Extracting Insights from Complex Data:

Constrained Multimodal Data Mining using Coupled Matrix and Tensor Factorizations

Evrim Acar

Simula Metropolitan Center for Digital Engineering

Oslo, Norway

Explainable AI for the Sciences: Towards Novel Insights

IPAM - January 9-13, 2022

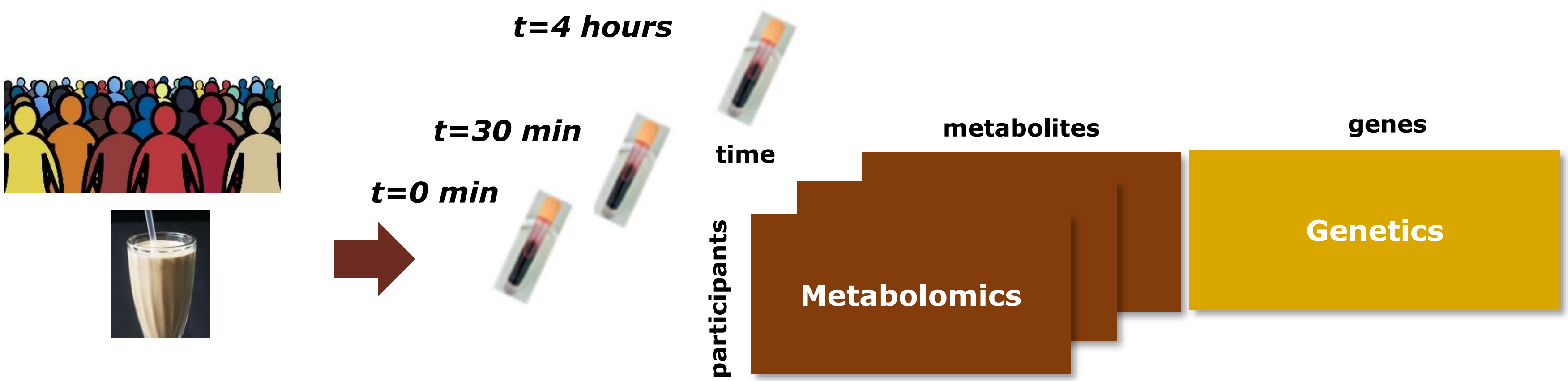
simula

Motivation: Understanding complex systems requires joint analysis of static and time-evolving data sets

A better understanding of how **complex systems evolve over time** can enable us to address important problems by capturing differences in **brain dynamics** or **metabolic responses,** potentially revealing **early signs of diseases**.

To gain such a level of understanding, we need to extract meaningful information from personal **data clouds**, where some data sets are **time-evolving** while others are **static**.

Other data sets: Gut microbiome, hormones…



Analysis of such data and capturing group differences is important for precision nutrition and medicine *[Price et al., Nature Biotechnology, 2017; Berry et al., Nature Medicine, 2020]*

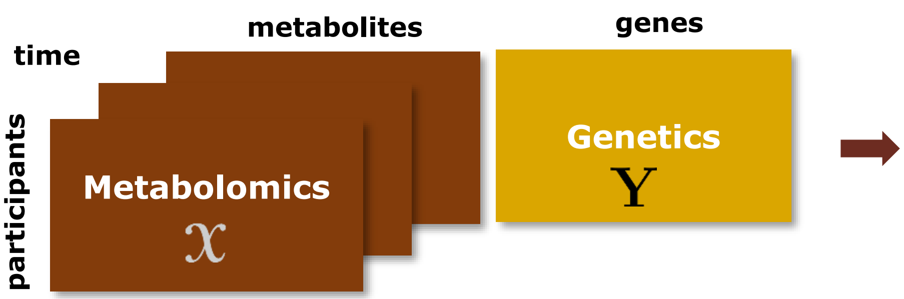
From multimodal heterogeneous data sets to interpretable patterns

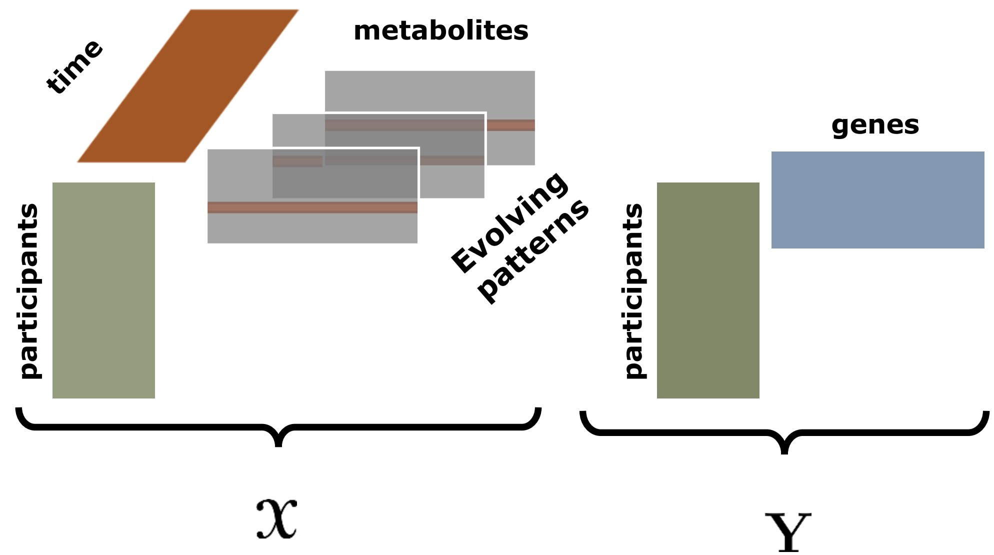
Our goal: Joint analysis of time-evolving and static data sets to capture underlying patterns as well as their evolution in time

Multimodal Heterogeneous Data Sets

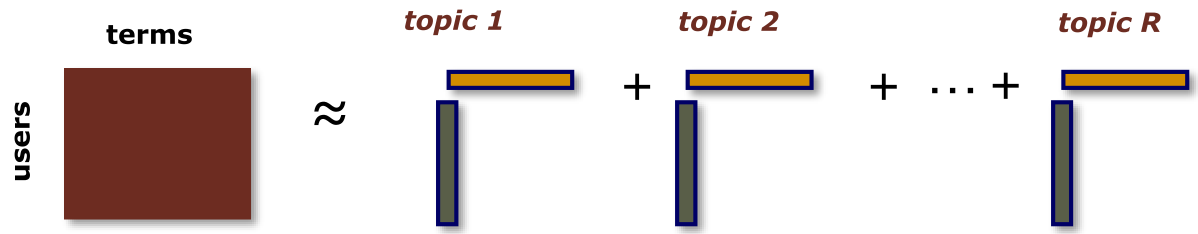
Interpretable Patterns and Their Evolution in Time

Other data sets: Gut microbiome, hormones…

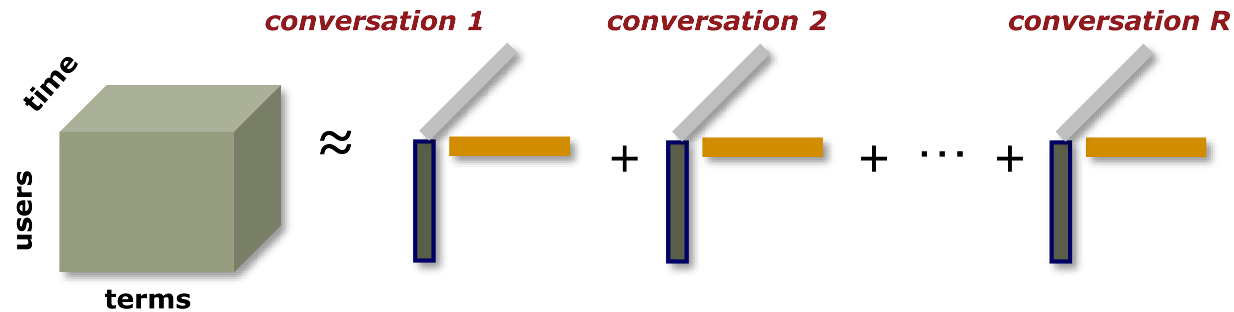




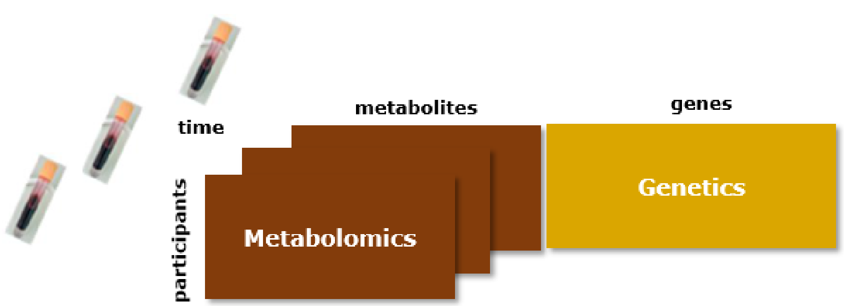
Matrix factorizations in data mining

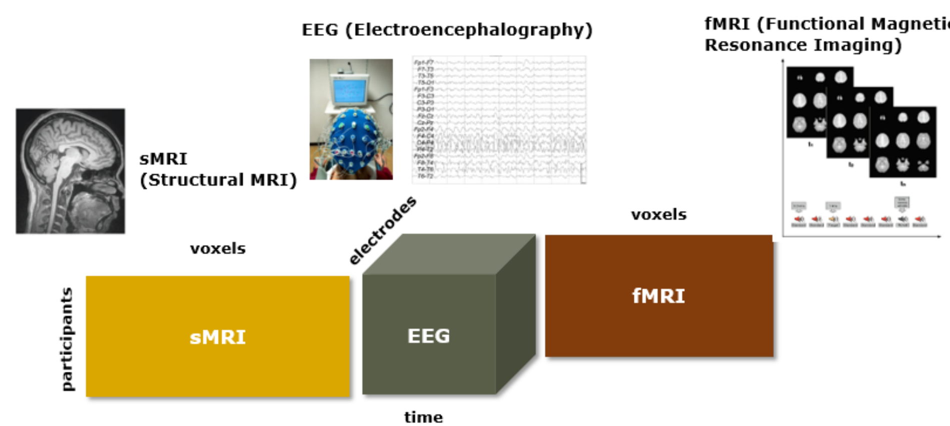


Data sets are often multi-way: Tensor factorizations



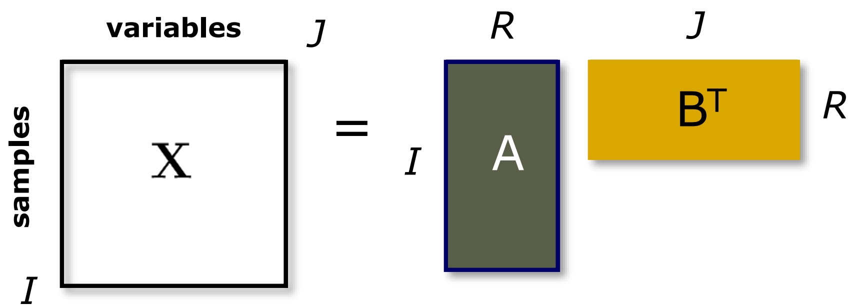
Data sets often come from multiple sources: Coupled matrix/tensor factorizations





Matrix Factorizations

Matrix Factorizations, e.g., Singular Value Decomposition (SVD), Non-negative Matrix Factorization (NMF), Independent Component Analysis (ICA), are commonly used in data mining to capture the underlying structures in data sets.



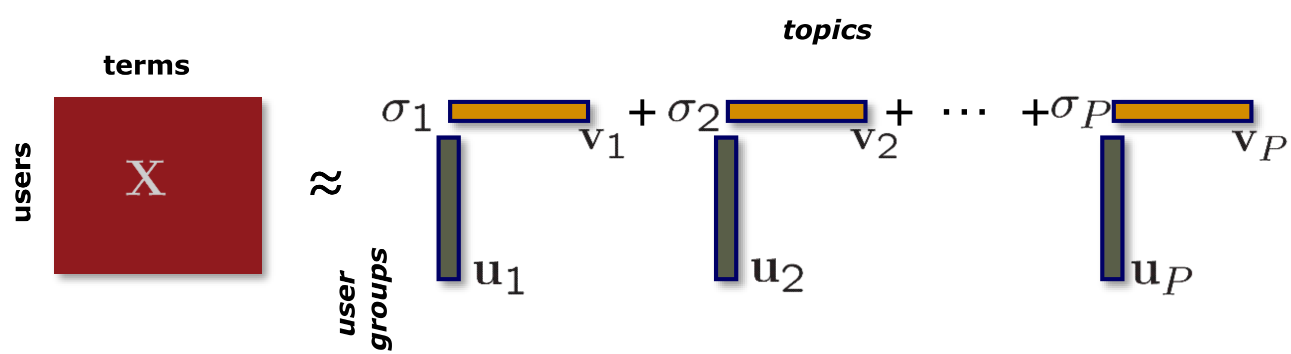
$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{a}_r \mathbf{b}_r^{\mathsf{T}} = \mathbf{A}\mathbf{B}^{\mathsf{T}}$$

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

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

For instance, we may compute truncated SVD of a *users* by *terms* matrix and identify the topics and user groups talking about those topics.

$$\mathbf{X} \approx \sum_{r=1}^{P} \sigma_r \mathbf{u}_r \mathbf{v}_r^{\mathsf{T}} \approx \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}$$



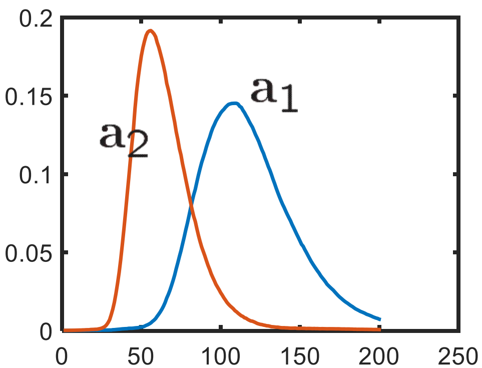
Uniqueness is an issue

$$\mathbf{X} = \mathbf{A}\mathbf{B}^\top = \mathbf{A}\mathbf{M}\mathbf{M}^{-1}\mathbf{B}^\top = \bar{\mathbf{A}}\bar{\mathbf{B}}^\top$$

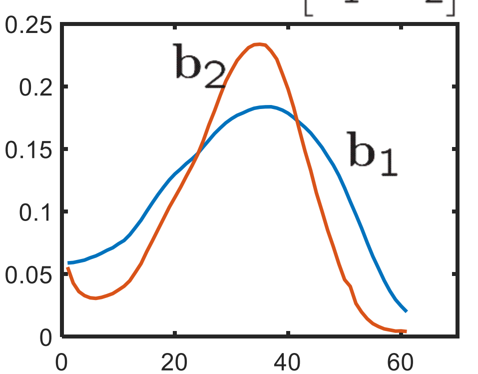
Constraints are used to deal with the uniqueness problem, e.g., SVD. However, factorizations with constraints may not be meaningful in terms of the application.

True factors

$$\mathbf{A} \in \mathbb{R}^{201\times 2} = [\mathbf{a}_1 \;\; \mathbf{a}_2]$$

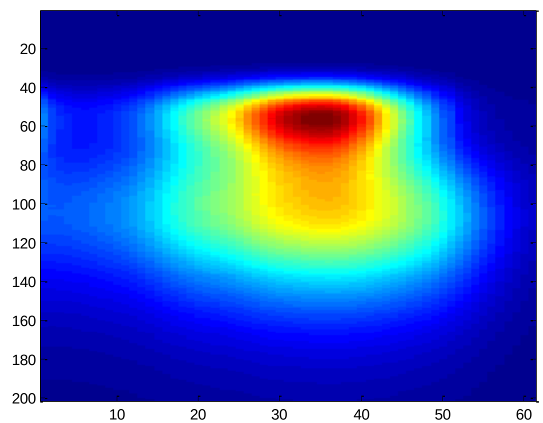


$$\mathbf{B} \in \mathbb{R}^{61\times 2} = [\mathbf{b}_1 \;\; \mathbf{b}_2]$$



Data matrix

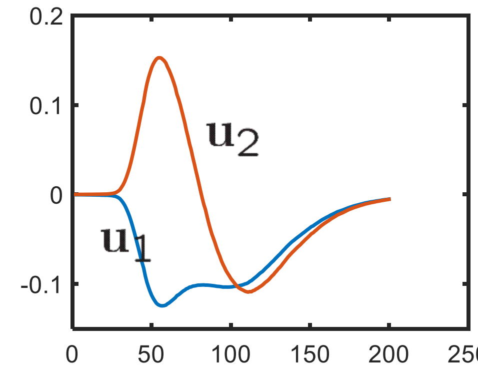
$$\mathbf{X} = \mathbf{A}\mathbf{B}^\top$$

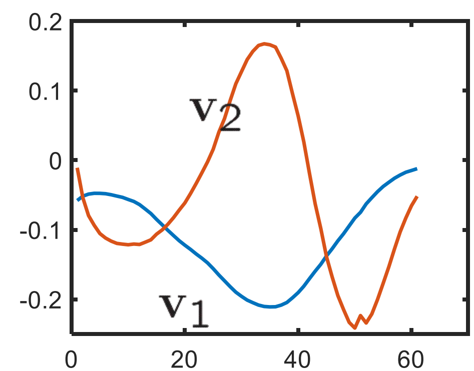


Given X, can we recover the true factors?

SVD captures...

