SAND2009-4210C

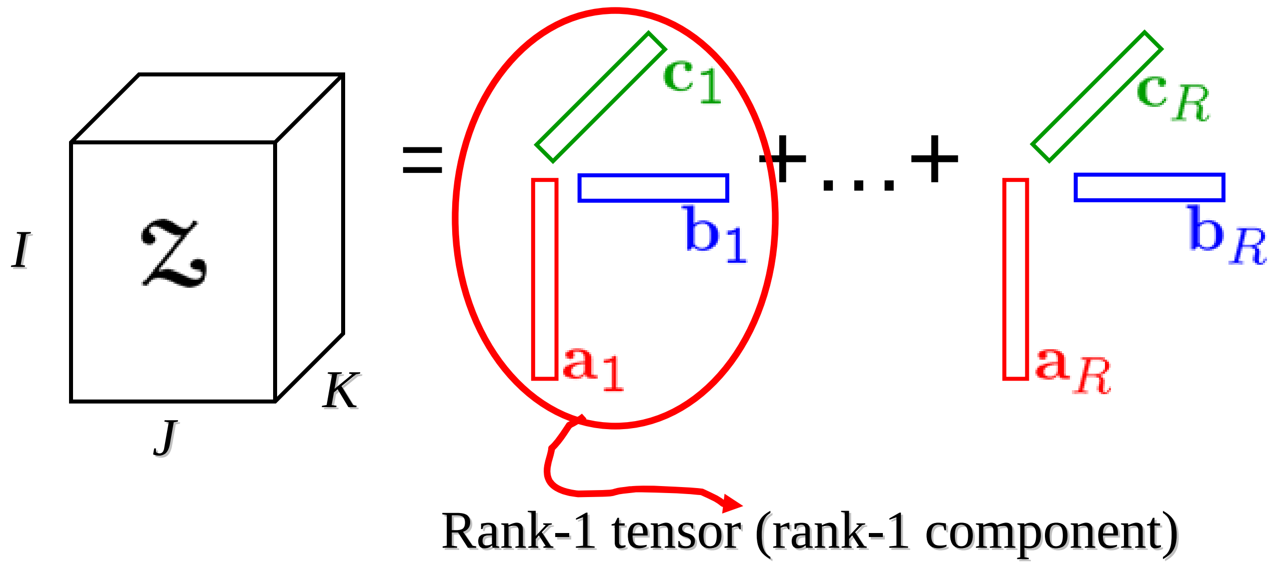
An Optimization Approach for Fitting Canonical Tensor Decompositions

Evrim Acar, Tamara G. Kolda and Daniel M. Dunlavy
Sandia National Labs

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

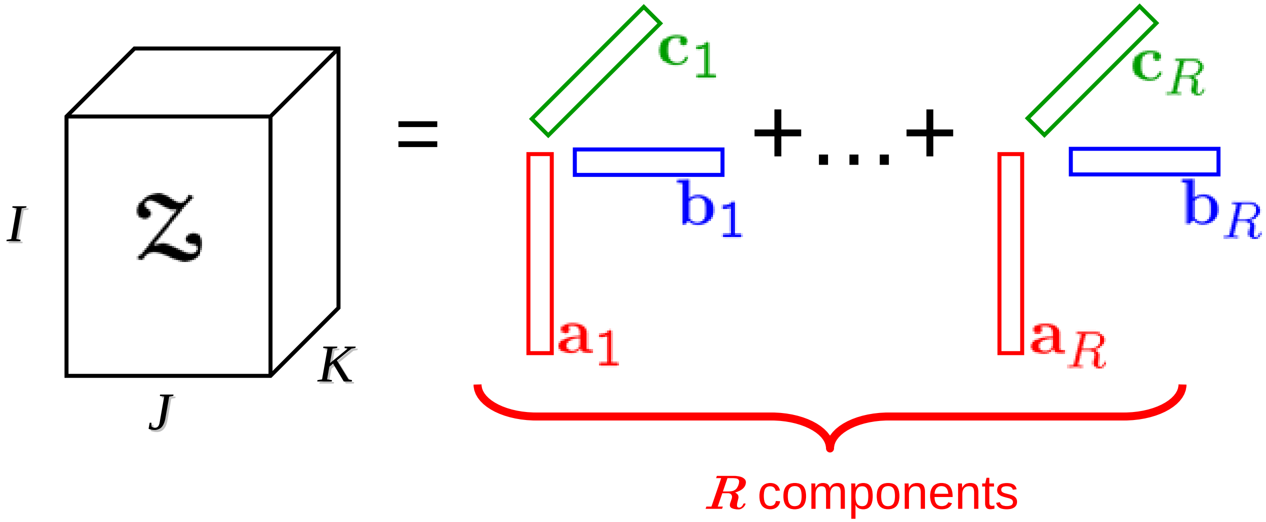
What is Canonical Tensor Decomposition?

CANDECOMP/PARAFAC (CP) model [Hitchcock'27, Harshman'70, Carroll & Chang'70]



What is Canonical Tensor Decomposition?

CANDECOMP/PARAFAC (CP) model [Hitchcock'27, Harshman'70, Carroll & Chang'70]



$$\mathcal{Z} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

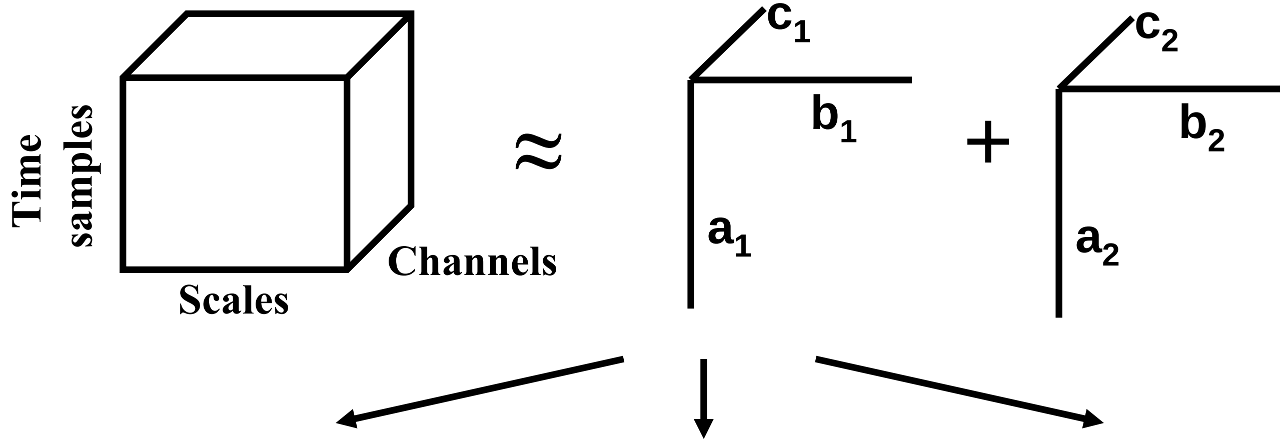
$$\mathbf{A} \in \mathbb{R}^{I \times R} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_R \end{bmatrix}$$

