UCLouvain

ACCELERATING NONNEGATIVE MATRIX FACTORIZATION OVER POLYNOMIAL SIGNALS WITH FASTER PROJECTIONS

Cécile Hautecoeur

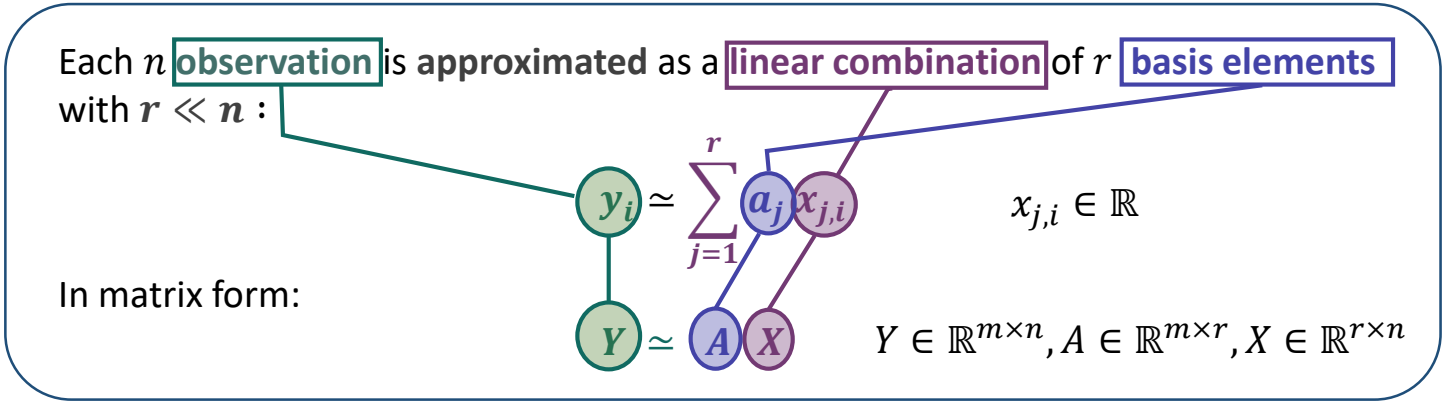
(joint work with François Glineur)

Université catholique de Louvain. ICTEAM/CORE Institute

icteam

EOS
THE EXCELLENCE OF SCIENCE

One approach : **Linear Dimensionality Reduction** (or low-rank matrix factorization)



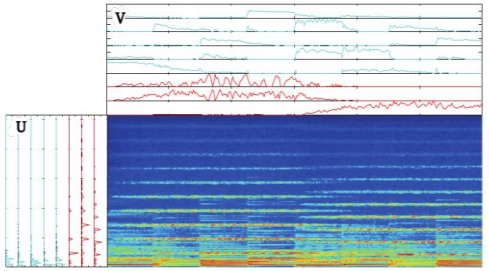
- ✓ Dimensionality reduction
- ✓ Identification of dominant features (A)
  - ➢ Filters redundancy and noise/errors
  - ➢ Helps user interpretation
- ✓ Compact representation
- ✓ Extraction of feature's coefficients (X)
  - ➢ Good for data analysis

Nonnegative Matrix Factorization (NMF)

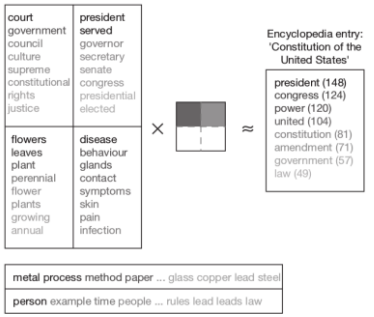
- Used for **nonnegative** data: images, amplitude spectrogram (sound), word frequencies (text), reflectance (hyper-spectral images), etc...

Examples in

Image processing,
[Buciu & al., 2008]



Music analysis,
[Kameoka & al., 2012]



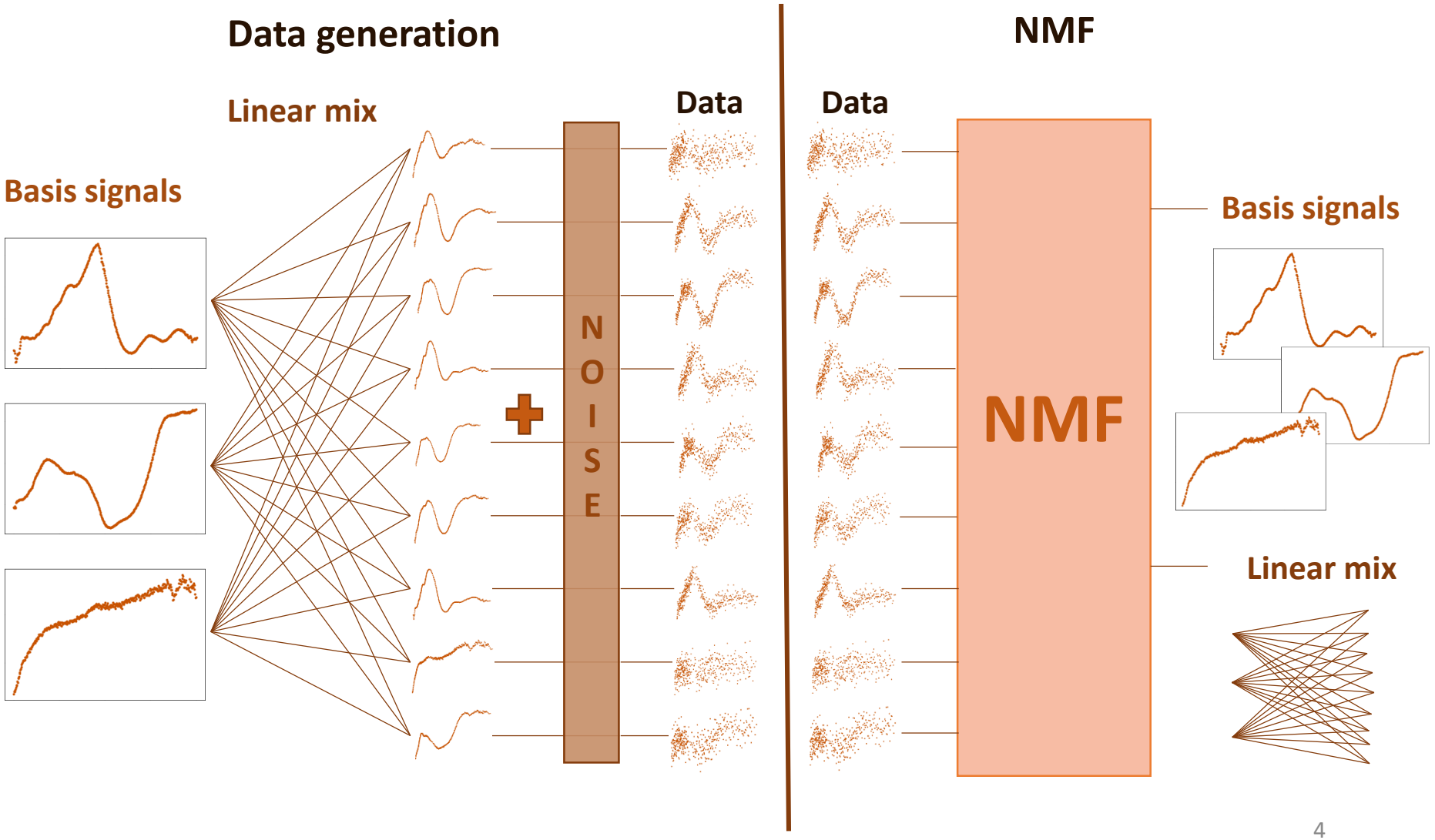
Text mining,
[Lee & Seung, 1999]

etc...

