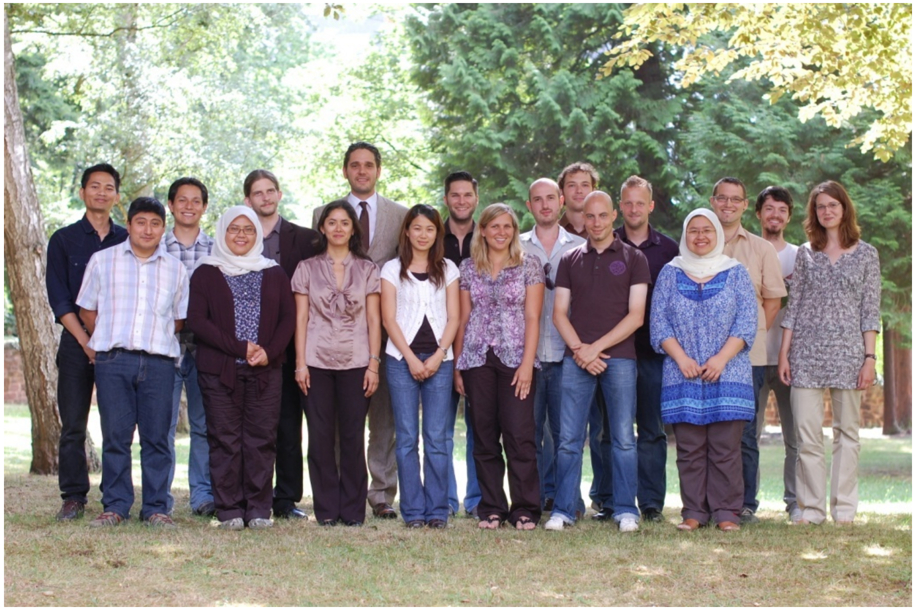
Systems Neuroscience & Neurotechnology Unit

Neurocenter, Saarland University Hospital

UNIVERSITÄT DES SAARLANDES

Hochschule für Technik und Wirtschaft des Saarlandes

HTW



Blind Source Separation - PCA, ICA, PARAFAC

What is PARAFAC?

- Multidimensional decomposition technique
- Decomposes three (or higher) -dimensional data into a series of distinct „atoms“ or components (**Smilde et al. (2004)**).
- Extension of factor analysis to higher orders

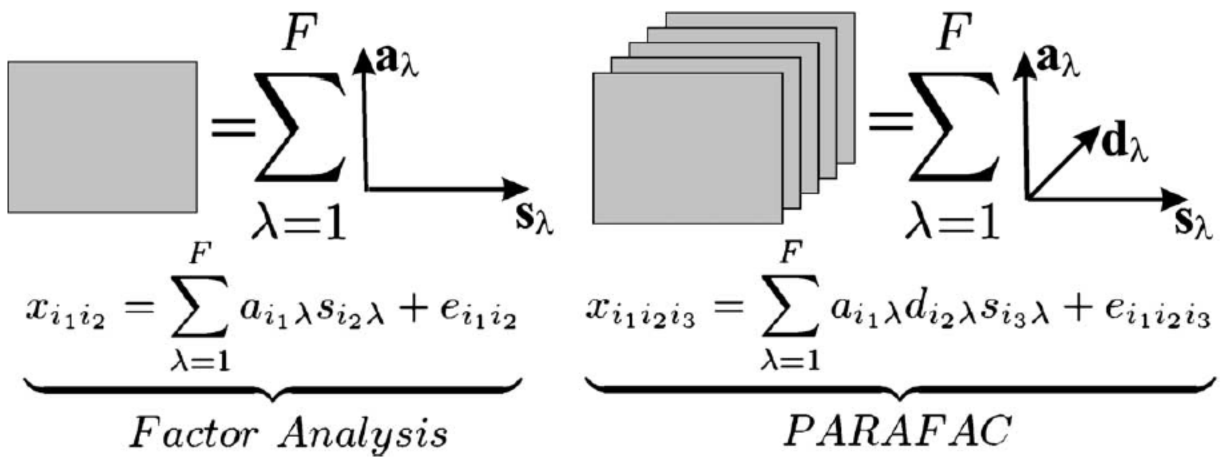


Fig. 1. Graphical representation of the factor analysis to the left and the PARAFAC decomposition of a 3-way array to the right. Like the factor analysis, PARAFAC decomposes the data into factor effects pertaining to each modality. *F* denotes the number of factors.