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$

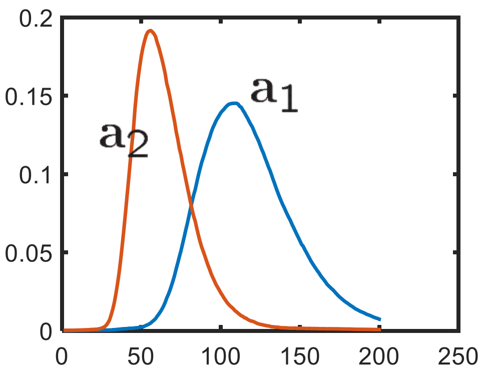




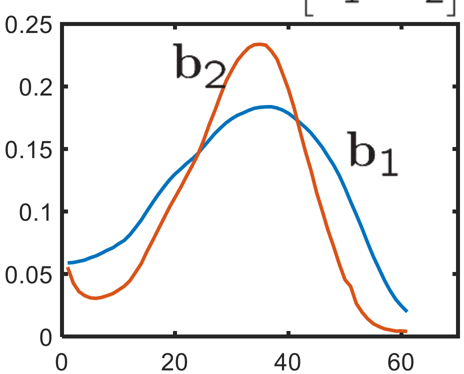
What if we have multiple matrices with the same underlying factors but in different proportions…

True factors

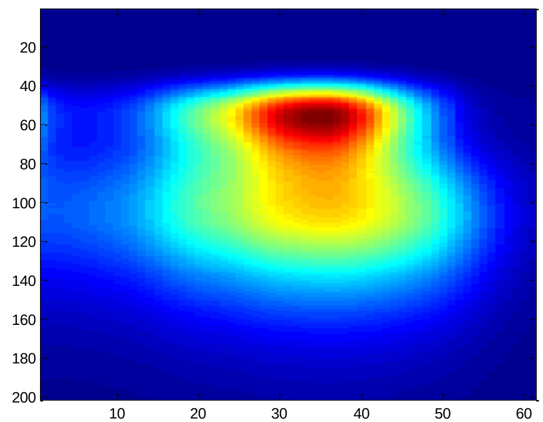
$$\mathbf{A} \in \mathbb{R}^{201\times 2} = [\mathbf{a}_1 \;\; \mathbf{a}_2]$$



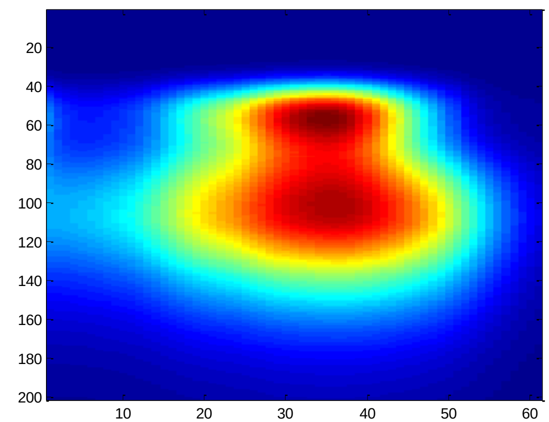
$$\mathbf{B} \in \mathbb{R}^{61\times 2} = [\mathbf{b}_1 \;\; \mathbf{b}_2]$$



$$\mathbf{X}_1 = \mathbf{A}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\mathbf{B}^\mathsf{T}$$



$$\mathbf{X}_2 = \mathbf{A}\begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}\mathbf{B}^\mathsf{T}$$

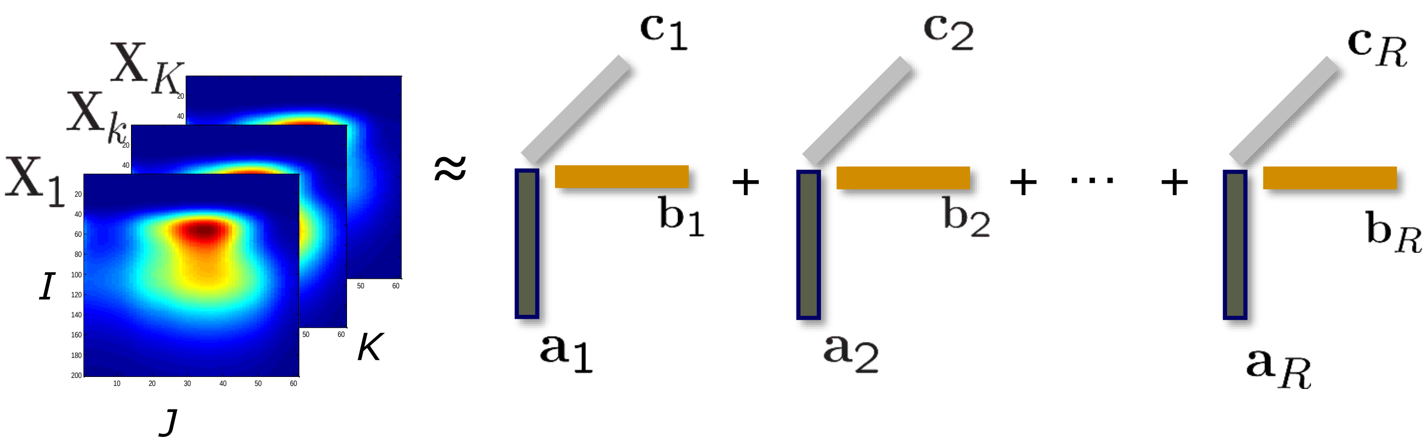


We can recover the true factors uniquely up to trivial indeterminacies, i.e., scaling and permutation.

Tensor Factorizations: CANDECOMP/PARAFAC (CP)

[Hitchcock, 1927; Harshman, 1970; Carroll & Chang, 1970]

As an extension of matrix factorizations to higher-order tensors (multi-way arrays), tensor factorizations are used to extract the underlying factors in higher-order data sets. In particular, we are interested in the CP model, which represents a tensor as a sum of rank-one tensors:



$$\mathbf{X}_k \approx \mathbf{A} \begin{bmatrix} c_{k1} & & 0 \\ & \ddots & \\ 0 & & c_{kR} \end{bmatrix} \mathbf{B}^{\mathsf{T}}$$

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

$$\approx [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

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

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

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

N-mode vector outer product

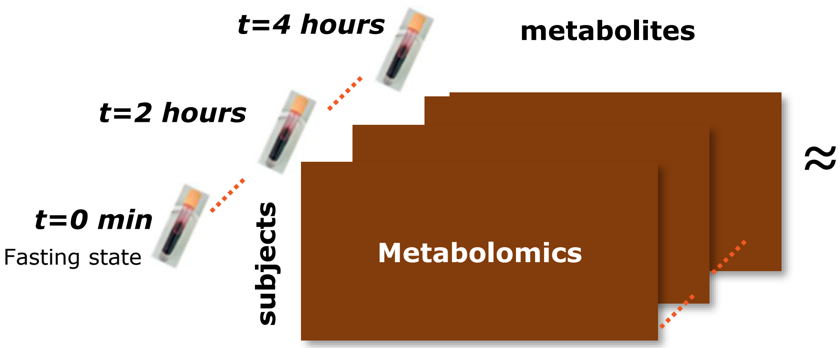
$$\mathcal{X} \in \mathbb{R}^{I \times J \times K}, \mathbf{a} \in \mathbb{R}^{I}, \mathbf{b} \in \mathbb{R}^{J}, \mathbf{c} \in \mathbb{R}^{K}$$

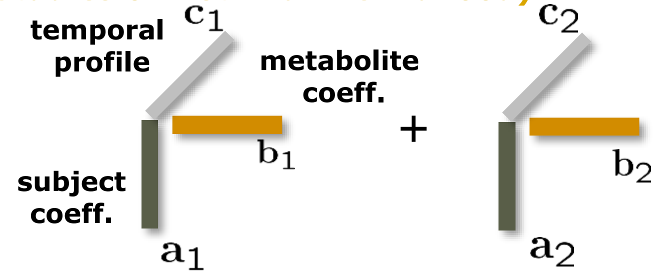
$$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \text{ iff } x_{ijk} = a_i b_j c_k$$

Omics: CP reveals differences among groups of subjects in terms of their response to the meal challenge test

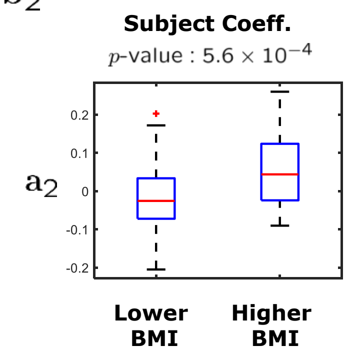
[Li et al., bioRxiv, 2022]

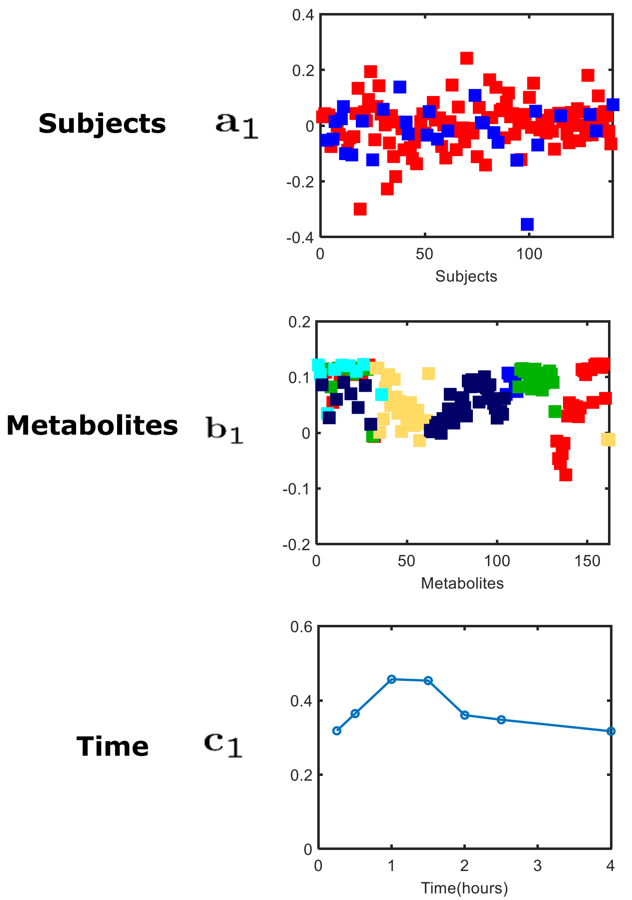
Ongoing work with COPSAC (Copenhagen Prospective Studies on Asthma in Childhood)

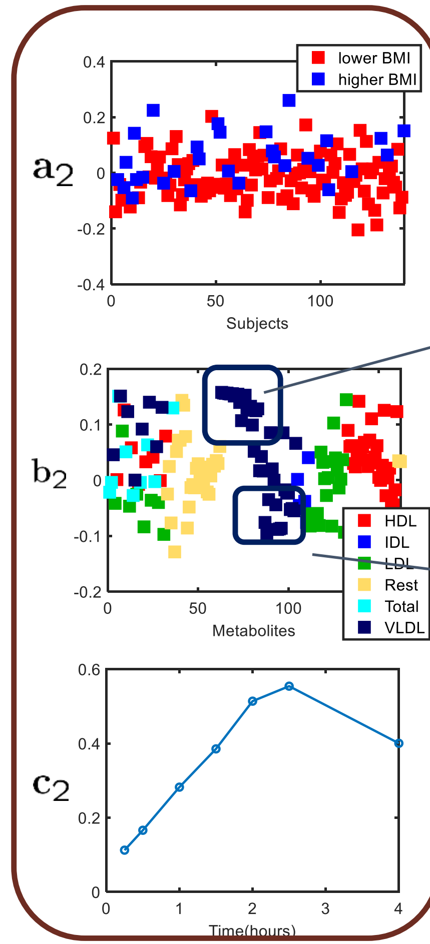




Nuclear Magnetic Resonance Spectroscopy measurements of plasma samples collected from the COPSAC$_{2000}$ cohort







Extremely large, very large and large VLDLs

BMI ↑

XXL/XL/L-VLDL↑

Medium, small VLDLs

BMI ↑

M/S-VLDL↓

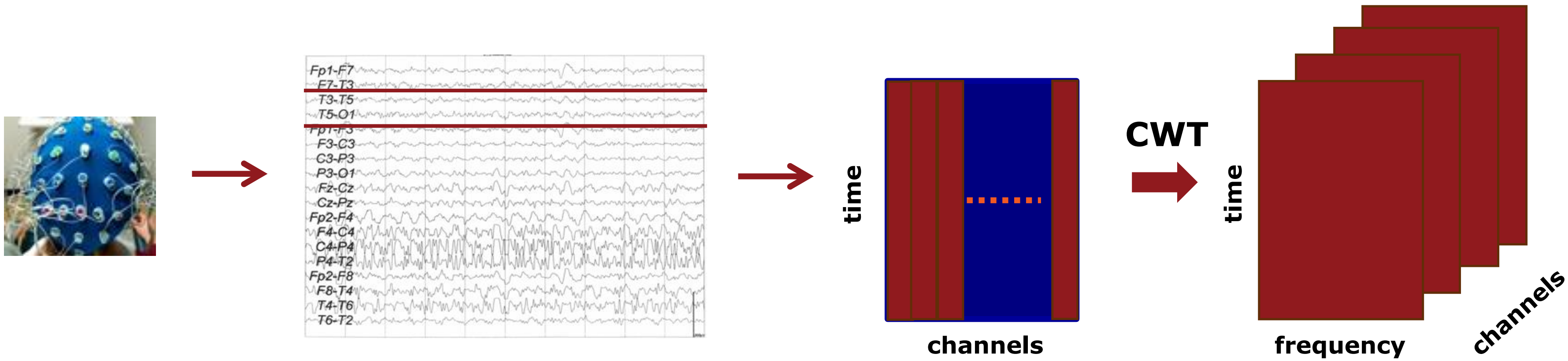
Not visible in the fasting state!

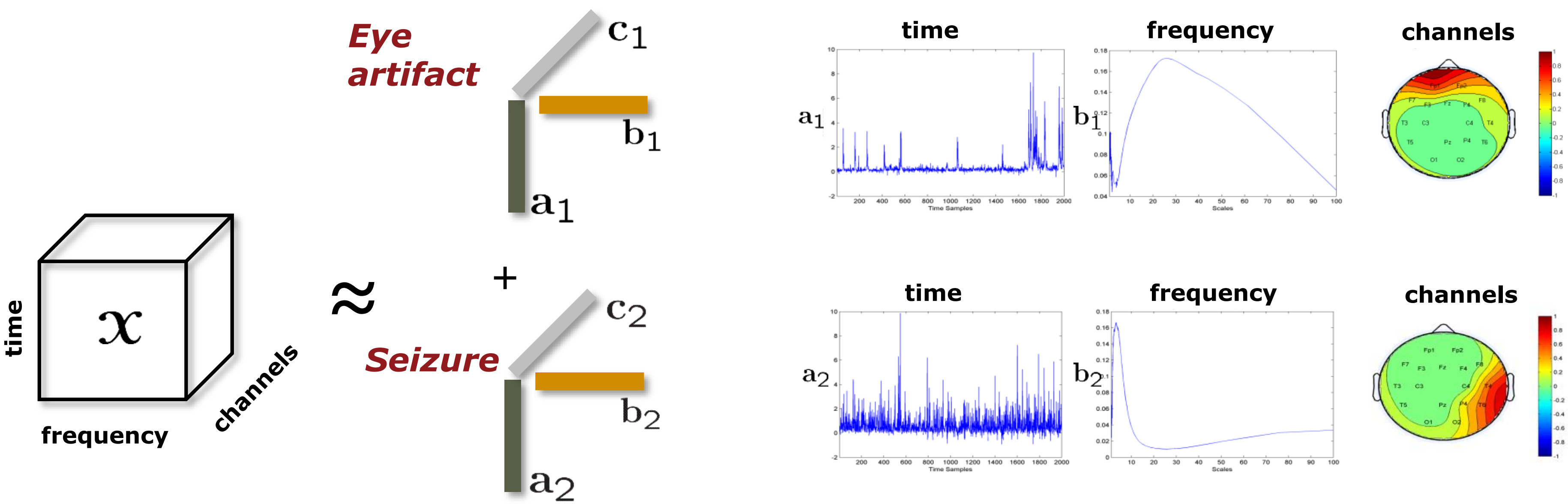
HDL (high density), IDL (intermediate density), LDL (low density), VLDL (very low density):

Lipoproteins carrying fat, e.g., cholesterol, to cells through blood

Neuroscience: CP components have been shown to localize epileptic seizures

[Acar et al., *Bioinformatics*, 2007; De Vos et al., *Neuroimage*, 2007]



