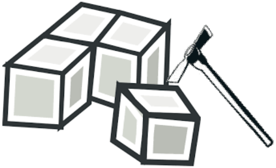
Environmental Low Rank Data Mining



Jérémy E. Cohen

UMONS, FNRS

16 June 2017

Background

PhD at GIPSA-lab/CNRS
Tensor Decompositions in data mining





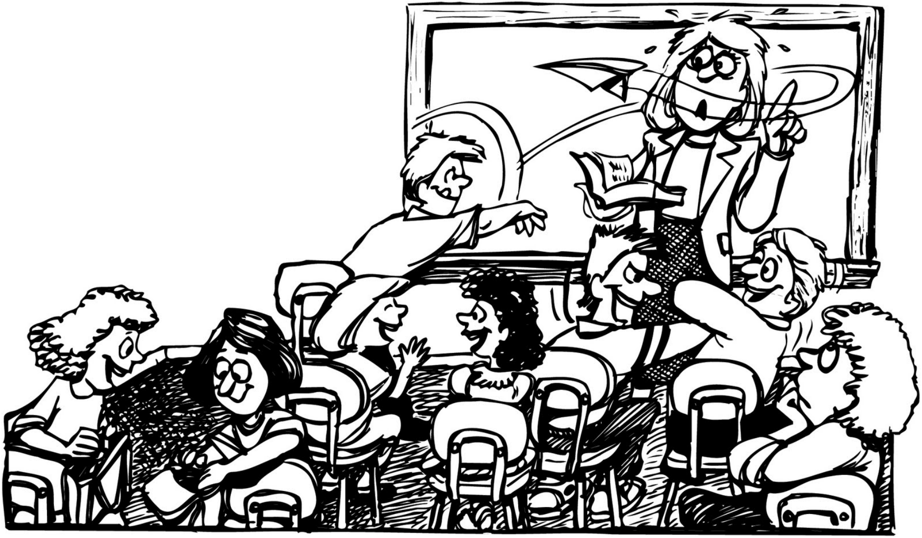
Post-doc at UMONS/FNRS
Matrix and Tensor low rank approximation



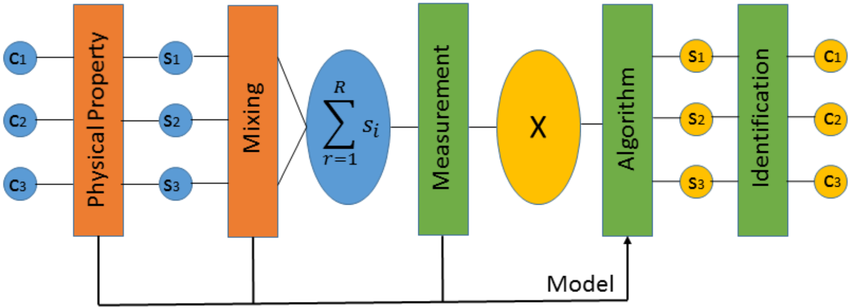

1. Introduction and Notations
   - Some environmental multiway data
   - Challenges in environmental data mining
   - The tensor decomposition approach

