Learning with Low Rank Approximations

Jérémy E. Cohen

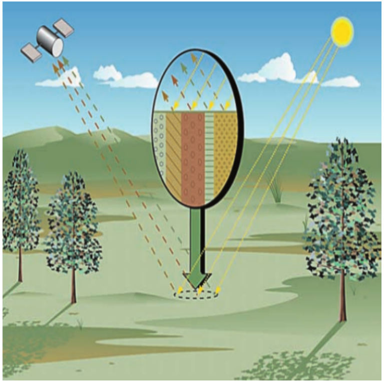
Team Panama, IRISA, CNRS

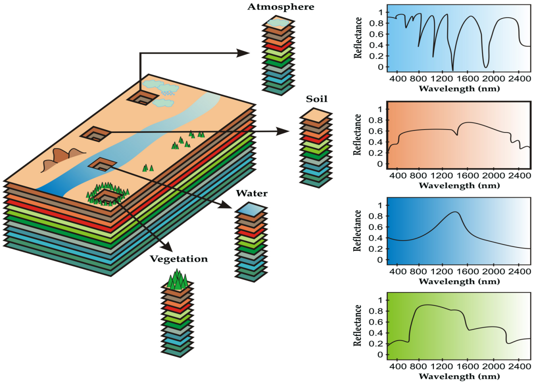
CREATIS Seminar, 2021 September 22

Roadmap

1. An introduction to nonnegative low-rank approximations
2. Nonnegative Tucker decomposition of music for automatic segmentation
3. Algorithms for constrained linearly coupled factorizations

Spectral Unmixing



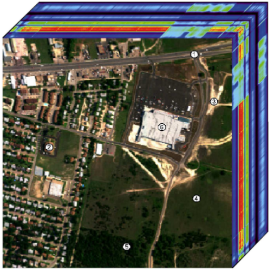


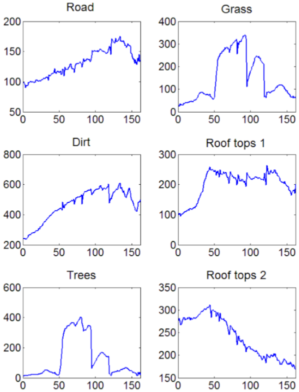
- Each pixel is a **mixture** of various materials.
- Each material has a unique spectral response.

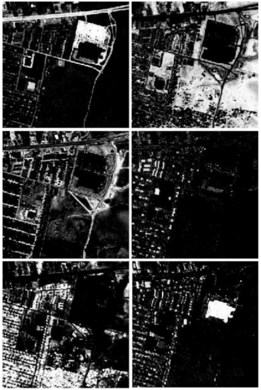
Credits for illustrations: Veganzones(left) and Bioucas(right)

Nonnegative Matrix Factorization for spectral unmixing

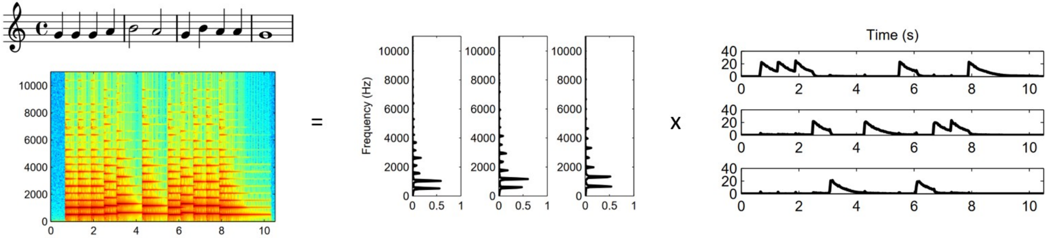
$$\underbrace{X(:,j)}_{\substack{\text{spectral signature} \\ \text{of } j\text{th pixel}}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{\substack{\text{spectral signature} \\ \text{of } k\text{th endmember}}} \underbrace{H(k,j)}_{\substack{\text{abundance of } k\text{th endmember} \\ \text{in } j\text{th pixel}}} .$$







Another example: Automatic Transcription



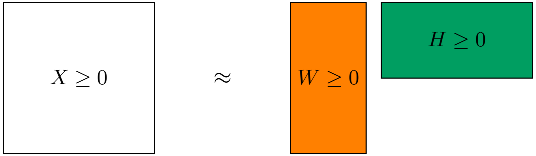
(Nonnegative) Low-rank approximation techniques are pattern mining tools!