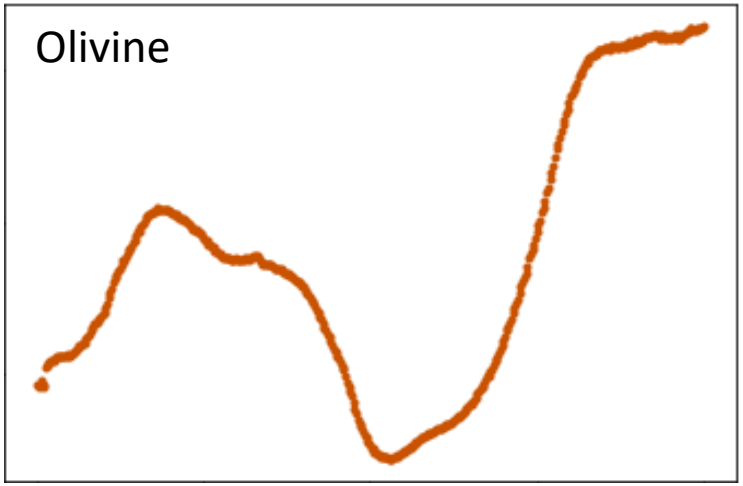
- The **basis elements** (vectors a_j) as well as the **coefficients** ($x_{j,i}$) are imposed to be nonnegative:

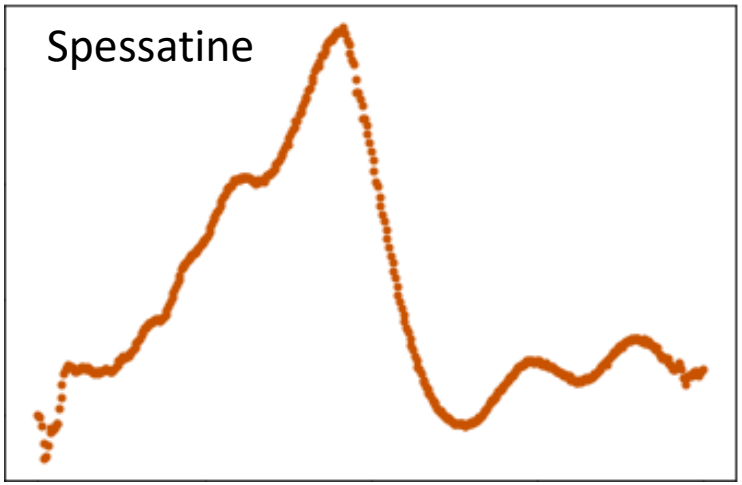
$$a_j \geq 0 \; and \; x_{j,i} \geq 0 \quad \forall i, j.$$

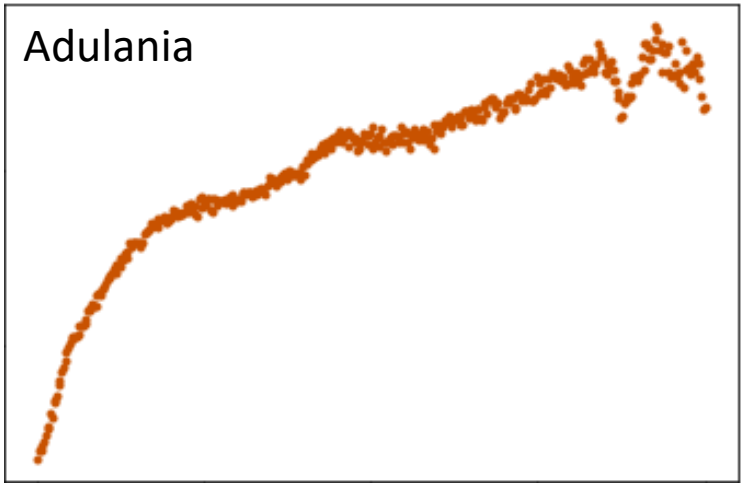
Nonnegative Matrix Factorization: illustration



Ground truth signals

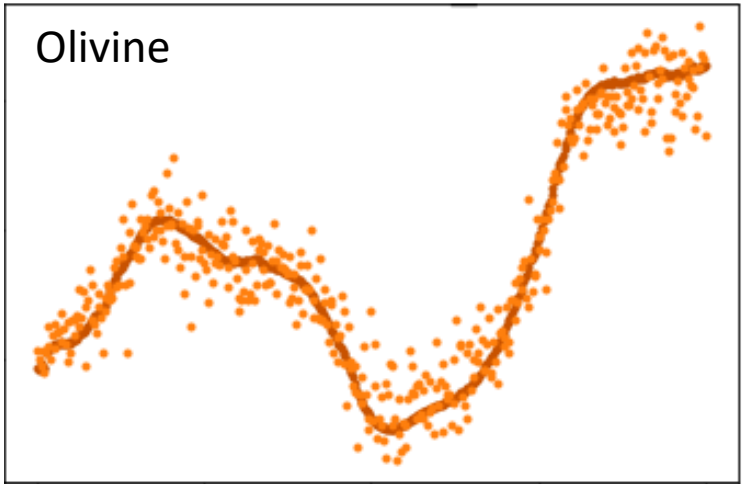


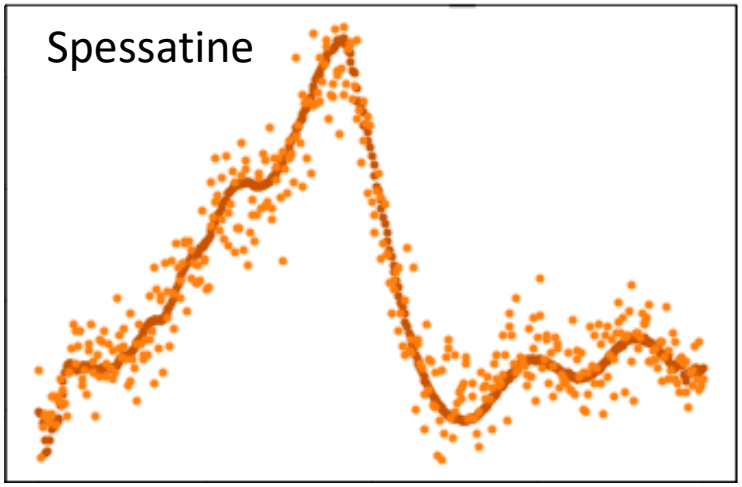


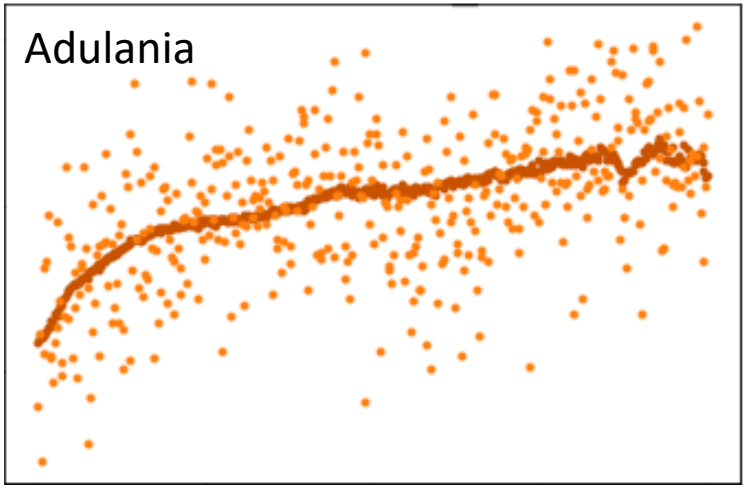


Reflectance signals from https://www.usgs.gov/labs/spec-lab/capabilities/spectral-library