Historical Background

- Independently proposed by
  - Harshman (1970) : Parallel Factor Analysis (**PARAFAC**)
  - Carrol and Chang (1970): Canonical Decomposition (**CANDECOMP**)
- Based on the principle of "*Parallel Proportional Profiles*" (**Cattell (1944)**)
  - Two data matrices with the same variables should contain the same components and in the second matrix each „component should be accentuated or reduced in influence"(**Cattel,1944,p.274**)
  - Principle constrains the PARAFAC method

$$\mathbf{X}^{(i)} = \mathbf{A}\mathbf{D}^{(i)}\mathbf{S} + \mathbf{E}^{(i)} \quad \text{, where} \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(M)} \end{bmatrix} \text{, and } \mathbf{D}^{(i)} \text{ is a diagonal matrix}$$

Fields of Application

- Psychometrics
  - Analyzing multiple (dis)similarity matrices of a variety of subjects (**Caroll at al. (1970)**)
- Chemometrics
  - Fluorescence emission spectra (**Bro et al. (1996)**)
- Neuroscience
  - Decomposition of ERP data (**Morup et al. (2005)**)
  - Identification of activity in specific frequency bands (**Miwakeichi et al. (2004)**)
  - Localization of the seizure onset zone in epileptic data (**Acar et al. (2007)**)
  - fMRI data (**Martinez-Montes et al. (2004)**)
- …

Tensors (I)

- Multidimensional array

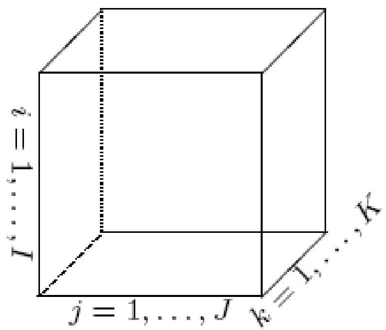
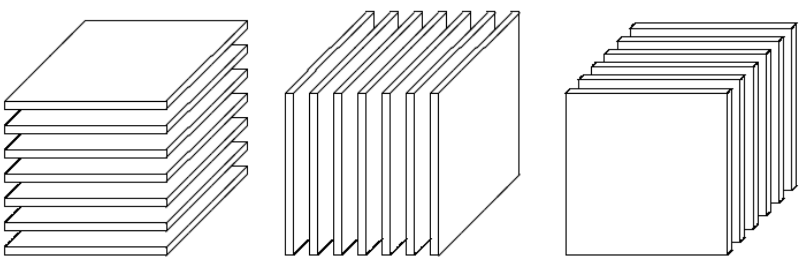


Fig. 1.1: A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I\times J\times K}$



(a) Horizontal slices: $\mathbf{X}_{i::}$ (b) Lateral slices: $\mathbf{X}_{:j:}$ (c) Frontal slices: $\mathbf{X}_{::k}$ (or $\mathbf{X}_k$)

- Tensor unfolding

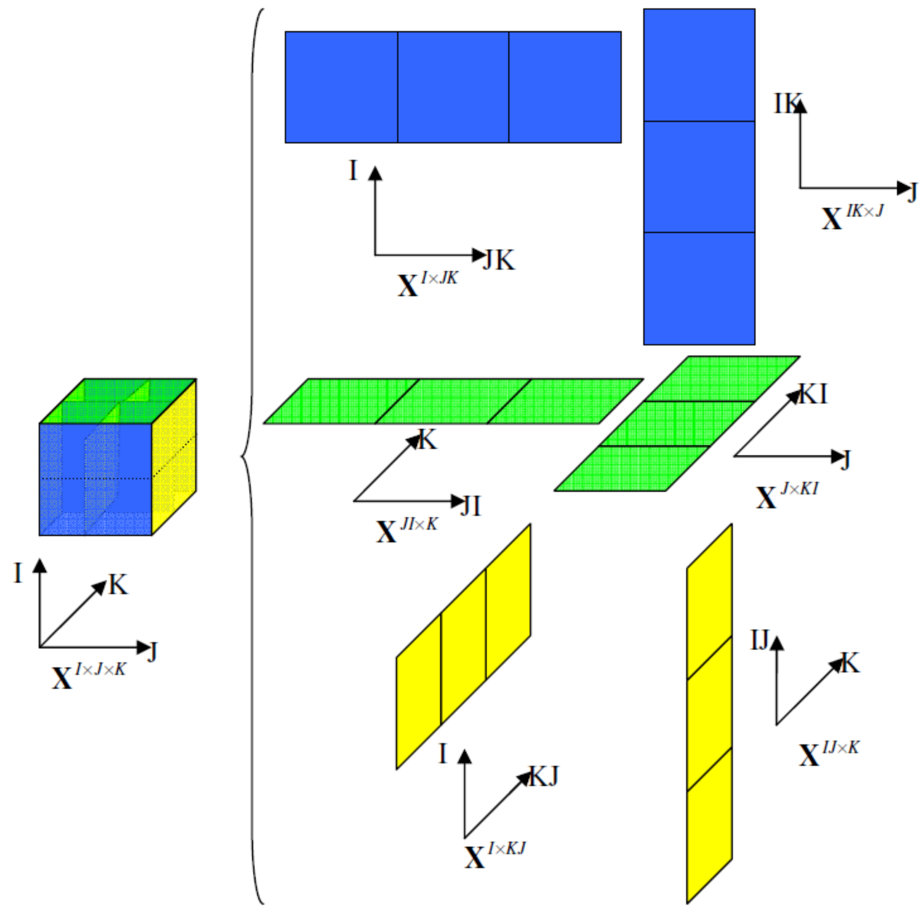


Figure 1.4: The six ways of unfolding the three-way array $\boldsymbol{\mathcal{X}}$ into a matrix.