Modeling Spectral unmixing

One material k has separable intensity:

$$X_k(\lambda, x, y) = w_k(\lambda) h_k(x, y)$$

where w_k is a spectrum characteristic to material k, and h_k is its abundance map.



Therefore, for an image M with r materials,

$$X(\lambda, x, y) = \sum_{k=1}^{r} X_k(\lambda, x, y) = \sum_{k=1}^{r} w_k(\lambda) h_k(x, y)$$

This means the measurement matrix $X_{i,j} = X(\lambda_i, \text{pixel}_j)$ is low rank!

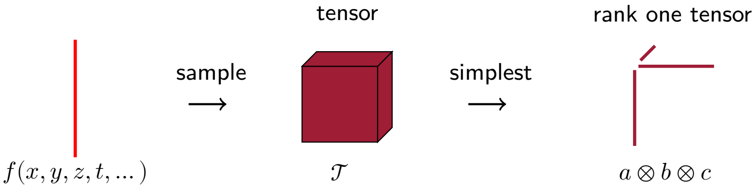
Nonnegative matrix factorization

$$\text{Find } W, H \text{ in } \underset{W \geq 0, H \geq 0}{\text{argmin}} \|X - \sum_{k=1}^{r} w_k h_k^t\|_F^2$$

where $X_{i,j} = X(\lambda_i, [x \otimes_K y]_j,)$ is the vectorized hyperspectral image.

Matrix and Tensor rank

What are tensors?



Definition: (nonnegative) rank-one matrix / tensor

A tensor $\mathcal{T}_{ijk} \in \mathbb{R}^{I \times J \times K}$ is said to be a [decomposable] [separable] [simple] [rank-one] tensor iff there exist $a \in \mathbb{R}^{I}_{(+)}, b \in \mathbb{R}^{J}_{(+)}, c \in \mathbb{R}^{K}_{(+)}$ so that

$$\mathcal{T}_{ijk} = a_i b_j c_k$$

or equivalently,

$$\mathcal{T} = a \otimes b \otimes c$$

where $\otimes$ is a multiway equivalent of the exterior product $a \otimes b = ab^t$.

ALL tensor decomposition models are based on separability

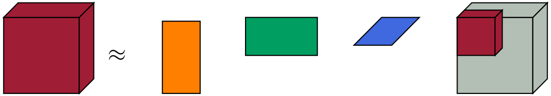
Canonycal Polyadic Decomposition (CPD):

$$\mathcal{T} = \sum_{k=1}^{r} a_k \otimes b_k \otimes c_k$$

$$\mathcal{T} = a_1 \otimes b_1 \otimes c_1 + \cdots + a_r \otimes b_r \otimes c_r$$

Tucker Decomposition:

$$\mathcal{T} = \sum_{k_1,k_2,k_3=1}^{r_1,r_2,r_3} g_{k_1k_2k_3} a_{k_1} \otimes b_{k_2} \otimes c_{k_3}$$



$$\mathcal{T} \approx (A \otimes B \otimes C)\ \mathcal{G}$$

Definition: tensor rank

$$\text{rank}(\mathcal{T}) = \min\{r \mid \mathcal{T} = \sum_{k=1}^{r} a_k \otimes b_k \otimes c_k\}$$

Tensor CP rank coincides with matrix “usual” rank!

Making use of low-rank representations

Let $A = [a_1, a_2, \ldots, a_r]$, B and C similarly built.

Uniqueness / Pattern mining

- CPD: Under mild conditions the CPD is essentially unique (i.e.) the rank-one terms are unique.
- NMF: Quite complicated, but in general requires additional regularizations.

