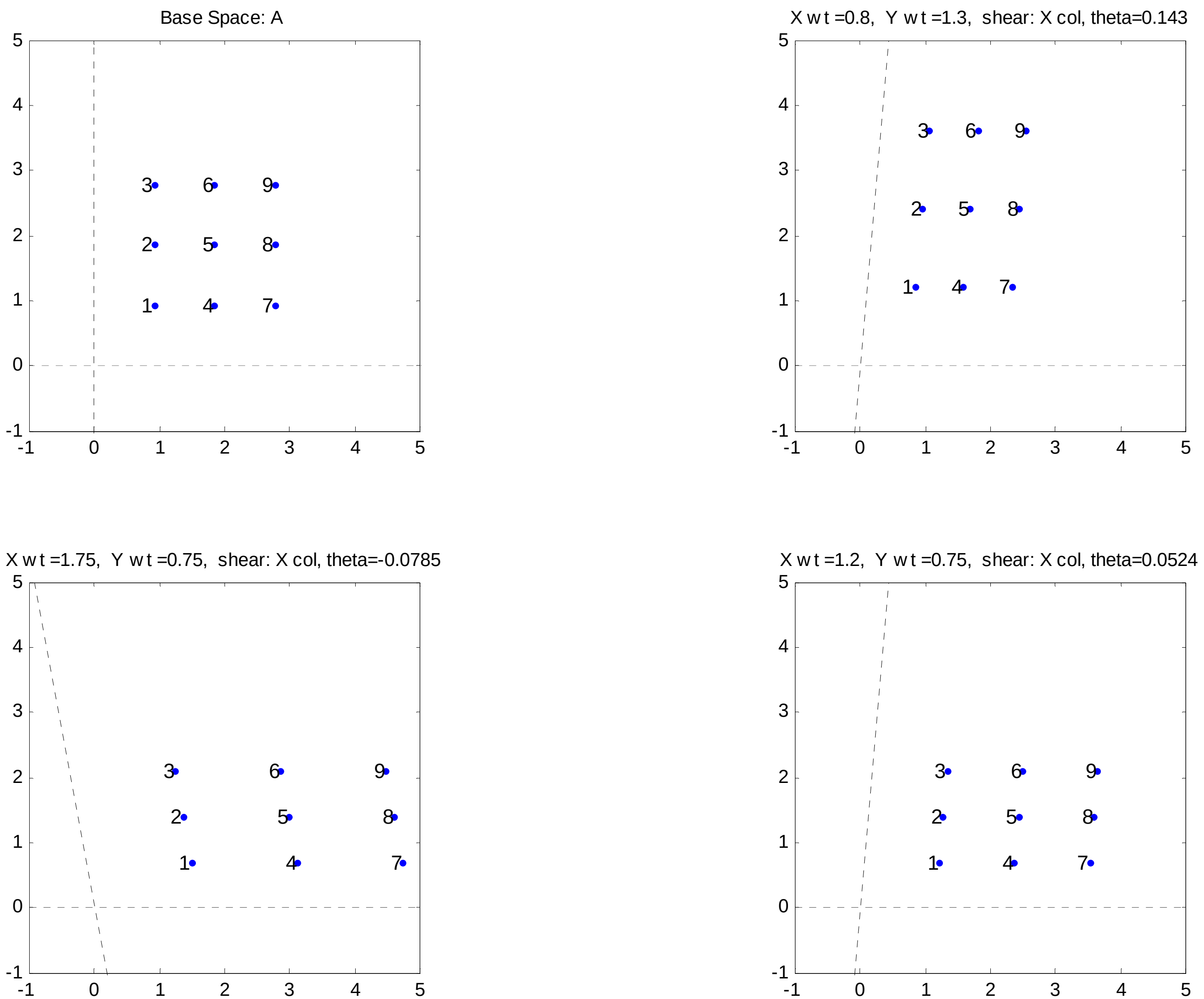
- In psychology, for example, changes in factor length would correspond to increases or decreases in psychological importance or impact of a given dimension, whereas changes in orientation would correspond to changes in "character" or "overtones of meaning" of a dimension.

- Next 2 slides illustrate how the models represent the two types of variation

Picture of Parafac variation: axis reweighting only for 2 factors X and Y**



Picture of "Tucker-variation": axis weight *plus* skew variation for 2 factors**



Enough for now...

Thank you.