Tensors (II)

- Rank-one tensors An N-way tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is *rank one* if it can be written as the outer product of N vectors, i.e.,

$$\mathcal{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}.$$

The symbol "$\circ$" represents the vector outer product. This means that each element of the tensor is the product of the corresponding vector elements:

$$x_{i_1 i_2 \cdots i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \cdots a^{(N)}_{i_N} \quad \text{for all } 1 \leq i_n \leq I_n.$$

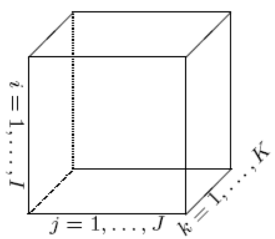


Fig. 1.1: A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

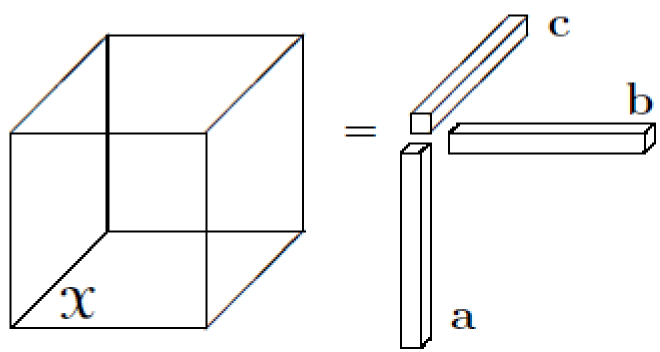
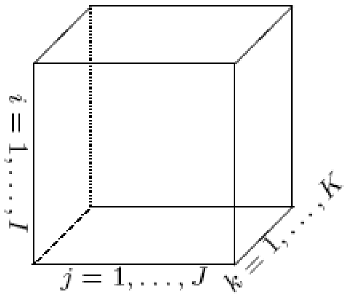


Fig. 2.3: Rank-one third-order tensor, $\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$. The (i, j, k) element of $\mathcal{X}$ is given by $x_{ijk} = a_i b_j c_k$.

Definition of CANDECOMP/PARAFAC (CP) (I)

The CP decomposition factorizes a tensor into a sum of component rank-one tensors. For example, given a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, we wish to write it as



$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \tag{3.1}$$

where R is a positive integer, and $\mathbf{a}_r \in \mathbb{R}^I$, $\mathbf{b}_r \in \mathbb{R}^J$, and $\mathbf{c}_r \in \mathbb{R}^K$, for $r = 1, \dots, R$. Elementwise, (3.1) is written as

$$x_{ijk} \approx \sum_{r=1}^{R} a_{ir}\, b_{jr}\, c_{kr}, \quad \text{for } i = 1, \dots, I,\ j = 1, \dots, J,\ k = 1, \dots, K.$$

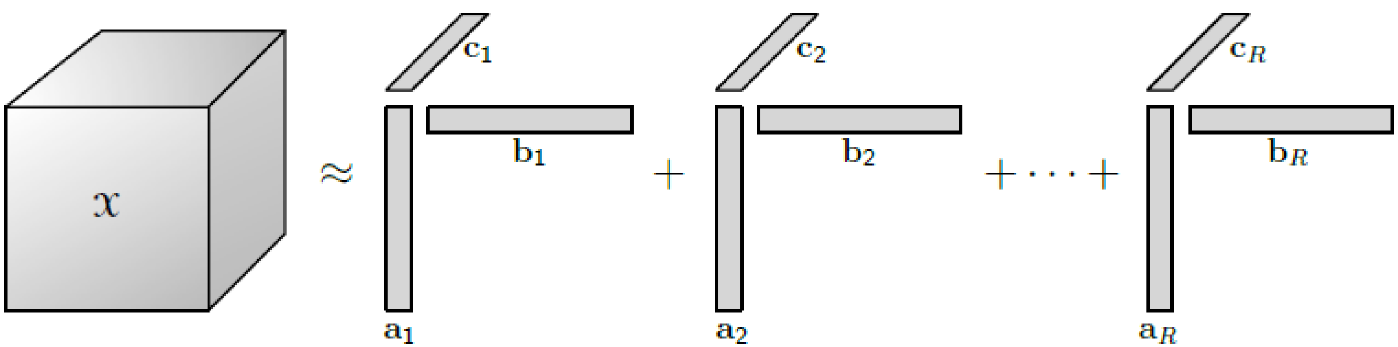


Fig. 3.1: CP decomposition of a three-way array.