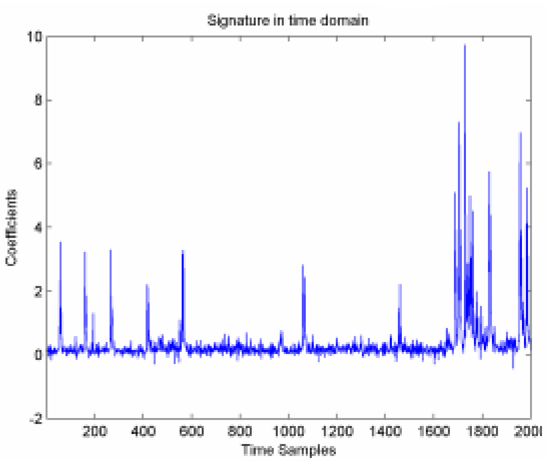
$$\mathbf{B} \in \mathbb{R}^{J \times R} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix}$$

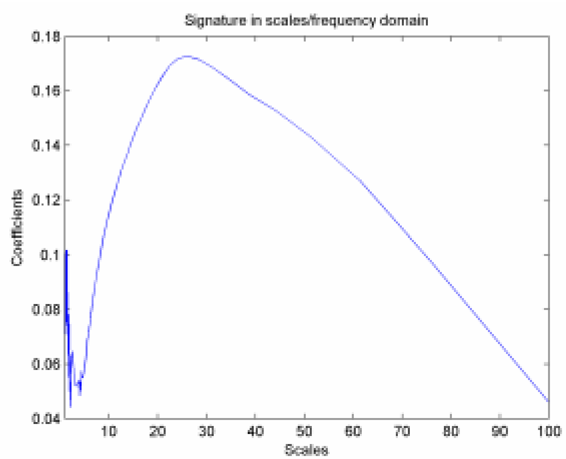
$$\mathbf{C} \in \mathbb{R}^{K \times R} = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_R \end{bmatrix}$$

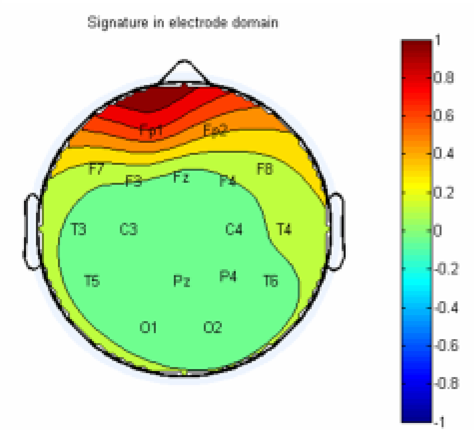
CP Application: Neuroscience

Epileptic Seizure Localization:



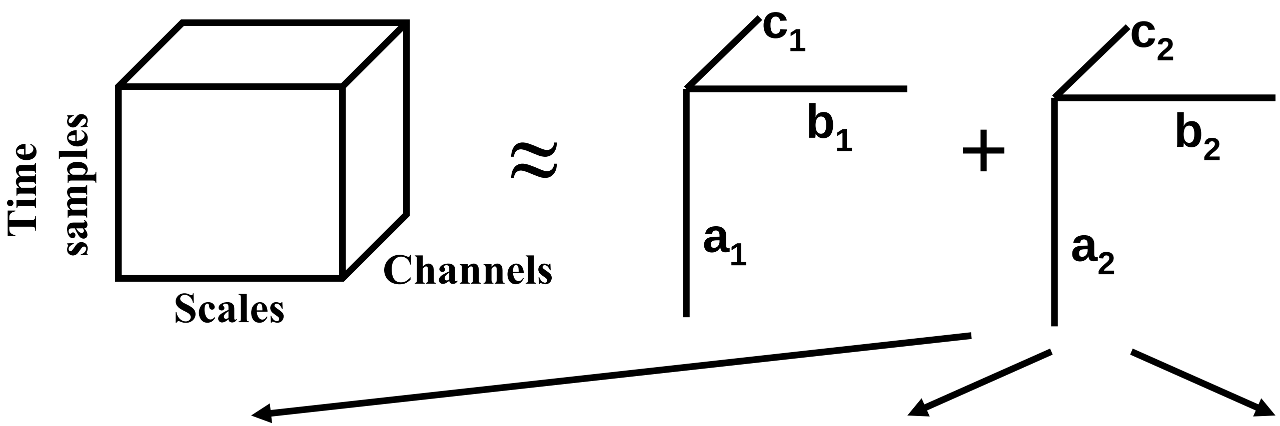


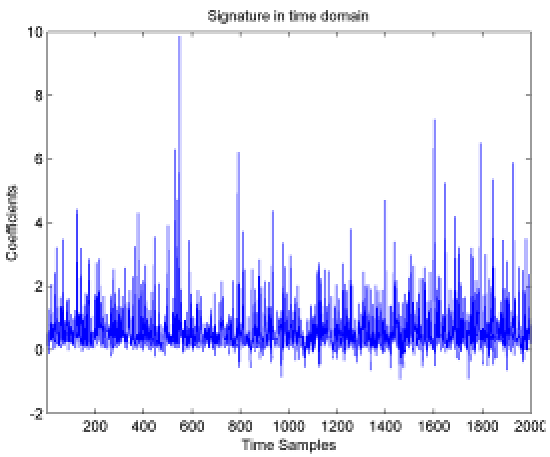


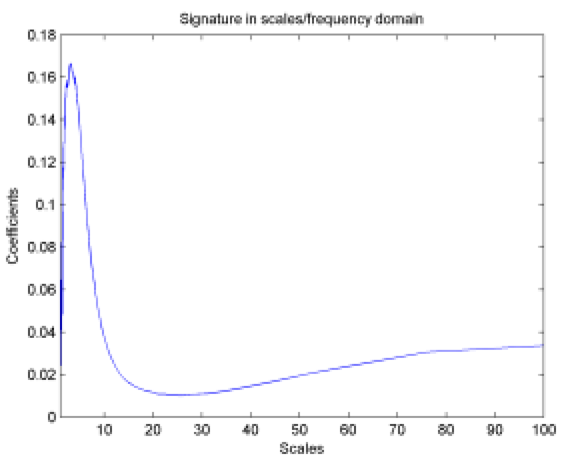


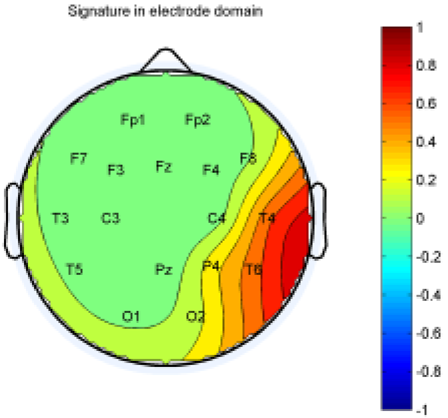
CP Application: Neuroscience

Epileptic Seizure Localization:





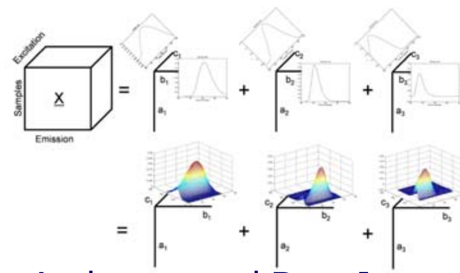




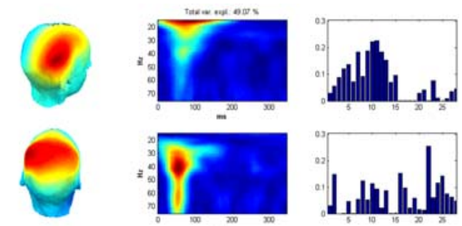
Acar et al, 2007, De Vos et al, 2007

CP has Numerous Applications!