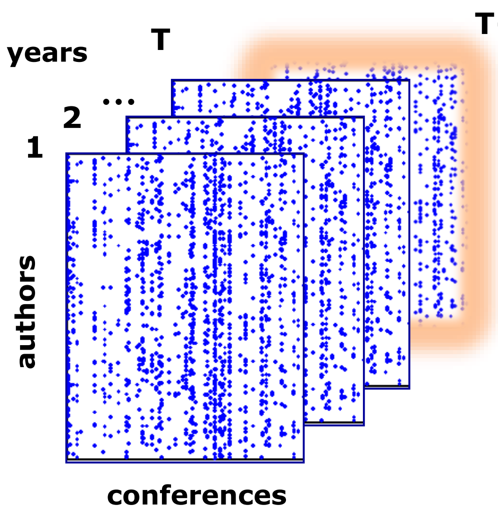


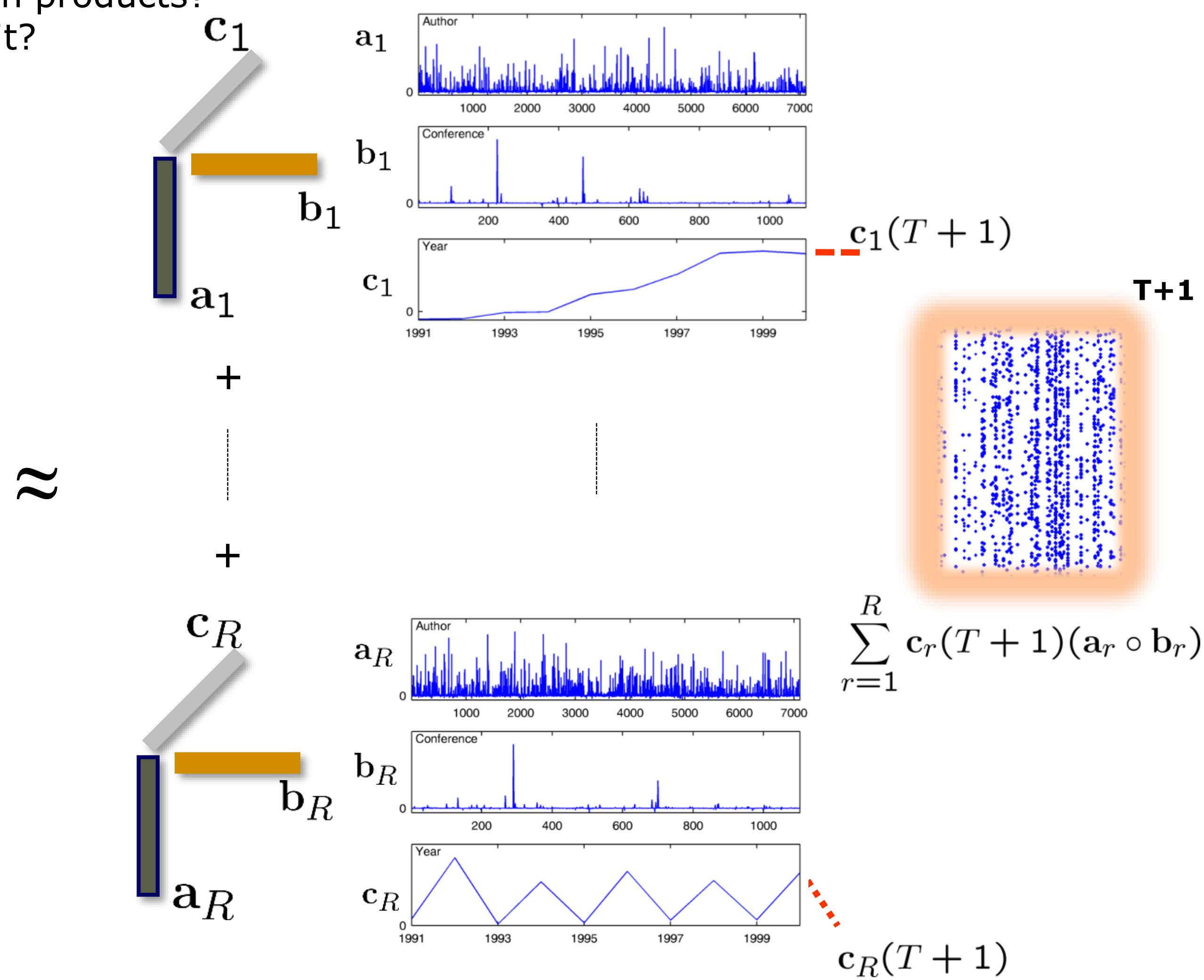
Recommender Systems: CP can capture temporal patterns useful for link prediction

[Dunlavy et al., *ACM TKDD,* 2011]

Temporal Link Prediction

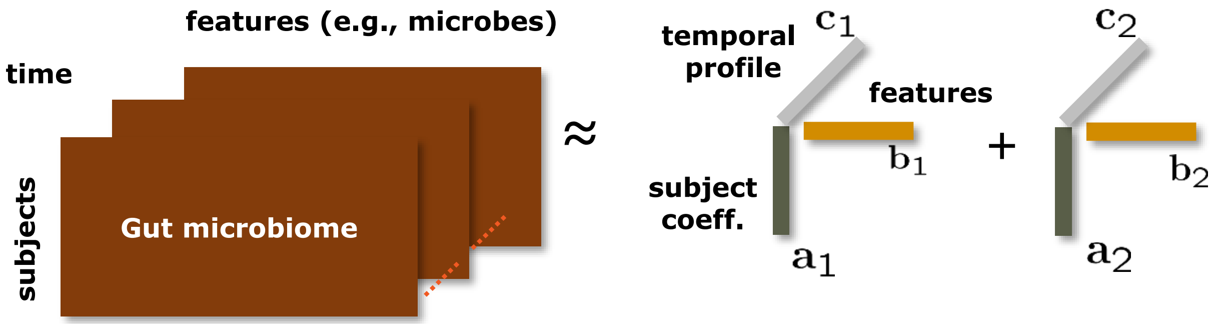
Which customers will buy which products?
Which webpages will users visit?



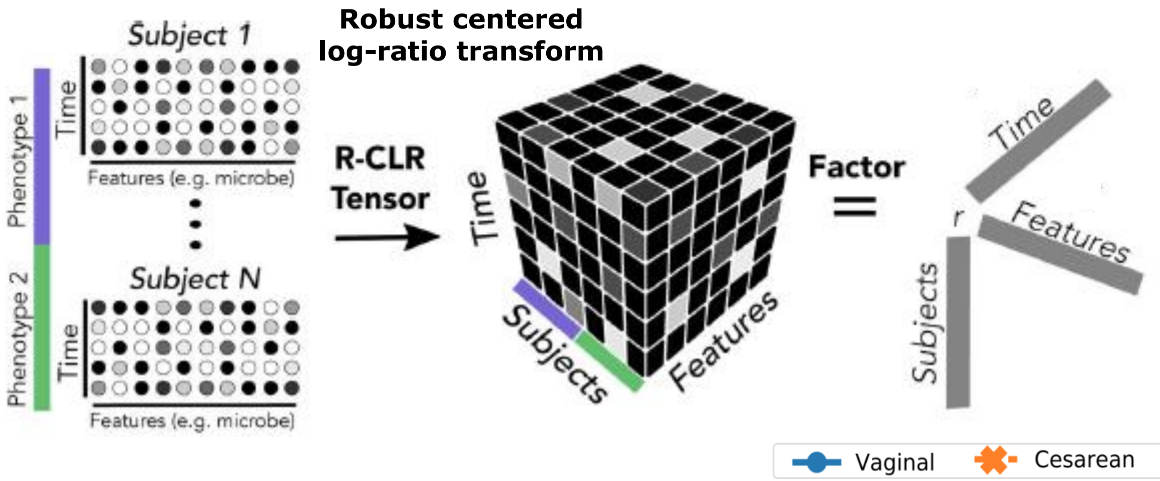


Microbiome: CP reveals gut microbial community dynamics

[Martino et al., *Nature Biotechnology*, 2021]



CP analysis of the gut microbiome data from infants (followed for the first few years of their life) reveals group differences (according to the birth mode, vaginal vs. cesarean) in terms of microbial community compositions.

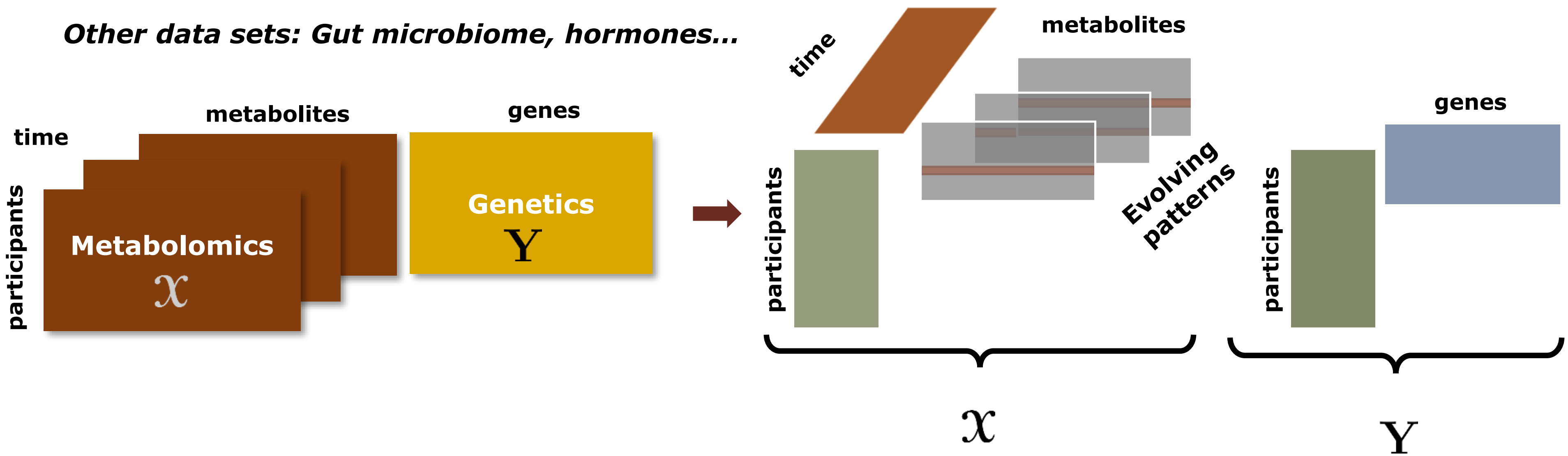


From multimodal heterogeneous data sets to interpretable patterns

Our goal: Joint analysis of time-evolving and static data sets to capture underlying patterns as well as their evolution in time

Multimodal Heterogeneous Data Sets

Interpretable Patterns and Their Evolution in Time



Many Challenges: Algorithmic and modelling

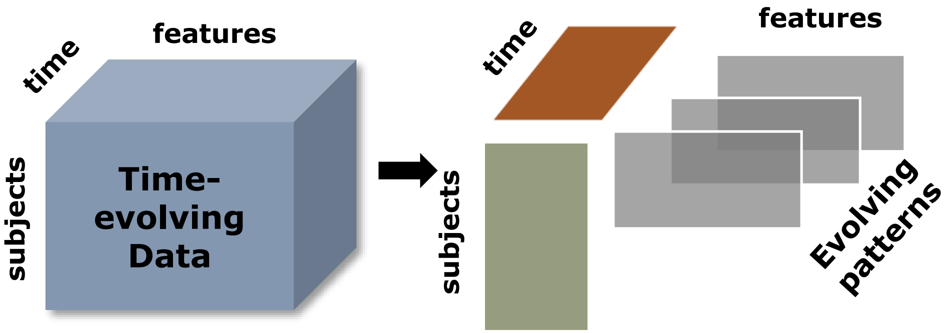
Data Fusion – how to jointly analyze heterogeneous data sets?

- Data in the form of matrices and higher-order tensors
- Different data distributions
- Shared patterns, and patterns visible only in one modality
- Need for interpretable and unique patterns

…

Part I:
A flexible algorithmic approach for regularized matrix-tensor factorizations with linear couplings

Evolving Patterns – how to extract evolving patterns from temporal data?



- Patterns evolving in time
- Need for interpretable and unique patterns

…

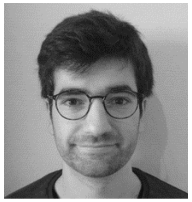
Part II:
Tracing evolving networks using PARAFAC2

Algorithmic framework for PARAFAC2 with constraints in all modes

Part III:
Putting the two building blocks together to jointly analyze dynamic and static data sets

Part I. A Flexible Algorithmic Approach for Regularized Matrix-Tensor Factorizations with Linear Couplings

joint work with



Carla Schenker
Simula & OsloMet

Jeremy Cohen
CNRS

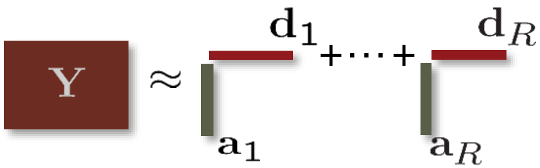
C. Schenker, J. E. Cohen, E. Acar. A Flexible Optimization Framework for Regularized Matrix-Tensor Factorizations with Linear Couplings, *IEEE Journal of Selected Topics in Signal Processing*, 15(3): 506-521, 2021

Coupled Matrix and Tensor Factorizations (CMTF)

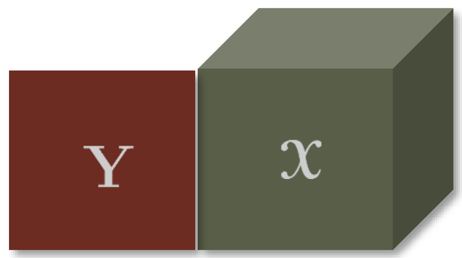
[Banerjee et al., *SDM*, 2007; Acar et al., *KDD Workshop MLG*, 2011]

Joint analysis of data sets in the form of matrices and higher-order tensors from multiple sources can be formulated as a coupled matrix and tensor factorization problem.

Matrix Factorization

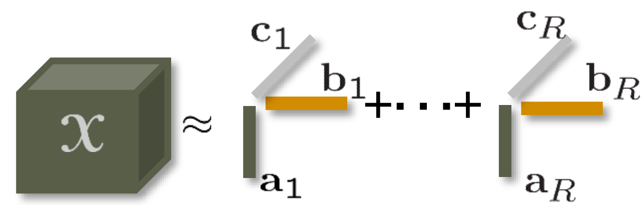


$$\mathbf{Y} \approx \sum_{r=1}^{R} \mathbf{a}_r \mathbf{d}_r^{\top} \approx \mathbf{A}\mathbf{D}^{\top}$$



$$\mathrm{Y} \approx \mathrm{A}\mathrm{D}^{\top} \quad \mathcal{X} \approx [\![\mathrm{A}, \mathrm{B}, \mathrm{C}]\!]$$

Tensor Factorization: CP

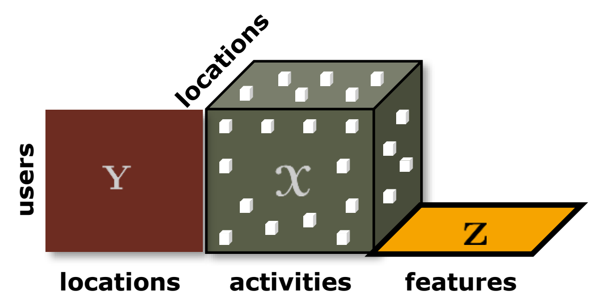


$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \approx [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

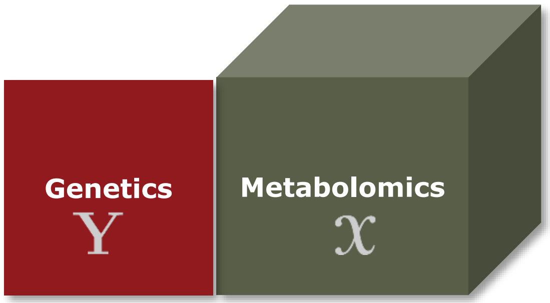
The problem can be formulated as:

$$\min_{\mathrm{A,B,C,D}} \| \mathcal{X} - [\![\mathrm{A}, \mathrm{B}, \mathrm{C}]\!] \|^2 + \| \mathrm{Y} - \mathrm{A}\mathrm{D}^{\top} \|^2$$

Many successful applications in recommender systems [Zheng et al., 2010; Ermis et al., 2015; Araujo et al., 2017]



There is a need for a flexible framework for data fusion



Various constraints on the factors

Various types of couplings between data sets

Data with different distributions (e.g., count data, binary data, real entries)➔ different loss functions

$$\mathbf{Y} \approx \mathbf{A}\mathbf{D}^{\mathsf{T}} \quad \mathcal{X} \approx [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$$

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} \| \mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \|^2 + \| \mathbf{Y} - \mathbf{A}\mathbf{D}^{\mathsf{T}} \|^2 \Rightarrow \min_{\mathbf{A}_1,\mathbf{A}_2,\mathbf{B},\mathbf{C},\mathbf{D}} L_x(\mathcal{X}, [\![\mathbf{A}_1, \mathbf{B}, \mathbf{C}]\!]) + L_y(\mathbf{Y}, \mathbf{A}_2\mathbf{D}^{\mathsf{T}}) + g(\mathbf{B})$$

$$\text{s.t.} \quad \mathbf{H}\mathbf{A}_1 = \mathbf{A}_2$$

State-of–the art in terms of CMTF algorithms

Alternating least squares (ALS)-based approaches **[Wilderjans et al., 2009; Bahargam and Papalexakis, 2019]**

Limited to Frobenius norm

--- with linear coupling **[Farias et al., 2016; Kanatsoulis et al., 2018]**

Limited in terms of constraints

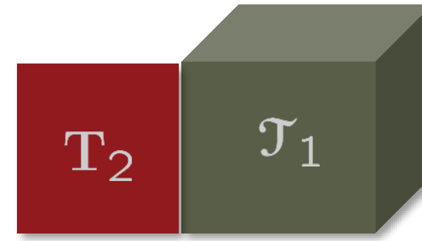
All-at-once optimization

- Unconstrained using gradient-based approaches **[Acar et al., 2011]**
- Nonlinear least squares **[Sorber et al., 2015; Vervliet et al., 2016]** ***Limited to Frobenius norm***
  --- with linear coupling and various constraints
- Constrained optimization using a general–purpose optimization solver **[Acar et al., 2014]**

There is a need for a flexible algorithmic framework that can handle various loss functions, incorporate different types of constraints and couplings

Alternating Optimization (AO) – Alternating Direction Method of Multipliers (ADMM) for Regularized CMTF with Linear Couplings

[Schenker et al., *EUSIPCO, 2020 & IEEE JSTSP*, 2021]



Previously, AO-ADMM has shown promising flexibility for constrained tensor factorizations [Huang et al., 2016]. We extend this framework to coupled matrix/tensor factorizations to incorporate various **constraints**, **loss functions** and **linear couplings**.

$$\min_{\{\mathbf{C}_{i,d},\boldsymbol{\Delta}_d\}_{d\le D_i, i\le N}} \mathcal{L}_1(\mathcal{T}_1, [\![\mathbf{C}_{1,1},\mathbf{C}_{1,2},\mathbf{C}_{1,3}]\!]) + \mathcal{L}_2(\mathbf{T}_2, \mathbf{C}_{2,1}\mathbf{C}_{2,2}^{\mathsf{T}}) + \sum_{i=1}^{N}\sum_{d=1}^{D_i} g_{i,d}\left(\mathbf{C}_{i,d}\right)$$
$$\text{s.t.} \quad \mathbf{H}_{i,d}\mathbf{C}_{i,d} = \boldsymbol{\Delta}_d$$

Fix all other modes, and solve for one mode using an alternating scheme (AO).

