# Hidden Parafac2

(in progress)

Sungjin Hong
University of Illinois at Urbana-Champaign

Acknowledgment: Marg Lundy and Richard Harshman
Heungsun Hwang

TRICAP 2006 Chania, Crete, Greece; June 8

# Overview



- Heterogeneity in component analysis
- Parafac2 for multiple groups
- Fuzzy c-lines (or clusterwise regression)
- Parafac2 with unknown grouping
- Weighted ALS for Hidden Parafac2
- A simulation
- Discussion

- Multiple-groups analysis, e.g., in Confirmatory Factor Analysis (CFA)
  - Different sets of loading matrices are inferred according to a priori known grouping:

    *e.g., distinctive loading patterns of components underlying political perception variables on voting, by groups of different political affiliations*

  - Some loadings might be constrained to be equal across groups

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

$\mathbf{A}_k$

| | | |
|---|---|---|
| x | 0 | 0 |
| x | 0 | 0 |
| x | 0 | 0 |
| 0 | x | 0 |
| 0 | x | 0 |
| 0 | x | 0 |
| 0 | 0 | x |
| 0 | 0 | x |
| 0 | 0 | x |

# Parafac2 for multiple groups



- Parafac2 can be considered as a "constrained" multiple-groups component model with

  – invariant angles between component "score" vectors, $\mathbf{\Phi} = \mathbf{A}_k' \mathbf{A}_k$

  – essentially invariant, but systematically reweighted loading matrix $\mathbf{B}\langle \mathbf{c}_k \rangle$, $\langle \mathbf{c}_k \rangle \equiv \mathrm{diag}\left( \mathbf{c}_k \right)$ -- weights for group $k$ in mode C;

    *e.g., loadings of "national security" component are weighted more by Republicans than by Democratics*