Definition of CANDECOMP/PARAFAC (CP) (II)

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \tag{3.1}$$

The *factor matrices* refer to the combination of the vectors from the rank-one components, i.e., $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_R \end{bmatrix}$ and likewise for $\mathbf{B}$ and $\mathbf{C}$. Using these definitions, (3.1) may be written in matricized form

dimension:

dimension	
I x JK	$\mathbf{X}_{(1)} \approx \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$,
J x KI	$\mathbf{X}_{(2)} \approx \mathbf{B}(\mathbf{C} \odot \mathbf{A})^{\mathsf{T}}$,
K x JI	$\mathbf{X}_{(3)} \approx \mathbf{C}(\mathbf{B} \odot \mathbf{A})^{\mathsf{T}}$.

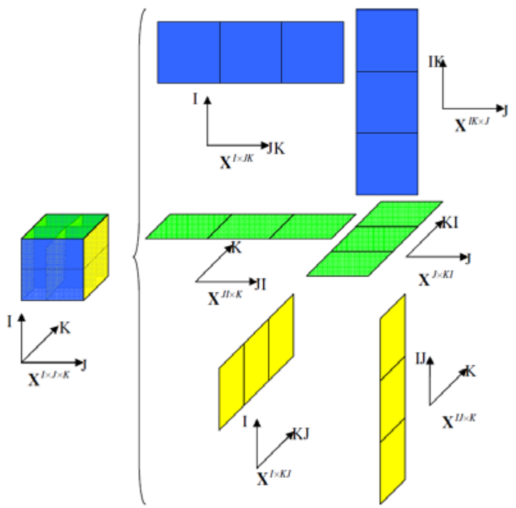


Figure 1.4: The six ways of unfolding the three-way array $\boldsymbol{\mathcal{X}}$ into a matrix.

The *Khatri-Rao product* [200] is the "matching columnwise" Kronecker product. Given matrices $\mathbf{A} \in \mathbb{R}^{I \times K}$ and $\mathbf{B} \in \mathbb{R}^{J \times K}$, their Khatri-Rao product is denoted by $\mathbf{A} \odot \mathbf{B}$. The result is a matrix of size $(IJ) \times K$ and defined by

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_K \otimes \mathbf{b}_K \end{bmatrix}.$$

The *Kronecker product* of matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{K \times L}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$. The result is a matrix of size $(IK) \times (JL)$ and defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_1 \otimes \mathbf{b}_2 & \mathbf{a}_1 \otimes \mathbf{b}_3 & \cdots & \mathbf{a}_J \otimes \mathbf{b}_{L-1} & \mathbf{a}_J \otimes \mathbf{b}_L \end{bmatrix}.$$

The three-way model is sometimes written in terms of the frontal slices of $\mathcal{X}$

$$\mathbf{X}_k \approx \mathbf{A}\mathbf{D}^{(k)}\mathbf{B}^{\mathsf{T}}$$

where $\mathbf{D}^{(k)} \equiv \operatorname{diag}(\mathbf{c}_{k:})$

for $k = 1, \ldots, K$.