Example: Coupling only in the first mode

while convergence criterion is not met **do**

mode 1

$$\min_{\{\mathbf{C}_{i,1}\}_{i\le N},\boldsymbol{\Delta}_1} \mathcal{L}_1(\mathcal{T}_1, [\![\mathbf{C}_{1,1},\mathbf{C}_{1,2},\mathbf{C}_{1,3}]\!]) + \mathcal{L}_2(\mathbf{T}_2, \mathbf{C}_{2,1}\mathbf{C}_{2,2}^{\mathsf{T}}) + \sum_{i=1}^{N} g_{i,1}\left(\mathbf{C}_{i,1}\right)$$
$$\text{s.t.} \quad \mathbf{H}_{i,1}\mathbf{C}_{i,1} = \boldsymbol{\Delta}_1$$

mode 2

$$\min_{\{\mathbf{C}_{i,2}\}_{i\le N}} \mathcal{L}_1(\mathcal{T}_1, [\![\mathbf{C}_{1,1},\mathbf{C}_{1,2},\mathbf{C}_{1,3}]\!]) + \mathcal{L}_2(\mathbf{T}_2, \mathbf{C}_{2,1}\mathbf{C}_{2,2}^{\mathsf{T}}) + \sum_{i=1}^{N} g_{i,2}\left(\mathbf{C}_{i,2}\right)$$

end while

ADMM
Alternating Direction Method of Multipliers

$$\min_{\mathbf{x},\mathbf{z}} \quad f(\mathbf{x}) + g(\mathbf{z})$$
$$\text{s.t.} \quad \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}$$

Algorithm 1 Skeleton of scaled-form ADMM

while convergence criterion is not met **do**

$$x^{(k+1)} = \underset{x}{\operatorname{argmin}}\ f(x) + \tfrac{\rho}{2}\|Ax + Bz^{(k)} - c + \mu^{(k)}\|_2^2$$
$$z^{(k+1)} = \underset{z}{\operatorname{argmin}}\ g(z) + \tfrac{\rho}{2}\|Ax^{(k+1)} + Bz - c + \mu^{(k)}\|_2^2$$
$$\mu^{(k+1)} = \mu^{(k)} + Ax^{(k+1)} + Bz^{(k+1)} - c$$
$$k = k + 1$$

end while

ADMM subproblem for regularized CMTF with linear couplings (mode 1)

Optimization Problem

$$\min_{\{\mathbf{C}_{i,1}\}_{i\leq N},\boldsymbol{\Delta}_1} \mathcal{L}_1(\mathcal{T}_1, [\![\mathbf{C}_{1,1},\mathbf{C}_{1,2},\mathbf{C}_{1,3}]\!]) + \mathcal{L}_2(\mathbf{T}_2, \mathbf{C}_{2,1}\mathbf{C}_{2,2}^\mathsf{T}) + \sum_{i=1}^{N} g_{i,1}\left(\mathbf{C}_{i,1}\right)$$

$$\text{s.t.} \quad \mathbf{H}_{i,1}\mathbf{C}_{i,1} = \boldsymbol{\Delta}_1$$

Introduce variable $\mathbf{Z}_{i,1}$ to separate the regularization from the factorization

$$\min_{\{\mathbf{C}_{i,1},\mathbf{Z}_{i,1}\}_{i\leq N},\boldsymbol{\Delta}_1} \mathcal{L}_1(\mathcal{T}_1, [\![\mathbf{C}_{1,1},\mathbf{C}_{1,2},\mathbf{C}_{1,3}]\!]) + \mathcal{L}_2(\mathbf{T}_2, \mathbf{C}_{2,1}\mathbf{C}_{2,2}^\mathsf{T}) + \sum_{i=1}^{N} g_{i,1}\left(\mathbf{Z}_{i,1}\right)$$

$$\text{s.t.} \quad \mathbf{H}_{i,1}\mathbf{C}_{i,1} = \boldsymbol{\Delta}_1$$

$$\mathbf{C}_{i,1} = \mathbf{Z}_{i,1}$$

Introduce dual variables and formulate the augmented Lagrangian

$$L(\mathbf{C}_{i,1},\mathbf{Z}_{i,1},\boldsymbol{\Delta}_1,\mu_{i,1_{(z)}},\mu_{i,1_{(\Delta)}}) = \mathcal{L}_1\left(\mathcal{T}_1, [\![\mathbf{C}_{1,1},\mathbf{C}_{1,2},\mathbf{C}_{1,3}]\!]\right) + \mathcal{L}_2\left(\mathbf{T}_2, \mathbf{C}_{2,1}\mathbf{C}_{2,2}^\mathsf{T}\right)$$
$$+ \sum_{i=1}^{N}\left[g_{i,1}(\mathbf{Z}_{i,1}) + \frac{\rho}{2}\left\|\mathbf{C}_{i,1} - \mathbf{Z}_{i,1} + \mu_{i,1_{(z)}}\right\|_F^2 + \frac{\rho}{2}\left\|\mathbf{H}_{i,1}\mathbf{C}_{i,1} - \boldsymbol{\Delta}_1 + \mu_{i,1_{(\Delta)}}\right\|_F^2\right]$$

Using alternating optimization, solve for $\{\mathbf{C}_{i,1}\}_{i\leq N}, \{\mathbf{Z}_{i,1}\}_{i\leq N}, \boldsymbol{\Delta}_1$ followed by dual updates.

In case of Frobenius norm-based loss:
Solution of a linear least squares problem or Sylvester eqn.

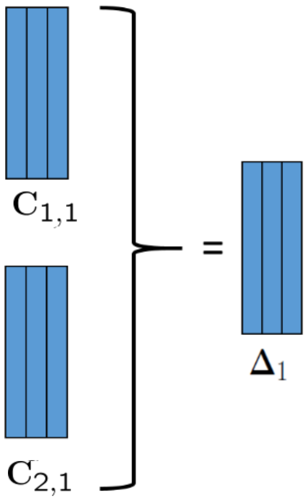
Proximal operators

$$\text{prox}_{\lambda,g}(\mathbf{x}) = \underset{\mathbf{u}}{\text{argmin}} \;\; g(\mathbf{u}) + \frac{1}{2\lambda}\|\mathbf{x}-\mathbf{u}\|_2^2$$

Other differentiable losses:
Numerical optimization using LBFGS-B

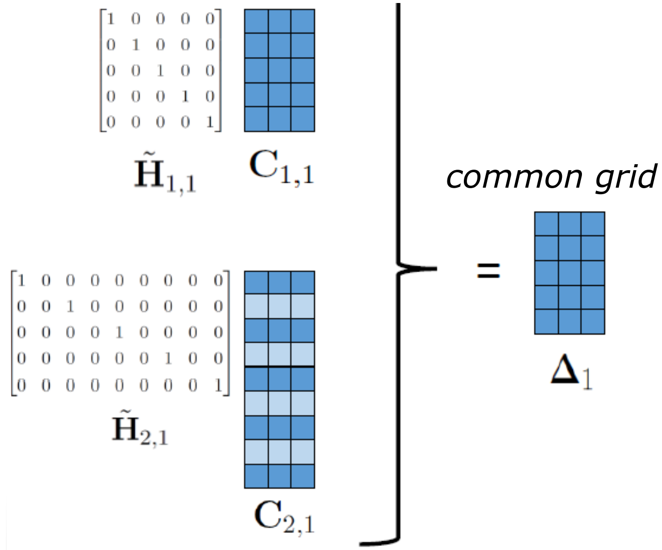
Efficient updates derived for the following types of linear coupling:

Case 1: Hard/exact coupling

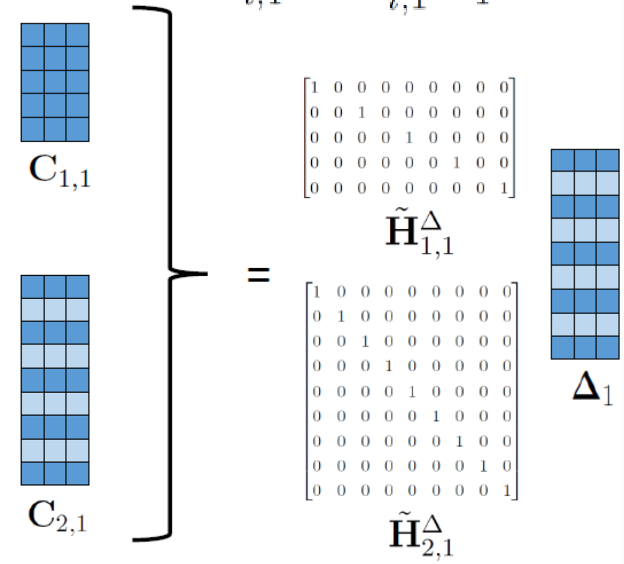


Case 2: Transformations in *mode* dimension

$$\tilde{\mathbf{H}}_{i,1}\mathbf{C}_{i,1} = \boldsymbol{\Delta}_1$$

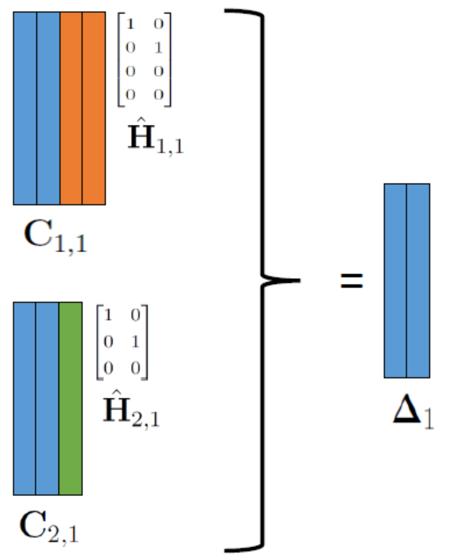


$$\mathbf{C}_{i,1} = \tilde{\mathbf{H}}^{\Delta}_{i,1}\boldsymbol{\Delta}_1$$

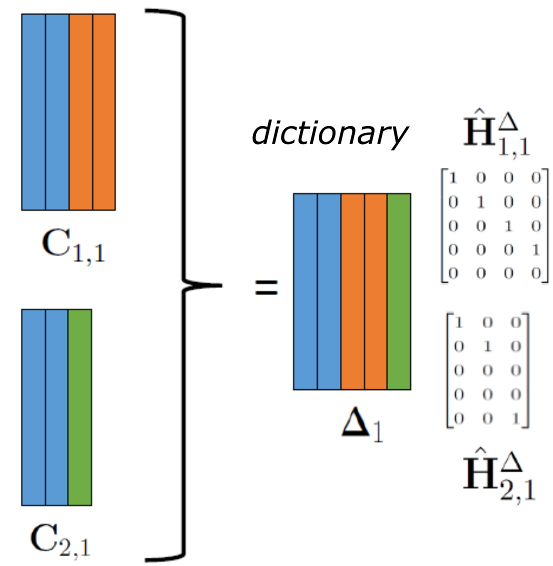


Case 3: Transformations in *component* dimension

$$\mathbf{C}_{i,1}\hat{\mathbf{H}}_{i,1} = \boldsymbol{\Delta}_1$$

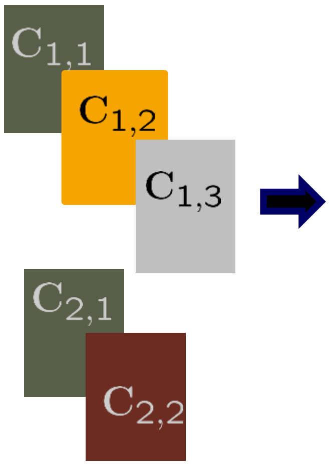


$$\mathbf{C}_{i,1} = \boldsymbol{\Delta}_1\hat{\mathbf{H}}^{\Delta}_{i,1}$$

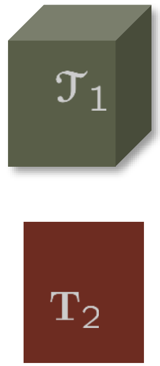


Exact coupling: AO-ADMM framework is accurate and efficient!

Generate factor matrices



Construct coupled data sets



$$= [\![C_{1,1}, C_{1,2}, C_{1,3}]\!] + \mathcal{N}$$

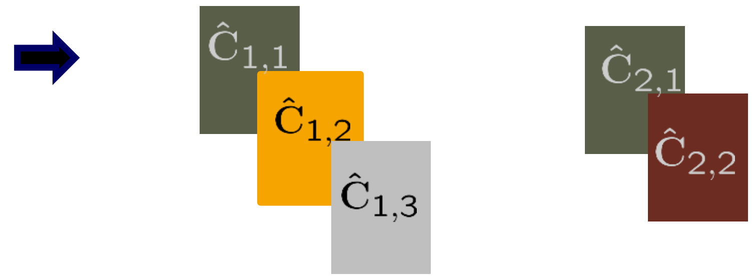
$$= C_{2,1} C_{2,2}^{T} + N$$

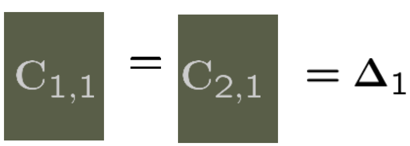
Solve using AO-ADMM

$$\min_{\{C_{i,d}\}_{d \le D_i, i \le N}, \Delta_1} \left\| \mathcal{T}_1 - [\![C_{1,1}, C_{1,2}, C_{1,3}]\!] \right\|_F^2 + \left\| T_2 - C_{2,1} C_{2,2}^{T} \right\|_F^2$$

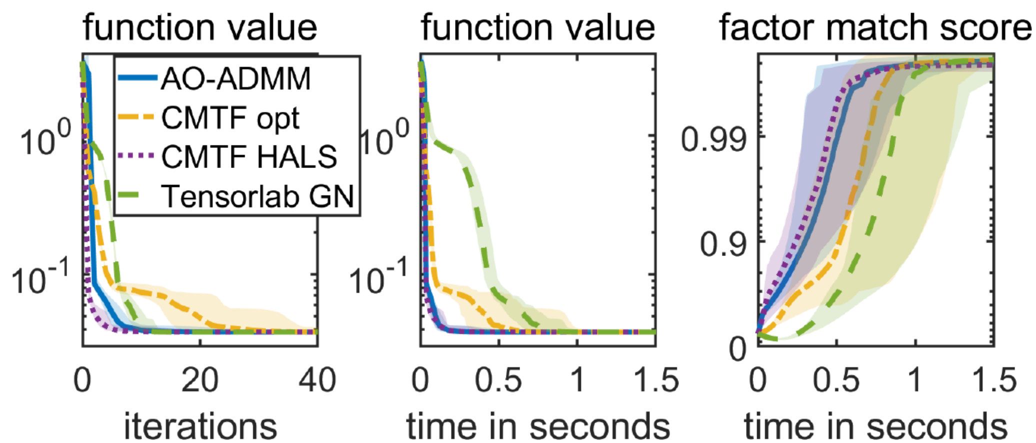
$$\text{s.t.} \quad C_{i,1} = \Delta_1$$

$$C_{i,d} \ge 0, i \le N, d \le D_i$$





50 random datasets,
best run out of multiple
random initializations reported



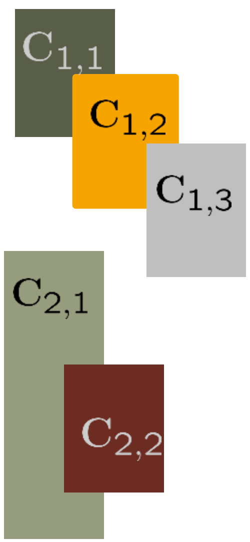
Factor match score:

$$\text{FMS} = \prod_{i=1}^{N} \frac{1}{R_i} \sum_{r=1}^{R_i} \left(\prod_{d=1}^{D_i} \frac{C_{i,d}(:,r)^{T} \hat{C}_{i,d}(:,r)}{\left\| C_{i,d}(:,r) \right\| \left\| \hat{C}_{i,d}(:,r \right\|} \right)$$

Comparable performance in terms of accuracy and computational efficiency!

Linear coupling: AO-ADMM framework is accurate and efficient!