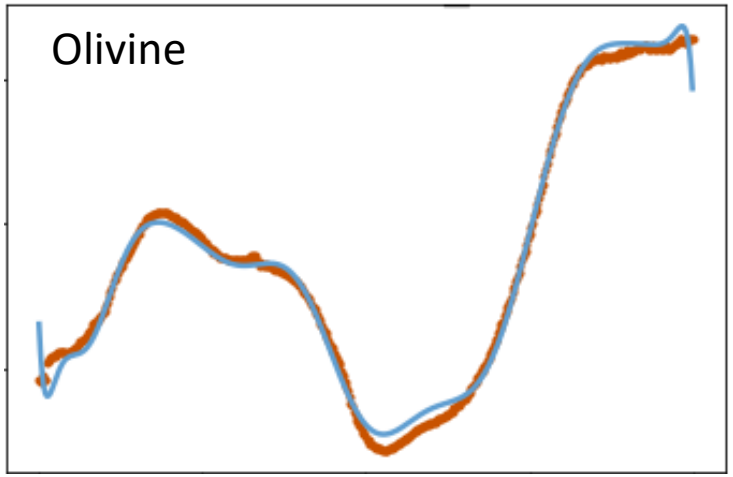
Example of recovered signals, with noise level = 20 dB and 50 observations

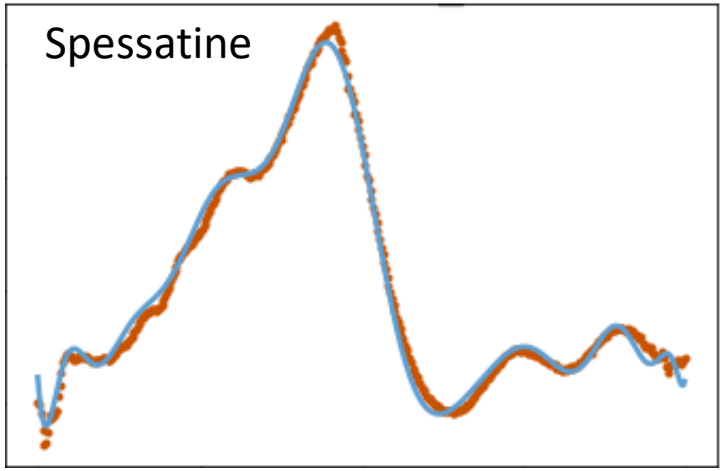


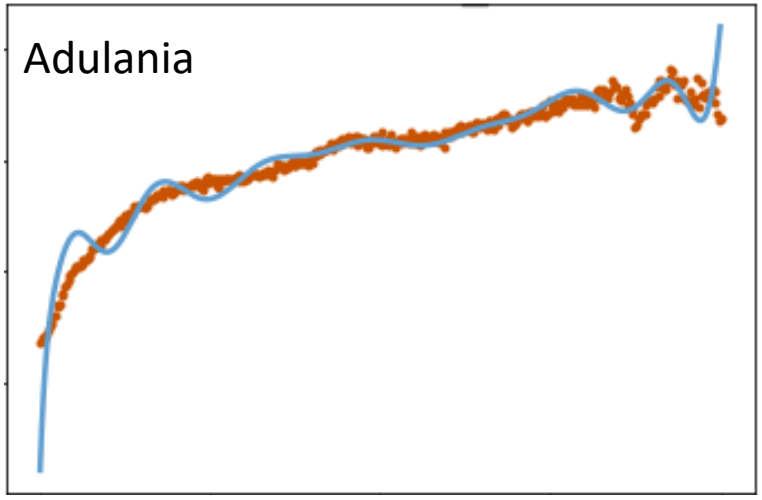




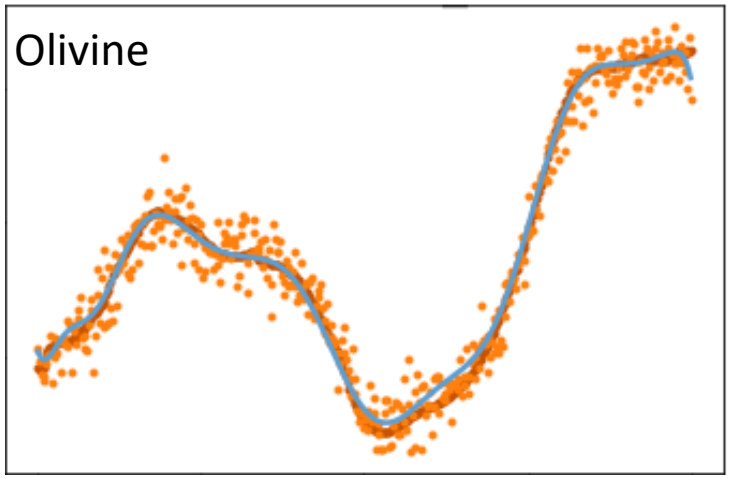
Example of recovered signals, imposing them to be polynomials