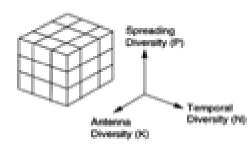
- **Chemometrics**
  - Fluorescence Spectroscopy
  - Chromatographic Data Analysis
- **Neuroscience**
  - Epileptic Seizure Localization
  - Analysis of EEG and ERP
- **Signal Processing**
- **Computer Vision**
  - Image compression, classification
  - Texture analysis
- **Social Network Analysis**
  - Web link analysis
  - Conversation detection in emails
  - Text analysis
- **Approximation of PDEs**



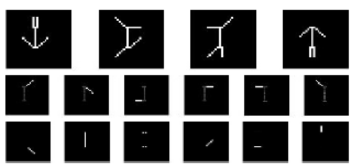
Andersen and Bro, *Journal of Chemometrics*, 2003.



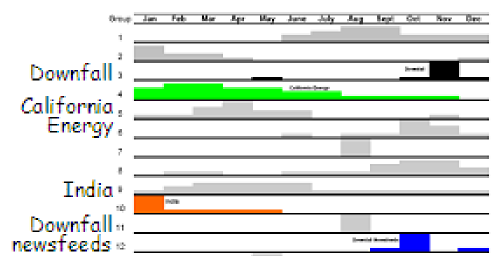
Mørup, Hansen and Arnfred, *Journal of Neuroscience Methods*, 2007.



Sidiropoulos, Giannakis and Bro, *IEEE Trans. Signal Processing*, 2000.



Hazan, Polak and Shashua, *ICCV 2005*.



Bader, Berry, Browne, *Survey of Text Mining: Clustering, Classification, and Retrieval*, 2nd Ed., 2007.

$$\mathscr{L}(x,t,\omega;u) = f(x,t,\omega)\ (x,t) \in \mathscr{D} \times [0,T]$$
$$u(x,t,\omega;u) = g(x,t)\ (x,t) \in \partial\mathscr{D} \times [0,T]$$
$$u(x,0,\omega) = h(x,\omega)\ x \in \mathscr{D},$$

Doostan and Iaccarino, *Journal of Computational Physics*, 2009.

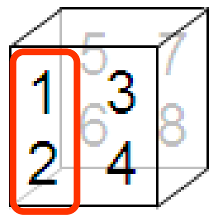
Algorithms: How Can We Compute CP?

Mathematical Details for CP

Unfolding (Matricization)

Columns: mode-1 fibers

$\mathcal{X} =$



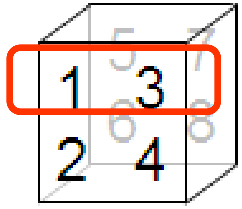
$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Mathematical Details for CP

Unfolding (Matricization)

$\mathcal{X} =$ 

Row: mode-2 fibers

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

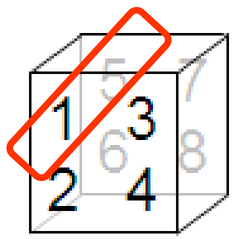
$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

Mathematical Details for CP

Unfolding (Matricization)

$\mathcal{X} =$

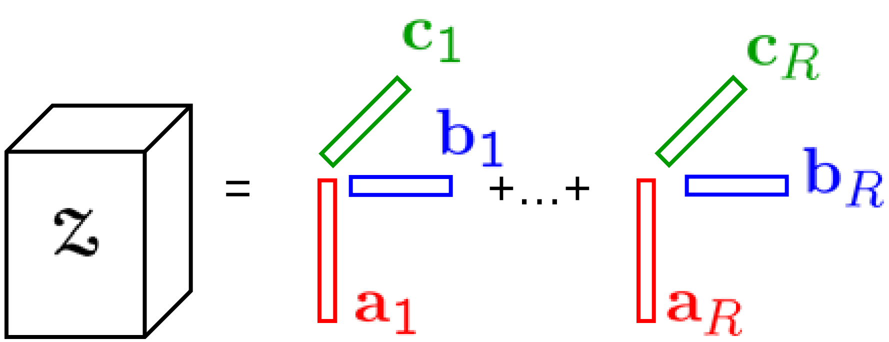


Tube: mode-3 fibers

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{bmatrix}$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$



$$\mathcal{Z} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

$$\mathcal{Z} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\mathbf{Z}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$$

$$\mathbf{Z}_{(2)} = \mathbf{B}(\mathbf{C} \odot \mathbf{A})^{\mathsf{T}}$$

$$\mathbf{Z}_{(3)} = \mathbf{C}(\mathbf{B} \odot \mathbf{A})^{\mathsf{T}}$$

Column-wise Khatri-Rao Product

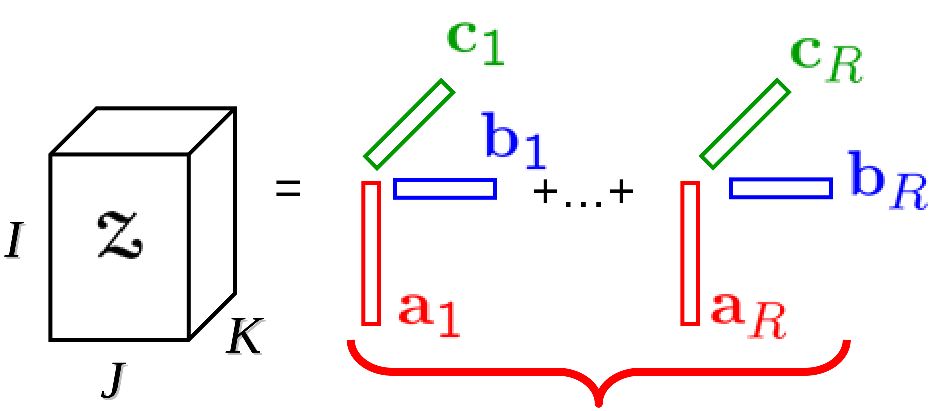
$$\mathbf{U} \odot \mathbf{V} = \begin{bmatrix} \mathbf{u}_1 \otimes \mathbf{v}_1 & \cdots & \mathbf{u}_R \otimes \mathbf{v}_R \end{bmatrix}$$

CP is a Nonlinear Optimization Problem

Given tensor $\mathfrak{Z}$ and R (# of components), find matrices **A, B, C** that solve the following problem:

Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$



$R(I+J+K)$
variables

Objective Function

$$f(\mathbf{x}) = \frac{1}{2} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$

where the vector **x** comprises the entries of A, B, and C stacked column-wise:

$$\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_R \\ \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_R \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_R \end{bmatrix}$$

Traditional Approach: CPALS

CPALS dating back to Harshman'70 and Carroll & Chang'70 solves for one factor matrix at a time.

Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathfrak{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|^2$$

Alternating Algorithm

for k = 1,...

$$\min_{\mathbf{A}} \| \mathfrak{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|^2$$

$$\min_{\mathbf{B}} \| \mathfrak{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|^2$$

$$\min_{\mathbf{C}} \| \mathfrak{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|^2$$

end

Each step can be converted to a matrix least squares problem:

$$\min_{\mathbf{A}} \left\| \mathbf{Z}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}} \right\|^2$$

$$\mathbf{A} = \mathbf{Z}_{(1)} \left((\mathbf{C} \odot \mathbf{B})^{\mathsf{T}} \right)^{\dagger}$$

$\mathbf{Z}_{(1)}$: I x *JK*; $(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}}$: *JK* x *R*

$$\mathbf{A} = \mathbf{Z}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}$$

$\mathbf{A}$: *I* x *R*; $\mathbf{Z}_{(1)}$: *I* x *JK*; $(\mathbf{C} \odot \mathbf{B})$: *JK* x *R*; $(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})$: *R* x *R* matrix

Traditional Approach: CPALS

Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$

Repeat the following steps until "convergence":

$$\mathbf{A} = \mathbf{Z}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}$$

$$\mathbf{B} = \mathbf{Z}_{(2)}(\mathbf{C} \odot \mathbf{A})(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{A}^{\mathsf{T}}\mathbf{A})^{\dagger}$$

$$\mathbf{C} = \mathbf{Z}_{(3)}(\mathbf{B} \odot \mathbf{A})(\mathbf{B}^{\mathsf{T}}\mathbf{B} * \mathbf{A}^{\mathsf{T}}\mathbf{A})^{\dagger}$$

Very fast, but not always accurate.

Not guaranteed to converge to a stationary point.

Other issues, e.g., cannot exploit symmetry.

Our Approach: CPOPT

Unlike CPALS, CPOPT solves for all factor matrices simultaneously using a gradient based optimization.

Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$

Define the objective function:

$$f(\mathbf{x}) = \frac{1}{2} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$

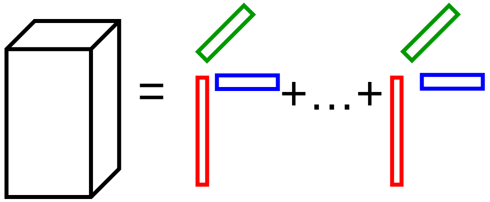
$$\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_R \\ \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_R \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_R \end{bmatrix} \qquad f : \mathbb{R}^{(I+J+K)R} \mapsto \mathbb{R}$$

$$\Rightarrow \quad \nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{a}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathbf{a}_R} \\ \frac{\partial f}{\partial \mathbf{b}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathbf{b}_R} \\ \frac{\partial f}{\partial \mathbf{c}_1} \\ \vdots \\ \frac{\partial f}{\partial \mathbf{c}_R} \end{bmatrix}$$

Objective and Gradient

Objective Function

$$f(\mathbf{x}) = \frac{1}{2} \| \mathcal{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$



Gradient (for $r = 1,\ldots,R$)

$$\frac{\partial f}{\partial \mathbf{a}_r}(\mathbf{x}) = -\mathcal{Z} \times_2 \mathbf{b}_r \times_3 \mathbf{c}_r + \sum_{k=1}^{R} \left(\mathbf{b}_r^{\mathsf{T}} \mathbf{b}_k \mathbf{c}_r^{\mathsf{T}} \mathbf{c}_k \right) \mathbf{a}_k$$

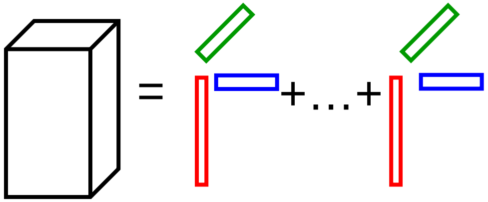
$$\frac{\partial f}{\partial \mathbf{b}_r}(\mathbf{x}) = -\mathcal{Z} \times_1 \mathbf{a}_r \times_3 \mathbf{c}_r + \sum_{k=1}^{R} \left(\mathbf{a}_r^{\mathsf{T}} \mathbf{a}_k \mathbf{c}_r^{\mathsf{T}} \mathbf{c}_k \right) \mathbf{b}_k$$

$$\frac{\partial f}{\partial \mathbf{c}_r}(\mathbf{x}) = -\mathcal{Z} \times_1 \mathbf{a}_r \times_2 \mathbf{b}_r + \sum_{k=1}^{R} \left(\mathbf{a}_r^{\mathsf{T}} \mathbf{a}_k \mathbf{b}_r^{\mathsf{T}} \mathbf{b}_k \right) \mathbf{c}_k$$

Gradient in Matrix Form

Objective Function

$$f(\mathbf{x}) = \frac{1}{2} \| \mathcal{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$



Gradient

$$\frac{\partial f}{\partial \mathbf{A}}(\mathbf{x}) = -\mathbf{Z}_{(1)}(\mathbf{C} \odot \mathbf{B}) + \mathbf{A}(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})$$

$$\frac{\partial f}{\partial \mathbf{B}}(\mathbf{x}) = -\mathbf{Z}_{(2)}(\mathbf{C} \odot \mathbf{A}) + \mathbf{B}(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{A}^{\mathsf{T}}\mathbf{A})$$

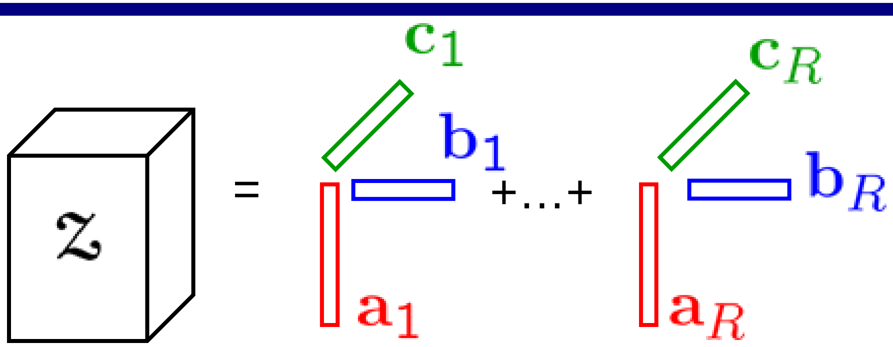
$$\frac{\partial f}{\partial \mathbf{C}}(\mathbf{x}) = -\mathbf{Z}_{(3)}(\mathbf{B} \odot \mathbf{A}) + \mathbf{C}(\mathbf{B}^{\mathsf{T}}\mathbf{B} * \mathbf{A}^{\mathsf{T}}\mathbf{A})$$

Note that this formulation can be used to derive the ALS approach!

$$\mathbf{A} = \mathbf{Z}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B})^{\dagger}$$

Indeterminacies of CP

- CP is often unique.
- However, CP has two fundamental indeterminacies
  - **Permutation** – The components can be reordered
    - Swap $\mathbf{a}_1$, $\mathbf{b}_1$, $\mathbf{c}_1$ with $\mathbf{a}_3$, $\mathbf{b}_3$, $\mathbf{c}_3$
  - **Scaling** – The vectors comprising a single rank-one factor can be scaled
    - Replace $\mathbf{a}_1$ and $\mathbf{b}_1$ with 2 $\mathbf{a}_1$ and ½ $\mathbf{b}_1$



Not a big deal. Leads to multiple, but separated, minima.