Generate factor matrices



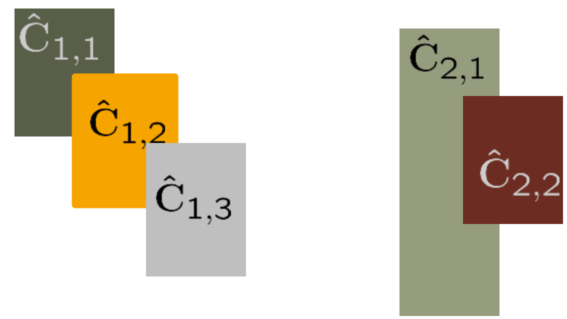
Construct coupled data sets

$$= [\![\mathbf{C}_{1,1}, \mathbf{C}_{1,2}, \mathbf{C}_{1,3}]\!] + \mathcal{N}$$

$$= \mathbf{C}_{2,1}\mathbf{C}_{2,2}^{\mathsf{T}} + \mathbf{N}$$

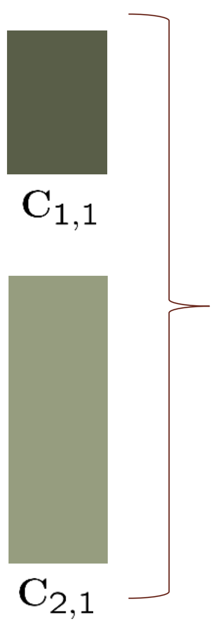
Solve CMTF with linear couplings using AO-ADMM

$$\min_{\{\mathbf{C}_{i,d}\}_{d \le D_i, i \le N}, \mathbf{\Delta}_1} \left\| \mathcal{T}_1 - [\![\mathbf{C}_{1,1}, \mathbf{C}_{1,2}, \mathbf{C}_{1,3}]\!] \right\|_F^2 + \left\| \mathbf{T}_2 - \mathbf{C}_{2,1}\mathbf{C}_{2,2}^{\mathsf{T}} \right\|_F^2$$
$$\text{s.t.} \quad \mathbf{H}_{i,1}\mathbf{C}_{i,1} = \mathbf{\Delta}_1$$

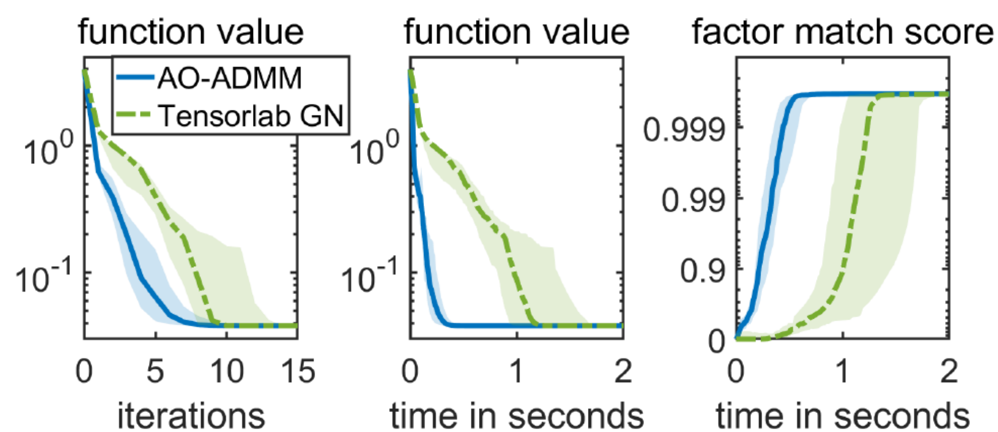


$$\mathbf{H}_{1,1} = \begin{pmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \end{pmatrix}$$

$$\mathbf{H}_{2,1} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&1&0 \end{pmatrix}$$



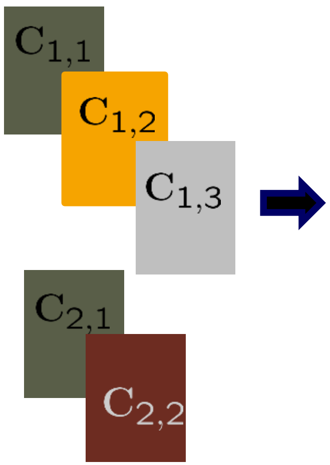
$$= \mathbf{\Delta}_1$$



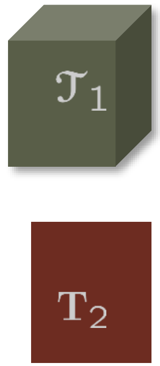
AO-ADMM is computationally competitive!

Kullback-Leibler (KL) loss: AO-ADMM framework is flexible!

Generate factor matrices



Construct coupled data sets

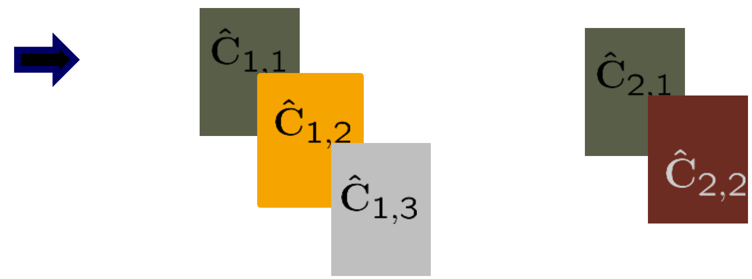


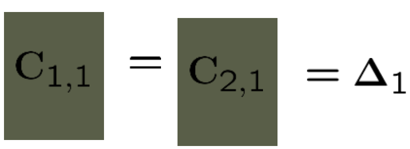
$\approx [\![C_{1,1}, C_{1,2}, C_{1,3}]\!]$

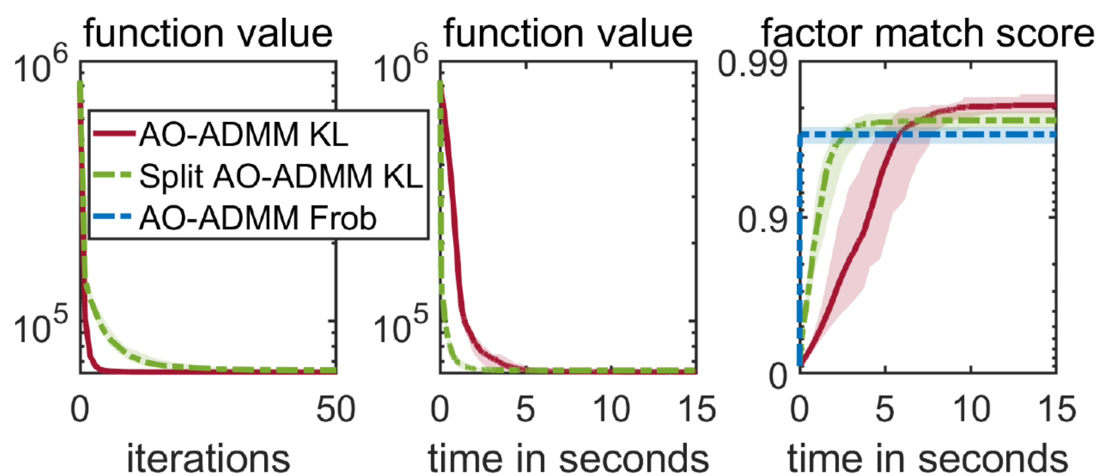
$\approx C_{2,1}C_{2,2}^{T}$

Solve using AO-ADMM

$$\min_{\{C_{i,d}\}_{d\le D_i, i\le N}, \Delta_1} \mathcal{L}(\mathcal{T}_1, [\![C_{1,1}, C_{1,2}, C_{1,3}]\!]) + \mathcal{L}(T_2, C_{2,1}C_{2,2}^{T})$$
$$\text{s.t.} \quad C_{i,1} = \Delta_1$$
$$C_{i,d} \ge 0, i \le N, d \le D_i$$

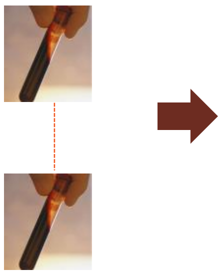


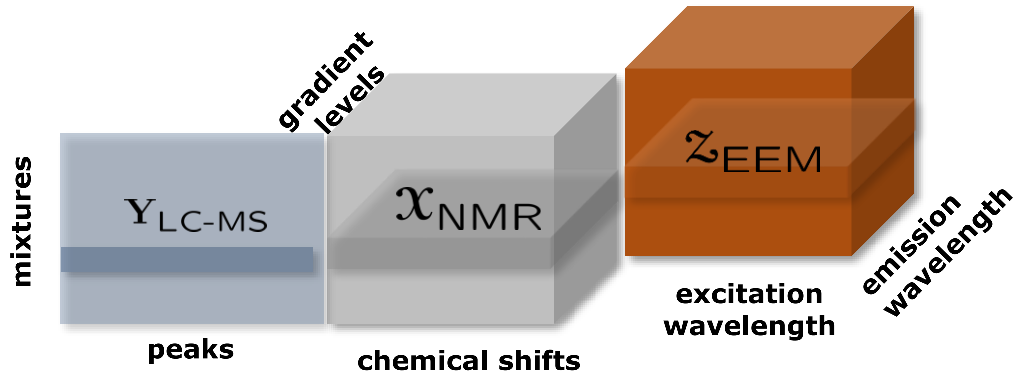




AO-ADMM is accurate with different loss functions as well!

Chemometrics: Underlying design and patterns captured accurately!





Mixtures prepared using five chemicals:

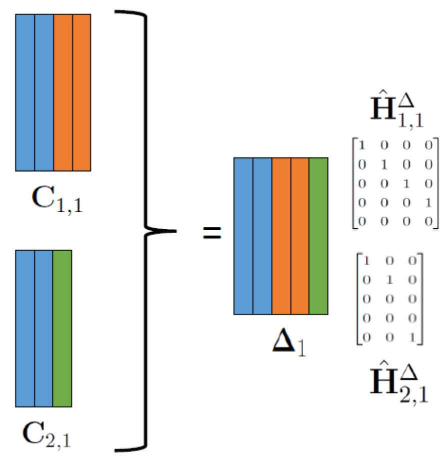
- Val-Try-Val
- Trp - Gly
- Phe
- Maltoheptaose
- Propanol

$$\min_{\substack{\{\mathbf{C}_{i,d}\}_{i=1,2,3} \\ d \le D_i}} \left\| \mathcal{Z}_{\text{EEM}} - [\![\mathbf{C}_{1,1}, \mathbf{C}_{1,2}, \mathbf{C}_{1,3}]\!] \right\|_F^2 + \left\| \mathcal{X}_{\text{NMR}} - [\![\mathbf{C}_{2,1}, \mathbf{C}_{2,2}, \mathbf{C}_{2,3}]\!] \right\|_F^2 + \left\| \mathbf{Y}_{\text{LCMS}} - \mathbf{C}_{3,1}\mathbf{C}_{3,2}^{\mathsf{T}} \right\|_F^2$$

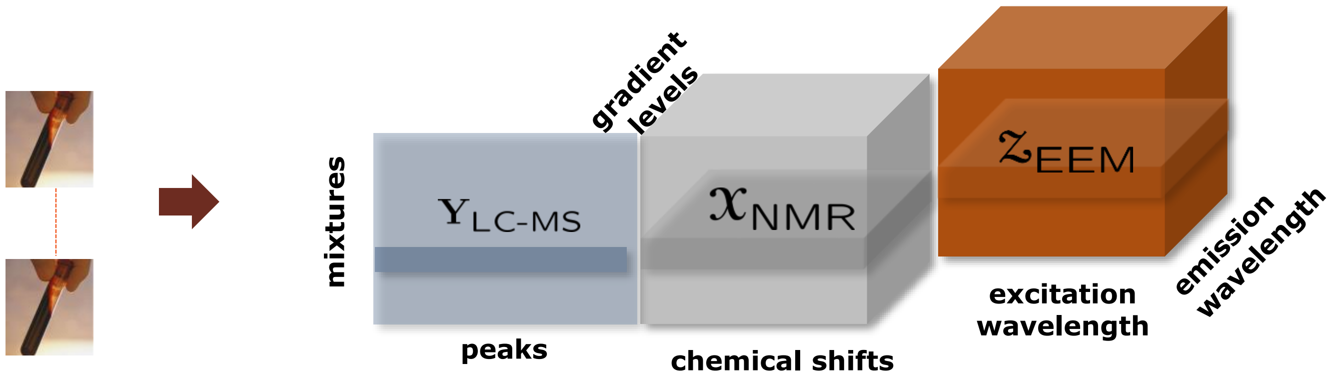
$$\texttt{s.t.} \quad \mathbf{C}_{i,d} \ge 0, i = 1,2,3, d \le D_i$$

$$\mathbf{C}_{1,1} = \boldsymbol{\Delta}_1 \hat{\mathbf{H}}^{\Delta}_{1,1}, \mathbf{C}_{2,1} = \boldsymbol{\Delta}_1 \hat{\mathbf{H}}^{\Delta}_{2,1}, \mathbf{C}_{3,1} = \boldsymbol{\Delta}_1 \hat{\mathbf{H}}^{\Delta}_{3,1}$$

$$\mathbf{C}_{i,1} = \boldsymbol{\Delta}_1 \hat{\mathbf{H}}^{\Delta}_{i,1}$$



Chemometrics: Underlying design and patterns captured accurately!



Mixtures prepared using five chemicals:

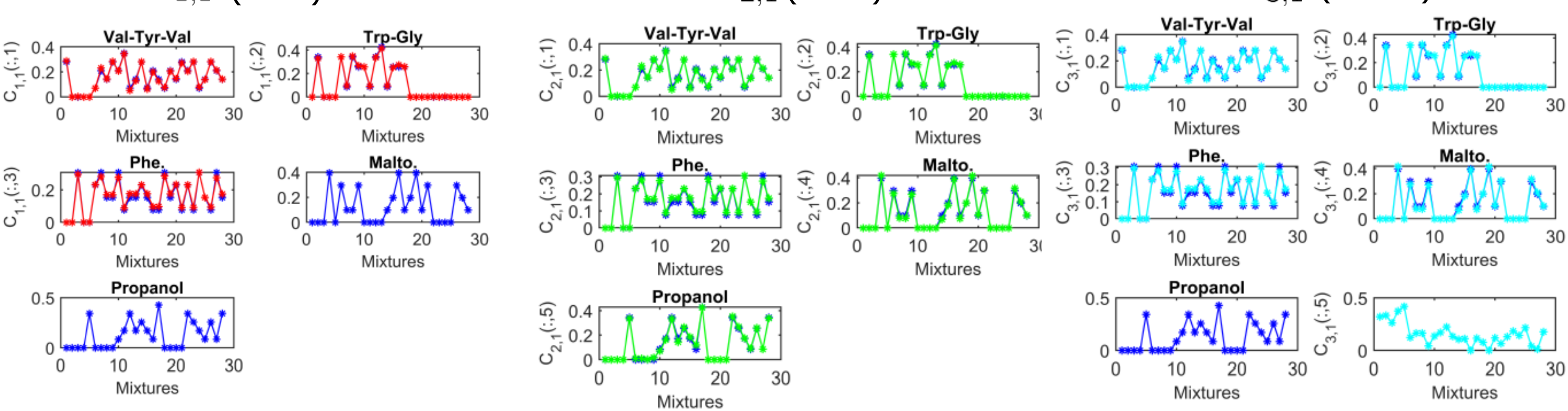
- Val-Try-Val
- Trp - Gly
- Phe
- Maltoheptaose
- Propanol

$$\min_{\substack{\{\mathbf{C}_{i,d}\}_{i=1,2,3} \\ d\le D_i}} \left\| \mathcal{Z}_{\text{EEM}} - [\![\mathbf{C}_{1,1},\mathbf{C}_{1,2},\mathbf{C}_{1,3}]\!] \right\|_F^2 + \left\| \mathcal{X}_{\text{NMR}} - [\![\mathbf{C}_{2,1},\mathbf{C}_{2,2},\mathbf{C}_{2,3}]\!] \right\|_F^2 + \left\| \mathbf{Y}_{\text{LCMS}} - \mathbf{C}_{3,1}\mathbf{C}_{3,2}^{\text{T}} \right\|_F^2$$

$$\texttt{s.t.} \quad \mathbf{C}_{i,d} \ge 0, i = 1,2,3, d \le D_i$$

$$\mathbf{C}_{1,1} = \mathbf{\Delta}_1 \hat{\mathbf{H}}_{1,1}^{\Delta}, \mathbf{C}_{2,1} = \mathbf{\Delta}_1 \hat{\mathbf{H}}_{2,1}^{\Delta}, \mathbf{C}_{3,1} = \mathbf{\Delta}_1 \hat{\mathbf{H}}_{3,1}^{\Delta}$$

$\mathbf{C}_{1,1}$ (EEM) $\mathbf{C}_{2,1}$ (NMR) $\mathbf{C}_{3,1}$ (LCMS)



Many Challenges: Algorithmic and modelling

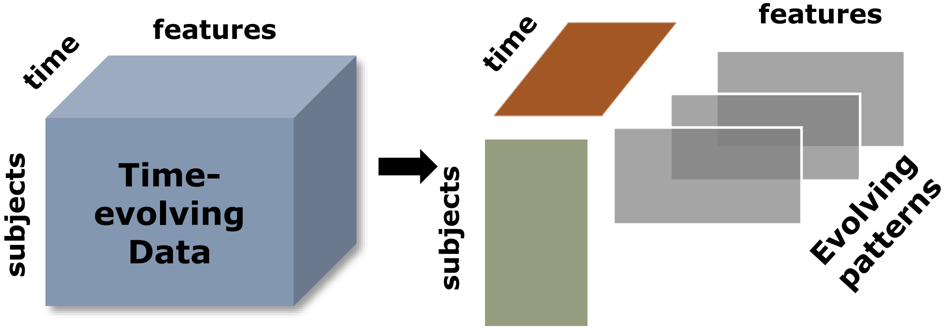
Data Fusion – how to jointly analyze such heterogeneous data?

- Data in the form of matrices and higher-order tensors
- Different data distributions
- Shared patterns, and patterns visible only in one modality
- Need for interpretable and unique patterns

…

Part I:
A flexible algorithmic approach for regularized matrix-tensor factorizations with linear couplings

Evolving Patterns – how to extract evolving patterns from temporal data?