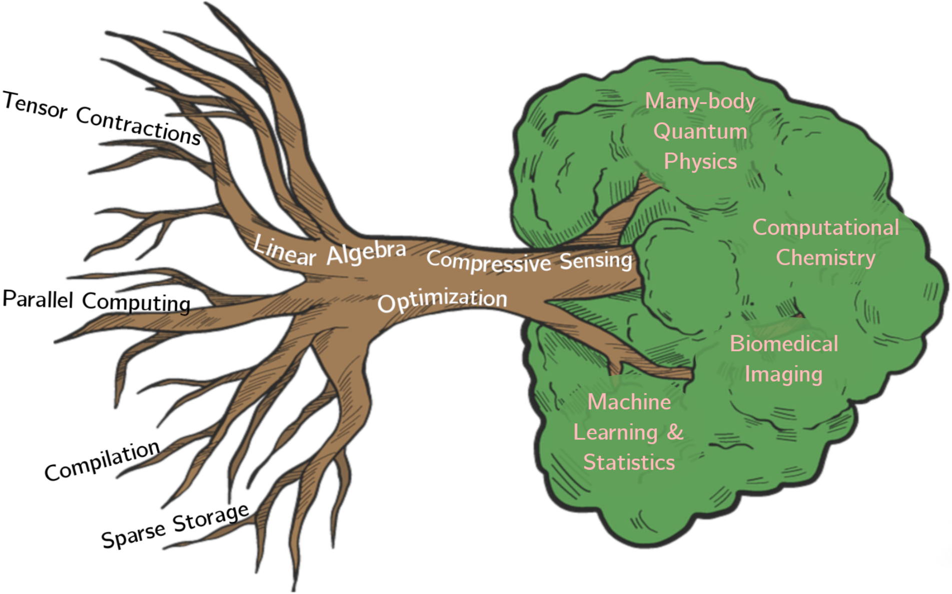
This means we can interpret the rank-one terms a_k, b_k, c_k
$\rightarrow$ Source Separation!

Compression

- The CPD involves $r(I + J + K - 2)$ free parameters, while $\mathcal{T}$ contains IJK entries.
- The NMF involves $r(I + J - 1)$ free parameters, while X has IJ entries.

If the rank is small, this means huge compression/dimentionality reduction!

The landscape of research on low-rank approximations



Computing LRA: an inverse problem

Low-rank r approximate NCPD

Given a tensor $\mathcal{T}$, find tensor $\mathcal{G}^* = \sum_{k=1}^{r} a_k \otimes b_k \otimes c_k$ that minimizes

$$\eta(A,B,C) = \|\mathcal{T} - \sum_{k=1}^{r} a_k \otimes b_k \otimes c_k\|_F^2 \text{ so that } a_k \geq 0, b_k \geq 0, c_k \geq 0$$

- Nonconvex, but convex w.r.t. each mode.
- The minimum is always attained (coercivity)!

My favorite class of algorithms to solve aNCPD: block-coordinate descent!

$$\underset{a_1,\dots,a_r}{\text{argmin}} \|\mathcal{T} - \sum_{k=1}^{r} a_k \otimes b_k \otimes c_k\|_F^2 \text{ so that } a_k \geq 0$$

is convex. It is a (Nonnegative) Least Squares problem, good algorithms are known.

My ongoing research projects

LoRAiA (ANR JCJC)

Semi-supervision and Tensors:

- Dictionaries/sparse coding
- Optimal Transport

with efficient implementations/algorithms!

Automatic Transcription

With semi-supervision and NMF.

Tensoptly (Inria)

Tensorly optimization layer:

- Constrained models
- Faster algorithms
- Customization

Music Segmentation

PhD of Axel Marmoret.

Sparse/Fast Optimization

Long-term collaboration with N. Gillis (UMONS).

Multimodality

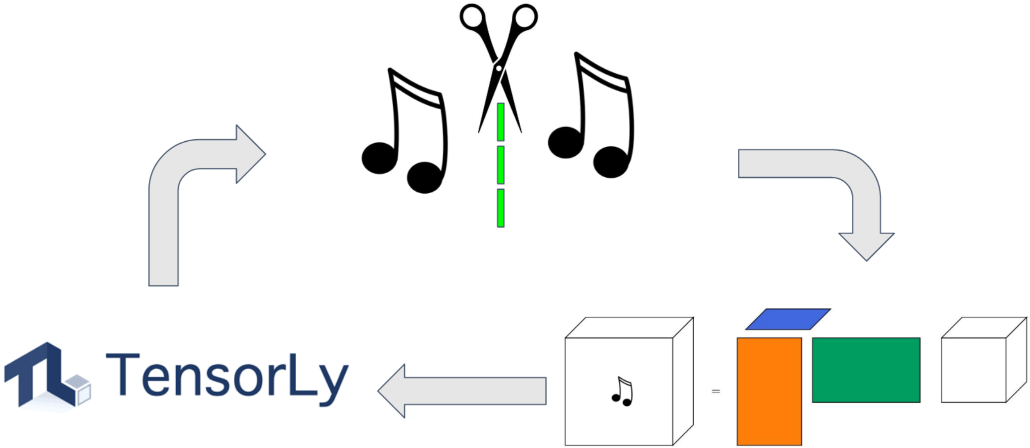
Long-term collaboration with E. Acar (SimulaMet).