2. Facing the challenges
   - Multiway Data Fusion
   - Fast nonnegative tensor decomposition
   - Dictionary-based CPD

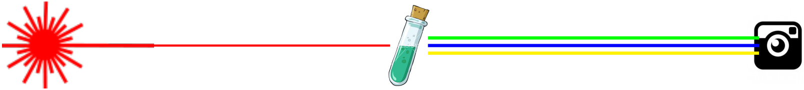
Source Separation

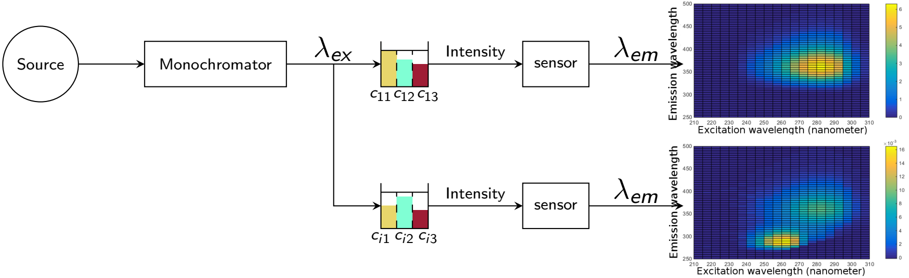


Source Separation



Fluorescence Spectroscopy

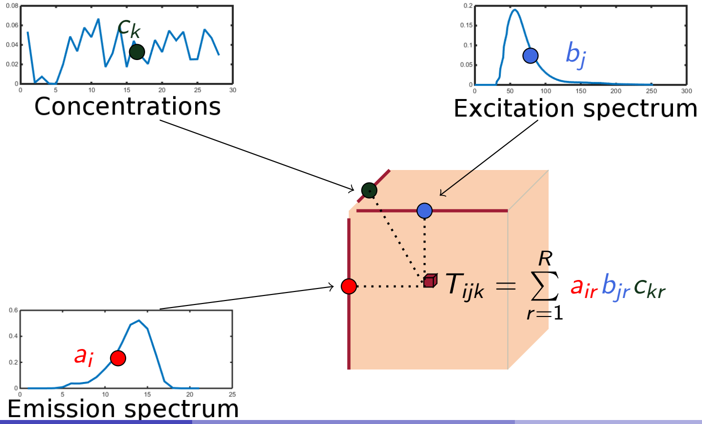




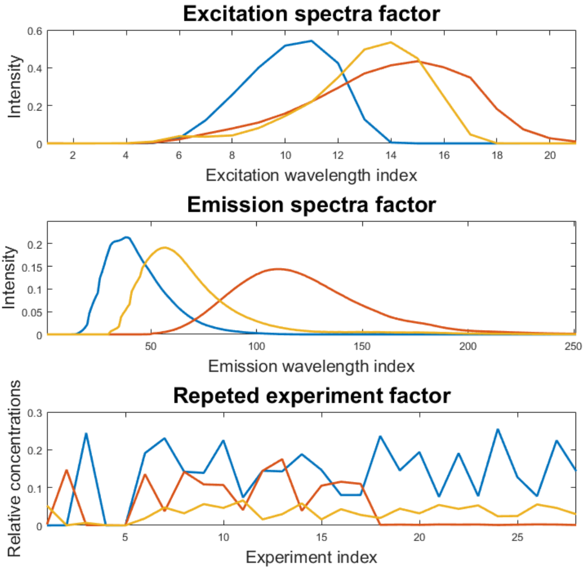
[Acar,2013]

Fluorescence is multilinear

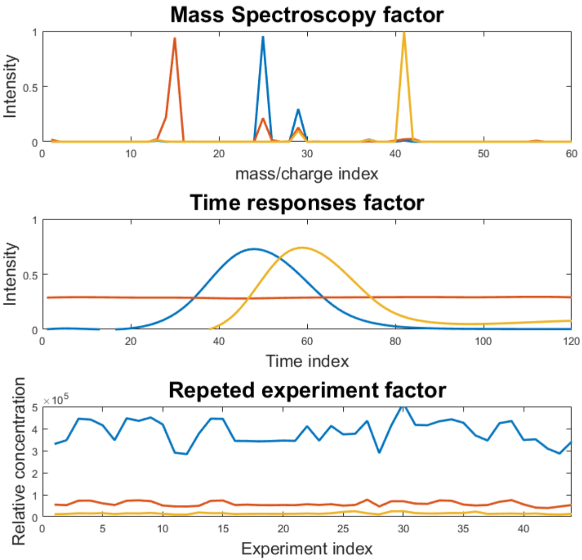
$$T(\lambda_{ex}, \lambda_{em}, k) = \sum_{r=1}^{R} a_r(\lambda_{ex}) b_r(\lambda_{em}) c_r(k)$$



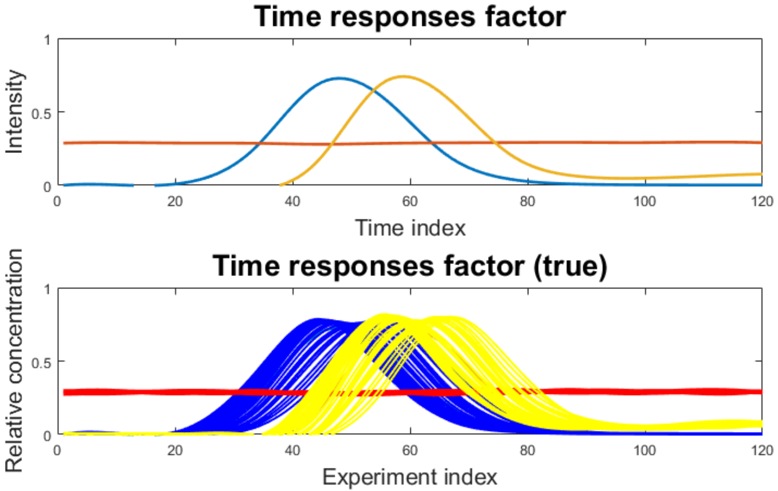
How should the component look like ?