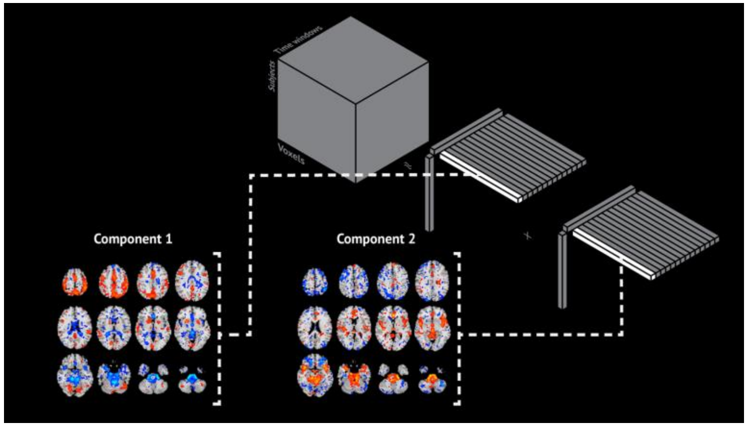
- Patterns evolving in time
- Need for interpretable and unique patterns

…

Part II:
Tracing evolving networks using PARAFAC2

Algorithmic framework for PARAFAC2 with constraints in all modes

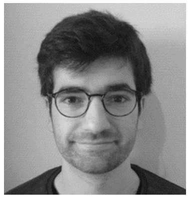
Part III:
Putting the two building blocks together to jointly analyze dynamic and static data sets



From Marie Roald's ICASSP 2020 talk

Part II. Tracing Evolving Networks using PARAFAC2

joint work with



Marie Roald
Simula & OsloMet

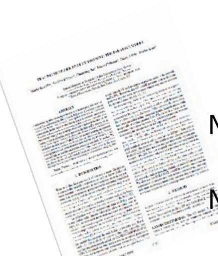
Carla Schenker
Simula & OsloMet

Vince Calhoun
Georgia State University

Tulay Adali
University of Maryland

Rasmus Bro
University of Copenhagen

Jeremy Cohen
CNRS



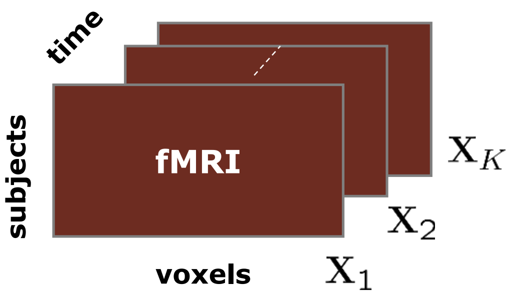
M. Roald, C. Schenker, V. Calhoun, T. Adali, R. Bro, J. Cohen, E. Acar. An AO-ADMM approach to constraining PARAFAC2 on all modes, *SIAM Journal on Mathematics of Data Science*, 4(3): 1191-1222, 2022

M. Roald, S. Bhinge, C. Jia, V. Calhoun, T. Adali, E. Acar. Tracing Network Evolution using the PARAFAC2 Model, *ICASSP*, pp. 1100-1104, 2020

How can we extract evolving patterns from time-evolving data?

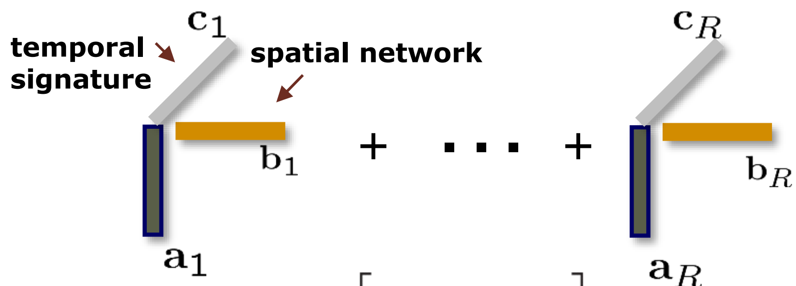
Higher-order tensors are natural data representations for temporal data in general. For instance, when studying **spatial dynamics** in the **brain**, we are interested in analyzing **time-evolving fMRI** data, which can be represented as a ***subjects* by *voxels* by *time windows*** tensor.

Static spatial component/network



CP

$\approx$



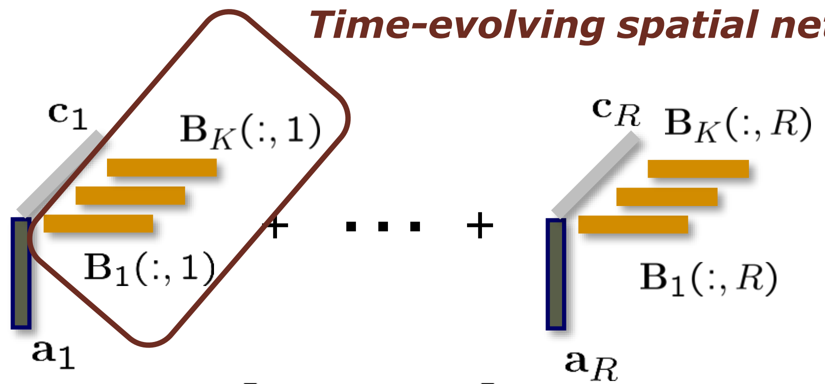
$$X_k \approx A \begin{bmatrix} c_{k1} & & 0 \\ & \ddots & \\ 0 & & c_{kR} \end{bmatrix} B^\top$$

Time-evolving spatial network

PARAFAC2

[Harshman, 1972]

$\approx$



$$X_k \approx A \begin{bmatrix} c_{k1} & & 0 \\ & \ddots & \\ 0 & & c_{kR} \end{bmatrix} B_k^\top$$

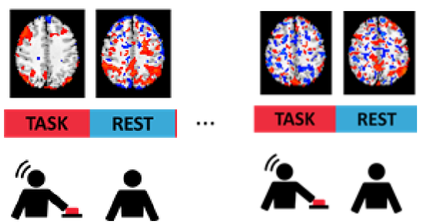
Can PARAFAC2 reveal evolving patterns?

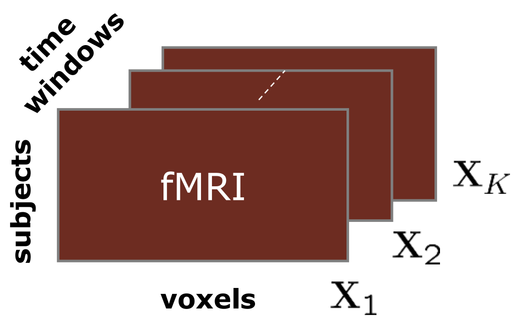
$$B_k^\top B_k = \Phi, \text{ for } k = 1, 2, ..K$$

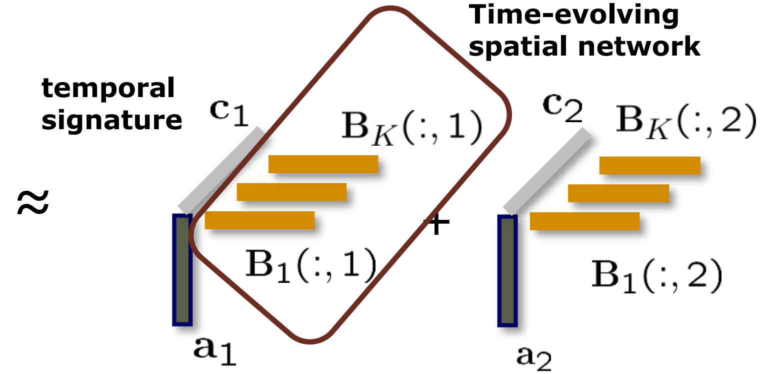
PARAFAC2 can reveal evolving spatial regions (spatial dynamics)

[Roald et al., *ICASSP*, 2020; Acar et al., *Frontiers in Neuroscience*, 2022]

The traditional approach in fMRI data analysis is to assume that underlying spatial regions of interest are static. We arrange time-evolving fMRI data as a ***subjects* by *voxels* by *time tensor*** and analyze using a **PARAFAC2 model** to reveal **spatial dynamics**.



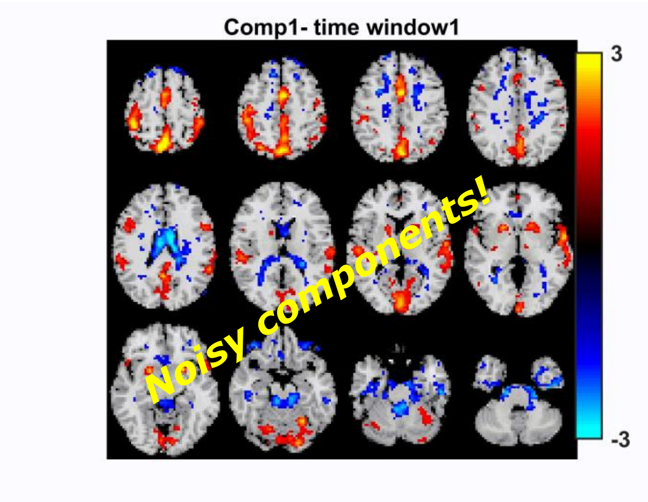


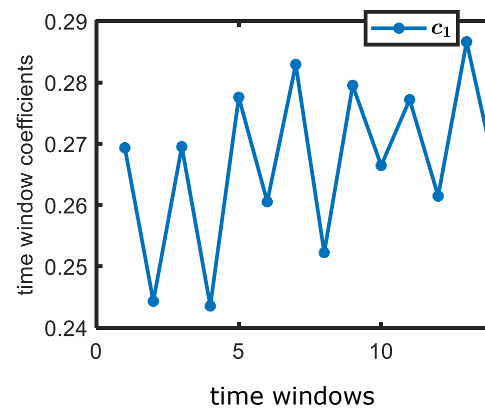


147 healthy controls
106 patients with schizophrenia
[Gollub et al., 2013]

The first component identified as a statistically significant component using a two-sample *t*-test on the columns of **A**

p-value : 7.8×10^{-6}





Challenging to impose constraints on evolving patterns using the traditional ALS - based algorithm for PARAFAC2

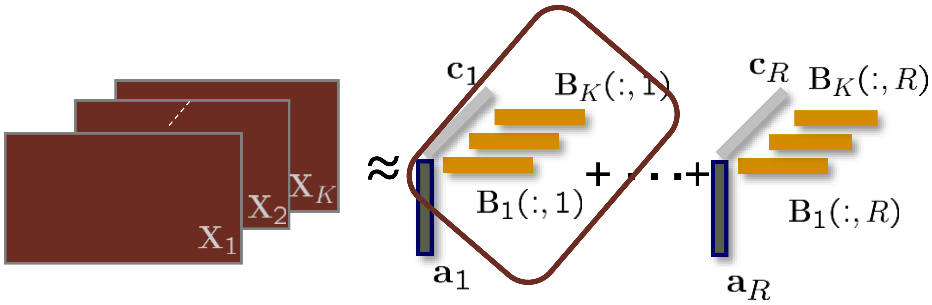
[Kiers et al., *J. Chemom.*, 1999]

Optimization Problem

$$\mathbf{D}_k = \text{diag}(\mathbf{c}(k,:))$$

$$\min_{\mathbf{A},\{\mathbf{B}_k\}_{k\le K},\mathbf{C}} \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\top \right\|_F^2$$

$$\text{s.t.} \quad \mathbf{B}_k^\top\mathbf{B}_k = \boldsymbol{\Phi}, \text{ for } k = 1, \ldots, K$$



$$\mathbf{B}_k = \mathbf{P}_k\mathbf{B}, \text{ and } \mathbf{P}_k^\top\mathbf{P}_k = \mathbf{I}$$

$$\min_{\mathbf{A},\mathbf{B},\{\mathbf{P}_k\}_{k\le K},\mathbf{C}} \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}^\top\mathbf{P}_k^\top \right\|_F^2$$

$$\text{s.t.} \quad \mathbf{P}_k^\top\mathbf{P}_k = \mathbf{I}, \text{ for } k = 1, \ldots, K$$

Challenging to impose constraints on $\mathbf{B}_k$

PARAFAC2-ALS

while convergence criterion is not met **do**

Solve for $\mathbf{P}_k$, for $k = 1, \ldots, K$

$$[\mathbf{U}_k, \boldsymbol{\Sigma}_k, \mathbf{V}_k] = \text{svd}(\mathbf{X}_k^\top\mathbf{A}\mathbf{D}_k\mathbf{B}^\top)$$

$$\mathbf{P}_k = \mathbf{U}_k\mathbf{V}_k^\top$$

Solve the following using regular CP-ALS updates

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \sum_{k=1}^{K} \left\| \mathbf{X}_k\mathbf{P}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}^\top \right\|_F^2$$

end while

PARAFAC2 AO-ADMM enables having constraints in all modes

[Roald et al., *EUSIPCO*, 2021 & *SIMODS*, 2022]

Optimization Problem (with regularization terms possibly in all modes):

$$\min_{\mathbf{A},\{\mathbf{B}_k,\mathbf{D}_k\}_{k\le K}} \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\mathsf{T} \right\|_F^2 + g_{\mathbf{A}}(\mathbf{A}) + \sum_{k=1}^{K} \left\{ g_{\mathbf{B}_k}(\mathbf{B}_k) + g_{\mathbf{D}_k}(\mathbf{D}_k) \right\}$$

$$\text{s.t.} \quad \{\mathbf{B}_k\}_{k\le K} \in \mathcal{P}$$

$$\mathcal{P} = \{\{\mathbf{B}_k\}_{k=1}^{K} \mid \mathbf{B}_{k_1}^\mathsf{T}\mathbf{B}_{k_1} = \mathbf{B}_{k_2}^\mathsf{T}\mathbf{B}_{k_2} \ \forall k_1, k_2 \le K\}$$

PARAFAC2 AO-ADMM

while convergence criterion is not met **do**

$$\min_{\{\mathbf{B}_k\}_{k\le K}} \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\mathsf{T} \right\|_F^2 + g_{\mathbf{B}_k}(\mathbf{B}_k) \quad \text{s.t.} \ \{\mathbf{B}_k\}_{k\le K} \in \mathcal{P}$$

$$\min_{\mathbf{A}} \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\mathsf{T} \right\|_F^2 + g_{\mathbf{A}}(\mathbf{A})$$

$$\min_{\mathbf{D}_k} \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\mathsf{T} \right\|_F^2 + g_{\mathbf{D}_k}(\mathbf{D}_k)$$

end while

ADMM Updates for $\mathbf{B}_k$

$$\min_{\{\mathbf{B}_k\}_{k\le K}} \sum_{k=1}^{K} \left[\left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\mathsf{T} \right\|_F^2 + g_{\mathbf{B}_k}\left(\mathbf{Z}_{\mathbf{B}_k}\right) \right] + \iota_{\mathcal{P}}\left(\left\{ \mathbf{Y}_{\mathbf{B}_k} \right\}_{k\le K} \right)$$

$$\text{s.t.} \quad \mathbf{B}_k = \mathbf{Z}_{\mathbf{B}_k} \quad \forall k \le K, \qquad \mathbf{B}_k = \mathbf{Y}_{\mathbf{B}_k} \quad \forall k \le K$$

while convergence criterion is not met **do**

for *k=1, …, K*

Solve for $\mathbf{B}_k$ using a least squares update

Update $\mathbf{Z}_{\mathbf{B}_k}$ using a proximal operator

end for

Update $\{\mathbf{Y}_{\mathbf{B}_k}\}_{k\le K}$

$$\min_{\mathbf{\Delta}_\mathbf{B},\{\mathbf{P}_k\}_{k\le K}} \left[\sum_{k=1}^{K} \frac{\rho_{\mathbf{B}_k}}{2} \left\| \mathbf{B}_k - \mathbf{P}_k\mathbf{\Delta}_\mathbf{B} + \mu_{\mathbf{\Delta}_{\mathbf{B}_k}} \right\|^2 \right] \quad \text{subject to} \quad \mathbf{P}_k^\mathsf{T}\mathbf{P}_k = \mathbf{I} \quad \forall k \le K.$$

Update the dual variables

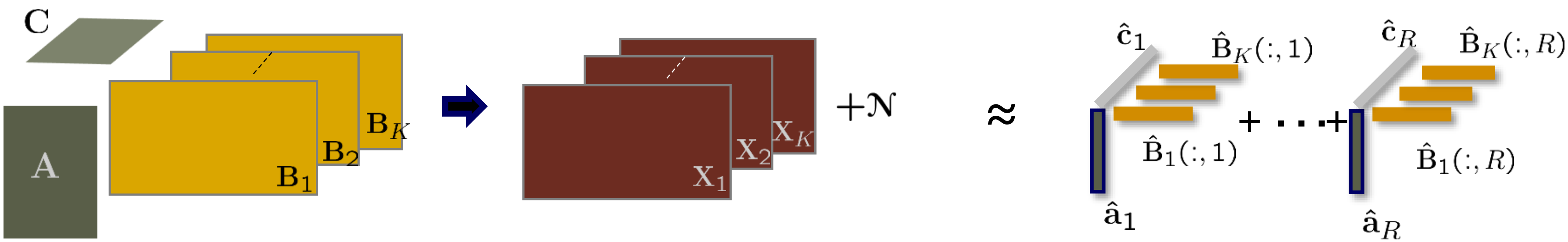
end while

PARAFAC2 AO-ADMM is accurate and computationally efficient!