This leads to a continuous space of equivalent solutions.

Sandia National Laboratories

Adding Regularization

Objective Function

$$f(\mathbf{x}) = \frac{1}{2} \| \mathcal{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \frac{\lambda}{2} \left(\| \mathbf{A} \|_F^2 + \| \mathbf{B} \|_F^2 + \| \mathbf{C} \|_F^2 \right)$$

Gradient

$$\frac{\partial f}{\partial \mathbf{A}}(\mathbf{x}) = -\mathbf{Z}_{(1)}(\mathbf{C} \odot \mathbf{B}) + \mathbf{A}(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{B}^{\mathsf{T}}\mathbf{B}) + \lambda \mathbf{A}$$

$$\frac{\partial f}{\partial \mathbf{B}}(\mathbf{x}) = -\mathbf{Z}_{(2)}(\mathbf{C} \odot \mathbf{A}) + \mathbf{B}(\mathbf{C}^{\mathsf{T}}\mathbf{C} * \mathbf{A}^{\mathsf{T}}\mathbf{A}) + \lambda \mathbf{B}$$

$$\frac{\partial f}{\partial \mathbf{C}}(\mathbf{x}) = -\mathbf{Z}_{(3)}(\mathbf{B} \odot \mathbf{A}) + \mathbf{C}(\mathbf{B}^{\mathsf{T}}\mathbf{B} * \mathbf{A}^{\mathsf{T}}\mathbf{A}) + \lambda \mathbf{C}$$

Our methods: CPOPT & CPOPTR

CPOPT: Apply derivative-based optimization method to the following objective function:

$$f(\mathbf{x}) = \frac{1}{2} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2$$

CPOPTR: Apply derivative-based optimization method to the following regularized objective function:

$$f(\mathbf{x}) = \frac{1}{2} \| \mathfrak{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \frac{\lambda}{2} \left(\| \mathbf{A} \|_F^2 + \| \mathbf{B} \|_F^2 + \| \mathbf{C} \|_F^2 \right)$$

Another competing method: CPNLS

CPNLS: Apply nonlinear least squares solver to the following equations:

$$F(\mathbf{x}) = \text{vec}\left(\mathcal{Z} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\right)$$

⬇

$$F : \mathbb{R}^{(I+J+K)R} \mapsto \mathbb{R}^{IJK}$$

⬇

Jacobian is of size $IJK \times (I+J+K)R$.

Proposed by **Paatero'**97 and also **Tomasi and Bro**'05.

Experimental Set-Up[Tomasi&Bro'06]

20 triplets

Step 1: Generate random factor matrices A, B, C with R_{true} = 3 or 5 columns each and collinearity set to 0.5,

i.e., $\mathbf{a}_r^{\mathsf{T}} \mathbf{a}_s = 0.5$

Step 2: Construct tensor from factor matrices and add noise. All combinations of:

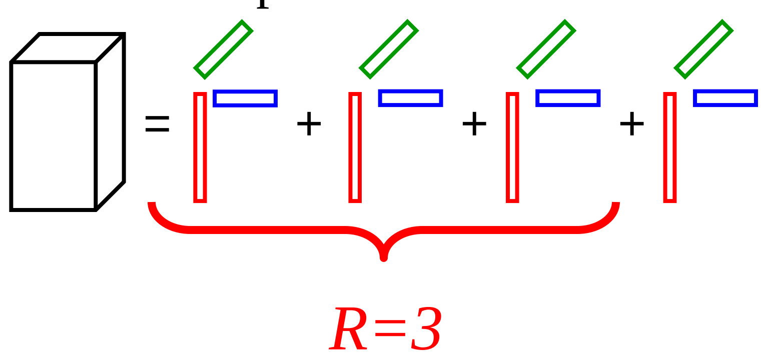
- Homoscedastic: 1%, 5%, 10%
- Heteroscedastic: 0%, 1%, 5%

$$\mathcal{Z} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] + \mathcal{N}$$

180 tensors

Step 3: Use algorithm to extract factors, using R_{true} and $R_{true}+1$ factors. Compare against factors in Step 1.

360 tests



Implementation Details

- All experiments were performed in MATLAB on a Linux workstation (Quad-Core Intel Xeon 2.50GHz, 9 GB RAM).

- Methods
  - **CPALS** – Alternating least squares. Used **parafac_als** in the Tensor Toolbox (Bader & Kolda)
  - **CPNLS** – Nonlinear least squares. Used **PARAFAC3W**, which implements Levenberg-Marquadt (necessary due to scaling ambiguity), by Tomasi and Bro.
  - **CPOPT** – Optimization. Used routines in the **Tensor Toolbox** in calculation of function values and gradients. Optimization via Nonlinear Conjugate Gradient (NCG) method with Hestenes-Stiefel update, using **Poblano** (in-house code to be released soon).
  - **CPOPTR** – Optimization with regularization. Same as above. (Regularization parameter = 0.02.)

CPOPT is Fast and Accurate

Generated 360 dense test problems (with ranks 3 and 5) and factorized with *R* as the correct number of components and one more than that. Total of 720 tests for each entry below.