What about Liquid Chromatography - Mass Spectroscopy ?

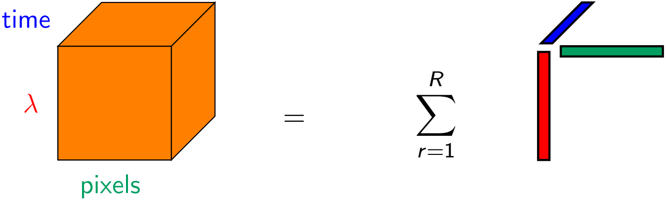


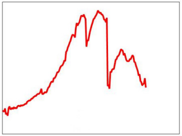
Retention shifts in LC-MS are challenging



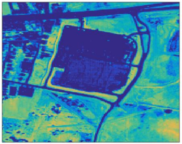
Time response also depends on experiment index ! !

Hyperspectral Data

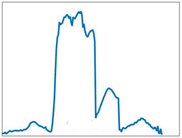




Spectre(λ)



Répartition spatiale(pixel)

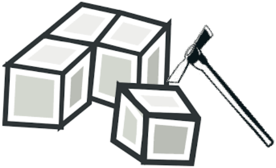


Signature temporelle(t)

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

Identification problem : Which materials generate estimated spectra ?

Challenges in environmental and biomedical data mining



Take into account *a priori* information into the factorization models.

- Data Fusion - Subject Variability
- Identification of the components
- Constrained Decompositions : non-negativity, sparsity, orthogonality

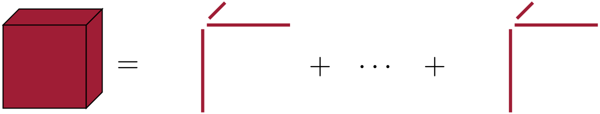
Main tool : Canonical Polyadic Decomposition

Canonical Polyadic Decomposition [Hitchcock,1927] aims at extracting all R components.

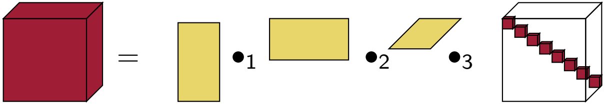
Tensor = first component + ⋯ + *Rth* component

- Unmixing in theory does not require additional knowledge.
- For matrices, rarely unique → SVD (orthogonality), NMF (non-negativity).

CPD



$$\mathcal{T} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \cdots + \mathbf{a}_R \otimes \mathbf{b}_R \otimes \mathbf{c}_R$$



$$\mathcal{T} = (\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}) \; \mathcal{I}_R$$

$\mathcal{T}$ has sizes $K \times L \times M$ $\quad$ $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_R]$ has sizes $K \times R$

$\otimes$ is the tensor product $\quad$ $\bullet_i$ is the contraction on mode i

R is the rank of $\mathcal{T}$, *i.e.* smallest number of rank-one tensors spanning $\mathcal{T}$.

Tensor decomposition as an approximation problem

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B},\mathbf{C}}. \quad & \|\mathcal{T} - (\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C})\mathcal{I}_R\| \\ \text{sub. to} \quad & \mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathcal{C}_{A,B,C} \end{aligned}$$

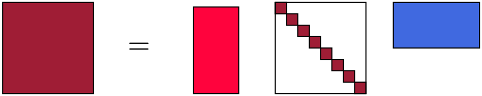
- Non-convex in the general case but convex with respect to each block $\mathbf{A}, \mathbf{B}, \mathbf{C}$.
- Example : Non-negative Matrix Factorization with Frobenius norm

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B}}. \quad & \|\mathbf{M} - \mathbf{A}\mathbf{B}^T\|_F^2 \\ \text{sub. to} \quad & \mathbf{A} \geq 0 \quad \mathbf{B} \geq 0 \end{aligned}$$

1. Introduction and Notations
   - Some environmental multiway data
   - Challenges in environmental data mining
   - The tensor decomposition approach