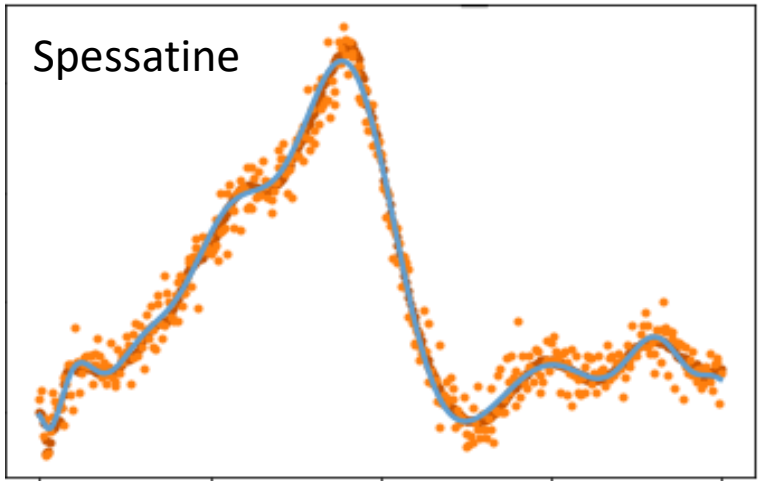


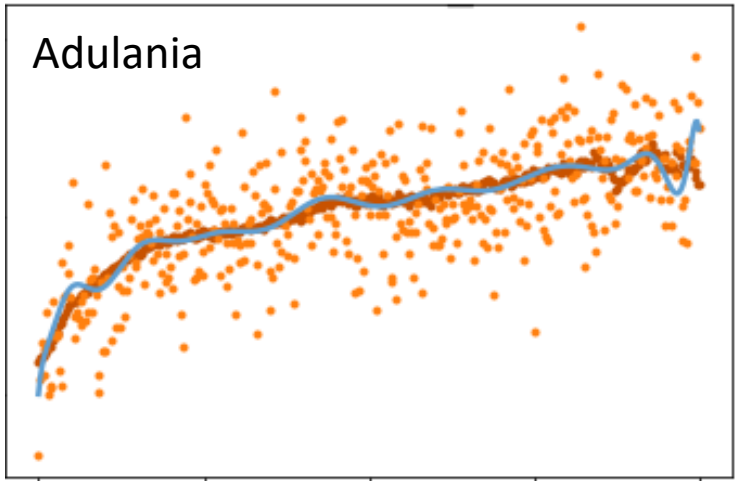


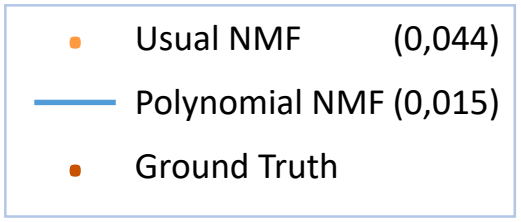


Comparison of signals recovered by usual NMF and polynomial NMF

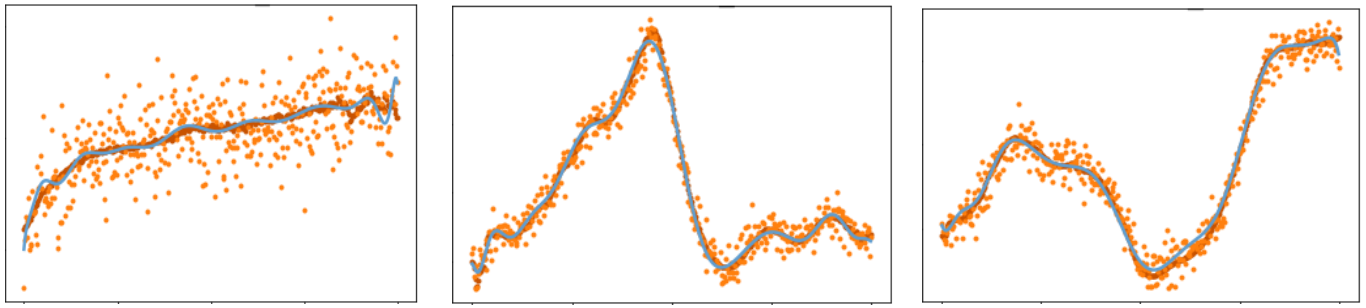








NMF using polynomials (Polynomial-NMF)



- An extension of NMF use polynomials in the factorization [Debals et al. 2017, Hautecoeur & Glineur 2019, Zdunek 2014 (using splines)]

  Each n **observation** is **approximated** as a **nonnegative linear combination** of r **nonnegative polynomials** with $r \ll n$:

$$y_i(t) \simeq \sum_{j=1}^{r} a_j(t)\, x_{j,i} \qquad a_j(t) \textbf{ nonnegative polynomial}, \; x_{j,i} \in \mathbb{R}^+$$

- Recovers smooth signals, and less sensitive to noise

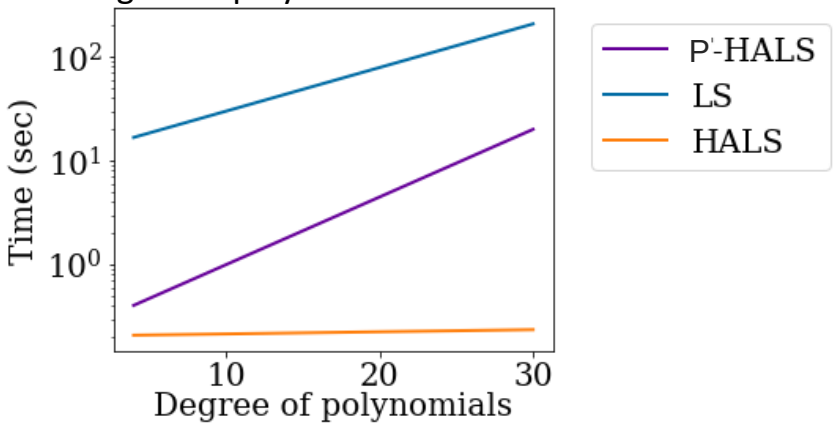
Polynomial NMF

- Recovers smooth signals, and less sensitive to noise

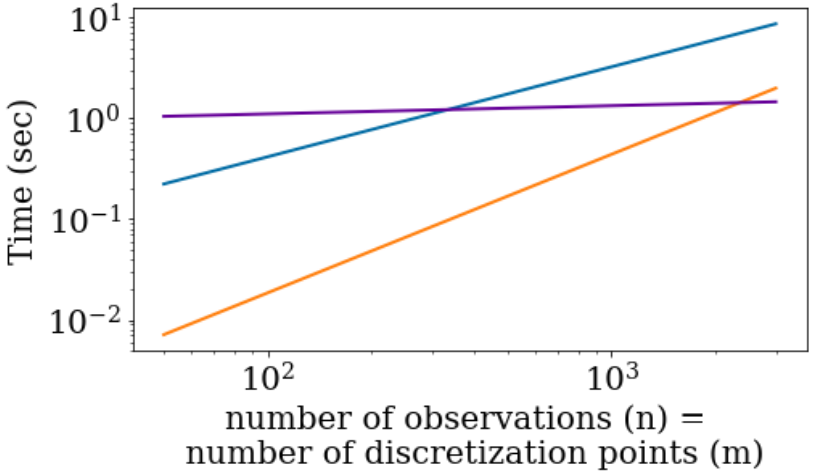
- But can be **slower**, especially when analyzing low-scale data or using **high-degree polynomials**

P-HALS and **LS** are 2 methods for NMF using polynomials.
HALS is a method for usual NMF.

Evolution of CPU time with respect to degree of polynomials used in NMF



Evolution of CPU time with respect to the size of the dataset



Polynomial NMF

- Recovers smooth signals, and less sensitive to noise
- But can be **slower**, especially when analyzing low-scale data or using **high-degree polynomials**

Can we **accelerate** these algorithms and **how**?

Outline

- Methods for Polynomial-NMF
- Improve polynomial projections
- Numerical results

LS algorithm [Debals & al., 2017]

Polynomial NMF: $\min_{A,X\geq 0} \left|\left|Y - AX\right|\right|_F^2$ where A contains polynomials

Can be formulated as an (unconstrained) nonlinear least-squares problem in B,H

$$\min_{B,H} \left|\left|Y - f(B,H)\right|\right|_F^2$$

Solved as a whole using least-squares solver.

- + unconstrained formulation
- + attains good accuracy solutions
- - depends on solver efficiency
- - slow

P-HALS algorithm [Hautecoeur & Glineur, 2019]

Polynomial NMF: $\min_{A,X\geq 0} ||Y - AX||_F^2$ where A contains polynomials

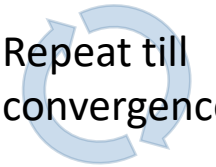
Let Π a Vandermonde-like matrix: $a_i(t) = \Pi b$, with b the coefficient vector of size d+1.