A common trait: nonnegativity!

Roadmap

1. An introduction to tensor methods
2. Nonnegative Tucker decomposition of music for automatic segmentation
3. Algorithms for constrained linearly coupled factorizations

The NTD project in a glance

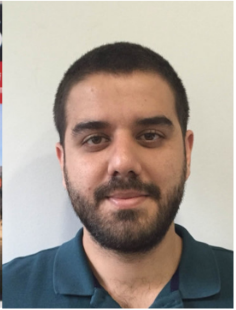


A team effort

Axel Marmoret
Doctorant UR1

Nancy Bertin
CR CNRS

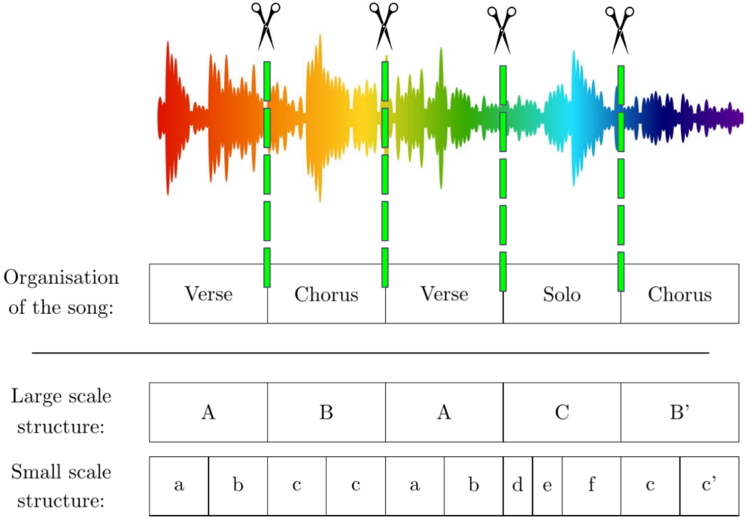
Frederic Bimbot
DR CNRS



Caglayan Tuna
Ingénieur Inria

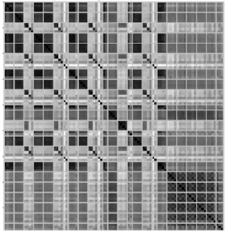
Axel Marmoret, Jérémy Cohen, Nancy Bertin, Frédéric Bimbot. Uncovering Audio Patterns in Music with Nonnegative Tucker Decomposition for Structural Segmentation. ISMIR 2020 - 21st International Society for Music Information Retrieval, Oct 2020, Montréal (Online), Canada. pp.1-7

Segmenting a song?



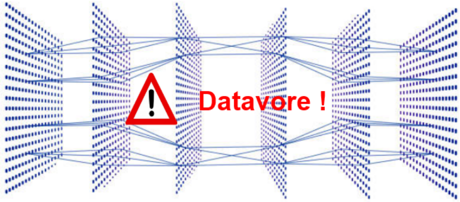
A word on the state-of-the-art

Unsupervised



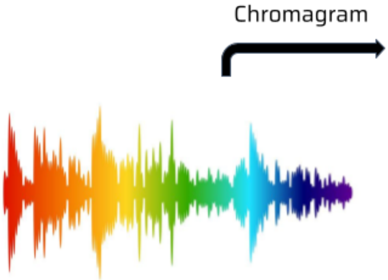
Signal Autosimilarity + post-processing

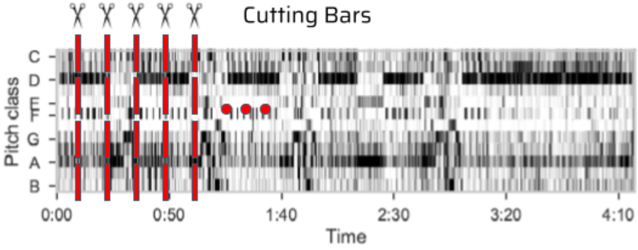
Supervised



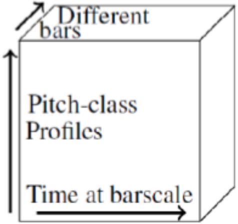
Deep learning

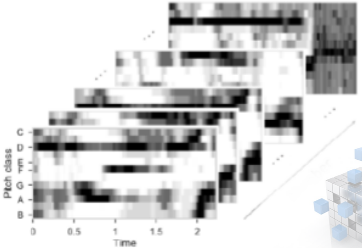
Our idea: a chromagram tensor...