Graphical Illustration

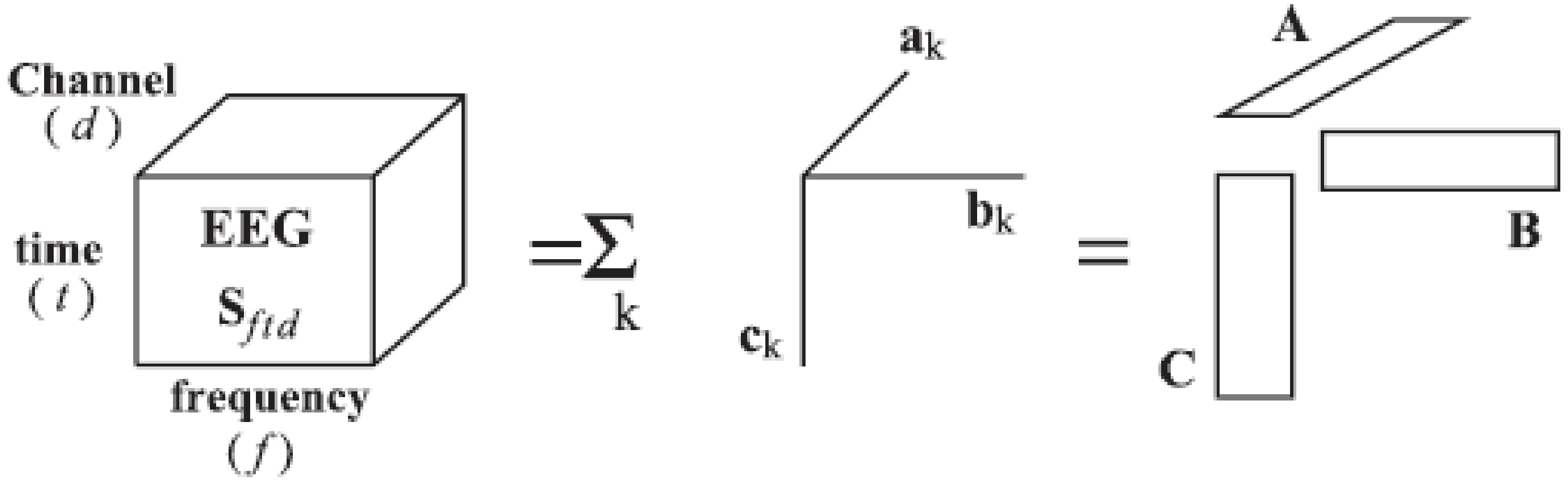
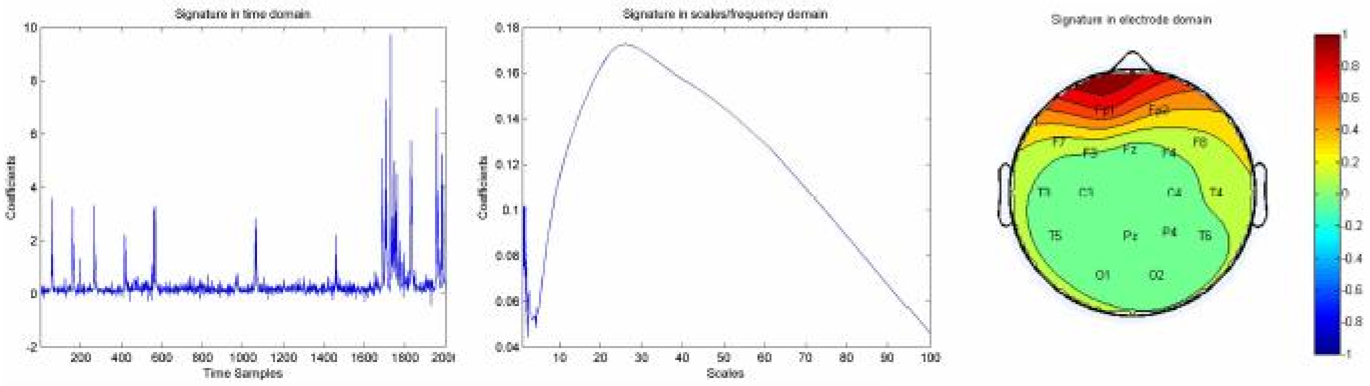


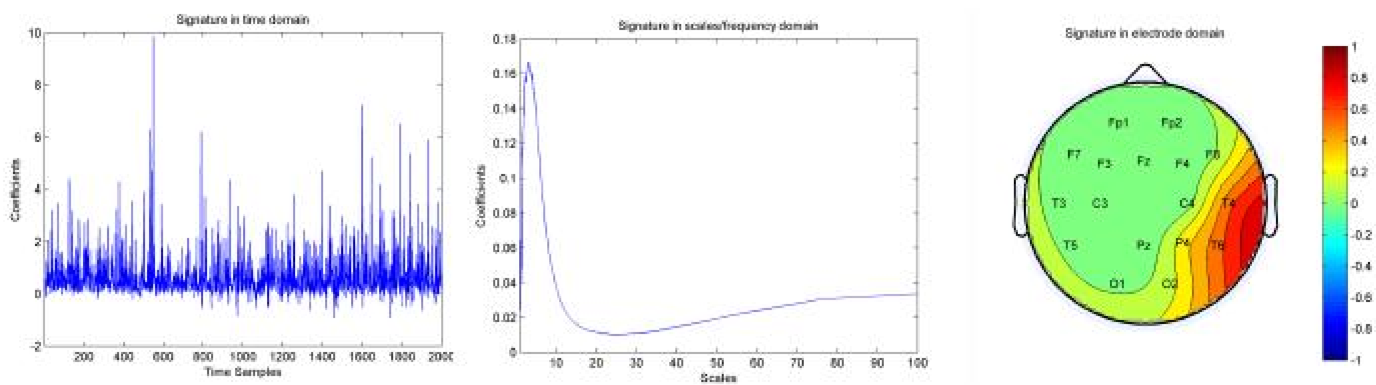
Fig. 1. Graphical explanation of the PARAFAC model. The multichannel EEG evolutionary spectrum **S** is obtained from a channel by channel wavelet transform. **S** is a three-way data array indicated by channel, frequency, and time. PARAFAC decomposes this array into the sum of "atoms". The *k*th atom is the tri-linear product of loading vectors representing spatial ($\mathbf{a}_k$), spectral ($\mathbf{b}_k$), and temporal ($\mathbf{c}_k$) "signatures". Under these conditions, PARAFAC can be summarized as finding the matrices $\mathbf{A} = \{\mathbf{a}_k\}$, $\mathbf{B} = \{\mathbf{b}_k\}$, and $\mathbf{C} = (\mathbf{c}_k)$, which explain **S** with minimal residual error.