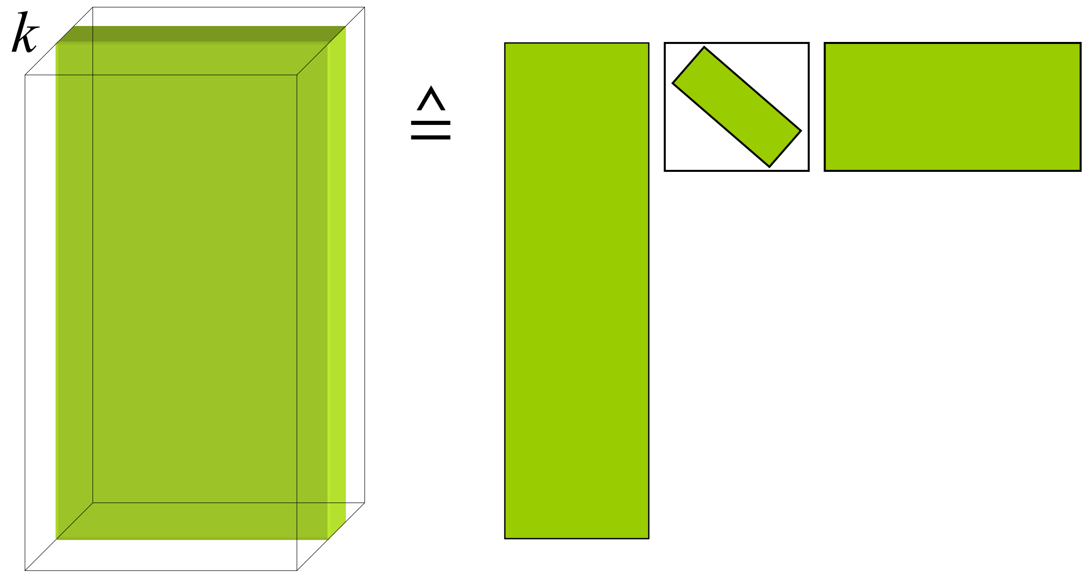
# Three-way Parafac2



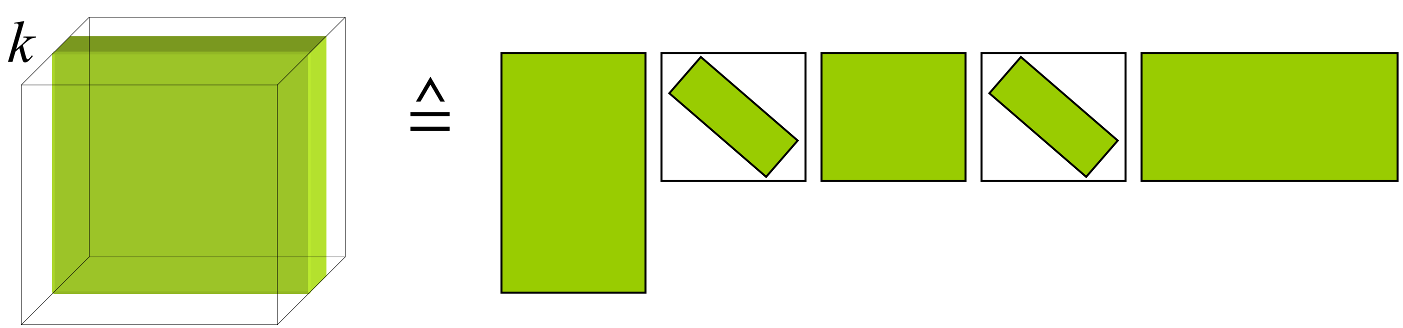
- Direct fitting form:

$$\mathbf{X}_k = \mathbf{A}_k \langle \mathbf{c}_k \rangle \mathbf{B}' + \mathbf{E}_k$$

$$\mathbf{A}_k' \mathbf{A}_k = \boldsymbol{\Phi}, \quad k = 1, \ldots, K$$




- Indirect fitting form: $\mathbf{X}_k' \mathbf{X}_k = \mathbf{B} \langle \mathbf{c}_k \rangle \boldsymbol{\Phi} \langle \mathbf{c}_k \rangle \mathbf{B}' + \mathbf{E}_k$



- Direct fitting form:
  - one grouping: $$\mathbf{X}_{kl} = \mathbf{A}_l \langle \mathbf{d}_l \rangle \langle \mathbf{c}_k \rangle \mathbf{B}' + \mathbf{E}_{kl}$$
    $$\mathbf{A}_l' \mathbf{A}_l = \mathbf{\Phi}, \quad l = 1, \ldots, L$$
  - two groupings: $$\mathbf{X}_{kl} = \mathbf{A}_{kl} \langle \mathbf{d}_l \rangle \langle \mathbf{c}_k \rangle \mathbf{B}' + \mathbf{E}_{kl}$$
    $$\mathbf{A}_{kl}' \mathbf{A}_{kl} = \mathbf{\Phi}, \quad k = 1, \ldots, K, \quad l = 1, \ldots, L$$
- Indirect fitting form:

$$\mathbf{X}_{kl}' \mathbf{X}_{kl} = \mathbf{B}_j \langle \mathbf{c}_k \rangle \langle \mathbf{d}_l \rangle \mathbf{\Phi} \langle \mathbf{d}_l \rangle \langle \mathbf{c}_k \rangle \mathbf{B}' + \mathbf{E}_{kl}$$

# Modeling for unknown heterogeneity

- As an analytic, descriptive approach, Bedzek's fuzzy c-lines (or clusterwise regression) identifies unknown heterogeneity in regression

  $K$ sets of parameters and "fuzzy" membership are alternately updated, continuously minimizing a weighted least-squares function; thus guaranteeing a local minimum

- Finite mixture approach models a set of scores as a mixture of $K$ distributions with unknown mixing probabilities

  These distributions are parametrically defined (e.g., Gaussian) and $K$ heterogeneous sets of model parameters and the mixing probabilities are estimated according to distributional properties (e.g., maximizing a joint likelihood function)

# ALS for fuzzy c-lines

- Part-worth regression weights are estimated per fuzzy cluster

$$\mathbf{y}_k = \mathbf{X}_k \mathbf{b}_k + \mathbf{e}_k, \qquad [\mathbf{y}_k \mid \mathbf{X}_k] = \langle \mathbf{u}_k \rangle^{0.5m} [\mathbf{y} \mid \mathbf{X}]$$

- Membership $\mathbf{U} = \{u_{ik}\}$ is updated, given regression weights $\mathbf{b}_k$ as

$$\hat{u}_{ik} = \left[ \sum_{k'=1}^{K} \left( \frac{e_{ik}}{e_{ik'}} \right)^{\frac{2}{m-1}} \right]^{-1}, \qquad e_{ik} = \| y_i - \mathbf{x}'_{ik} \mathbf{b}_k \|$$