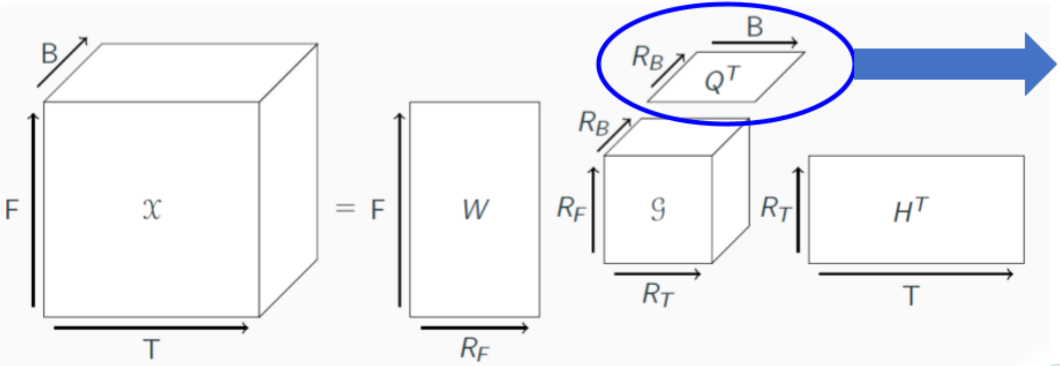
Chromagram of "Come Together", by The Beatles.





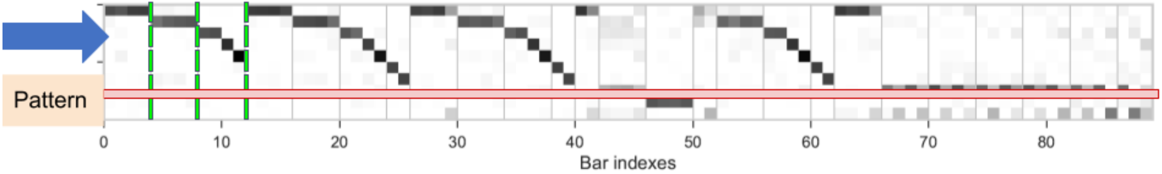
...decomposed to find redundancies!

Approximate Nonnegative Tucker Decomposition $\mathcal{X} \approx (W \otimes H \otimes Q)\mathcal{G}$

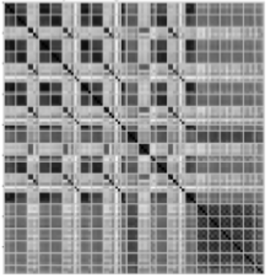


$\mathcal{X} \in \mathbb{R}_+^{F \times T \times B}, W \in \mathbb{R}_+^{F \times R_F}, H \in \mathbb{R}_+^{T \times R_T}, Q \in \mathbb{R}_+^{B \times R_B}, \mathcal{G} \in \mathbb{R}_+^{R_F \times R_T \times R_B}$

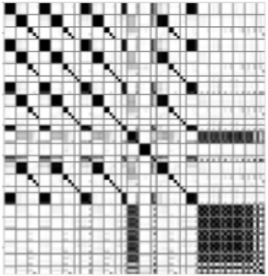
Back to segmentation



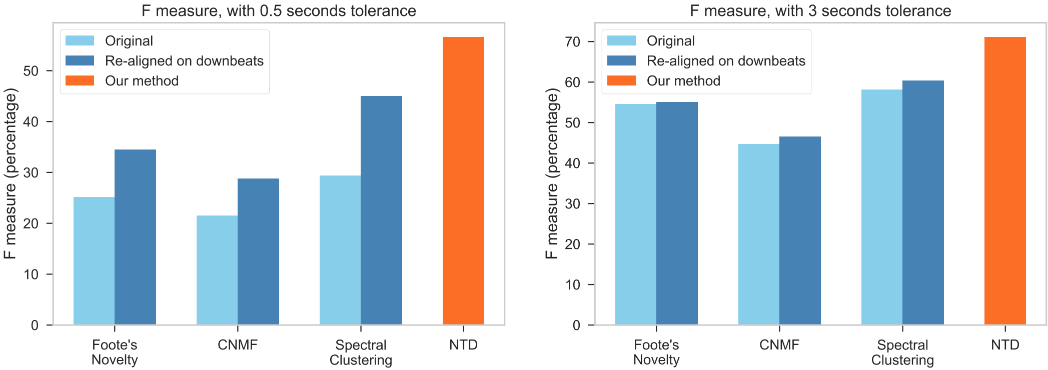
Signal Autosimilarity



Patterns autosimilarity



State-of-the-art unsupervised results!