The problem becomes

$$\min_{B,X} ||Y - \Pi BX||_F^2$$

$$such\ that\ X \geq 0, B \geq_{pol} 0$$

Solved alternatively using hierarchical alternating least squares (HALS):



Repeat till convergence

- Find optimal row X_i, with all other variable fixed
- Project X_i on its feasible set [repeat for all X_i]
- Find optimal column B_i, with all other variable fixed
- **Project B_i on its feasible set** [repeat for all B_i]

P-HALS algorithm

- Find optimal row X_i, with all other variable fixed
- Project X_i on its feasible set [repeat for all X_i]
- Find optimal column B_i, with all other variable fixed
- **Project B_i on its feasible set** [repeat for all B_i]

All steps are straigthforward except the projection of B_i

Projection dominates total algorithmic cost
(e.g. >97% for n=m=1000, d=12)

- + convex in B and X, potentially fast updates
- + guaranteed convergence to stationary point
- - update in B is costly (because of projection)

Objective

Accelerate the projection of polynomials on the nonnegative set

To improve the promising P-HALS algorithm.

Outline

- Methods for Polynomial-NMF
- Improve polynomial projections
- Numerical results

Projection over nonnegative polynomials using *Sum-Of-Squares*

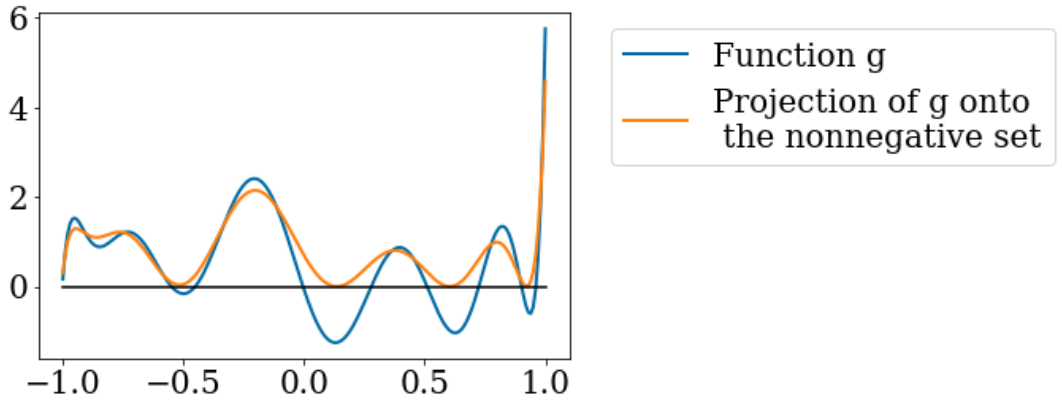
Projection computed as **Semidefinite optimization** problem:

To the best of our knowledge, **only way to compute exact projection**

Suppose g is a polynomial of degree d.

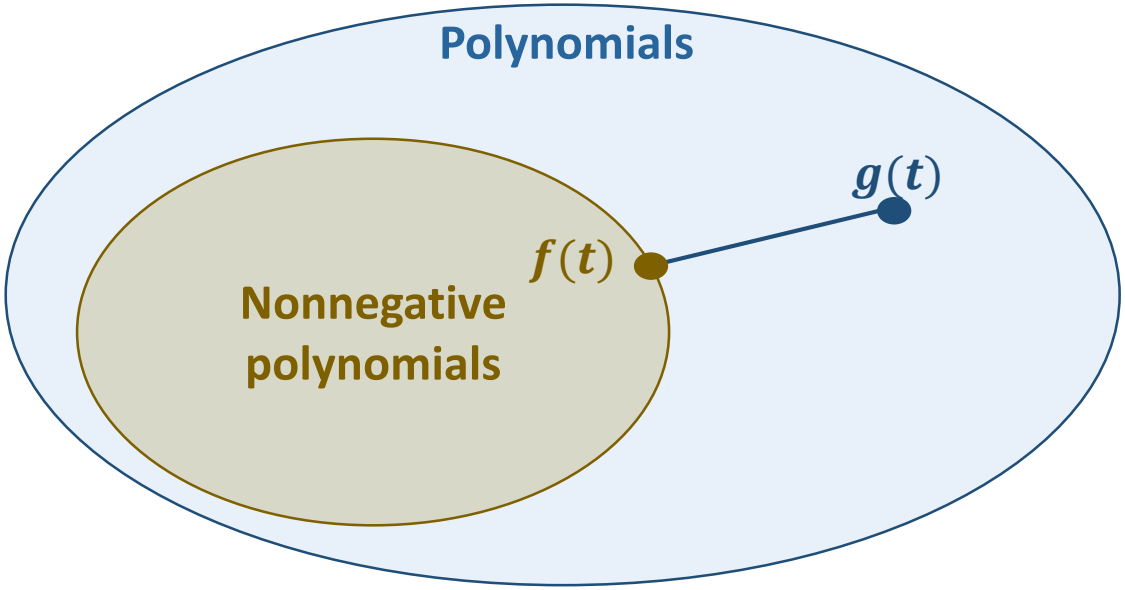
Can be expressed using positive **semidefinite cone**

- $g(t)$ is nonnegative iff it is Sum-Of-squares: $g(t) \geq 0 \; \forall t \;\Leftrightarrow g(t) = \sum_i f_i(t)^2$
- $g(t)$ is nonnegative over [-1,1] iff $g(t) = f(t) + (1 - t^2)h(t) \quad f(t), h(t) \geq 0 \; \forall t$



Three faster (approximate) projections

Projection: Given $g(t)$ find $f(t)$ the closest nonnegative polynomial to $g(t)$.

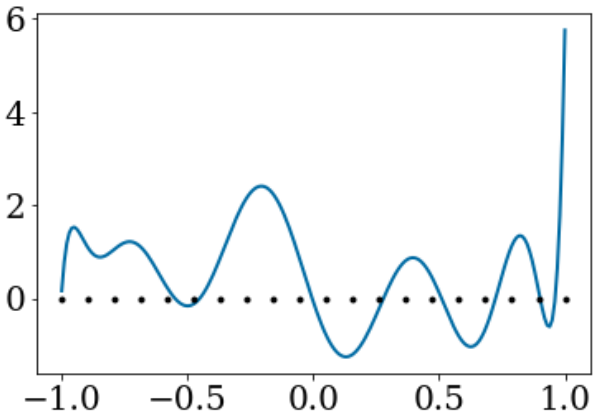


3 ideas:

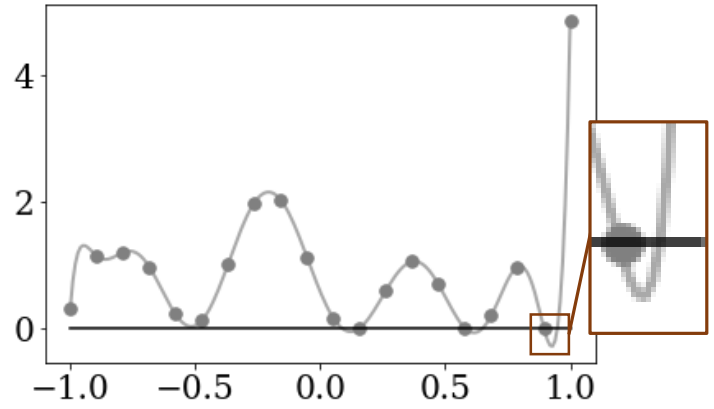
- **Discretization (Discr)**: discretizes signals
- **Proximal (Prox):** use known projection of a nearby polynomial
- **Iterative heuristic (H1 and H2):** uses iteratively polynomial curve fitting