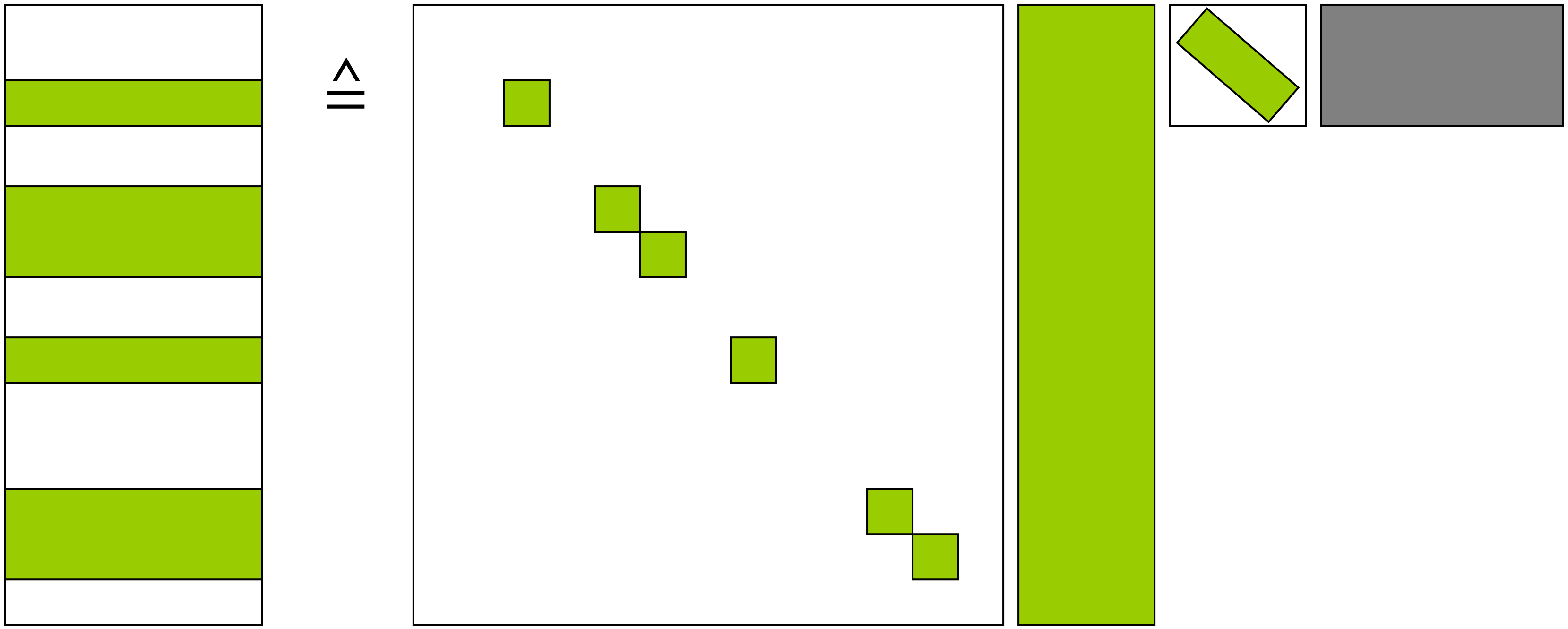
- These steps minimize a weighted LS function: $f = \sum_{i=1}^{I} \sum_{k=1}^{K} u_{ik}^{m} e_{ik}^{2}$

- A priori known “fuzzy weight” $m$ $(1 < m < \infty)$ determines <u>fuzziness</u> of clustering and the number of clusters $K$ is also to be provided

# Three-way Hidden Parafac2

- Suppose one suspects heterogeneous subgroups embedded in a data mode (e.g., those who like G.W.B. vs. don’t), over which there exists Parafac-type systematic factor variation