Algorithm	$P_{0.5}$	$R_{0.5}$	$F_{0.5}$	P_3	R_3	F_3
NTD, with "oracle ranks" for each song	67.1%	78.2%	71.5%	78.5%	90.2%	83.1%
Neural Networks[Grill2015]	80.4%	62.7%	69.7%	91.9%	71.1%	79.3%

Table: Averaged segmentation scores in the "oracle ranks" condition, compared to the current state-of-the-art (non-blind) method.

An algorithmic road

HALS principles
~2008
Nonnegative Matrix
factorization

Nicolas
Gillis,
UMONS

MATLAB

Implementation and
acceleration
~2012

PARAFAC
decomposition
~2019

Nonnegative
Tucker
~2020

Packages nnfac and MusicNTD

jupyter

GitLab

nnfac

Inria

TensorLy

TensorFlow

PyTorch

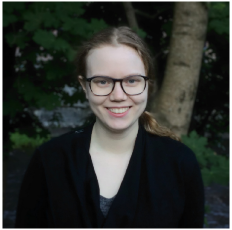
NVIDIA CUDA

Roadmap

1. An introduction to tensor methods
2. Nonnegative Tucker decomposition of music for automatic segmentation
3. Algorithms for constrained linearly coupled factorizations

Another team effort

Carla Schenker
PhD Student



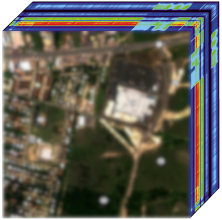
Marie Roald
PhD Student

Evrim Acar
Senior Researcher

SimulaMet, Oslo

- C. Schenker, J. E. Cohen, E. Acar, "A Flexible Optimization Framework for Regularized Matrix-Tensor Factorizations with Linear Couplings", IEEE Journal on Selected Topics in Signal Processing, 2020.
- M. Roald, C. Schenker, J. E. Cohen, Evrim Acar, "PARAFAC2 AO-ADMM: Constraints in all modes", EUSIPCO2021

Hyperspectral super-resolution: a motivating example



- High spatial resolution
- Low spectral resolution

- Low spatial resolution
- High spectral resolution

Hyperspectral super-resolution in equations

$$\underset{W_h \geq 0, H_h \geq 0, W_m \geq 0, H_m \geq 0}{\text{argmin}} \|X_h - W_h H_h\|_F^2 + \|X_m - W_m H_m\|_F^2 \text{ such that } RW_h = W_m, \; H_h = SH_m \quad (1)$$

Hyperspectral super-resolution in equations

$$\underset{W_h \geq 0, H_h \geq 0, W_m \geq 0, H_m \geq 0}{\text{argmin}} \|X_h - W_h H_h\|_F^2 + \|X_m - W_m H_m\|_F^2 \text{ such that } RW_h = W_m,\ H_h = SH_m \quad (1)$$

Low-rank approximations (NMFs)

Hyperspectral super-resolution in equations

$$\underset{W_h \geq 0, H_h \geq 0, W_m \geq 0, H_m \geq 0}{\text{argmin}} \|X_h - W_h H_h\|_F^2 + \|X_m - W_m H_m\|_F^2 \text{ such that } RW_h = W_m, \; H_h = SH_m \quad (1)$$

Low-rank approximations (NMFs)

Spectral reduction

Blur and downsampling

Hyperspectral super-resolution in equations

$$\underset{W_h \geq 0, H_h \geq 0, W_m \geq 0, H_m \geq 0}{\text{argmin}} \|X_h - W_h H_h\|_F^2 + \|X_m - W_m H_m\|_F^2 \text{ such that } RW_h = W_m,\ H_h = SH_m \quad (1)$$

Low-rank approximations (NMFs)

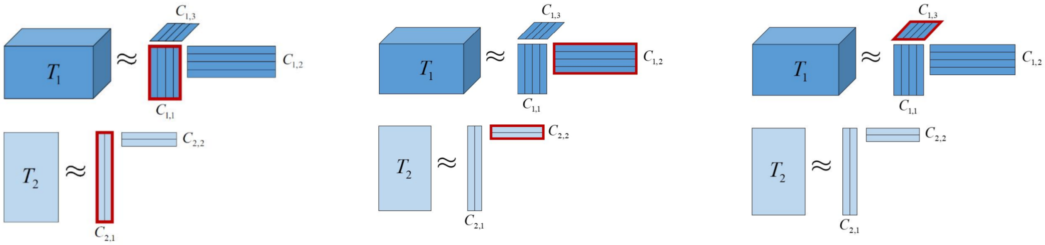
Spectral reduction

Blur and downsampling

In fact we deal with a more general problem:

- Inputs are low-rank tensors, there can be more than two.
- Constraints are versatile (l1 norm, smoothness, total variation...).
- Loss functions need not be Euclidean.