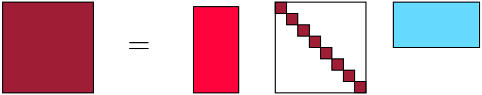
2. Facing the challenges
   - Multiway Data Fusion
   - Fast nonnegative tensor decomposition
   - Dictionary-based CPD

Subject Variability

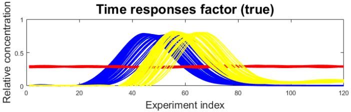


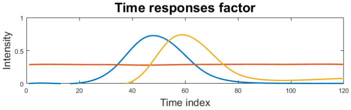
$$\mathbf{M}_1 = \mathbf{A} \quad \boldsymbol{\Sigma}_1 \quad \mathbf{B}_1^T$$

$\vdots$



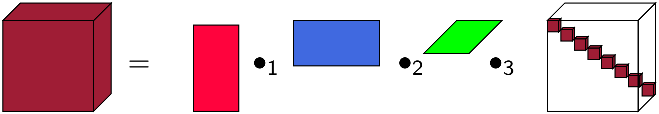
$$\mathbf{M}_i = \mathbf{A} \quad \boldsymbol{\Sigma}_i \quad \mathbf{B}_i^T$$





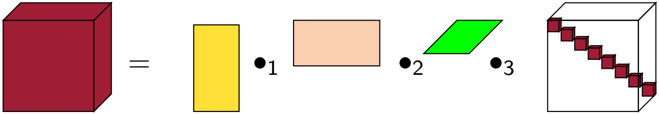
Example : LC-MS data.

Data Fusion with tensors



$$= \quad \bullet_1 \quad \bullet_2 \quad \bullet_3$$

$$\mathcal{T}_1 = (\mathbf{A}_1 \otimes \mathbf{B}_1 \otimes \mathbf{C}_1) \quad \mathcal{I}_R$$



$$= \quad \bullet_1 \quad \bullet_2 \quad \bullet_3$$

$$\mathcal{T}_2 = (\mathbf{A}_2 \otimes \mathbf{B}_2 \otimes \mathbf{C}_2) \quad \mathcal{I}_R$$

Example : Fluorescence and NMR data. Often $\mathbf{C}_1 := \mathbf{C}_2$. But :

- the sampling rates can be different ?
- the relation may not be trivial ? Can it be learned ?
- how does coupling affect the cost function ?

General Framework using a Bayesian approach [Cabral Farias, Cohen,2015]

- Parameters $\boldsymbol{\theta}_i = \begin{bmatrix} \text{vec}(\mathbf{A}_i) \\ \text{vec}(\mathbf{B}_i) \\ \text{vec}(\mathbf{C}_i) \end{bmatrix}$ are random

- Known prior distribution $p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)$ and likelihoods $p(\mathcal{Y}_i | \boldsymbol{\theta}_i)$

MAP estimation under conditionnal independance

$$\underset{\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_N}{\arg\max}\, p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N | \mathcal{Y}_1, \dots, \mathcal{Y}_N) = \underset{\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_N}{\arg\min}\, \Upsilon(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)$$

$$\begin{aligned} \Upsilon(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N) &= -\sum_{i=1}^{N} \log p(\mathcal{Y}_\mathbf{i} | \boldsymbol{\theta}_i) - \log p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N) \\ &= \text{ data fitting terms } + \text{ coupling} \end{aligned}$$

Some well-known coupling models

Coupled Tensor Factorization [Harshman, 1984]

$$\forall\ i \in [1, N], \quad \begin{cases} \mathcal{T}_i & = (\mathbf{A}_i \otimes \mathbf{B}_i \otimes \mathbf{C}_i)\mathcal{I}_R + \mathcal{E}_i \\ \mathbf{C}_i & = \mathbf{C}^* \end{cases}$$

$$\Upsilon(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, \mathbf{C}^*) = -\sum_{i=1}^{N} \frac{1}{\sigma_1^2} \|\mathcal{T}_i - (\mathbf{A}_i \otimes \mathbf{B}_i \otimes \mathbf{C}^*)\mathcal{I}_R\|_F^2$$