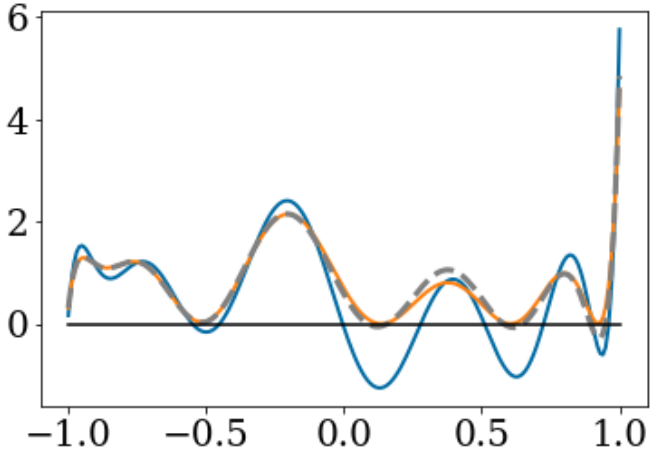
Discretization (Discr)

Discretize signals and impose nonnegativity at the discretization points
➡ gives an (easier) linear optimization problem



Linear optimization problem





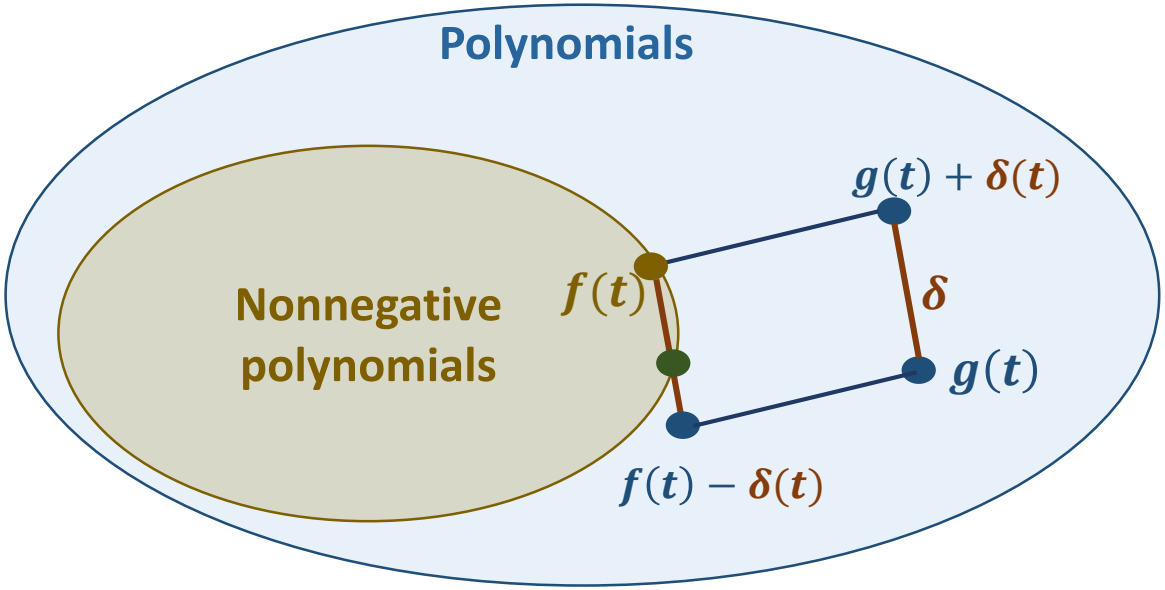
Proximal projection (prox)

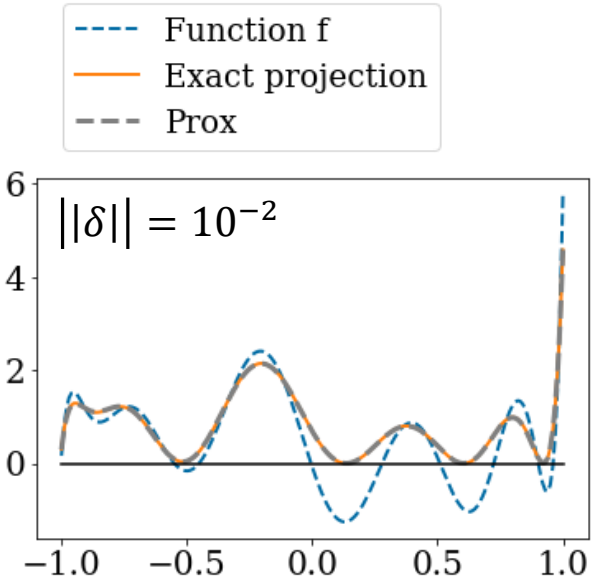
$$If \;\; Projection(\boldsymbol{g} + \boldsymbol{\delta}) = \boldsymbol{f}, \qquad Projection(\boldsymbol{g}) \simeq \boldsymbol{f} - \boldsymbol{\delta}$$

But $\boldsymbol{f} - \boldsymbol{\delta}$ not always nonnegative.

Find maximal $\gamma \in [0,1]$ such that $\boldsymbol{f} - \gamma\boldsymbol{\delta}$ nonnegative.

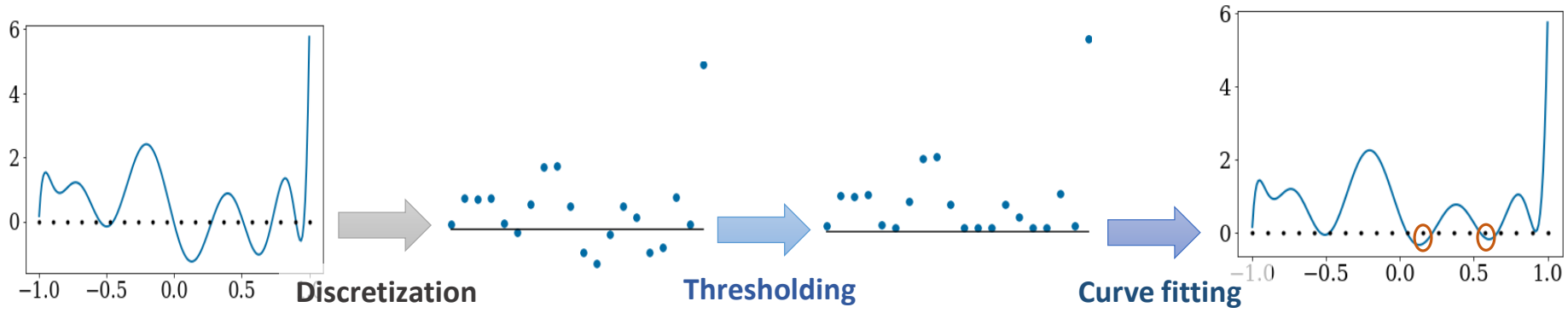
As convex can use bisection search, using discretization to check nonnegativity.