The proposed AO-ADMM framework



Algorithm 2 ADMM for subproblem w.r.t. mode 1 of regularized linearly coupled CPD

while convergence criterion is not met **do**

  for $i = 1, ..., N$ **do**

    $$\mathbf{C}_{i,1}^{(k+1)} = \underset{\mathbf{X}}{\text{argmin}}\ w_i \mathcal{L}_i\left(\boldsymbol{\mathcal{T}}_i, [\![\mathbf{X}, \mathbf{C}_{i,2}, ..., \mathbf{C}_{i,D_i}]\!]\right) + \frac{\rho}{2}\left(\left\|\mathbf{X} - \mathbf{Z}_{i,1}^{(k)} + \boldsymbol{\mu}_{i,1}{}_{(z)}^{(k)}\right\|_F^2 + \left\|\mathbf{H}_{i,1}\,\text{vec}(\mathbf{X}) - \mathbf{H}_{i,1}^{\Delta}\boldsymbol{\delta}_1^{(k)} + \boldsymbol{\mu}_{i,1}{}_{(\delta)}^{(k)}\right\|_2^2\right)$$

  end for

  $$\boldsymbol{\delta}_1^{(k+1)} = \underset{\mathbf{z}}{\text{argmin}} \sum_{i=1}^{N}\left\|\mathbf{H}_{i,1}\,\text{vec}(\mathbf{C}_{i,1}^{(k+1)}) - \mathbf{H}_{i,1}^{\Delta}\mathbf{z} + \boldsymbol{\mu}_{i,1}{}_{(\delta)}^{(k)}\right\|_2^2$$

  for $i = 1, ..., N$ **do**

    $$\mathbf{Z}_{i,1}^{(k+1)} = \underset{\mathbf{Z}}{\text{argmin}}\ g_{i,1}(\mathbf{Z}) + \frac{\rho}{2}\left\|\mathbf{C}_{i,1}^{(k+1)} - \mathbf{Z} + \boldsymbol{\mu}_{i,1}{}_{(z)}^{(k)}\right\|_F^2 = \text{prox}_{\frac{1}{\rho}g_{i,1}}\left(\mathbf{C}_{i,1}^{(k+1)} + \boldsymbol{\mu}_{i,1}{}_{(z)}^{(k)}\right)$$

    $$\boldsymbol{\mu}_{i,1}{}_{(z)}^{(k+1)} = \boldsymbol{\mu}_{i,1}{}_{(z)}^{(k)} + \mathbf{C}_{i,1}^{(k+1)} - \mathbf{Z}_{i,1}^{(k+1)}$$

    $$\boldsymbol{\mu}_{i,1}{}_{(\delta)}^{(k+1)} = \boldsymbol{\mu}_{i,1}{}_{(\delta)}^{(k)} + \mathbf{H}_{i,1}\,\text{vec}(\mathbf{C}_{i,1}^{(k+1)}) - \mathbf{H}_{i,1}^{\Delta}\boldsymbol{\delta}_1^{(k+1)}$$