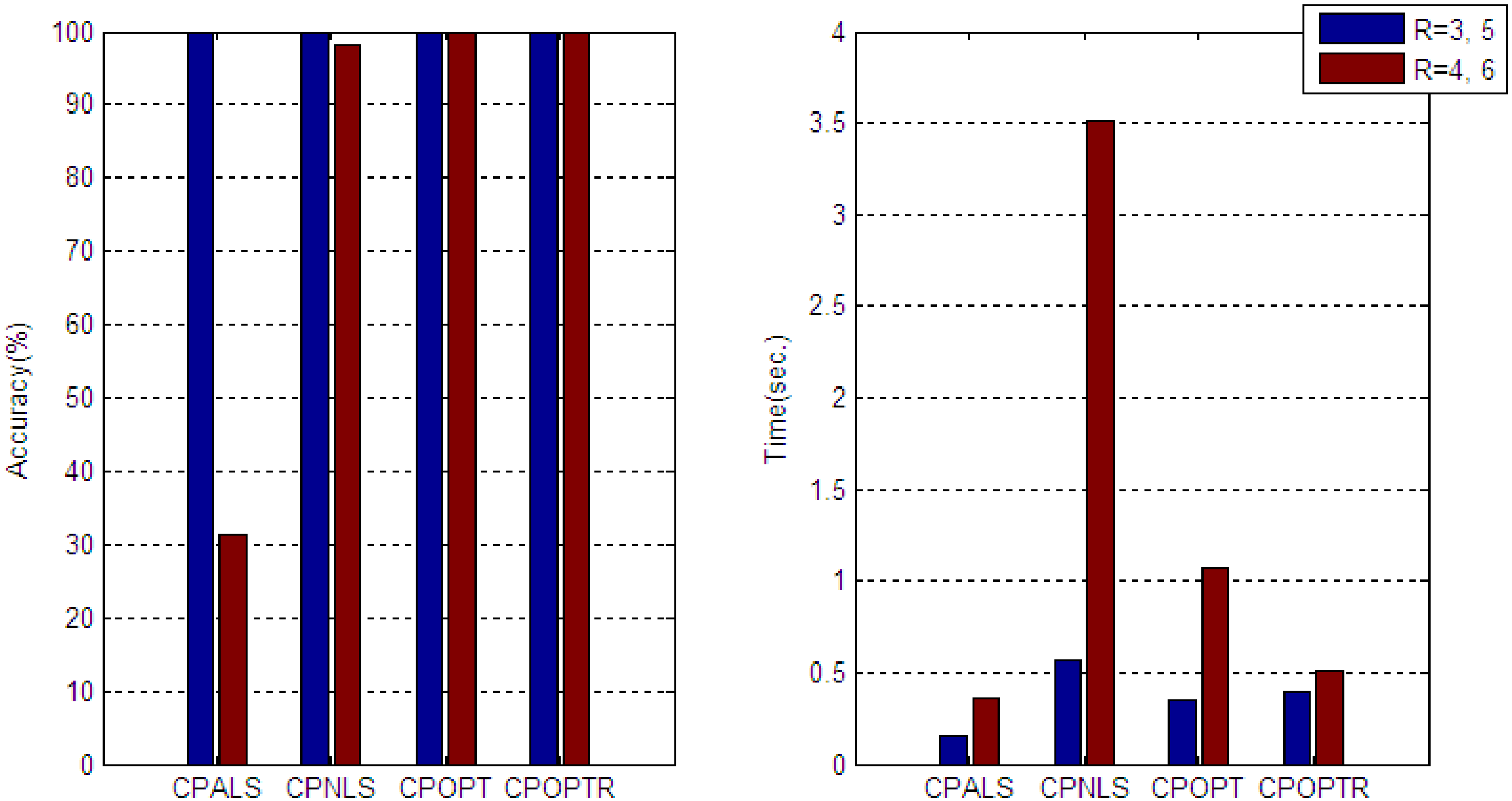
	Time (sec)			
Size	**CPALS**	**CPNLS**	**CPOPT**	**CPOPTR**
$20 \times 20 \times 20$	0.5 ± 1.0	0.3 ± 0.3	0.3 ± 0.2	0.2 ± 0.1
$50 \times 50 \times 50$	0.3 ± 0.3	2.0 ± 2.6	0.7 ± 0.5	0.5 ± 0.1
$100 \times 100 \times 100$	1.7 ± 1.1	11.5 ± 11.5	5.6 ± 3.6	4.3 ± 1.3
$250 \times 250 \times 250$	26.6 ± 9.1	143.9 ± 125.0	83.5 ± 35.2	81.9 ± 22.8

	Accuracy (%)			
Size	**CPALS**	**CPNLS**	**CPOPT**	**CPOPTR**
$20 \times 20 \times 20$	78.8	99.0	99.9	100.0
$50 \times 50 \times 50$	65.7	99.0	100.0	100.0
$100 \times 100 \times 100$	63.5	97.9	100.0	100.0
$250 \times 250 \times 250$	62.2	99.0	100.0	100.0

K x *K* x *K* *R* = # components	$O(RK^3)$	$O(R^3K^3)$	$O(RK^3)$	$O(RK^3)$

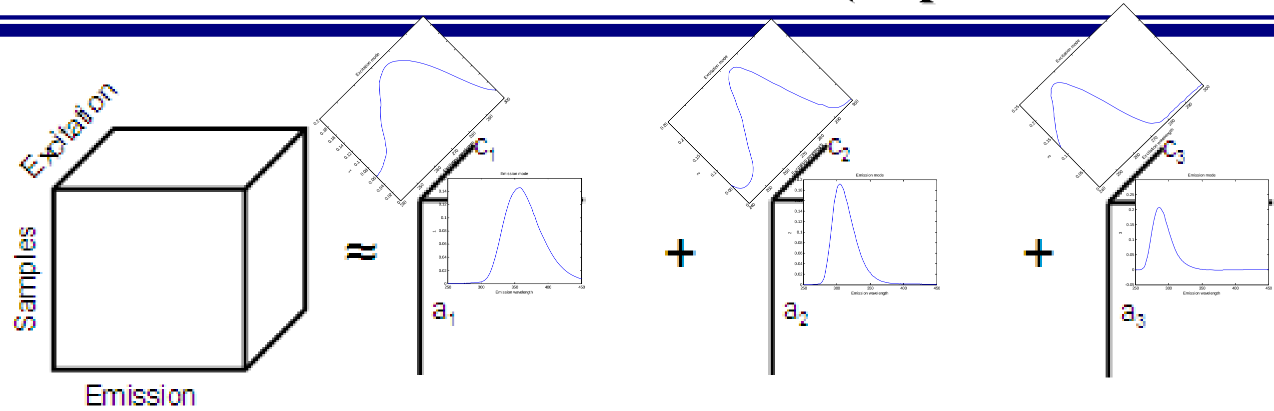
Overfactoring has a significant impact

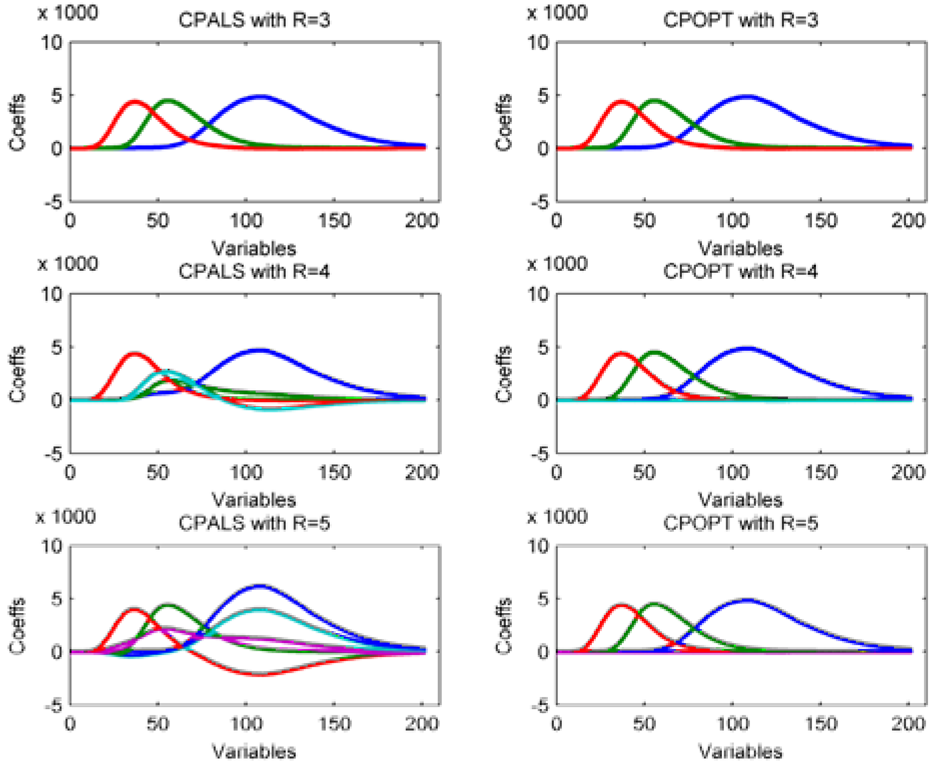
$50 \times 50 \times 50$



CPOPT is robust to overfactoring

Amino (http://www.models.life.ku.dk/)