Example

$$\underline{X} = a_1 \circ b_1 \circ c_1 + a_2 \circ b_2 \circ c_2 + \underline{E}$$



$$\underline{X} = a_1 \circ b_1 \circ c_1 + a_2 \circ b_2 \circ c_2 + \underline{E}$$



Acar et al. (2007)

Uniqueness of the CP components

- Kruskal‘s condition for essential uniqueness
  - k-rank of matrix A (size: n x r): maximum number of columns of the matrix A that are linearly independent (**Kruskal (1977)**)
  - Condition:

$$2R + 2 \leq k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}}$$

Computing the CP model

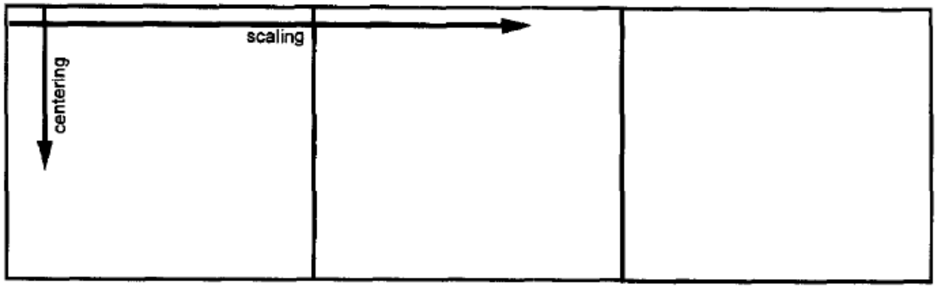
1. Preprocessing
   - Centering

Centering the first mode can be done by unfolding the calibration array to an $I \times JK$ matrix, and then center this matrix as in ordinary PCA:

$$x_{ijk}^{\text{cent}} = x_{ijk} - \overline{x_{jk}}$$

where

$$\overline{x_{jk}} = \frac{\sum_{i=1}^{I} x_{ijk}}{I}$$



   - Scaling

$$x_{ijk}^{\text{scal}} = \frac{x_{ijk}}{s_i}$$

where s_i can be defined as

$$s_i = \sqrt{\left(\sum_{j=1}^{J} \sum_{k=1}^{K} x_{ijk}^2\right)}$$

Computing the CP model

2. Determine the number of (rank-one) components
   – Fit multiple CP decompositions with different numbers of components

$$\text{Fit percentage} = \frac{\|\underline{\mathbf{X}}\|^2 - \|\underline{\mathbf{E}}\|^2}{\|\underline{\mathbf{X}}\|^2} \cdot 100$$

- Choose R such that adding more components does not significantly increase the fit in percentage

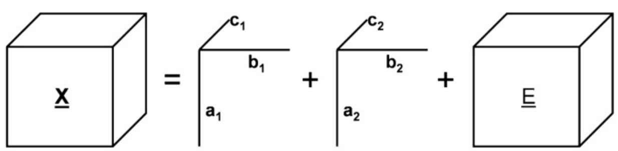
- Example:

R	1	2	3	4	5
Fit %	10.2	16.4	18.7	18.9	19.1

Computing the CP model (II)

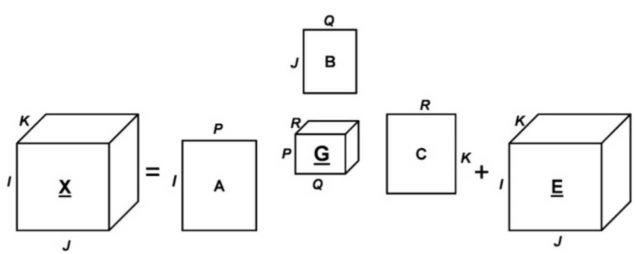
2. Determine the number of (rank-one) components
   – Core Consistency Diagnostic (CORCONDIA) (Bro et al., 2003)
     - The core consistency quantifies the resemblance between a Tucker3 and a PARAFAC core
     - PARAFAC core T is zero apart from along the superdiagonal which has ones
     - Tucker3 core G is calculated by inserting the obtained factor matrices A,B, and C of the PARAFAC model

PARAFAC



$$Core\,Consistency = 100 \cdot \left(1 - \frac{\sum_{d=1}^{F}\sum_{e=1}^{F}\sum_{f=1}^{F}\left(t_{def} - g_{def}\right)^2}{\sum_{d=1}^{F}\sum_{e=1}^{F}\sum_{f=1}^{F} g_{def}^{\;2}} \right)$$

Tucker3



Core consistency > 90%: PARAFAC model with the specified number of components would be an appropriate model for the data.