PARAFAC2 [Harshman,1972][Bro,1999]

$$\forall\ i \in [1, N], \quad \begin{cases} \mathbf{M}_i & = \mathbf{A}_i \boldsymbol{\Sigma}_i \mathbf{B}_i^T + \mathbf{E}_i \\ \mathbf{A}_i & = \mathbf{A}^* \\ \mathbf{B}_i & = \mathbf{P}_i \mathbf{B}^* \\ \mathbf{P}_i^T \mathbf{P}_i & = \mathbf{I} \end{cases}$$

Examples of flexible coupling models

Noisy exact coupling on $\mathbf{C}_i$

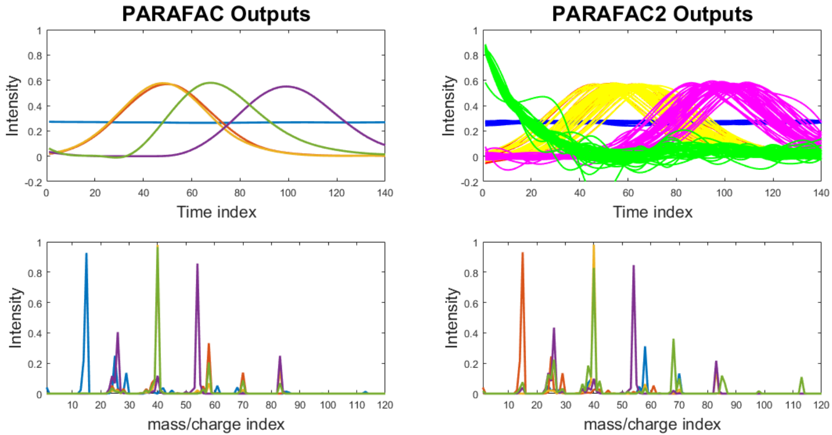
$$\forall\ i \in [1, N], \quad \begin{cases} \mathcal{T}_i & = (\mathbf{A}_i \otimes \mathbf{B}_i \otimes \mathbf{C}_i)\mathcal{I}_R + \mathcal{E}_i \\ \mathbf{C}_i & = \mathbf{C}^* + \mathbf{\Gamma}_i \\ \mathbf{\Gamma}_i & \sim \mathcal{AN}\left(\mathbf{0}, \frac{1}{\sigma_{c,i}^2}\mathbf{I} \otimes \mathbf{I}\right) \end{cases}$$

$$\Upsilon(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N, \mathbf{C}^*) = -\sum_{i=1}^{N} \frac{1}{\sigma_1^2} \|\mathcal{T}_i - (\mathbf{A}_i \otimes \mathbf{B}_i \otimes \mathbf{C}_i)\mathcal{I}_R\|_F^2 - \sum_{i=1}^{N} \frac{1}{\sigma_{ci}^2} \|\mathbf{C}_i - \mathbf{C}^*\|_F^2$$

Linear coupling on $\mathbf{C}_i$

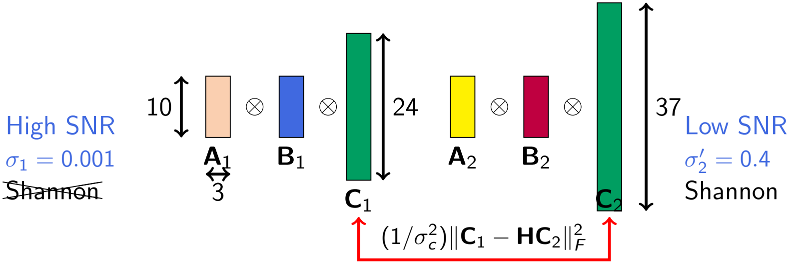
$$\forall\ i \in [1, N], \quad \begin{cases} \mathcal{T}_i & = (\mathbf{A}_i \otimes \mathbf{B}_i \otimes \mathbf{C}_i)\mathcal{I}_R + \mathcal{E}_i \\ \mathbf{H}_i\mathbf{C}_i & = \mathbf{H}_j\mathbf{C}_j + \mathbf{\Gamma}_{ij} \\ \mathbf{\Gamma}_{ij} & \sim \mathcal{AN}\left(\mathbf{0}, \frac{1}{\sigma_{ij}^2}\mathbf{I} \otimes \mathbf{I}\right) \end{cases}$$

PARAFAC2 vs PARAFAC on LC-MS data



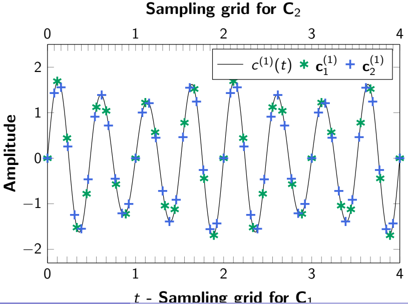
Many other solutions can be thought of to tackle subject variability !