CP-WOPT: Handling Missing Data

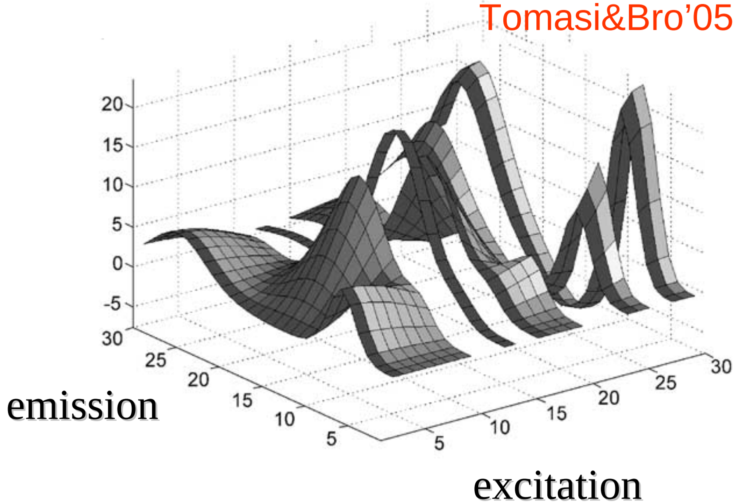
Missing Data Examples

Missing data in different disciplines due to loss of information, machine failures, different sampling frequencies or experimental-set ups.

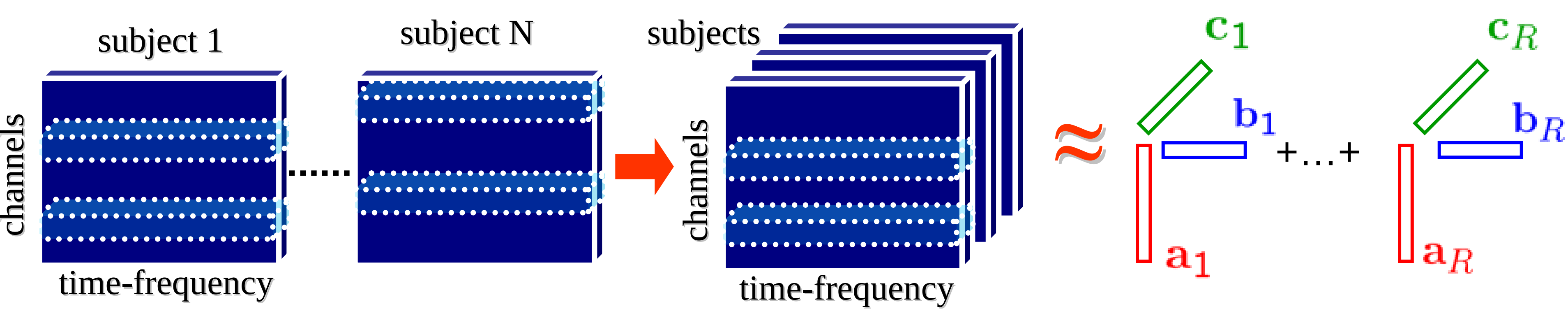
- Chemometrics
- Biomedical signal processing (e.g., EEG)
- Network traffic analysis (e.g., packet drops)
- Computer vision (e.g., occlusions)
- …

CHEMISTRY

Tomasi&Bro'05



EEG



Modify the objective for CP

NO MISSING DATA

Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathcal{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!] \|^2$$

FOR HANDLING MISSING DATA

Optimization Problem

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \| \mathcal{W} * (\mathcal{Z} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!]) \|^2$$

$$w_{ijk} = \begin{cases} 1 & \text{if } z_{ijk} \text{ is known,} \\ 0 & \text{if } z_{ijk} \text{ is missing.} \end{cases}$$

Objective Function

$$f_{\mathcal{W}}(\mathbf{A},\mathbf{B},\mathbf{C}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left\{ w_{ijk} \left(z_{ijk} - \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} \right) \right\}^2$$

Our approach: CP-WOPT

Objective Function

$$f_{\mathcal{W}}(\mathbf{A},\mathbf{B},\mathbf{C}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left\{w_{ijk}\left(z_{ijk} - \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr}\right)\right\}^2$$

$$f_{\mathcal{W}} : \mathbb{R}^{(I+J+K)R} \mapsto \mathbb{R}$$

$$\mathbf{x} = \begin{bmatrix} a_{11} \\ \vdots \\ a_{IR} \\ b_{11} \\ \vdots \\ b_{JR} \\ c_{11} \\ \vdots \\ c_{KR} \end{bmatrix} \quad \Rightarrow \quad \nabla f_{\mathcal{W}}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_{\mathcal{W}}}{\partial a_{11}} \\ \vdots \\ \frac{\partial f_{\mathcal{W}}}{\partial a_{IR}} \\ \frac{\partial f_{\mathcal{W}}}{\partial b_{11}} \\ \vdots \\ \frac{\partial f_{\mathcal{W}}}{\partial b_{JR}} \\ \frac{\partial f_{\mathcal{W}}}{\partial c_{11}} \\ \vdots \\ \frac{\partial f_{\mathcal{W}}}{\partial c_{KR}} \end{bmatrix}$$

Objective and Gradient

Objective Function

$$f_{\mathbf{W}}(\mathbf{A},\mathbf{B},\mathbf{C}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left\{w_{ijk}\left(z_{ijk} - \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr}\right)\right\}^2$$

Gradient (for *r* = 1,…,*R*; *i*=1,…*I*; *j*=1,..*J*; *k*=1,..*K*)

$$\frac{\partial f_{\mathbf{W}}}{\partial a_{ir}} = -2\sum_{k=1}^{K}\sum_{j=1}^{J} w_{ijk}z_{ijk}b_{jr}c_{kr} + 2\sum_{k=1}^{K}\sum_{j=1}^{J} w_{ijk}\left(\sum_{l=1}^{R} a_{il}b_{jl}c_{kl}\right)b_{jr}c_{kr}$$

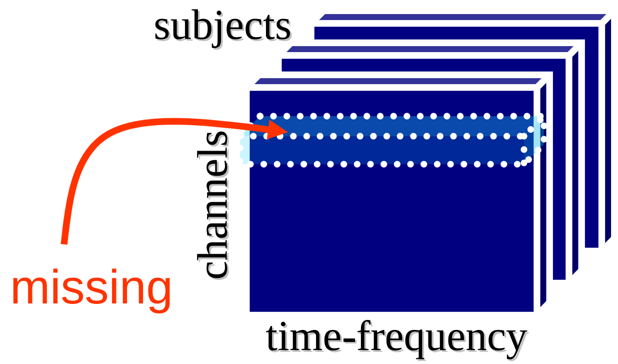
$$\frac{\partial f_{\mathbf{W}}}{\partial b_{jr}} = -2\sum_{k=1}^{K}\sum_{i=1}^{I} w_{ijk}z_{ijk}a_{ir}c_{kr} + 2\sum_{k=1}^{K}\sum_{i=1}^{I} w_{ijk}\left(\sum_{l=1}^{R} a_{il}b_{jl}c_{kl}\right)a_{ir}c_{kr}$$