Computing the CP model (III)

3. Compute CP decomposition via Alternating Least Squares (besides others)

- Initialize all model parameters randomly (or eigenvalue decompositions)
- Update each parameter by minimizing a cost function with respect to the parameter while holding all other parameters fixed
- PARAFAC model $\mathbf{X}^{I \times JK} = \mathbf{A}\left(\mathbf{S}|\otimes|\mathbf{B}\right)^T$
- Cost function

$$\min \left\| \mathbf{X}^{I \times JK} - \mathbf{A}\left(\mathbf{S}|\otimes|\mathbf{B}\right)^T \right\|^2$$

- Stopping criterion
  - Number of iterations
  - Little change in factor matrices, objective function

ALSPARAFAC

Initialize B and S

iter=0, $\Delta SSE > \varepsilon$, $SSE_0 = 0$

While iter<Criterion & $\Delta SSE > \varepsilon$

iter=iter+1

$$\mathbf{Z} = \left(\mathbf{S}|\otimes|\mathbf{B}\right)$$

$$\mathbf{A} = \mathbf{X}^{(I \times JK)}\mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^+$$

$$\mathbf{Z} = \left(\mathbf{S}|\otimes|\mathbf{A}\right)$$

$$\mathbf{B} = \mathbf{X}^{(J \times IK)}\mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^+$$

$$\mathbf{Z} = \left(\mathbf{B}|\otimes|\mathbf{A}\right)$$

$$\mathbf{S} = \mathbf{X}^{(K \times IJ)}\mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^+$$

$$SSE_{iter} = \left\| \mathbf{X} - \mathbf{A}\left(\mathbf{B}|\otimes|\mathbf{S}\right)^T \right\|^2$$

$$\Delta SSE = \left| SSE_{iter} - SSE_{iter-1} \right|$$

Example

Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG (Mørup et al., 2005)

- PARAFAC used to decompose wavelet transformed event-related EEG
- 64 – Electrode – EEG
- Visual presentation of black white drawings (Objects, Non-Objects)
- Subjects respond by mouseclick if picture is Object or not (Ob, Nob)

Decompositions of EEG

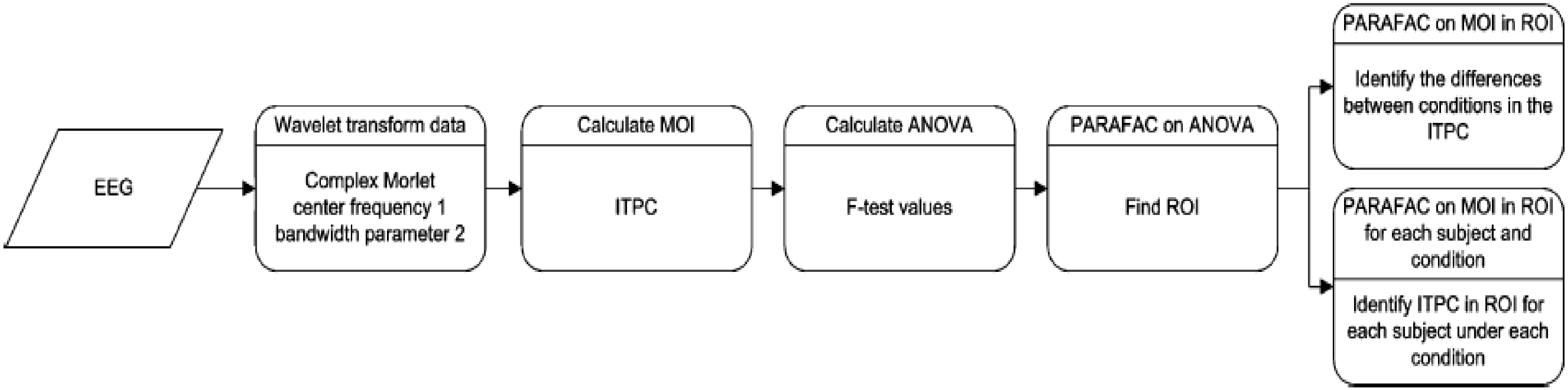
- multi-way decompositions:
  - Multi-way array of F test values from analysis of variance (ANOVA)
  - 3-way array (c,f,t) of inter-trial phase coherence (ITPC)

$$\mathrm{ITPC}\,(c,f,t) = \left| \frac{1}{n} \sum_{e=1}^{n} \frac{\mathbf{X}_e(c,f,t)}{|\mathbf{X}_e(c,f,t)|} \right|$$

Also referred to as phase locking factor, measure of evoked activity
- ITPC value of 1 indicates perfect phase coherence in all epochs
- random noise on the average has a coherence of $\frac{1}{\sqrt{n}}$