Simulation : Resampling Bandlimited Signals



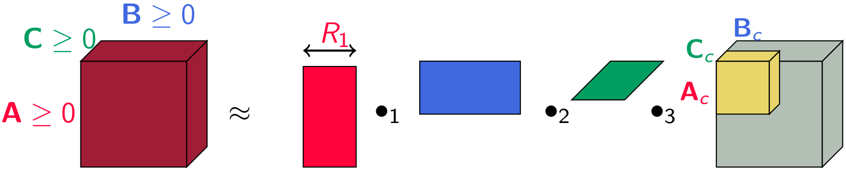
Total MSE on the continuous functions (numerical integration)

	$\mathbf{C}_1$ ~~Shannon~~	$\mathbf{C}_2$ noisy
Uncoupled	33.4968	2.6581
Coupled	33.4968	1.0375

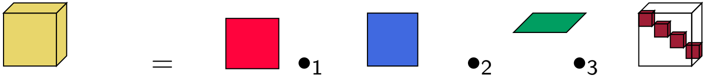


I S B N 2-205-00921-4

Unconstrained compression ...



$$\mathcal{T} \approx (\mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W})\, \mathcal{G}$$



$$\mathcal{G} = (\mathbf{A_c} \otimes \mathbf{B_c} \otimes \mathbf{C_c})\, \mathcal{I}_R$$

$$\mathcal{T} \approx (\mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W})\, \mathcal{G} = (\mathbf{UA_c} \otimes \mathbf{VB_c} \otimes \mathbf{WC_c})\, \mathcal{I}_R$$

... but constrained decomposition !

Compressed domain NN CP :

$$\begin{aligned} \min_{\mathbf{A}_c,\mathbf{B}_c,\mathbf{C}_c} \quad & \|\mathcal{G} - (\mathbf{A}_c \otimes \mathbf{B}_c \otimes \mathbf{C}_c)\mathcal{I}\| \\ \text{sub. to} \quad & \widehat{\mathbf{U}}\mathbf{A}_c, \widehat{\mathbf{V}}\mathbf{B}_c, \widehat{\mathbf{W}}\mathbf{C}_c \geqslant 0 \end{aligned}$$

Issue	Solution
Easy unconstrained/difficult constrained	Unconstrained solution→ projection
Difficult exact projection $\widehat{\mathbf{U}}\mathbf{A}_c$	Approximate projection

Approximate projection and PROCO-ALS

Approximate projection $\mathbf{\Pi}$:

Given Least Squares update $\widehat{\mathbf{A}}_c$

1. Decompression : $\widehat{\mathbf{A}} := \widehat{\mathbf{U}}\widehat{\mathbf{A}}_c$
2. Projection : $\widehat{\mathbf{A}} := \left[\widehat{\mathbf{A}}\right]^+$
3. Compression : $\widehat{\mathbf{A}}_c := \widehat{\mathbf{U}}^\intercal\widehat{\mathbf{A}}$

$$\mathbf{\Pi}\left[\widehat{\mathbf{A}}\right] = \mathbf{U}^\intercal\left[\mathbf{U}\mathbf{A}_c\right]^+$$

Projected and compressed framework (PROCO) [Cohen,2014]

Other possible algorithms and related problems

- PROCO-ALS [Cohen,2014], Compressed-AOADMM [Cohen,2016]

$$\begin{aligned}&\text{minimize } \|\widehat{\mathcal{G}} - (\mathbf{A}_c \otimes \mathbf{B}_c \otimes \mathbf{C}_c)\mathcal{I}_R\|_F^2\\ &\text{w.r.t. } \mathbf{A}_c, \mathbf{B}_c, \mathbf{C}_c\\ &\text{s.t. } \widehat{\mathbf{U}}\mathbf{A}_c \succeq 0\end{aligned}$$

- Tensorlab 3.0 [Vervliet,2016]

$$\begin{aligned}&\text{minimize } \|\left(\widehat{\mathbf{U}} \otimes \widehat{\mathbf{V}} \otimes \widehat{\mathbf{W}}\right)\widehat{\mathcal{G}} - (\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C})\mathcal{I}_R\|_F^2\\ &\text{w.r.t. } \mathbf{A}, \mathbf{B}, \mathbf{C}\\ &\text{s.t. } \mathbf{A} \succeq 0\end{aligned}$$