Generate factor matrices

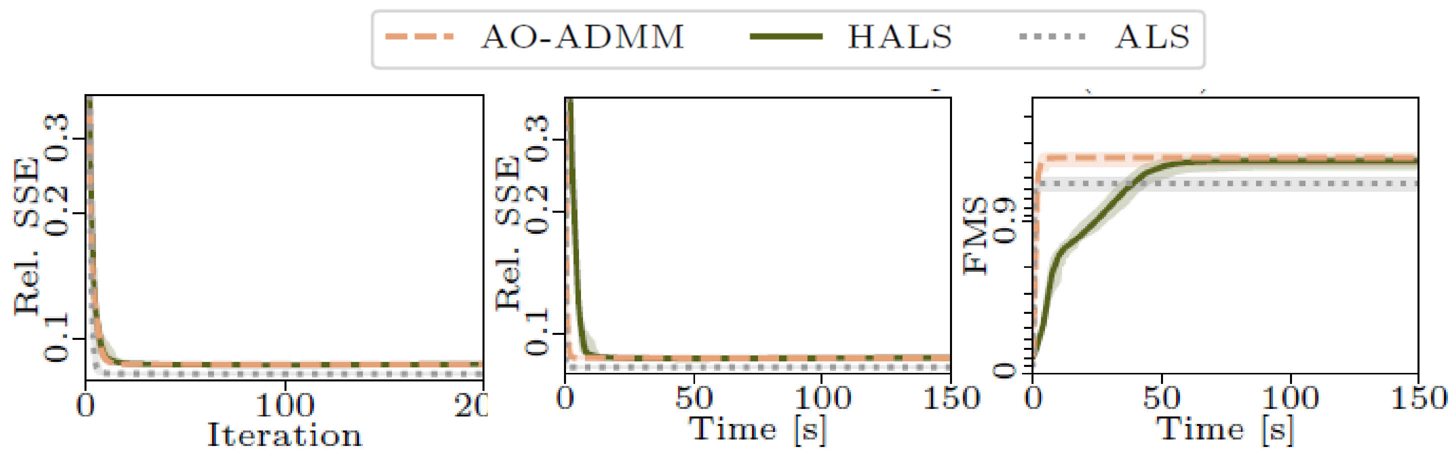
Construct a tensor following a PARAFAC2 structure

Fit a PARAFAC2 model with non-negativity constraints



Non-negative factor matrices

50 random datasets, best run out of multiple random initializations reported



PARAFAC2 with flexible coupling
[Cohen and Bro, 2018]

$$\min_{\mathbf{A},\{\mathbf{D}_k\}_{k\leq K},\{\mathbf{B}_k\}_{k\leq K},\boldsymbol{\Delta}_\mathbf{B},\{\mathbf{P}_k\}_{k\leq K}} \left[\sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\mathsf{T} \right\|_F^2 + \mu_t \left\| \mathbf{B}_k - \mathbf{P}_k \boldsymbol{\Delta}_\mathbf{B} \right\|_F^2\right]$$

$$\begin{aligned} \text{subject to} \quad & \mathbf{P}_k^\mathsf{T}\mathbf{P}_k = \mathbf{I} \qquad \forall k \leq K, \\ & \mathbf{A} \geq 0, \\ & \mathbf{B}_k, \mathbf{D}_k \geq 0 \qquad \forall k \leq K \end{aligned}$$

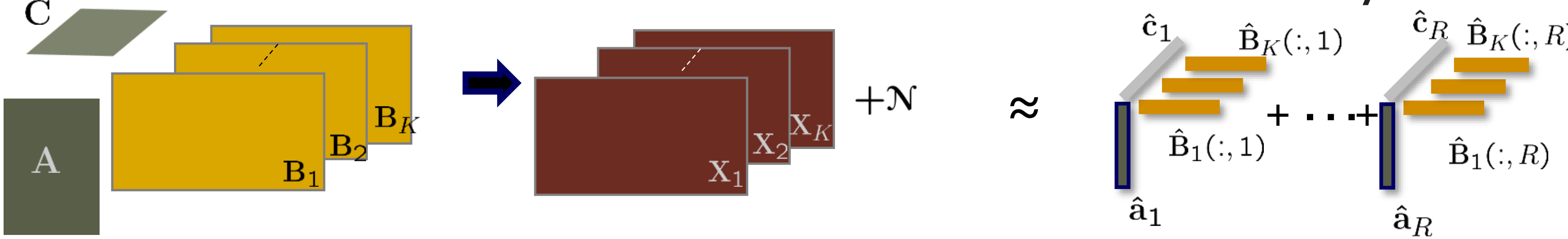
AO-ADMM is as accurate as the flexible coupling approach using hierarchical ALS (HALS), and computationally more efficient!

PARAFAC2 AO-ADMM is flexible!

Generate factor matrices

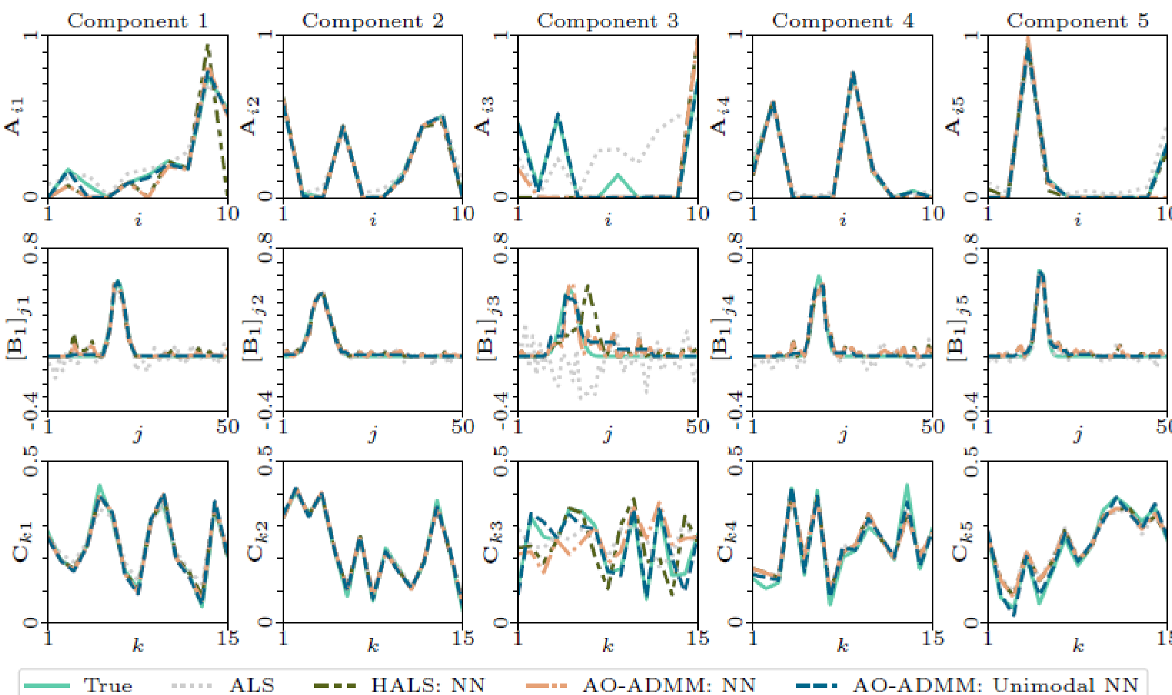
Construct a tensor (almost) following a PARAFAC2 structure

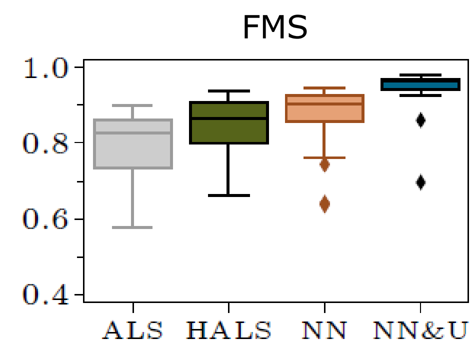
Fit a PARAFAC2 model with non-negativity and/or unimodality constraints



Non-negative A & C

Non-negative and unimodal B_k

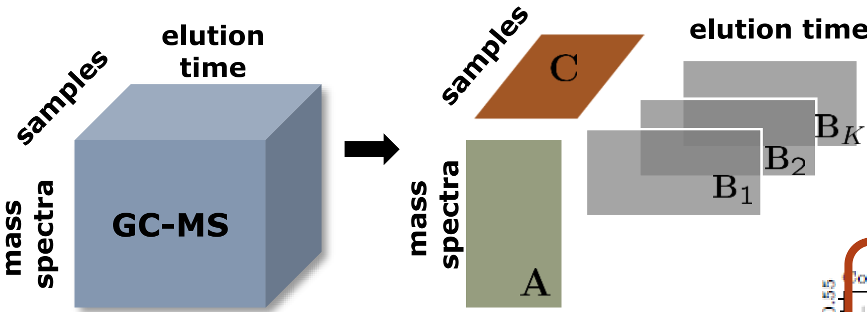




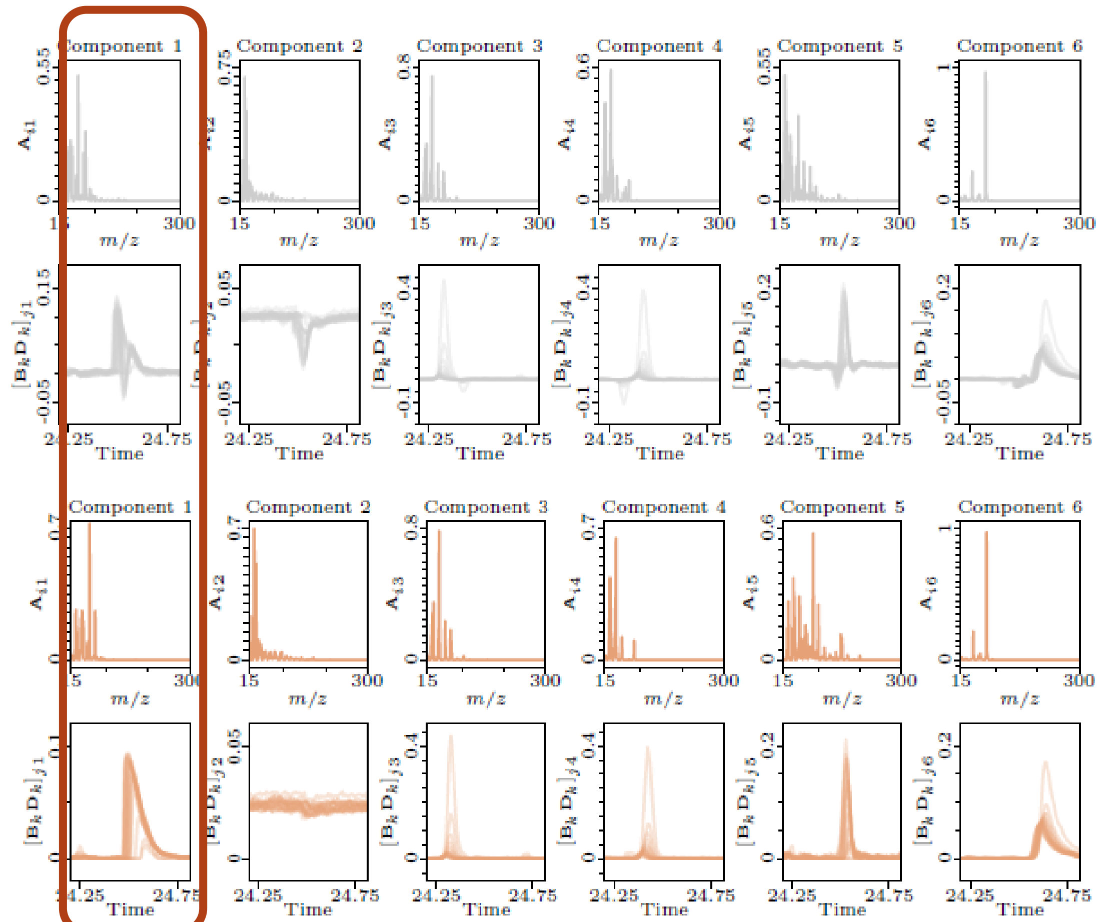
AO-ADMM with non-negativity and unimodality constraints is more accurate!

Chemometrics: PARAFAC2 with non-negativity constraints in all modes reveals more accurate patterns

PARAFAC2 is a useful tool for analyzing gas chromatography - mass spectrometry (GC-MS) measurements of mixtures **[Bro et al., 1999; Amigo et al., 2008].**



Wine samples arranged as a 286 (***mass spectra***) by 95 (***elution time***) by 19 (***samples***) tensor, and analyzed using a 6-component PARAFAC2 model.

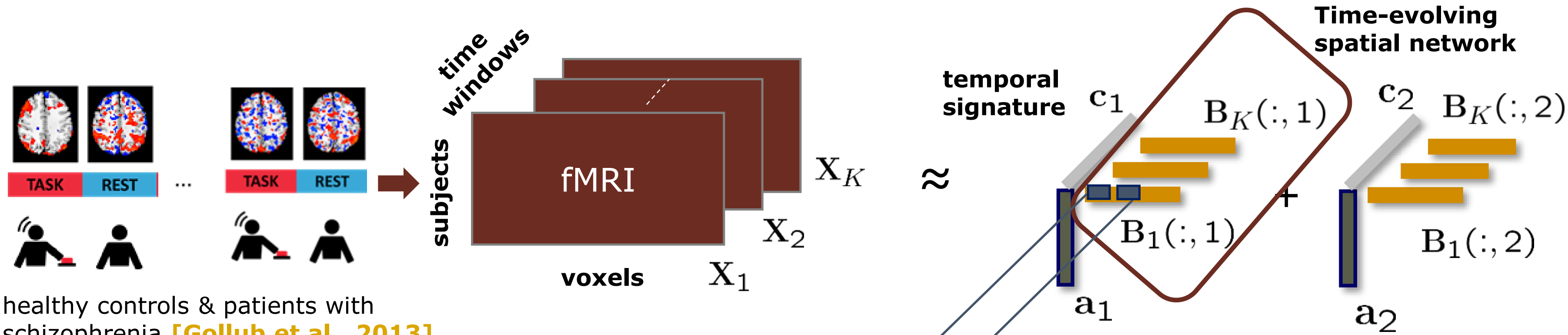


ALS with non-negativity constraints in **A** and **C**

AO-ADMM with non-negativity constraints in all modes

More interpretable components are captured using AO-ADMM with non-negativity constraints in all modes!

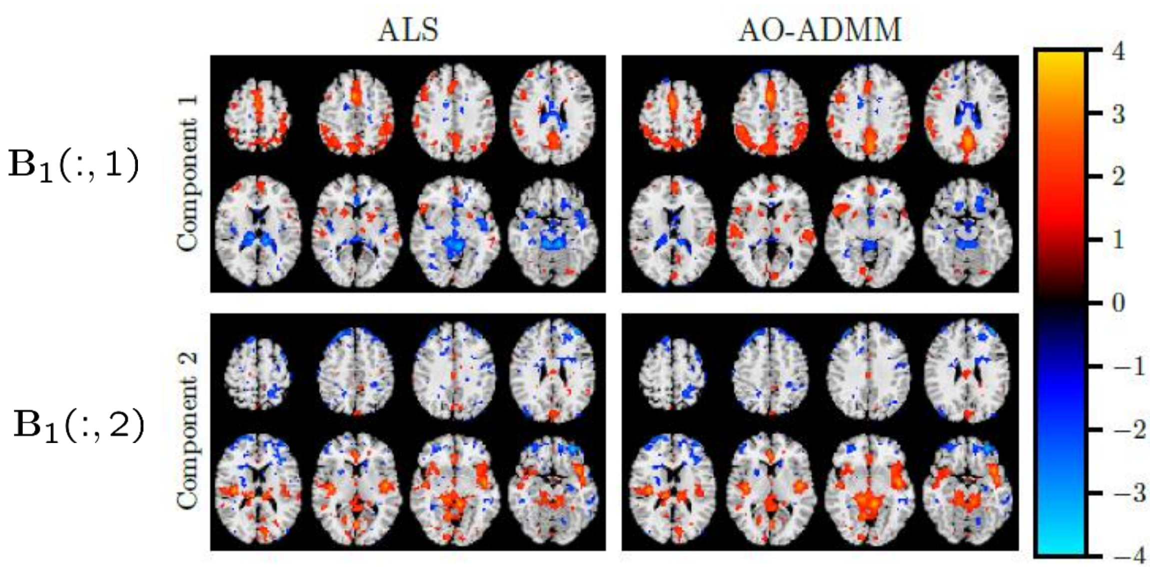
Neuroscience: PARAFAC2 with smoothness constraints reveals components with more defined regions of activation and less noise

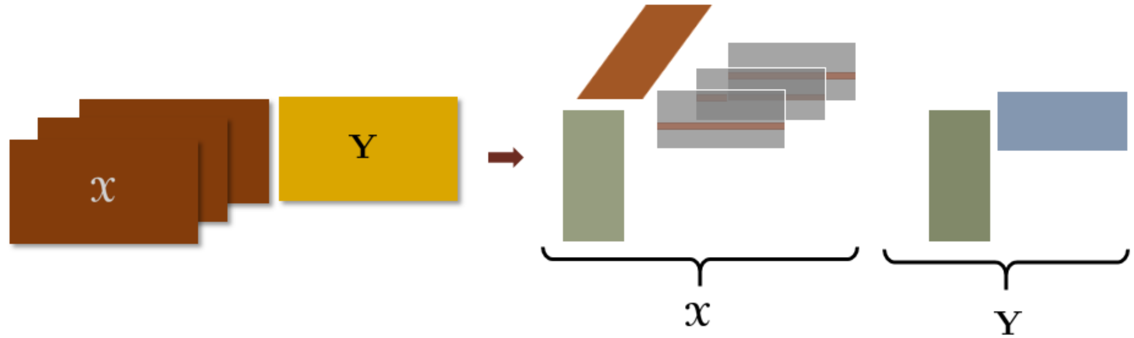


healthy controls & patients with schizophrenia **[Gollub et al., 2013]**

PARAFAC2 is fitted to the task fMRI data using a graph Laplacian-based regularizer penalizing pairwise differences between neighboring voxels.

$$g(\mathbf{x}) = \sum_{i,j} w_{ij}(x_i - x_j)^2 = \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}$$