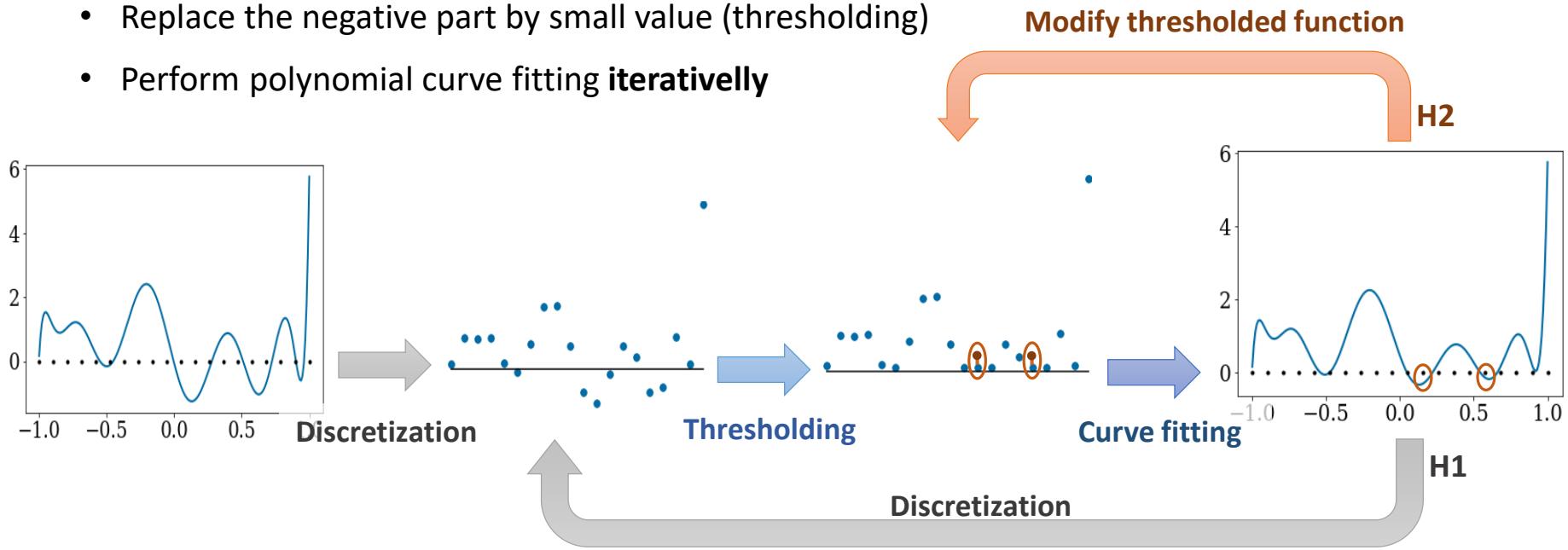
Iterative approaches (H1 and H2)

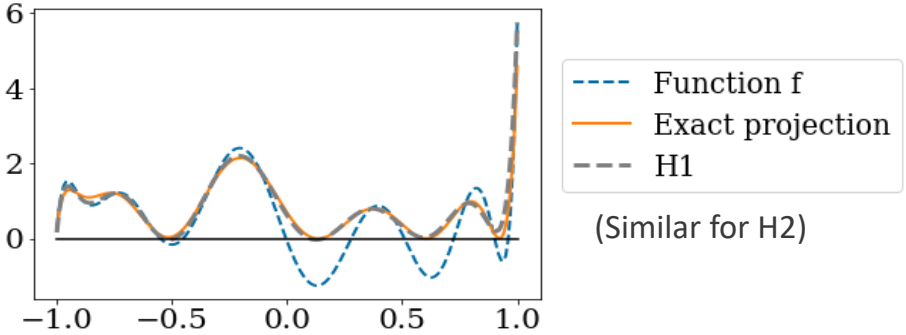
- Discretize signals
- Replace the negative part by small value (thresholding)
- Perform polynomial curve fitting



Iterative approaches (H1 and H2)

- Discretize signals
- Replace the negative part by small value (thresholding)
- Perform polynomial curve fitting **iteratively**





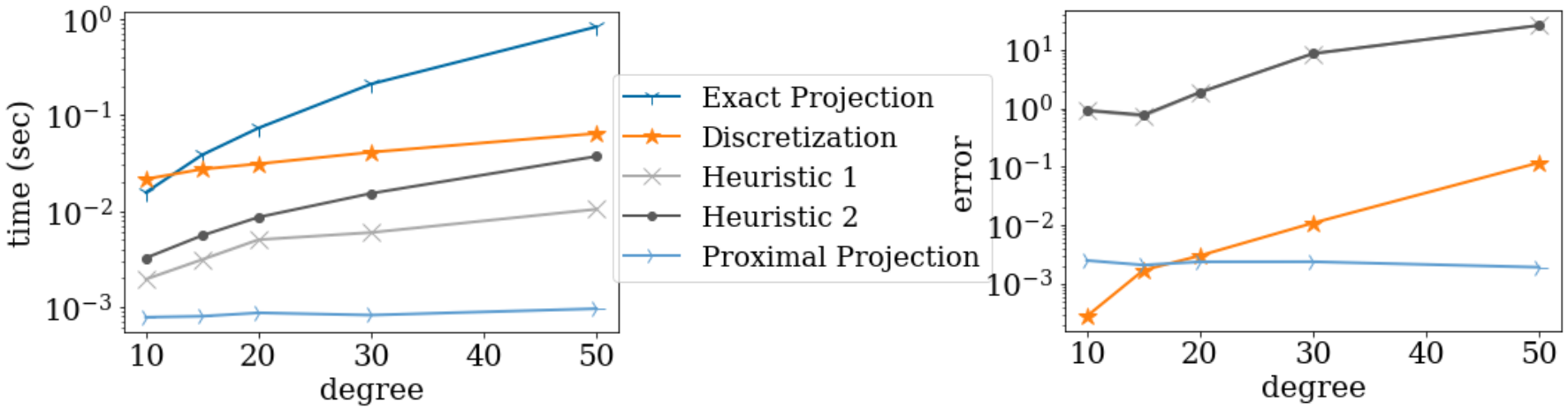
(Similar for H2)

Outline

- Methods for Polynomial-NMF
- Improve polynomial projections
- Numerical results

Comparison of approximate projections

Projection of random polynomials with various degrees.
Assess both CPU time and accuracy (error) for each approximate projection technique



- **Discr:** Precise, faster than Exact projection but still relatively slow (especially for low-degree polynomials)
- **H1/2**: Less precise but (much) faster than Exact projection (H1 better than H2)
- **Prox**: Fast and precise, but requires knowing projection of close polynomial

Numerical results for solving Polynomial NMF

- Test on real signals with synthetic mixing (ground truth is known)
- Generation:
  - $A = 5$ real reflectance spectra with 414 discretization points[1]
  - $X \sim N(0,1)$, with negative values replaced by 0
  - $Y = AX + N_o$ where N_o = Gaussian additive noise with SNR = 20dB