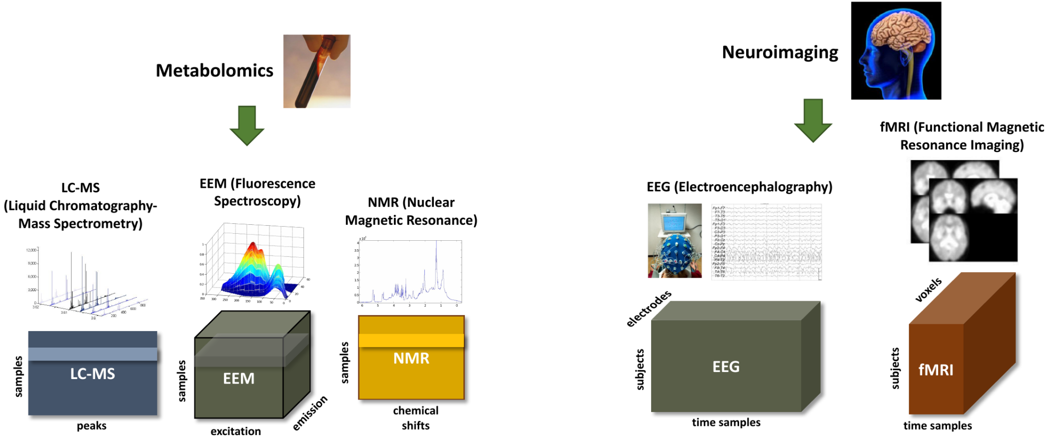
  end for

  $k = k + 1$

end while

ADMM algorithm for each mode

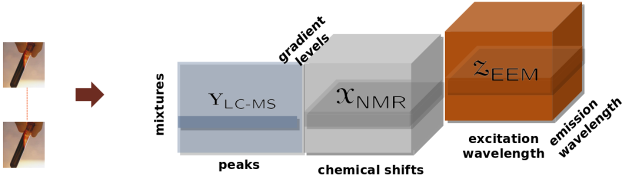
What is this good for?



Also used in

- Super-resolution in hyperspectral imaging (mostly remote sensing)
- Chemometrics/Metabolomics: Bypassing time-retention shifts in GC-MS
- Neuro-imaging: EEG and Oculometry
- Many more...

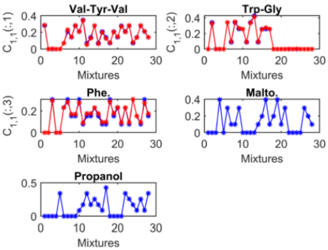
Chemometrics: Underlying design and patterns captured accurately!



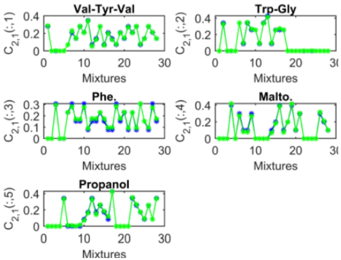
Mixtures prepared using five amino-acids/chemicals

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

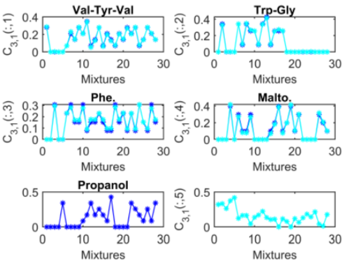
$C_{1,1}$ (EEM)



$C_{2,1}$ (NMR)

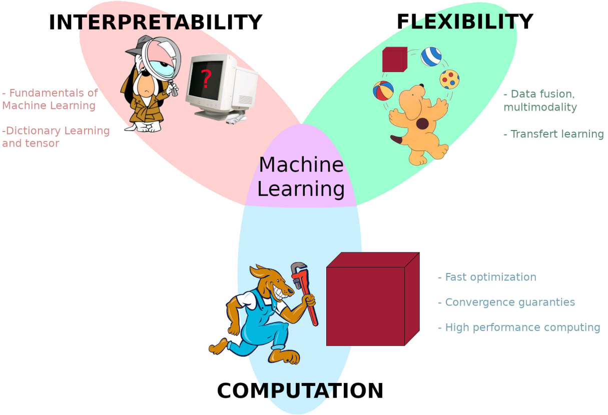


$C_{3,1}$ (LCMS)



Conclusion: Low-rank approximations are versatile

- Vector and Matrix dataset can be tensorized (cf Audio project) and processed with tensor decompositions.
- Tensor dataset can be matricized and treated with matrix factorization (cf Hyperspectral Imaging)
- Domain-specific constraints can be accounted for with some work on the algorithms (cf Data fusion using AO-ADMM)



My ongoing research projects

LoRAiA (ANR JCJC)

Semi-supervision and Tensors:

- Dictionaries/sparse coding
- Optimal Transport

with efficient implementations/algorithms!

Automatic Transcription

With semi-supervision and NMF.

Tensoptly (Inria)

Tensorly optimization layer:

- Constrained models
- Faster algorithms
- Customization

Music Segmentation

PhD of Axel Marmoret.

Sparse/Fast Optimization

Long-term collaboration with N. Gillis (UMONS).

Multimodality

Long-term collaboration with E. Acar (SimulaMet).

Thank you for your attention