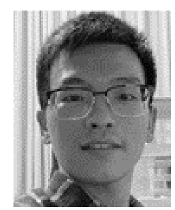




Part III. Joint Analysis of Dynamic and Static Data

joint work with

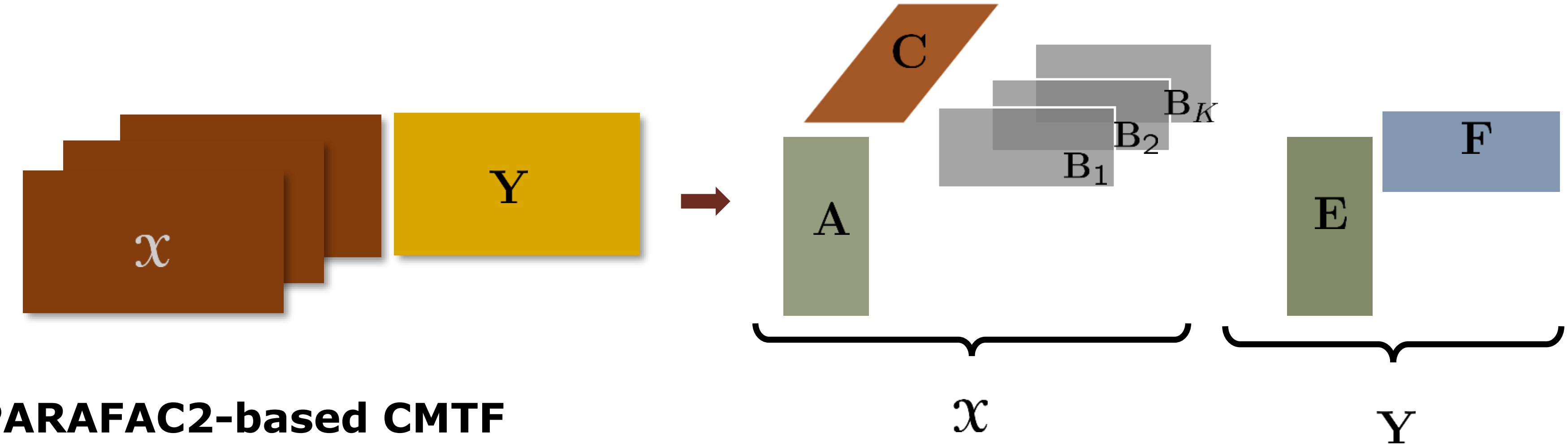
Carla Schenker
OsloMet &
SimulaMet



Xiulin Wang
Affiliated Zhongshan
Hospital of Dalian
University

C. Schenker, X. Wang, E. Acar. PARAFAC2-based Coupled Matrix and Tensor Factorizations, *arXiv:2210.13054v1*, 2022

We extend our flexible regularized coupled matrix/tensor factorizations with linear coupling framework by incorporating the PARAFAC2 model



PARAFAC2-based CMTF

$$\min_{\mathbf{A},\mathbf{B},\mathbf{E},\mathbf{F},\Delta_1,\{\mathbf{D}_k,\mathbf{B}_k,\mathbf{P}_k\}_{k\leq K}} \left\| \mathbf{Y} - \mathbf{E}\mathbf{F}^\top \right\|_F^2 + \sum_{k=1}^{K} \left[\left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\top \right\|_F^2 + g_D(\mathbf{D}_k) + g_B(\mathbf{B}_k) \right] + g_A(\mathbf{A}) + g_E(\mathbf{E}) + g_F(\mathbf{F})$$

$$\text{s.t.} \quad \mathbf{B}_k = \mathbf{P}_k\mathbf{B}, \ \mathbf{P}_k^\top\mathbf{P}_k = \mathbf{I} \quad k = 1, ..., K,$$
$$\mathbf{H}_1\mathbf{E} = \Delta_1,$$
$$\mathbf{H}_2\mathbf{A} = \Delta_1$$

Solved using AO-ADMM

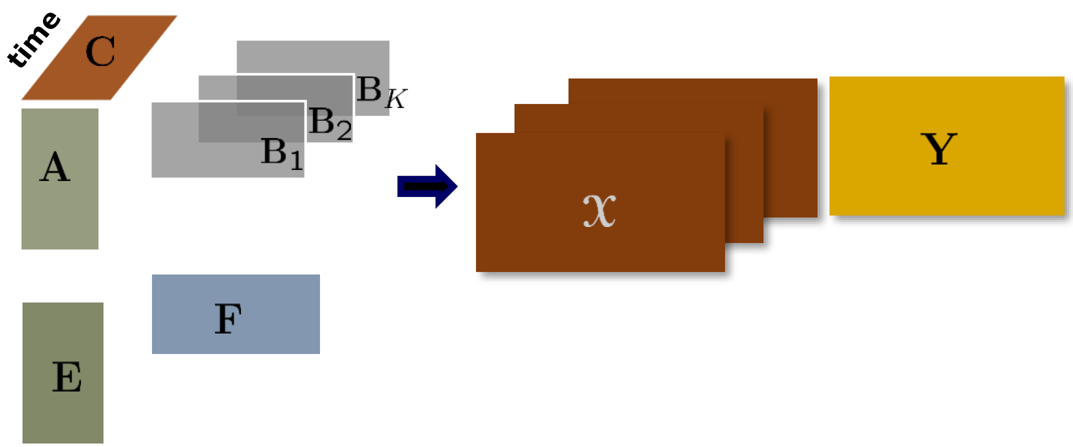
Fusion methods using a PARAFAC2 model:

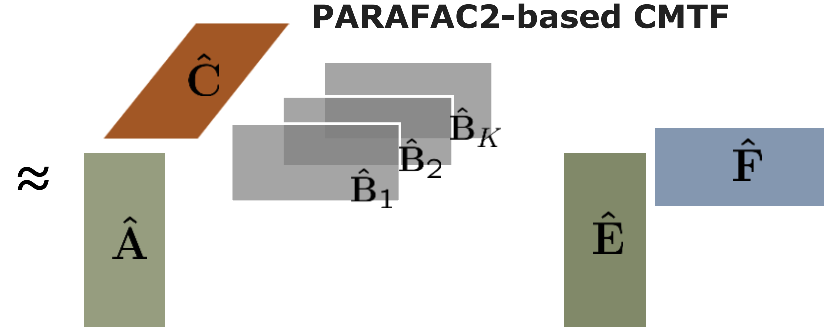
PARAFAC2 has also been incorporated in other data fusion methods:
(i) PARAFAC2 coupled with a matrix factorization for EHR-based phenotyping **[Afshar et al., 2020]**
(ii) PARAFAC2 coupled with CP for joint analysis of EEG and fMRI data **[Chatzichristos et al., 2018]**

Limited in terms of constraints and/or coupling structure

PARAFAC2-based CMTF model captures evolving patterns accurately while improving the clustering performance through fusion!

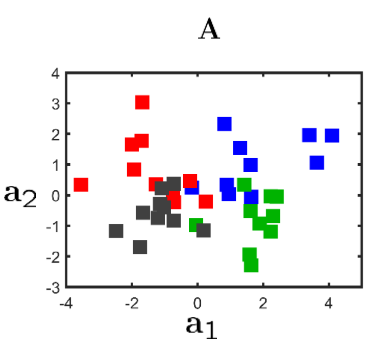
Generate factor matrices | **Construct coupled data sets** | **Jointly analyze using PARAFAC2-based CMTF**



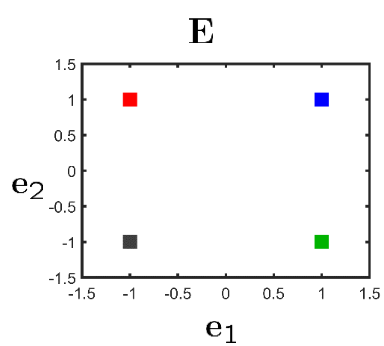


$$\min_{\mathbf{A},\mathbf{B},\mathbf{E},\mathbf{F},\Delta_1,\{\mathbf{D}_k,\mathbf{B}_k,\mathbf{P}_k\}_{k\le K}} \left\|\mathbf{Y}-\mathbf{E}\mathbf{F}^\top\right\|_F^2 + \sum_{k=1}^{K} \left\|\mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^\top\right\|_F^2$$

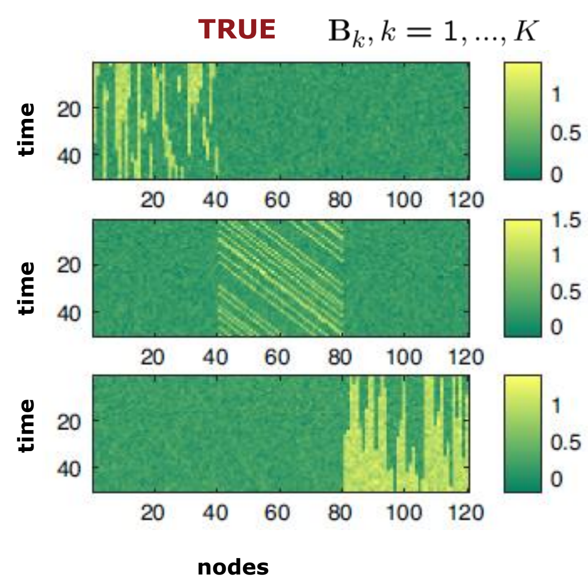
$$\text{s.t.} \quad \mathbf{B}_k = \mathbf{P}_k\mathbf{B},\ \mathbf{P}_k^\top\mathbf{P}_k = \mathbf{I} \quad k = 1, ..., K,$$
$$\mathbf{E} = \mathbf{A} = \Delta_1,$$
$$\mathbf{F}, \mathbf{C} \geq 0$$



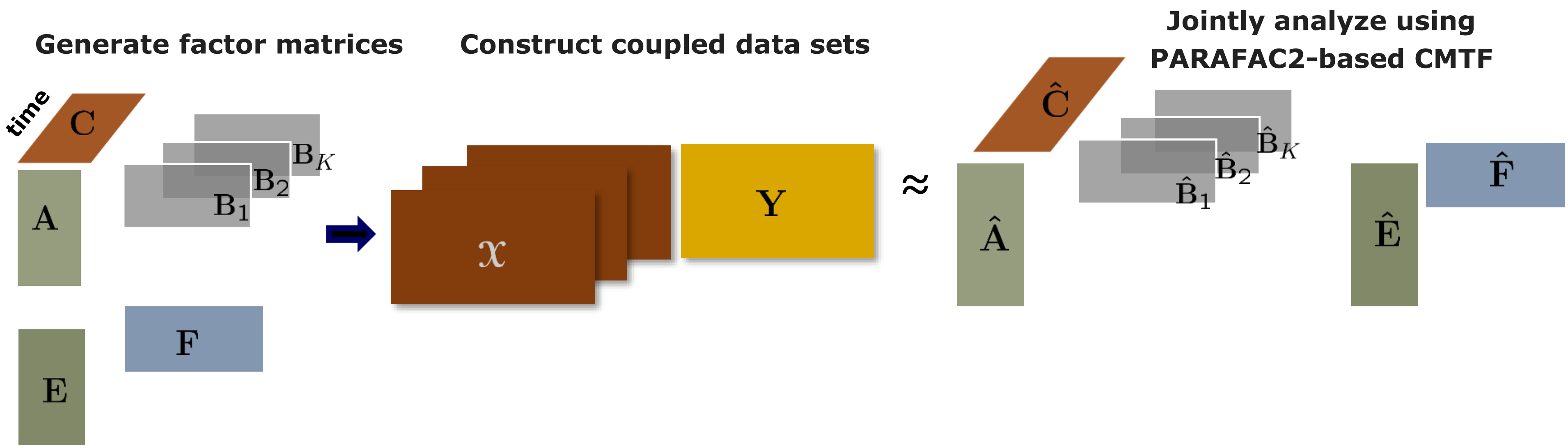
Dynamic data: noisy clusters



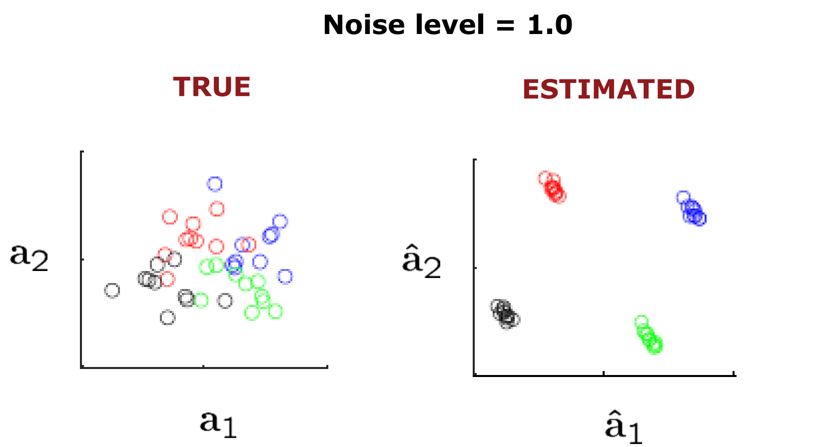
Static data: clean clusters



PARAFAC2-based CMTF model captures evolving patterns accurately while improving the clustering performance through fusion!



Coupling	Noise	Fit (%)		FMS				
		PAR2	Matrix	A	B	C	E	F
no	0	99.98	100	1 (1)	0.99	1	-	-
	0.5	99.98	100	1 (0.89)	0.99	1	-	-
	1	99.98	100	1 (0.71)	0.99	1	-	-
yes	0	99.75	100	1(1)	0.99	0.99	1	1
	0.5	84.51	99.98	0.90(0.98)	0.99	1	0.98	0.99
	1	56.67	99.95	0.72(0.96)	0.98	0.99	0.96	0.99

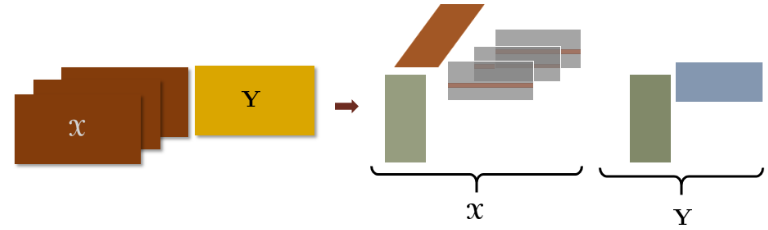


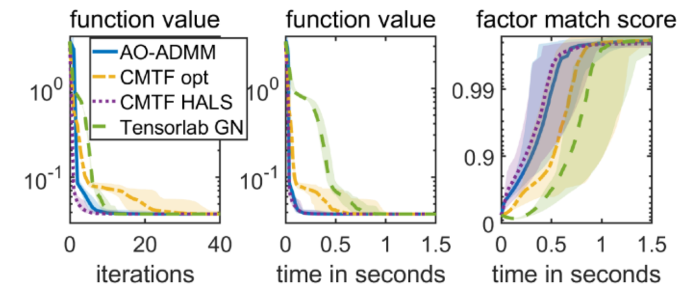
Summary

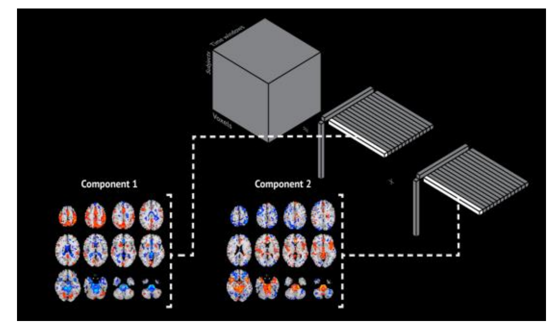
Goal: Extracting insights from complex data sets - often complementary data sets collected from multiple sources, some are time-evolving while some are static

Our approach: Unsupervised interpretable pattern discovery

- **Tensor factorizations** revealing unique patterns
- **Flexible data fusion framework based on coupled matrix and tensor factorizations** incorporating various constraints, linear coupling, loss functions and evolving patterns
- **Applications of interest:** Neuroscience and omics

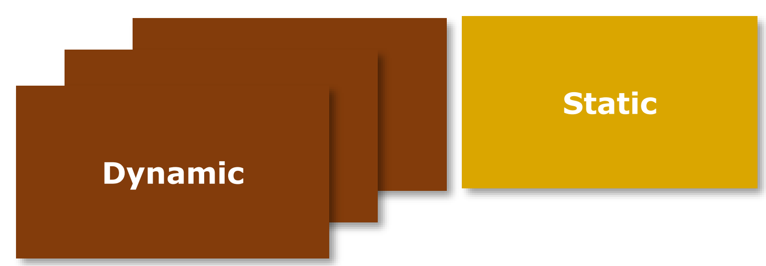






Ongoing work

- Analysis of metabolomics measurements collected during the challenge test - together with other omics measurements
- Time-aware tensor factorizations



Thank you!

Flexible data fusion framework

C. Schenker, J. E. Cohen, E. Acar. A Flexible Optimization Framework for Regularized Matrix-Tensor Factorizations with Linear Couplings*. IEEE Journal of Selected Topics in Signal Processing*, 15(3): 506-521, 2021

PARAFAC2-based Coupled Matrix and Tensor Factorizations

C. Schenker, X. Wang, E. Acar. PARAFAC2-based Coupled Matrix and Tensor Factorizations*, arXiv:2210.13054v1,* 2022

PARAFAC2 AO-ADMM algorithm

M. Roald, C. Schenker, V. Calhoun, T. Adali, R. Bro, J. E. Cohen, E. Acar. An AO-ADMM approach to constraining PARAFAC2 on all modes, *SIAM Journal on Mathematics of Data Science*, 4(3): 1191-1222, 2022

Evolving patterns

M. Roald, S. Bhinge, C. Jia, V. Calhoun, T. Adali, E. Acar. Tracing Network Evolution using the PARAFAC2 model. *ICASSP*, pp. 1100-1104, 2020

E. Acar, M. Roald, K. Hossain, V. Calhoun, T. Adali. Tracing Evolving Networks Using Tensor Factorizations vs. ICA-Based Approaches. *Frontiers in Neuroscience*, 16: 861402, 2022

Software

AO-ADMM for CMTF & PARAFAC2 **https://github.com/AOADMM-DataFusionFramework/**

AO-ADMM for PARAFAC2 **https://github.com/MarieRoald/PARAFAC2-AOADMM-SIMODS**

python

TrACEr

Time-Aware ConstrainEd Multimodal Data Fusion

https://tracer.simulamet.no

novo nordisk fonden

The Research Council of Norway

simulamet