  - 5-way array of ITPC, additional modalities: subject and condition

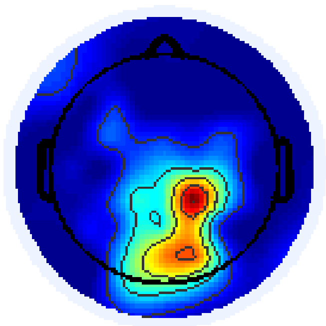
Flow chart of Analysis

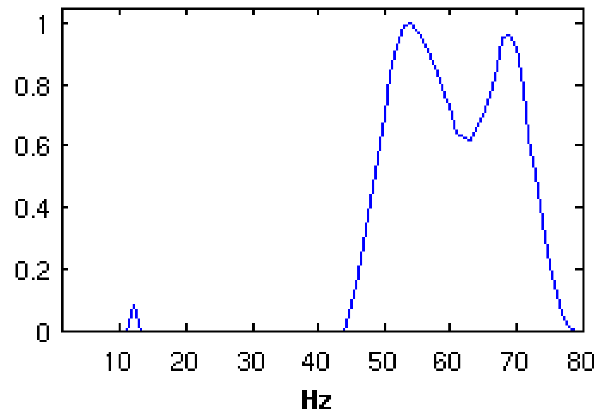


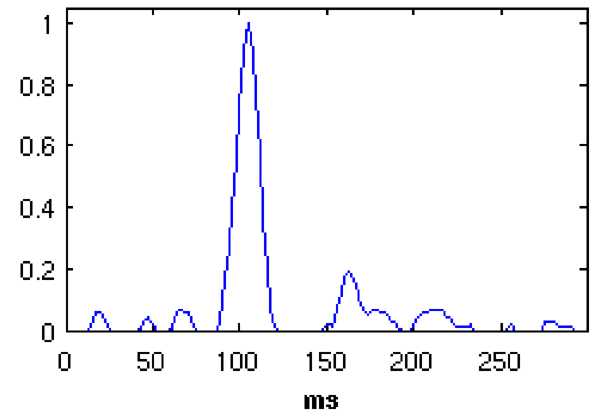
1. Wavelet transform of EEG-Data
2. MOI = Measure of interest (i.e. ITPC) calculated for each subject and each condition
3. 3-way array of ANOVA F test value calculated
4. F test arrays analyzed using PARAFAC to find region of interest (ROI) , region of most difference between 2 conditions
   (here time-frequency domain)
   - ROI analyzed using PARAFAC on 5-way array
     (channel x frequency x time x subject x condition) of ITPC
   - ROI analyzed using PARAFAC on 3-way array
     (channel x frequency x time) of ITPC

Results

- The 3-way array PARAFAC analysis of ANOVA





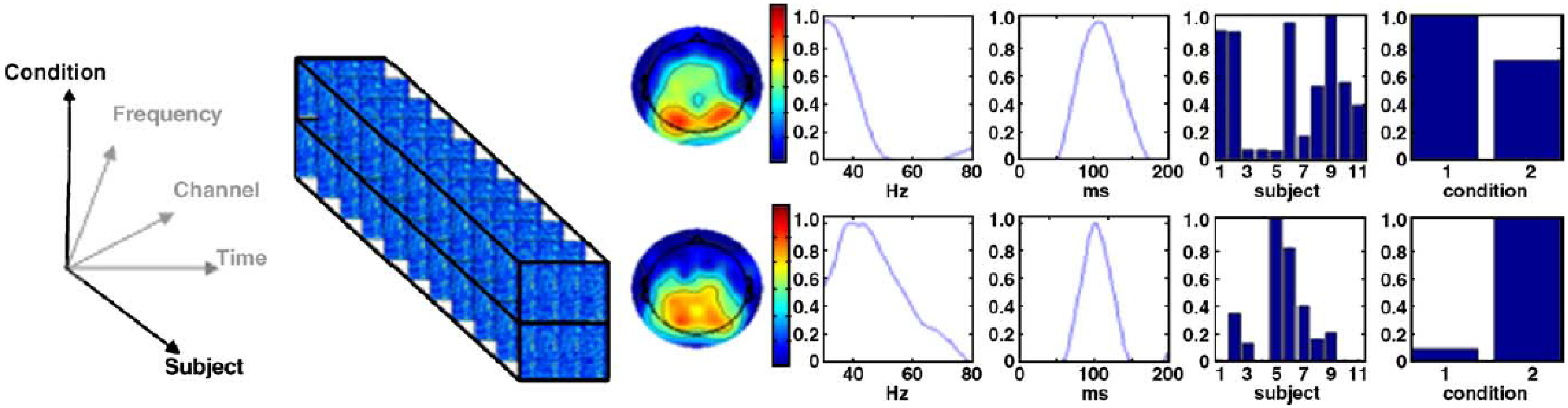


- testing the difference between the two conditions (Ob and Nob) for all subjects
- mainly in the gamma band around 40– 80 Hz
- occipital region
- at about 100 ms

  → fits expected data from previous studies

Results

- PARAFAC for ITPC 5-way array *ITPC*



- Result: two-component PARAFAC model
- <u>*First component (first row):*</u>

  occipital activity at ~ 30Hz and 100ms, in all subjects (not strong in subjects 3,4,5), in both conditions (stronger in 1 = Ob)

  → fits expected results
- <u>*Second component (second row):*</u>

  More anteriorly, higher frequency and 100ms, not present in 1,4,10,11, almost only in condition 2 (= Nob) → Doesn't fit expected results

Summary

- PARAFAC is a multidimensional decomposition technique
- Tensor can be decomposed into a sum of rank-one components
- Number of components can be determined via CORCONDIA or via percentage of fit
- PARAFAC model can be calculated via Alternating Least Squares Algorithm

Thank you for your attention!