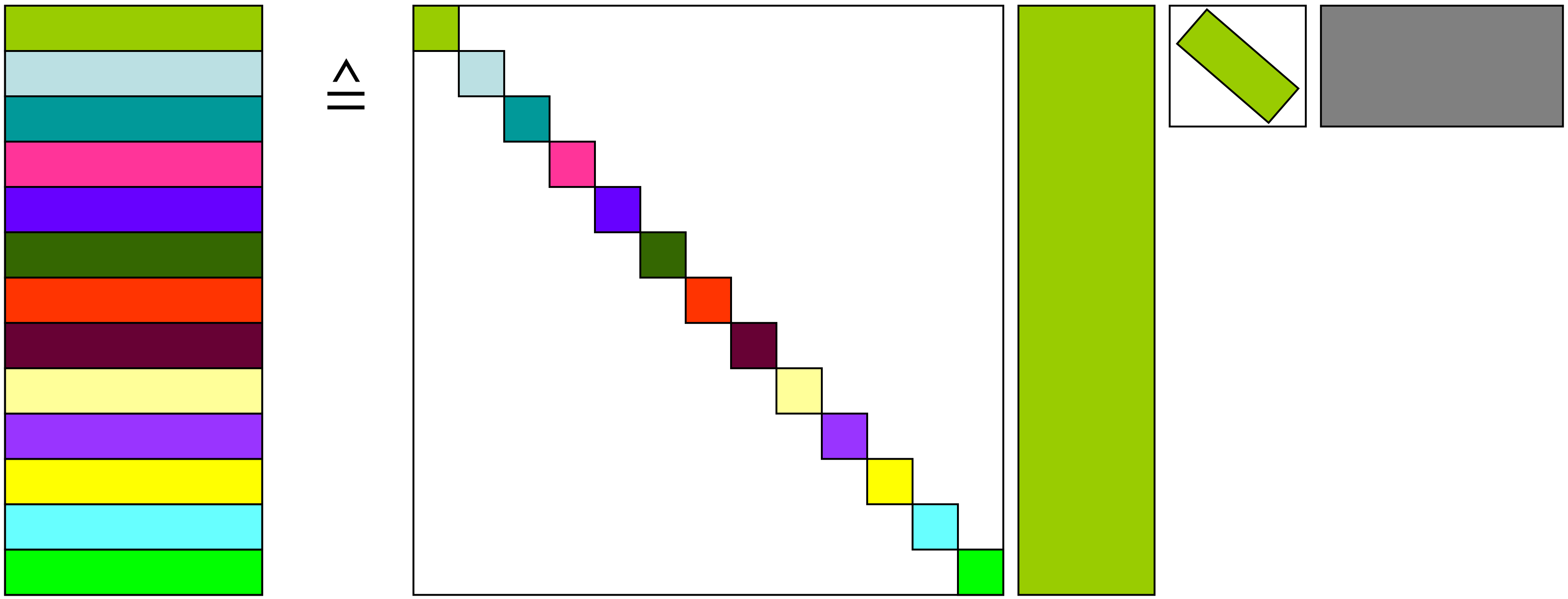
- Three-way Hidden Parafac2 fits Parafac2 to an optimal fuzzy clusters of two-mode data



$$\mathbf{X}_k = \mathbf{A}_k \langle \mathbf{c}_k \rangle \mathbf{B}' + \mathbf{E}_k, \quad \mathbf{A}'_k \mathbf{A}_k = \mathbf{\Phi}, \quad k = 1, \ldots, K$$

$$\mathbf{X}_k = \langle \mathbf{u}_k \rangle^{0.5m} \mathbf{X} \quad \text{subject to} \quad \sum_{k=1}^{K} u_{ik} = 1, \quad \sum_{i=1}^{I} u_{ik} > 0$$

- Clustering is hard or crisp if $u_{ik} = 0/1$ & fuzzy if $0 \leq u_{ik} \leq 1$



- Like the three-way case, one more mode is created by an optimal clustering of the disappearing mode; generating multiple partitions of a three-way data array

- Factor weights in two modes can easily be estimated by fitting three-mode Parafac to the original data, i.e., “stacked” data of the hidden four-mode data (if hard clustering assumed) as

$$\mathbf{X}_{(JK\times I)} = \left(\mathbf{C}\odot\mathbf{B}\right)\tilde{\mathbf{A}}', \quad \tilde{\mathbf{A}}' = \left[\, \tilde{\mathbf{A}}_1' \mid \cdots \mid \tilde{\mathbf{A}}_L' \,\right]$$

$$= \left[\langle \mathbf{d}_1 \rangle \mathbf{A}_1' \mid \cdots \mid \langle \mathbf{d}_L \rangle \mathbf{A}_L' \right]$$

$$\mathbf{A}_l'\mathbf{A}_l = \mathbf{\Phi}, \quad l = 1,\ldots,L$$

# Other methods of clustering in reduced space



- Factorial K-means -- Vichi & Kiers
- Similar clustering in the reduced space by Tucker3 (**A** and **G**) – Rocci & Vichi
- Candclus: Candecomp + binary constraints on component weights in any subset of modes – Carroll and colleagues
- Clusterwise GSCA (Generalized Structured Component Analysis) – Hwang & Takane
- And more…

# Weighted ALS for Hidden Parafac2

Step 1: Given a fixed $\mathbf{U}$, all weight matrices are updated by the directing fitting ALS algorithm for Parafac2 (Kiers, et al)

Step 2: Given all weight matrices fixed, membership is updated as in the fuzzy-clines step

- These steps minimize a weighted LS function:

$$f = \sum_{i=1}^{I} \sum_{k=1}^{K} u_{ik}^{m} e_{ik}^{2}, \qquad e_{ik}^{2} = \sum_{j=1}^{J} \left\| x_{ijk} - \mathbf{a}'_{ik} \langle \mathbf{c}_{k} \rangle \mathbf{b}_{j} \right\|^{2}$$