- AOADMM [Huang,2015], FastNNLS [Bro,1997], ANLS

$$\begin{aligned}&\text{minimize } \|\mathcal{T} - (\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C})\mathcal{I}_R\|_F^2\\ &\text{w.r.t. } \mathbf{A}, \mathbf{B}, \mathbf{C}\\ &\text{s.t. } \mathbf{A} \succeq 0\end{aligned}$$

Application in Fluorescence Spectroscopy

Fluorescence spectroscopy data : excitation spectra
emission spectra
mixtures

multimodal chemometrics data set from Acar *et al*[1]

Description

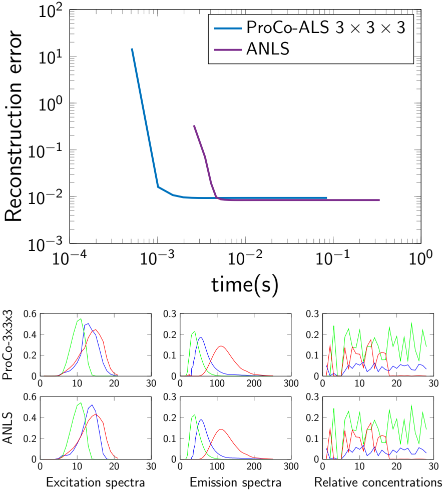
5 compounds : Valine-Tyrosine-Valine (Val), Tryptophan- Glycine (Gly), Phenylalanine (Phe), Maltoheptaose (Mal) and Propanol (Pro)

Nb. of excitation wave lengths	21 (**A**)
Nb. of emission wave lengths	251 (**B**)
Nb. of Mixtures	28 (**C**)
Missing values	30% (replaced by zeros)

1. E. Acar, A.J. Lawaetz, M.A. Rasmussen, and R. Bro. Structure-revealing data fusion model with applications in metabolomics. In Conf. Proc. IEEE Eng. Med. Biol. Soc., pages 6023– 6026. IEEE, 2013

Application to Fluorescence Spectroscopy

ANLS (nonnegative) and ProCo-ALS



Application to Spectral Unmixing

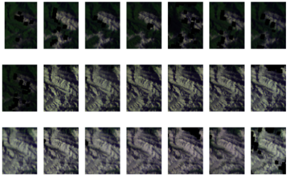


FIGURE – Multiple Snapshots of the Alps along time [Meteo France]

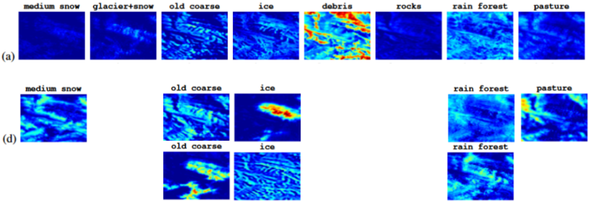
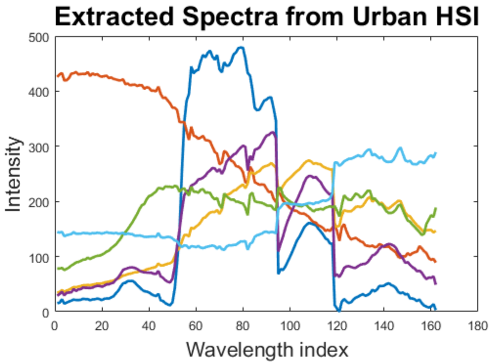
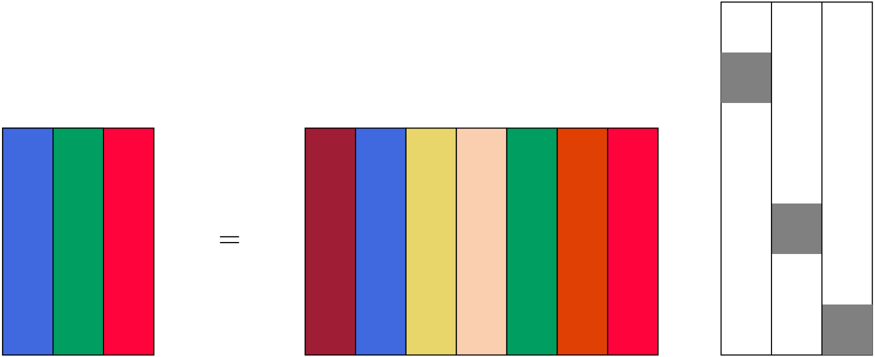