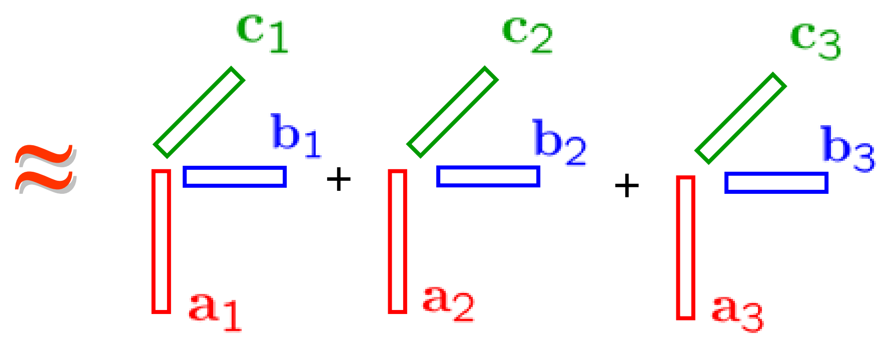
$$\frac{\partial f_{\mathbf{W}}}{\partial c_{kr}} = -2\sum_{j=1}^{J}\sum_{i=1}^{I} w_{ijk}z_{ijk}a_{ir}b_{jr} + 2\sum_{j=1}^{J}\sum_{i=1}^{I} w_{ijk}\left(\sum_{l=1}^{R} a_{il}b_{jl}c_{kl}\right)a_{ir}b_{jr}$$

CP-WOPT is useful for real data!

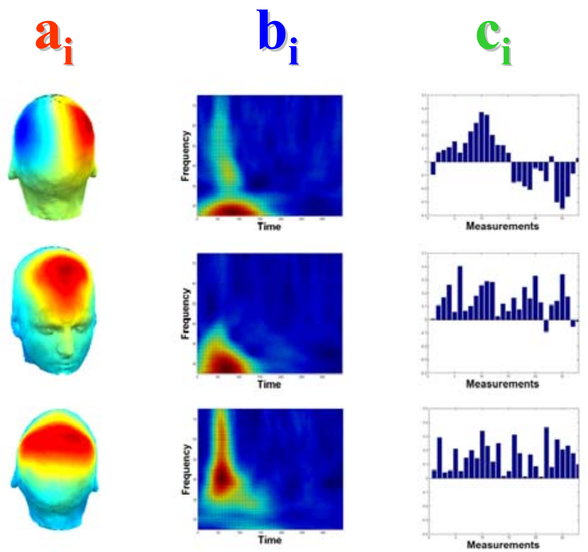
Thanks to Morten Mørup!

GOAL: To differentiate between left and right hand stimulation

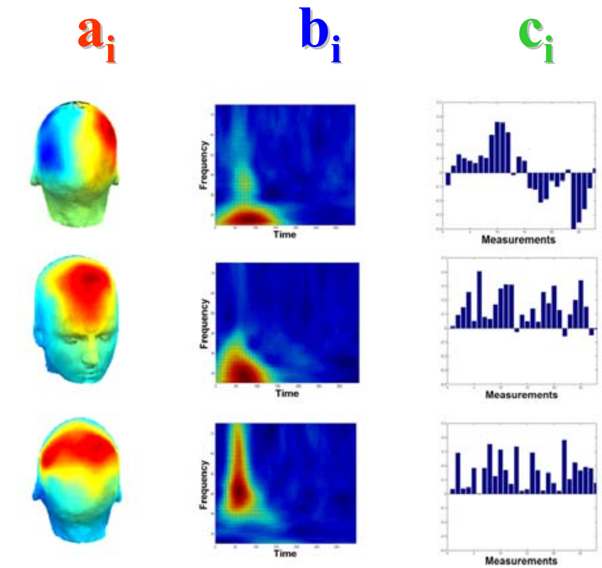




COMPLETE

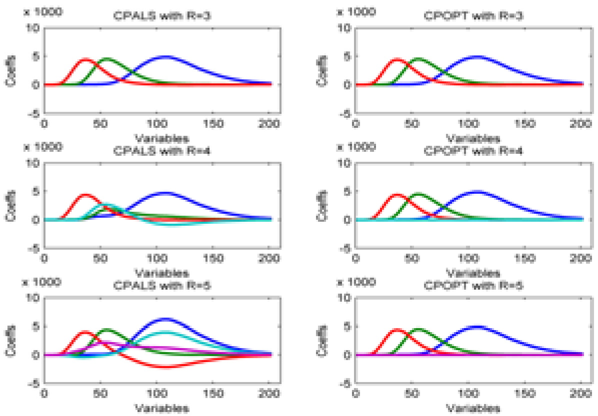


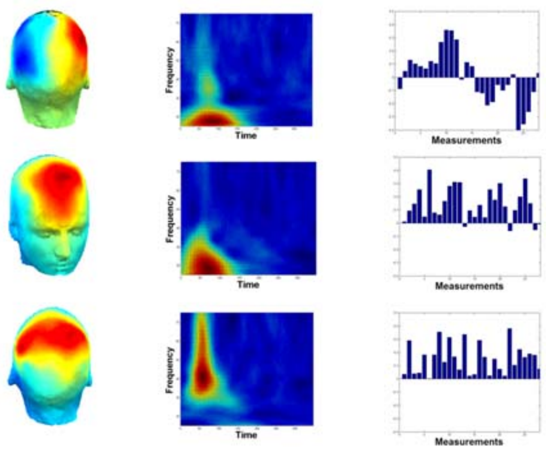
INCOMPLETE



Summary & Future Work

- New CPOPT method
  - Accurate & scalable
- Extend CPOPT to CP-WOPT to handle missing data
  - Accurate & scalable
- More work…
  - Starting point?
  - Tuning the optimization
  - Regularization
  - Exploiting sparsity
  - Nonnegativity





Thank you!

- **More on tensors and tensor models:**
  - **Survey :** E. Acar and B. Yener, Unsupervised Multiway Data Analysis: A Literature Survey, *IEEE Transactions on Knowledge and Data Engineering*, 21(1): 6-20, 2009.
  - **CPOPT :** E. Acar, T. G. Kolda and D. M. Dunlavy, An Optimization Approach for Fitting Canonical Tensor Decompositions, *Submitted for publication.*
  - **CP-WOPT :** E. Acar, T.G. Kolda, D. M. Dunlavy and M. Mørup, Tensor Factorizations with Missing Data, *Submitted for publication*.

- **Contact:**

  - Evrim Acar, **eacarat@sandia.gov**
  - Tamara G. Kolda, **tgkolda@sandia.gov**
  - Daniel M. Dunlavy, **dmdunla@sandia.gov**

MS31 Tensors and Tensor-based Computations (Part II of II) @4pm