- Parameters:

r	5
# observations	100
# discretization points	414
Degree (polynomials)	12

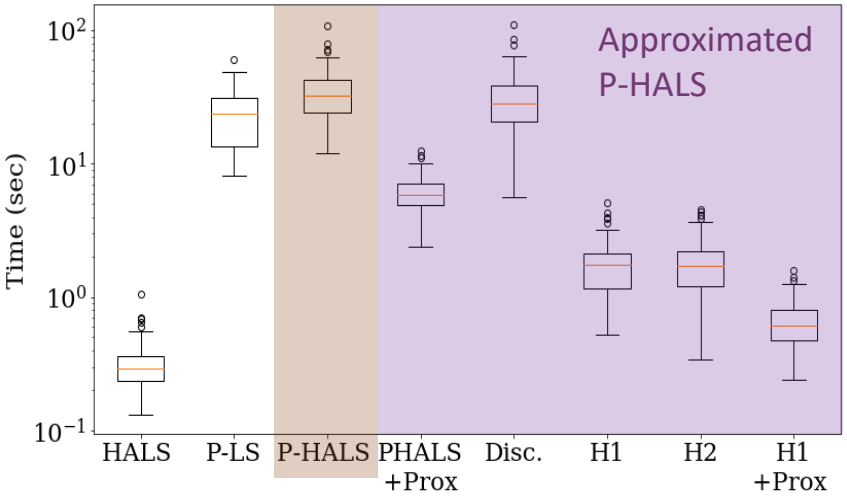
- Evaluation with $\tilde{B}, \tilde{X}$ the obtained matrices:
  - Residual = $dist(AX, \Pi\tilde{B}\tilde{X})$
  - SIR_A = Signal to interference ratio $(A, \bar{A})$
    with $\bar{A}$ the best nonnegative linear combination of $\Pi\tilde{B}$ to obtain A $(A \simeq \bar{A} = \Pi\tilde{B}\,Q, \quad Q \in \mathbb{R}_+)$
  - SIR_X = Signal to interference ratio $(X, \bar{X})$
    with $\bar{X} = Q^{-1}\tilde{X}$

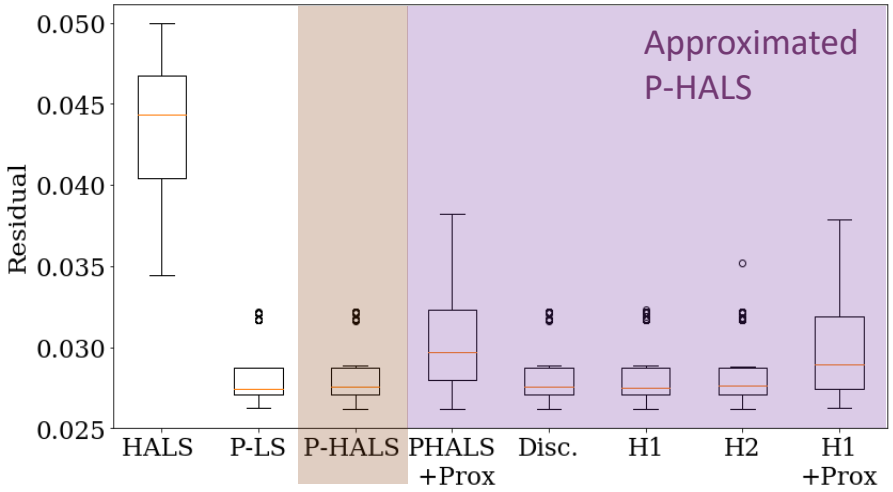
The lower the residual the better, and the higher the SIR, the better.

1. https://www.usgs.gov/labs/spec-lab/capabilities/spectral-library

Testing algorithms for P-NMF with/without approximate projections

Box Plot of the CPU time and accuracy (error) for different NMF algorithms.





H1 and H2 perform quite well although they rely on much less precise projections

.

➡ Using an approximate but much accelerated projection leads to final solutions with accuracy similar to exact P-HALS

Take home messages

➢**Nonnegative Matrix Factorization** given a nonnegative dataset recovers characteristic and interpretable features in an unsupervised way.

➢ Improvable using **polynomial signals in the factorization** (LS and P-HALS)

➢P-HALS can be **accelerated** using **approximate projection** of polynomials**, without loss of accuracy**

➢Experimentally, the best heuristic is an iterative approach using polynomial curve fitting

Approximate projections are also effective on P-HALS using splines, using spline-specific heuristics.

➡ Look for a general heuristic for all functions.

P-HALS converges to a stationary point

➡ Look for convergence properties of approximated P-HALS

Future work

Thank you for your attention

Related work:

• **LS:** Otto Debals, Marc Van Barel, and Lieven De Lathauwer. "Nonnegative matrix factorization using nonnegative polynomial approximations". *IEEE Signal Processing Letters*, 24(7):948–952, 2017.

Our work:

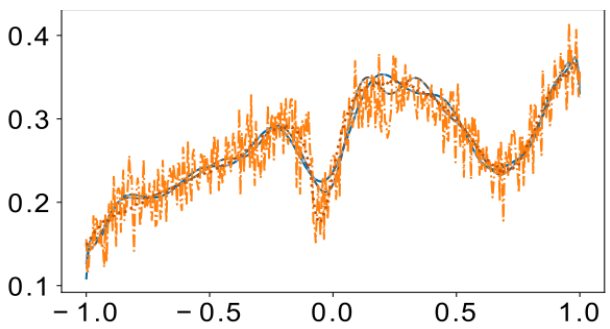
Cécile Hautecoeur and François Glineur, "Nonnegative matrix factorization with polynomial signals via hierarchical alternating least squares", in *European Symposium on Artificial Neural Networks (ESANN)*, pages 125–130, 2019.

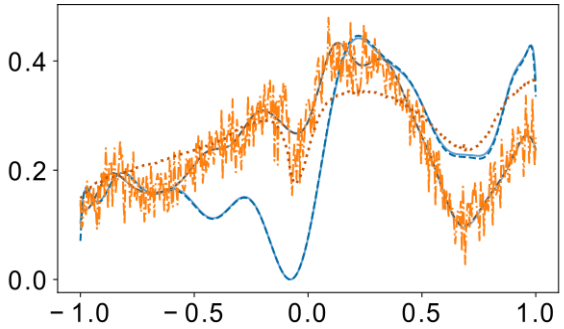
And its extention: "Nonnegative Matrix Factorization over continuous signals using parametrizable functions" (Neurocomputing, accepted with minor revision)

Cécile Hautecoeur and François Glineur, "Accelerating Nonnegative Matrix Factorization over polynomial signals with faster projections". In *2019 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2019.

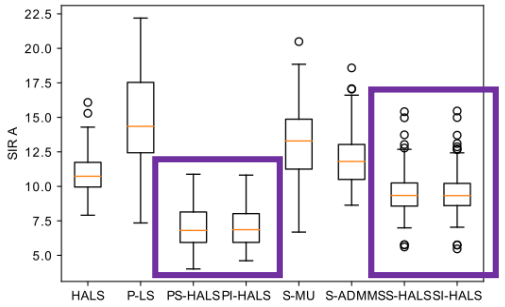
Comment about the recovered signals

- NMF is insensitive to permutations and scaling of matrices A and X
- Until now, we observed the best nonnegative linear combination of the obtained signals to recover the original basis





- In general, F-HALS obtains less similar signals than the other approaches when considering only permutations and scaling



F-HALS

Comparison

Method	Penalty term	Exact parametrization $(f(B), LS)$	Nonnegative functions $(f(B) \geq 0$, HALS)
Pro	• **Simple** implementation • Numerical stability • Fast	• Superlinear convergence • **Nonnegative feature functions**	• **Convergence to stationary point** • **Nonnegative feature functions** • **Avoid discretization** • **Extendable to other functions**
Cons	• Data-driven parameters • **Non continuous features**	• Nonlinear parametrization • **Slowest** (even though good asymptotic complexity)	• **Slower (due to projection, but can be accelerated ...)**