FIGURE – Unmixed Abundances (subset) for a)FLSU d)Proco-ALS [Veganzones,Cohen 2016]

Identification may be an issue



Let's choose **A** from a dictionary



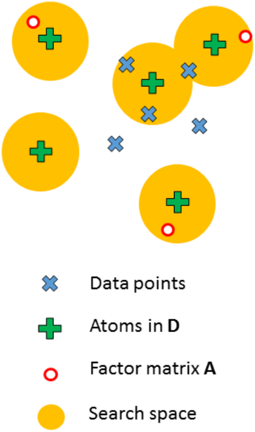
A = **D** **S**

$$\mathbf{X} = \mathbf{D}\mathbf{S}\mathbf{B}^T \text{ or } \mathcal{T} = (\mathbf{D}\mathbf{S} \otimes \mathbf{B} \otimes \mathbf{C})\mathcal{I}_R$$

where $\|\mathbf{S}\|_{col,0} = 1$ (Variation of Sparse Coding).

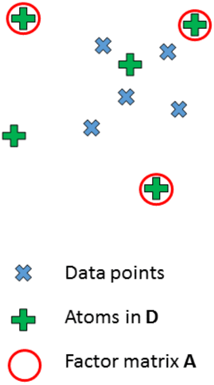
Flexibility and Separability

Flexibility



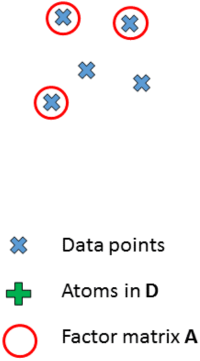
$$\mathbf{A} \approx \mathbf{DS}$$

Standard case



$$\mathbf{A} = \mathbf{DS}$$

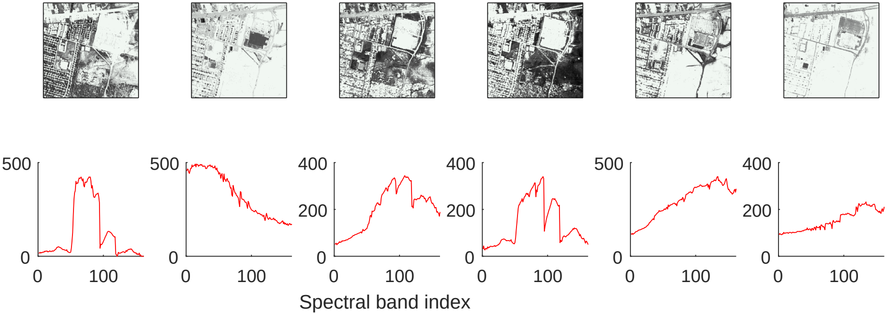
Separability



$$\mathbf{A} = \mathbf{XS}$$

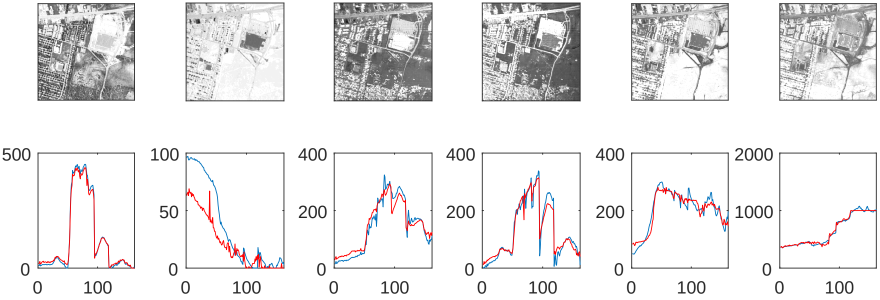
Application to Spectral Unmixing with Pure pixels

Spectra extracted exactly from the data (in red)



Application to Spectral Unmixing with Pure pixels

Spectra (in blue) close the data (in red)



Toolbox available on my personal webpage `jeremy-e-cohen.jimdo.com`
[Cohen Gillis, 2017]

Thank you for your attention !