# Simulated data



- $\mathbf{A}_k \sim N(\mathbf{0}, \mathbf{\Phi})$; $\phi_{ii} = 1$, $\phi_{ii'} = 0$ or $0.5$, $I = 50$ for $k = 1, \ldots, 5$
- # of factors = 3
- $\mathbf{U}$: binary (i.e., hard clustering), 250 × 5
- factor weights in known other modes ($\mathbf{B}$ in three-way and $\mathbf{B}$ and $\mathbf{C}$ in four-way case): $\sim N(\mathbf{0}, \mathbf{I})$
- For fallible case, random noise (30%) added to the error-free data
- 5 replications per data condition, generating 2 × 2 × 5 sets of data

# Analytic conditions



- The current ALS algorithm needs to know at start at least some partial information; thus random numbers sampled from a uniform distribution (0,1) were added to the true membership with varying weights as

$$\mathbf{U}_{\mathrm{s}} = \mathbf{U}_{\mathrm{t}} + w\mathbf{U}_{\mathrm{e}}, \quad w = 0,\ 1 \text{ or } 2$$

- All other parameters were initialized at 10 sets of random numbers
- All fitting used $m$ = 1.3
- The algorithm stopped at 1000 iterations or parameters not changing more than $10^{-7}$ when scaled to unit norm

# Three-way Hidden Parafac2 results ($n$ = 5 per cell)



| | $\phi = 0$ | | | $\phi = 0.5$ | | |
|---|---|---|---|---|---|---|
| | *$w = 0$ | 1 | 2 | $w = 0$ | 1 | 2 |
| | error-free | | | | | |
| fit ($R^2$) | 1.000 | 0.999 | 0.997 | 1.000 | 0.997 | 0.999 |
| $\phi$ (MAD) | 0.146 | 0.365 | 0.423 | 0.041 | 0.089 | 0.212 |
| **B** ($r$)** | 0.993 | 0.907 | 0.912 | 0.998 | 0.985 | 0.940 |
| **C** ($r$) | 0.992 | 0.904 | 0.852 | 0.998 | 0.926 | 0.871 |
| **U** ($r$) | 0.873 | 0.796 | 0.550 | 0.940 | 0.784 | 0.629 |
| | error = 30% | | | | | |
| fit | 0.771 | 0.771 | 0.771 | 0.775 | 0.775 | 0.775 |
| $\phi$ | 0.118 | 0.497 | 0.139 | 0.190 | 0.132 | 0.189 |
| **B** | 0.996 | 0.944 | 0.860 | 0.981 | 0.935 | 0.915 |
| **C** | 0.991 | 0.899 | 0.895 | 0.983 | 0.951 | 0.909 |
| **U** | 0.807 | 0.602 | 0.492 | 0.749 | 0.643 | 0.492 |

* $w$ = weight of random numbers added to true membership values at start

** $r$ = congruence coefficient

# Four-way Hidden Parafac2 results ($n$ = 5 per cell) 

| | $\phi = 0$ | | | $\phi = 0.5$ | | |
|---|---|---|---|---|---|---|
| | $w = 0$ | 1 | 2 | $w = 0$ | 1 | 2 |
| | error-free | | | | | |
| fit ($R^2$) | 1.000 | 1.000 | 0.996 | 1.000 | 1.000 | 0.997 |
| $\phi$ (MAD) | 0.000 | 0.000 | 0.007 | 0.000 | 0.000 | 0.122 |
| **D** ($r$) | 1.000 | 1.000 | 0.960 | 1.000 | 1.000 | 0.972 |
| **U** ($r$) | 1.000 | 1.000 | 0.725 | 0.992 | 1.000 | 0.716 |
| **B** ($r$) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **C** ($r$) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | error = 30% | | | | | |
| fit | 0.743 | 0.743 | 0.743 | 0.740 | 0.740 | 0.739 |
| $\phi$ | 0.015 | 0.015 | 0.013 | 0.008 | 0.016 | 0.121 |
| **D** | 0.994 | 0.994 | 0.965 | 0.997 | 0.995 | 0.957 |
| **U** | 0.786 | 0.786 | 0.600 | 0.773 | 0.767 | 0.588 |
| **B** | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| **C** | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

# Discussion



- The current WALS algorithm works when some fallible information available for the hidden membership
- A rational start of membership for cases when no information whatsoever available for the optimal grouping?
- What if a preprocessing necessary according to the hidden membership?

  Optimal rescaling and centering might be incorporated into the model